



Autónoma
Universidad Autónoma del Perú

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
CARRERA PROFESIONAL DE INGENIERÍA DE
SISTEMAS**

TESIS

“APLICACIÓN DEL ALGORITMO PARA PREDECIR EL RIESGO DE
CÁNCER DE MAMA EN LAS PACIENTES DEL HOSPITAL DE LA
SOLIDARIDAD DE VILLA EL SALVADOR”

**PARA OBTENER EL TÍTULO DE
INGENIERO DE SISTEMAS**

AUTORES

CYNTIA LISBETH GAMARRA CONDO
YADIRA VICTORIA SANTOS MELO

ASESOR

MG. JOSÉ LUIS HERRERA SALAZAR

LIMA, PERÚ, MARZO DE 2019

DEDICATORIA

A Dios, que me ha dado la vida, fortaleza y persistencia. A mis padres, que día a día luchan para darme lo mejor y sacarme adelante a pesar de las adversidades que se nos presentan. A mis hermanos, que siempre me apoyan y me dan ánimos de levantarme cada día.

Cyntia Lisbeth Gamarra Condo

A mis padres (Marlene y Gregorio), mis tíos, mi hermana, mis profesores y a mis verdaderas amistades porque creyeron en mí, dándome ejemplos dignos de superación y entrega, porque en gran parte gracias a ustedes, ya que siempre estuvieron impulsándome en los momentos más difíciles de mi carrera.

A una de mis mejores amigas y colega de esta gran investigación Cyntia Gamarra que, gracias a la perseverancia, se pudo tener éxito con el mismo objetivo trazado.

Así mismo quiero agradecer a Dios porque siempre me brindó la fortaleza, actitud, fuerzas para seguir adelante y poder culminar con una de mis metas.

Yadira Victoria Santos Melo

AGRADECIMIENTOS

La presente tesis es el resultado del esfuerzo y dedicación constante que sin la participación de muchas personas no hubiese sido posible llevarla a un feliz término. Por ello es para nosotras un verdadero placer utilizar este espacio para agradecer de manera justa y consecuente expresándoles nuestros respectivos y sinceros agradecimientos.

En primera instancia agradecemos a Dios, por habernos brindado salud, fortaleza y la sabiduría necesaria, para llevar a cabo este deseo y reto tan anhelado, que hoy en día se vuelve una realidad.

A nuestros padres: Gregorio Javier Reyes, Marlene Melo Gutiérrez (Santos Melo, Yadira Victoria); y Carlos Gamarra Rosales, Bertha Condo Antonio (Gamarra Condo, Cyntia) quienes nunca dudaron que alcanzaríamos este logro tan importante, asumiendo nuevos retos y decisiones. Gracias por su apoyo incondicional y motivación para continuar con nuestros estudios y dándonos palabras de aliento en todo momento de nuestra formación profesional.

A la Universidad Autónoma del Perú y de manera muy especial al cuerpo de docentes de la Facultad de Ingeniería, cuyos conocimientos impartidos nos permitieron contar con las herramientas necesarias para desarrollar este trabajo.

Un agradecimiento especial a nuestros asesores de tesis, el Ing. Percy de la Cruz Velez de Villa e Ing. José Luis Herrera Salazar, quien nos han brindado su tiempo, dedicación y orientación continua en el desarrollo de presente proyecto y a su vez por habernos facilitado siempre los medios necesarios para llevar a cabo todas las actividades propuestas durante la realización de la presente tesis.

RESUMEN

El cáncer de mama es una de las enfermedades más comunes que causan muertes en las mujeres en un porcentaje elevado. Por ello existe una amplia certeza de que la detección temprana es un factor muy importante en la reducción de la mortalidad de esta misma.

Sin embargo el realizar una mamografía a tiempo, es menos dañino y facilita una información útil sobre la presencia de esta enfermedad. Para muchos especialistas es complejo interpretar con exactitud este tipo de estudios (mamografías), es por ello que es necesario múltiples lecturas de un único examen. Por consiguiente en esta tesis proponemos aumentar la fiabilidad de un diagnóstico con la aplicación del algoritmo para predecir el riesgo de cáncer de mama en las pacientes del Hospital de la Solidaridad de Villa El Salvador para extraer información esencial de los signos o síntomas de las pacientes y convertirlas en patrones.

Con el tiempo estos patrones se clasificarán en sub-grupos para la sinergia de familias mediante la homogeneidad e incrementar los índices de coincidencia. El estudio propuesto disminuirá la intervención humana (el error médico) y optimizará la precisión de las respuestas computacionales y sea fácil de obtener los datos.

Palabras clave: Modelo Predictivo, Minería de Datos, Metodología CRISP – DM, Técnica del árbol.

ABSTRACT

Breast cancer is one of the most common diseases that cause deaths in women in a high percentage. For this reason, there is a wide certainty that early detection is a very important factor in reducing mortality.

However, performing a mammogram on time is less harmful and provides useful information about the presence of this disease. For many specialists it is complex to accurately interpret these types of studies (mammograms), which is why multiple readings from a single exam are necessary. Therefore in this thesis we propose to increase the reliability of a diagnosis with the application of the algorithm to predict the risk of breast cancer in the patients of the Hospital de la Solidaridad de Villa El Salvador to extract essential information from the signs or symptoms of the patients and turn them into patterns.

Over time these patterns will be classified into subgroups for family synergy by homogeneity and increasing coincidence rates. The proposed study will decrease human intervention (medical error) and optimize the precision of computational responses and make data easy to obtain.

Keywords: Predictive Model, Data Mining, Methodology CRISP - DM, Tree Technique.

ÍNDICE DE CONTENIDO

DEDICATORIA

AGRADECIMIENTOS

RESUMEN

ABSTRACT

INTRODUCCIÓN

CAPÍTULO I: PLANTEAMIENTO METODOLÓGICO

1.1. Planteamiento del problema.....	2
1.1.1. Descripción de la realidad problemática.....	2
1.1.2. Definición del problema.....	5
1.1.3. Enunciado del problema.....	11
1.2. Tipo y nivel de investigación.....	11
1.2.1. Tipo de investigación.....	11
1.2.2. Nivel de investigación.....	11
1.3. Justificación de la investigación.....	12
1.3.1. Justificación teórica.....	12
1.3.2. Justificación tecnológica.....	12
1.3.3. Justificación metodológica.....	13
1.3.4. Justificación práctica.....	14
1.4. Objetivos de la investigación.....	14
1.4.1. Objetivo general.....	14
1.4.2. Objetivos específicos.....	14
1.5. Hipótesis.....	15
1.6. Variable e indicadores.....	15
1.6.1. Variable independiente.....	15
1.6.2. Variable dependiente.....	16
1.7. Limitaciones de la investigación.....	17
1.8. Diseño de la investigación.....	18
1.9. Técnicas e instrumentos para recolección de información.....	19
1.9.1. Técnicas.....	19

1.9.2. Instrumentos	20
---------------------------	----

CAPÍTULO II: MARCO REFERENCIAL

2.1. Antecedentes de la investigación.....	22
2.2. Marco teórico.....	34
2.2.1. Algoritmo J48 aplicando minería de datos	34
2.2.2. El proceso de la minería de datos	34
2.2.3. Estadísticas vs minería de datos	35
2.2.4. OLAP vs minería de datos.....	36
2.2.5. Técnicas usadas en minería de datos	36
2.2.6. Árbol de decisión.....	37
2.2.7. Metodología para minería de datos	39
2.2.8. Fases de la metodología.....	40
2.2.9. Tareas y salidas de las fases de la metodología Crisp-Dm.....	48
2.2.10. Estudio comparativo de metodologías para minería de datos	54
2.2.11. Estudio comparativo de metodologías para minería de datos	55
2.2.12. Software weka	56
2.2.13. Cáncer de mama	58
2.2.14. Factores de riesgo	59
2.2.15. Estadíos del cáncer de mama.....	60
2.2.16. Supervivencia frente al cáncer de mama	66
2.2.17. Bases teóricas – científicas	67

CAPÍTULO III: DESARROLLO DE LA SOLUCIÓN

3.1. Estudio de factibilidad.....	69
3.1.1. Factibilidad técnica.....	69
3.1.2. Factibilidad operativa	69
3.1.3. Factibilidad económica.....	71
3.2. Fase 1 - comprensión del negocio	73
3.2.1. Determinar los objetivos del negocio	73
3.2.2. Valoración de la situación.....	91
3.2.3. Determinar los objetivos de minería de datos.....	95
3.2.4. Realizar el plan de trabajo	96

3.3.	Fase 2- Comprensión de los datos	97
3.3.1.	Recolectar los datos iniciales	97
3.3.2.	Descripción de los datos	98
3.3.3.	Exploración de los datos	100
3.3.4.	Verificar la calidad de los datos.....	101
3.4.	Fase 3 - Preparación de los datos	104
3.4.1.	Seleccionar los datos.....	104
3.4.2.	Limpiar los datos	105
3.4.3.	Estructurar los datos.....	105
3.4.4.	Integrar los datos.....	106
3.4.5.	Formateo de los datos	107
3.5.	Fase 4 - Modelado	112
3.5.1.	Seleccionar la técnica del modelado.....	112
3.5.2.	Generar el plan de prueba	113
3.5.3.	Construir el modelo	122
3.6.	Fase 5 - Evaluación	123
3.6.1.	Evaluar los resultados	123
3.7.	Fase 6 - Implantación	124
3.7.1.	Plan de implantación.....	124
3.7.2.	Plan de monitoreo y mantenimiento	124
3.7.3.	Informe final	125
3.7.4.	Revisión del proyecto	127

CAPÍTULO IV: ANÁLISIS DE RESULTADOS Y CONTRASTACIÓN

DE LA HIPÓTESIS

4.1.	Población y Muestra	129
4.1.1.	Población	129
4.1.2.	Muestra	129
4.2.	Nivel de confianza y grado de significancia.....	129
4.3.	Resultados genéricos	129
4.4.	Resultados específicos.....	131
4.5.	Análisis e interpretación de resultados	133

4.6. Contrastación de hipótesis	146
---------------------------------------	-----

CAPÍTULO V: CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones.....	156
------------------------	-----

5.2. Recomendaciones.....	157
---------------------------	-----

REFERENCIAS BIBLIOGRÁFICAS

ANEXOS

GLOSARIO DE TÉRMINOS

ÍNDICE DE TABLAS

Tabla 1	Cuadro de indicadores y datos de prueba	8
Tabla 2	Cuadro comparativo entre la situación actual y la situación propuesta.....	9
Tabla 3	Cuadro de indicadores de la variable independiente. presencia-ausencia.....	15
Tabla 4	Cuadro de indicadores de la variable independiente. presencia – ausencia	15
Tabla 5	Cuadro de definición de los indicadores de la variable dependiente.....	16
Tabla 6	Indicador variable dependiente.....	17
Tabla 7	Diseño de la investigación.....	18
Tabla 8	Cuadro de técnicas para la recolección de información	19
Tabla 9	Cuadro de instrumento para la recolección de información	20
Tabla 10	Estadísticas vs minería de datos	35
Tabla 11	Olap vs minería de datos	36
Tabla 12	Cuadro comparativo de diversas metodologías de minería de datos.....	54
Tabla 13	Análisis comparativo de cada enfoque	55
Tabla 14	Presupuesto mensual	71
Tabla 15	Presupuesto de costo único – total del proyecto.....	72
Tabla 16	Hospital de Solidaridad de Villa el Salvador	74
Tabla 17	Lista To Do.....	75
Tabla 18	Roles de equipo	76
Tabla 19	Lista de principales proveedores de suministro.....	83
Tabla 20	Lista de principales proveedores de servicio.....	83
Tabla 21	Procesos del negocio	86
Tabla 22	Matriz de selección de proceso.....	87
Tabla 23	Inventario de recursos.....	96
Tabla 24	Plan de trabajo	96
Tabla 25	Reporte de datos	98

Tabla 26 Descripción de los atributos	99
Tabla 27 Resultados de pre prueba y post prueba para los KPI1, KPI2, KPI3 y KPI4.....	131
Tabla 28 Resultados de pre-prueba y post-prueba para el KPI1	133
Tabla 29 Resultados de pre-prueba y post-prueba para el KPI2	137
Tabla 30 Resultados de pre-prueba y post-prueba para el KPI3	140
Tabla 31 Valores de pre-prueba.....	143
Tabla 32 Valores de estado y frecuencia de pre-prueba.....	143
Tabla 33 Valores de estado y frecuencia de pre-prueba.....	144
Tabla 34 Valores de post-prueba.....	145
Tabla 35 Valores de estado y frecuencia post-prueba	145
Tabla 36 Valores de estado y frecuencia post-prueba	145
Tabla 37 Recurso de indicadores.....	147
Tabla 38 Valores de pre-prueba. KPI1	148
Tabla 39 Valores de post-prueba. KPI1.....	148
Tabla 40 Estadísticas descriptivas	149
Tabla 41 Valores de pre-prueba. KPI2	150
Tabla 42 Valores de post-prueba. KPI2.....	150
Tabla 43 Estadísticas descriptivas	152
Tabla 44 Valores de pre-prueba. KPI3	153
Tabla 45 Valores de post-prueba.KPI3.....	153
Tabla 46 Estadísticas descriptivas	154

ÍNDICE DE FIGURAS

Figura 1	Mapa de ubicación del Hospital de la Solidaridad de Villa el Salvador.....	5
Figura 2	Flujograma del proceso de diagnóstico de cáncer de mama (AS-IS).	7
Figura 3	Flujograma del proceso de diagnóstico de cáncer de mama (TO-BE).....	10
Figura 4	Proceso de minería de datos.	35
Figura 5	Estadística de metodología de minería de datos.	39
Figura 6	Representación de la metodología crisp-dm.	40
Figura 7	Fase 1 - Comprensión del negocio.	41
Figura 8	Esquema de la primera fase.....	41
Figura 9	Fase 2 - Compresión de los datos.....	42
Figura 10	Esquema de la segunda fase.....	42
Figura 11	Secuencia de la segunda fase - comprensión de los datos.	43
Figura 12	Fase 3 - Preparación de los datos	43
Figura 13	Esquema de la tercera fase.	44
Figura 14	Secuencia de la tercera fase - preparación de los datos.	44
Figura 15	Fase 4 - modelado	45
Figura 16	Esquema de la cuarta fase.	45
Figura 17	Fase 5 – evaluación	46
Figura 18	Esquema de la quinta fase.....	46
Figura 19	Fase 6 – implantación.....	47
Figura 20	Esquema de la sexta fase.....	47
Figura 21	Ilustración del cáncer de mama.....	59
Figura 22	Estadio 0 del cáncer de mama.....	61
Figura 23	Estadio 1 del cáncer de mama.....	62
Figura 24	Estadio IIA del cáncer de mama.	62
Figura 25	Estadio IIB del cáncer de mama.....	63

Figura 26	Estadio IIIA del cáncer de mama.....	64
Figura 27	Estadio IIIB del cáncer de mama.....	64
Figura 28	Estadio IIIC del cáncer de mama.....	65
Figura 29	Estadio IV del cáncer de mama.....	65
Figura 30	Gráfico de la tasa de supervivencia del cáncer de mama.....	67
Figura 31	Organigrama del Hospital de la Solidaridad de Villa el Salvador.....	76
Figura 32	Cadena de valor del área de oncología del Hospital de la Solidaridad.....	85
Figura 33	Modelo de contexto.....	88
Figura 34	Diagrama de descomposición de funciones.....	88
Figura 35	Flujograma de actividades del Hospital de la Solidaridad de Villa el Salvador.....	89
Figura 36	Exploración de los datos parte I.....	100
Figura 37	Exploración de los datos parte II.....	100
Figura 38	Muestra de selección de síntomas del paciente.....	101
Figura 39	Datos generales de las pacientes.....	103
Figura 40	Características del paciente.....	104
Figura 41	Clínica y diagnóstico del paciente.....	104
Figura 42	Base de datos.....	105
Figura 43	Tabla síntoma.....	106
Figura 44	Tabla triaje.....	107
Figura 45	Reporte de datos del paciente: edad, talla, peso, frecuencia cardíaca, edad menarquía, cantidad embarazos, cantidad primer embarazo, cantidad abortos, cantidad hijos.....	108
Figura 46	Reporte de datos del paciente: ausencia de menstruación y dolor de la mama.....	109
Figura 47	Características generales del paciente: retracción de los pezones, secreción, endurecimiento de la mama, cambio en el tamaño de la mama e irritación.....	110
Figura 48	Características generales del paciente: secreción de sangre, mamas con heridas.....	111
Figura 49	Base de datos de nuevos pacientes.....	113

Figura 50	Conversión de archivos.....	114
Figura 51	Interfaz del software weka	114
Figura 52	Síntomas de los pacientes.....	115
Figura 53	Características generales de los pacientes.....	116
Figura 54	Número de grado del paciente.....	116
Figura 55	Clasificador de atributos de las pacientes.	118
Figura 56	Formación del árbol con atributos de las pacientes	118
Figura 57	Formación del árbol con atributos de las pacientes del modelo 1.....	119
Figura 58	Formación del árbol con atributos de las pacientes del modelo 2.....	119
Figura 59	Formación del árbol con atributos de las pacientes del modelo 3.....	120
Figura 60	Formación del árbol con atributos de las pacientes del modelo 4.....	120
Figura 61	Formación del árbol con atributos de las pacientes del modelo 5.....	121
Figura 62	Formación del árbol con atributos de las pacientes del modelo 6.....	121
Figura 63	Tiempo de demora del KPI1	122
Figura 64	Tiempo de demora del KPI2	122
Figura 65	Tiempo de demora de KPI3.	123
Figura 66	Informe del resumen KPI1 pre prueba.....	134
Figura 67	Informe del resumen KPI1 post prueba.	135
Figura 68	Comparación de pruebas.....	135
Figura 69	Informe del resumen KPI 2 pre prueba.....	138
Figura 70	Informe del resumen KPI2 post prueba.	138
Figura 71	Comparación de pruebas.....	139
Figura 72	Informe del resumen KPI3 pre prueba.....	141
Figura 73	Informe del resumen KPI3 post prueba	141
Figura 74	Comparación de las pruebas.....	142
Figura 75	Porcentaje de pre prueba del KPI4.....	144

Figura 76 Porcentaje de post prueba del KPI4.	146
Figura 77 Gráfica de distribución KPI1.	149
Figura 78 Gráfica de distribución KPI2.	151
Figura 79 Gráfica de distribución KPI3.	154

INTRODUCCIÓN

El actual trabajo de tesis es para recibir el Grado de Título de Ingeniero de Sistemas por la Universidad Autónoma del Perú. Es un estudio que trata concretizar las ideas, las teorías, los esfuerzos, opiniones y anhelos de aprendizaje, además de las valiosas enseñanzas de todos los maestros que compartieron sus conocimientos.

Esta investigación tiene como principal propósito de estudio la implementación de un modelo predictivo para predecir el riesgo del cáncer de mama aplicado en minería de datos las pacientes del Hospital de la Solidaridad de Villa El Salvador.

El cáncer de mama constituye un importante problema de salud en los países occidentales por su elevada incidencia y mortalidad. En los últimos veinte años, la probabilidad de desarrollar cáncer de mama en una mujer del mundo occidental ha pasado de 1 sobre 20 a 1 sobre 11. En cuanto a la incidencia, es el primer tumor en la mujer con tasas de incidencia que oscilan entre 30 - 60 por cada cien mil mujeres.

Con relación a la mortalidad, el cáncer de mama es la primera causa de muerte tumoral en la mujer, siendo también la primera causa absoluta de muerte en mujeres de 35 - 54 años. La supervivencia de este cáncer está en razón directa con el estudio diagnóstico y con el tratamiento efectuado. Así, el tratamiento de los cánceres no invasivos proporciona una mayor supervivencia y es menos mutilante. Si el cáncer se diagnostica en un estudio localizado, sin afectación ganglionar, la supervivencia a los cinco años alcanza un 87%; sin embargo, si existe afectación ganglionar, solo alcanza el 47%

La predicción es un intento de diagnosticar con precisión la evolución de un sistema específico usando para ello la información obtenida a partir de un conjunto concreto de variables que describen dicho sistema.

Con la finalidad de que la presente tesis sea más comprensible, es por ello que se ha dividido en cinco capítulos, siendo los siguientes:

Capítulo I, Planteamiento metodológico. - En este capítulo se mostrará todo sobre el planteamiento metodológico, ya que se define la realidad problemática, nivel de investigación, justificación, objetivos, hipótesis, variables e indicadores, diseño de la investigación y métodos de recolección de datos.

Capítulo II, Marco referencial. - En este capítulo se explica los antecedentes teniendo como referencia libros, artículos científicos, tesis y el contenido teórico de la tesis, el cual nos permite relacionar con los modelos y la metodología que usaremos para desarrollar la tesis.

Capítulo III, Desarrollo de la solución. - En este capítulo se realiza la implementación del producto a mostrar (Minería de datos), usando la metodología CRISP – DM con sus respectivas etapas ya mencionadas en el marco teórico.

Capítulo IV, Análisis e interpretación de resultados. - Se realiza la prueba empírica para la recopilación, análisis para la recopilación, análisis e interpretación de los resultados. En primer lugar, se describe la población y muestra, posteriormente el tipo de muestra y el nivel de confianza. De igual manera se muestra el análisis de pre prueba y post prueba. Posteriormente, los datos obtenidos se mostrarán en sus respectivas tablas, el cual serán analizadas para finalmente realizar la contrastación de hipótesis.

Capítulo IV, Conclusiones y recomendaciones. - En este capítulo se mostrarán las conclusiones y recomendaciones del proyecto que hemos realizados. Y para finalizar se presentarán las referencias bibliográficas, anexos, apéndices y glosario de términos.

CAPÍTULO I
PLANTEAMIENTO METODOLÓGICO

1.1. Planteamiento del problema

1.1.1. Descripción de la realidad problemática

Realidad mundial

La Organización Mundial de la Salud (2013) señala: “Desde hace varias décadas, las tasas de cáncer de mama han aumentado más rápidamente en los países ricos que en los pobres. Los científicos empiezan a conocer más acerca de sus causas, pero aún quedan muchos interrogantes por despejar” (p. 25).

El cáncer de mama es una de las enfermedades que no hacen distinción entre la población de países desarrollados y en desarrollo. Por otro lado, este es el tipo de cáncer que se encuentra con mayor presencia en las mujeres a nivel mundial.

En cuanto a la mortalidad por esta enfermedad, sí hay diferencias: En países de bajos ingresos ocurren la mayoría de los decesos, ya que generalmente el diagnóstico se realiza en fases avanzadas de la enfermedad, debido a la falta de acceso a servicios de salud y a la poca sensibilización para la detección precoz (conocimiento de signos, de síntomas iniciales y la autoexploración mamaria).

Según datos de la Organización Panamericana de la Salud – OPS (2013) indica:

En el continente americano, la tendencia es similar, es decir, el cáncer de mamá es el más común entre las mujeres (29% respecto del total de casos de cáncer) y es la segunda causa de muerte por tumores malignos para este grupo de población, superada únicamente por el cáncer de pulmón. (p. 57).

La incidencia de cáncer de mama está aumentando en el mundo en desarrollo debido a la mayor esperanza de vida, el aumento de la urbanización y la adopción de modos de vida occidentales.

La American Society of Clinical Oncology (2017) afirma:

Se diagnostica a más mujeres con cáncer de mama que ningún otro cáncer. Se estima que este año se diagnosticará cáncer de mama invasivo a 252,710 mujeres en los Estados Unidos y se diagnosticará cáncer de mama in situ a 63,410 mujeres. En los Estados Unidos, se les diagnosticará cáncer de mama a aproximadamente 2,470 hombres. Se calcula que este año se producirán unas 41.070 muertes (40,610 mujeres y 460 hombres) a causa del cáncer de mama. (p. 45).

La Society American Cancer (2018) señala:

Mientras que para el año 2018, los cálculos de la Sociedad Americana Contra El Cáncer indican que para este cáncer en los Estados Unidos se diagnosticarán alrededor de 266,120 nuevos casos de cáncer de seno invasivo en mujeres de este país. Se diagnosticarán alrededor de 63,960 nuevos casos de carcinoma in situ (CIS), (el CIS no es invasivo y es la forma más temprana de cáncer de seno) y alrededor de 40,920 mujeres morirán de cáncer de seno. (p. 87).

Aunque reducen en cierta medida el riesgo, las estrategias de prevención no pueden eliminar la mayoría de los casos de cáncer de mama que se dan en los países de ingresos bajos y medios, donde el diagnóstico del problema se hace en fases muy avanzadas, además la detección precoz con vistas a mejorar el pronóstico y la supervivencia de esos casos sigue siendo la piedra angular del control del cáncer de mama. (Organización mundial de la salud, 2018).

Para el año 2030, la OPS estima más de 596 000 casos nuevos y más de 142 100 muertes en la región, principalmente en la zona de América Latina y el Caribe.

Realidad nacional:

Según Velarde (2017) menciona:

El Perú está a la par de los países occidentales más avanzados, que son justamente los que tienen más incidencia y más mortalidad por cáncer de mama y se ha convertido en la segunda causa de muerte a partir del año 2000, con 17% de las muertes (el doble de lo que se tenía en la década de los 80), como consecuencia de un incremento del número de casos y, además por la alta frecuencia de detección de la enfermedad en

estados muy avanzados, por lo cual es más costoso su manejo y menor el éxito esperado. (p. 55).

El Comercio (2016) afirma:

Asimismo, la frecuencia del cáncer va en aumento anualmente, ocasiona daños profundos en las mujeres al causar muertes y discapacidades, pese a que muchos de esos casos pueden ser prevenidos o detectados y tratados tempranamente.

“Hace años, cuando no conocíamos los tipos de cáncer de mama, el tratamiento era igual para todos y la efectividad se reducía, lo que generaba que las pacientes se debilitaran”, explicó a El Comercio Fiorella Pando, oncóloga investigadora del Instituto Oncológico de Lima. (p. 96).

Según Wong (2014) sostiene: “En el Perú no hay una política estratégica para afrontar esta enfermedad. El problema es nacional, no estamos bien en lo que es tema de educación todavía. Entonces no dan bien la información de salud” (p. 72).

Según el cirujano oncológico Marco Velarde, en el Perú la incidencia ha aumentado y afecta a 42 de cada 100,000 habitantes, pero ya no sólo se presenta entre las mujeres a partir de los 40 años, sino que se detecta a edades más tempranas, desde los 30 años o incluso desde los 25 años.

Según la Liga Contra el Cáncer (2017) señala: “En el país, cada año se presentan unos 5,000 casos nuevos y la mayoría llega en estadios avanzados, cuando las probabilidades de curación son de solo el 50%. Se estima que cada año 2000 mujeres mueren por esta enfermedad” (p. 24).

Realidad Empresarial:

El diagnóstico clínico que brindan los médicos del hospital Solidaridad de Villa El Salvador, es un estudio tecnológico y complejo que es exacto en reconocimiento de patrones, cálculo de probabilidad condicional de capacitación y experiencia.

Diferentes especialistas de la salud relacionados con estudios a la mamografía con el tiempo han alcanzado una intuición que sea más sencilla la tarea; es por ello que es el

resultado de la vivencia de haber tenido casos similares de extensa carrera profesional.

Se tiene conocimiento que existen diferentes casos con diagnósticos errados por factores humanos hasta calcificaciones de varias morfologías, tamaño y distribución, que pueden ser representativos de maligno; por lo que es necesaria múltiple lectura de un mismo examen.

Ubicación

La investigación, se realizará en el Hospital de la Solidaridad ubicada en la Av. Pastor Sevilla S/N Ovalo Cocharcas – Villa El Salvador

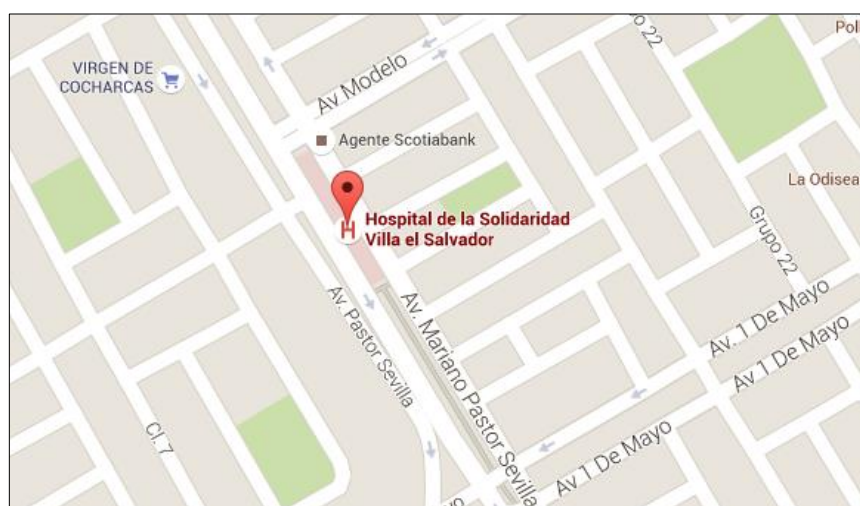


Figura 1. Mapa de ubicación del Hospital de la Solidaridad de Villa el Salvador.

Fuente: Google Maps (2016).

1.1.2. Definición del problema

Los médicos menos experimentados del Hospital Solidaridad de Villa El Salvador pueden disminuir los errores al momento de realizar el diagnóstico a una paciente, realizando un estudio tecnológico y sistemático; teniendo en cuenta la integridad, la confidencialidad y disponibilidad de sus diagnósticos y las fuentes más recurrente de error.

Si bien es cierto el médico puede no estar familiarizado con la enfermedad de la paciente o estar influenciado por la opinión diagnóstica de su colega que lo evaluó anteriormente; y como consecuencia de ello no considera un diagnóstico apropiado. Otro error común es la relación del conjunto de la sistematización de la paciente que presenta en un estado inicial; siendo uso de ello para reafirmar lo que se pensó inicialmente y descartando otros síntomas que no están relacionados.

Sub-problema:

Sin duda unos de los grandes problemas que no son ajenos a ello de la insatisfacción de los pacientes que son los tiempos de espera y respuesta antes los resultados de análisis ante cualquier tipo de enfermedad ya que por lo general están en un promedio de días, semanas hasta incluso meses para recibir dichos resultados para que finalmente el doctor pueda realizar las revisiones a las pacientes.

Esta investigación se centra en un antes y un después del progreso del Hospital de la Solidaridad de Villa El Salvador, y que se note como el proceso de automatización de mejora sobre la prestación de servicios, sobre todo en la disminución del tiempo de respuesta del diagnóstico de cáncer de mama, además de la mayor precisión de los resultados.

Es importante saber que el abuso de las pruebas de laboratorio para infinidad de pruebas en el hospital causa gastos mayores de los necesarios por utilización inapropiada de los recursos, uso indiscriminado de los equipos, carga adicional de trabajo que atenta contra la calidad y el tiempo de respuesta de los resultados, sin imaginar que el gran aumento de la demanda en los servicios ocasiona tiempos de espera prolongados que afectan negativamente la atención de los pacientes más que probablemente tengan el cáncer de mama; avanzado y retrasando un posible tratamiento que requiera con urgencia.

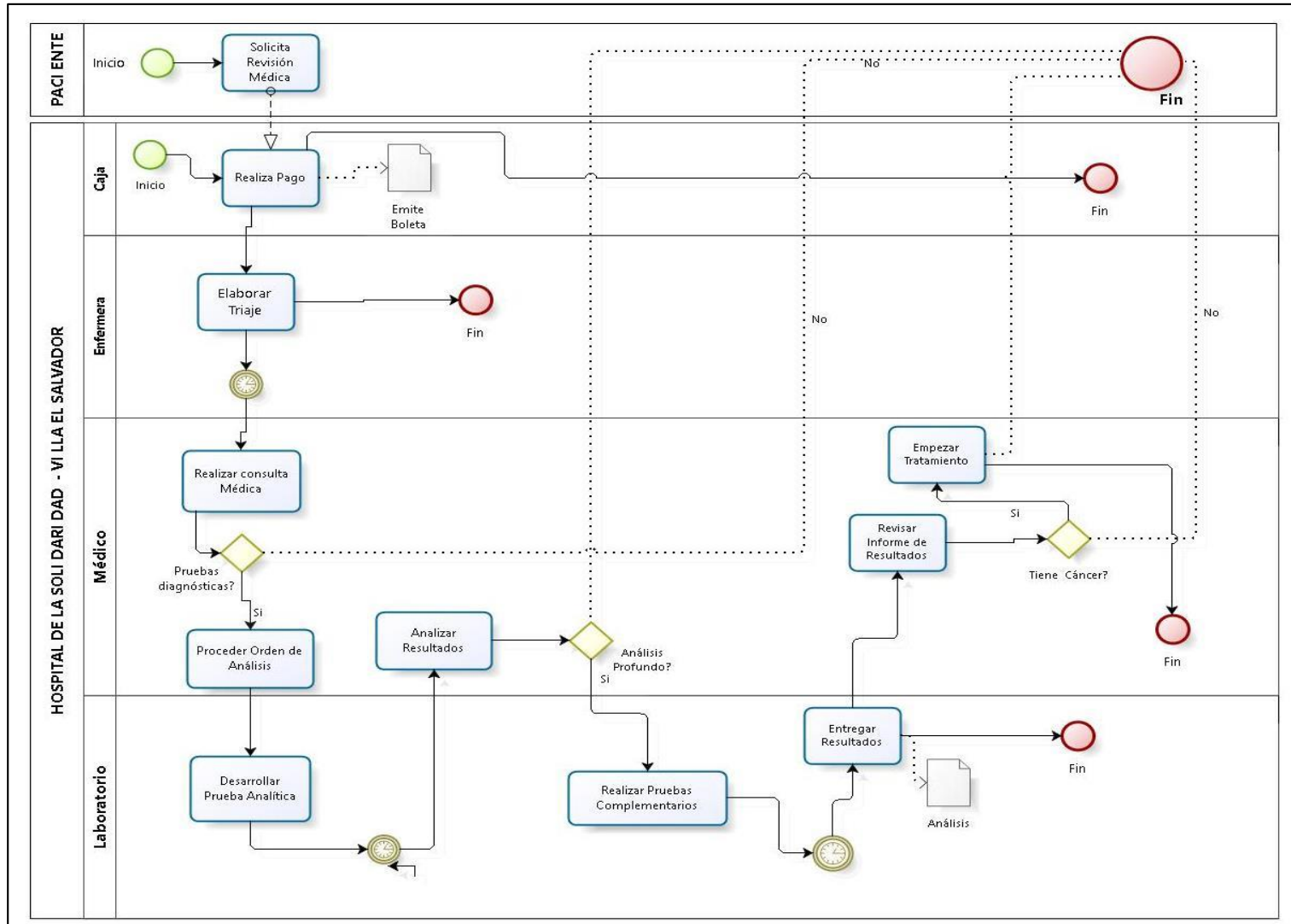


Figura 2. Flujograma del proceso de diagnóstico de cáncer de mama (AS-IS).

Los problemas que presenta el flujograma del Hospital de la Solidaridad de Villa El Salvador son:

- Tiempo para procesar el registro de los datos del paciente.
- Tiempo para procesar la data. (guardar)
- Tiempo para predecir el cáncer de mama.
- Nivel de satisfacción del paciente.

Tabla 1

Cuadro de indicadores y datos de prueba

Indicadores	Datos de prueba (Promedio)
Tiempo para verificar los datos del paciente.	7.6 minutos
Tiempo para procesar la data.	8.5 minutos
Tiempo para predecir el cáncer de mama.	4080 minutos
Nivel de satisfacción del paciente.	Poco satisfecho

Para solucionar los problemas que existen en el Hospital Solidaridad de Villa El Salvador, la solución más factible es desarrollar un modelo predictivo para la identificación de las pacientes propensas a tener cáncer de mama para reducir errores al momento de realizar el diagnóstico a una paciente, realizando un estudio tecnológico y sistemático; teniendo en cuenta la integridad, la confidencialidad y disponibilidad de sus diagnósticos y las fuentes más recurrente de error.

Tabla 2

Cuadro comparativo entre la situación actual y la situación propuesta

AS-IS	TO-BE
Tiempo inadecuado para verificar los datos del paciente.	Tiempo adecuado para verificar los datos del paciente.
Tiempo inadecuado para procesar la data.	Tiempo adecuado para procesar el registro de los datos del paciente.
Tiempo inadecuado para predecir el cáncer de mama.	Tiempo adecuado para predecir el cáncer de mama.
Nivel regular de satisfacción del paciente.	Alto nivel de satisfacción del paciente.

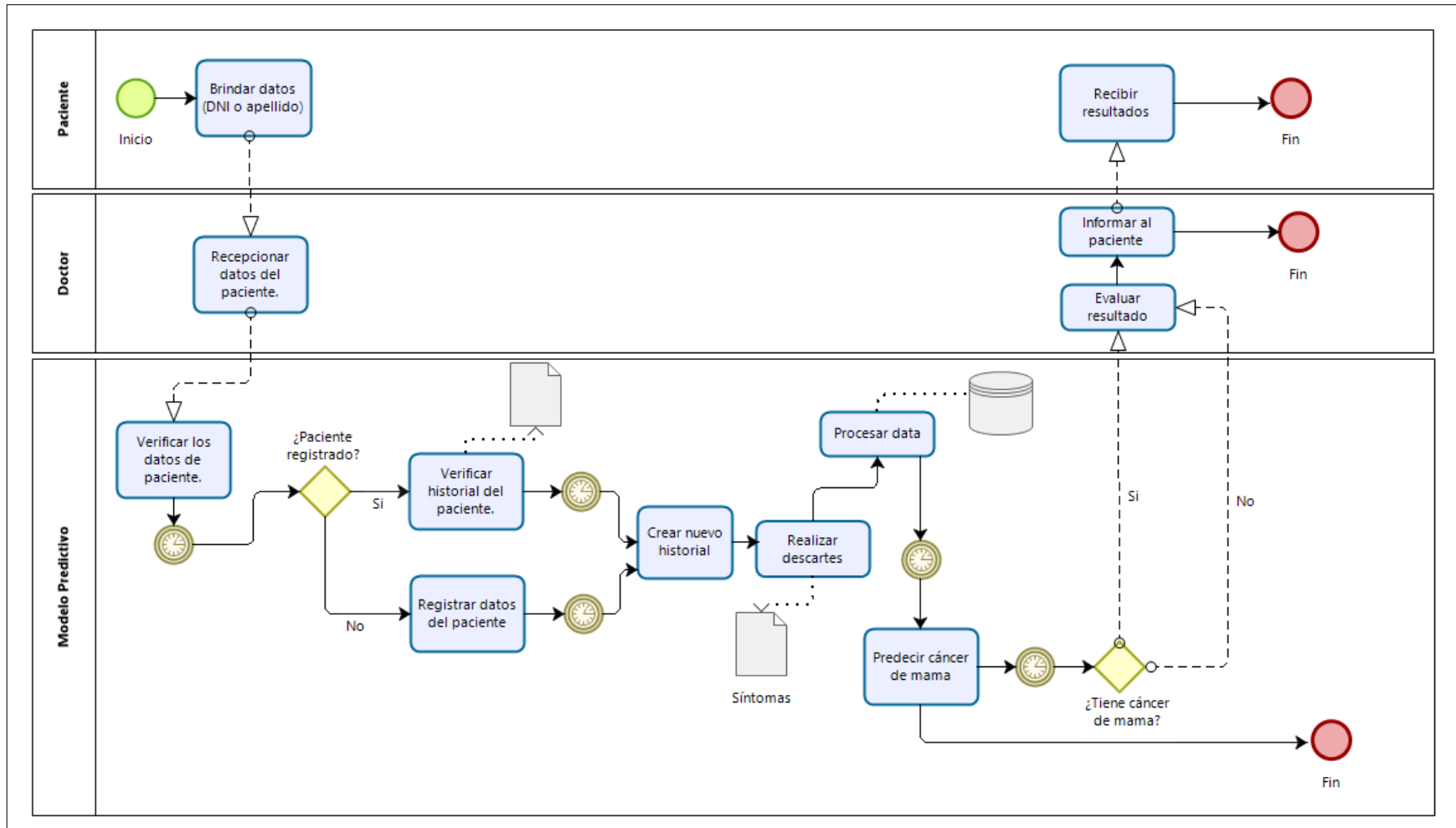


Figura 3. Flujograma del proceso de diagnóstico de cáncer de mama (TO-BE).

1.1.3. Enunciado del problema

¿En qué medida la aplicación del algoritmo mejorará el proceso de predecir el riesgo del cáncer de mama de las pacientes del Hospital de la Solidaridad de Villa El Salvador?

1.2. Tipo y nivel de investigación

1.2.1. Tipo de investigación

El tipo de investigación de la presente tesis es aplicada, debido a que aplicará una solución a la realidad problemática que se está presentando, en este caso se desarrollará un modelo predictivo para diagnosticar el cáncer de mama de las pacientes del Hospital de la Solidaridad de Villa El Salvador.

Vargas (2009) sostiene: “La investigación aplicada la relaciona a la utilización de los conocimientos en la práctica, para aplicarlos en provecho de los grupos que participan en esos procesos y en la sociedad en general” (p. 159).

1.2.2. Nivel de investigación

De acuerdo a la naturaleza del estudio de la investigación, reúne por su nivel las características de un estudio:

- Explicativa

Dado que nuestra investigación aplicará las causas del enfoque de relaciones causa-efecto, la investigación desarrollará las causas del tiempo proceso de demora para el diagnóstico de cáncer de mama, así como también los efectos que aportará el modelo predictivo para mejorar los tiempos de respuesta en el diagnóstico de cáncer de mama.

Rodríguez (2012) afirma:

La investigación explicativa está dirigida a responder a las causas de los eventos físicos o sociales y su interés se centra en explicar por qué y en qué condiciones ocurre un fenómeno o por qué dos o más variables se relacionan. Hay predominio de explicación, descripción y correlación. Es aquella que tiene relación causal, no solo persigue describir o acercarse a un problema, sino que intenta encontrar las causas del mismo. Son aquellas que parten de una situación problema o conocimiento presente para luego indagar posibles

causas o factores asociados que permiten interpretarla. (p. 39).

1.3. Justificación de la investigación

Para poder implementar la minería de datos en el Hospital de la Solidaridad de Villa El Salvador o en cualquier otro hospital, es necesario analizar las fases trazados de cada hospital y sobre todo analizar los procesos del diagnóstico de síntomas de cada paciente obtenida por el hospital.

1.3.1. Justificación teórica

El uso de un modelo predictivo es aplicable para la identificación de las pacientes propensas a tener cáncer de mama, logrando simplificar los tiempos de desarrollo, obteniendo resultados de mejor calidad y facilitando las tareas de actualización y mantenimiento de la misma. Así mismo mejorar los procesos de salud de las pacientes del Hospital Solidaridad de Villa El Salvador.

López (2013) afirma:

La importancia de desarrollar el proyecto basado en la metodología CRISP DM es conseguir un correcto orden y alineamiento con los objetivos propuestos presentando una herramienta importante para diagnosticar la enfermedad en su etapa inicial. Este modelo aparte de reducir el tiempo de espera, ayudará a predecir con mayor precisión el cáncer de mama. (p. 131).

Becerra (2014) sostiene:

En su revista “Prevención de Cáncer de mama”; nos detallan lo siguiente: Teniendo en cuenta el carácter multifactorial e incierto de las etapas de la enfermedad, hacemos énfasis en la prevención de la enfermedad; es decir; la búsqueda de un diagnóstico temprano de la enfermedad, el cual, junto con el tratamiento oportuno, puede ser definitivo para la supervivencia para las mujeres. (p. 56).

1.3.2. Justificación tecnológica

Las nuevas tecnologías están cambiando todos los ámbitos de nuestras vidas, afectando la forma en que se hacen las cosas: trabajar, aprender, desarrollar y cambiar sutilmente nuestra

forma de pensar y de vida, puesto que han desempeñado un papel fundamental en la configuración de la sociedad y la cultura.

Trabajar con esta tecnología implica poder reaccionar de una manera rápida y efectiva a los cambios de demanda del mercado, además hay que comprender todos los factores que influyen el negocio, observar los factores claves y proveer a los que toman decisiones de información relevante y actual.

Castro (2016) afirman:

Los avances en las tecnologías y en los sistemas computacionales han facilitado el rápido desplazamiento de recursos, bienes y servicios, lo que ha generado interrelaciones estrechas entre la economía mundial y mejoras a nivel de nuestros hospitales, efectuándose una mejor atención y un eficaz resultado. (p. 16).

1.3.3. Justificación metodológica

Hoy en día el análisis predictivo es una herramienta imprescindible que permite sustituir las corazonadas y las apreciaciones personales por proyecciones científicas capaces de eliminar o disminuir las incertidumbres a la hora de elaborar estrategias o toma de decisiones.

El análisis predictivo es una sub disciplina del análisis de datos que usa técnicas de estadística, como aprendizaje computacional o minería de datos, para desarrollar modelos que predicen eventos futuros o conductas. Estos modelos predictivos permiten aprovechar los patrones de comportamiento encontrados en los datos actuales e históricos para identificar riesgos y oportunidades.

Calleja (2010) afirma:

Hay empresas de importancia que han desarrollado potentes herramientas de Análisis Predictivo, sin embargo, el objetivo de esta tesis se centra en el estudio importancia de realizar el modelo predictivo para la mejora de los tiempos de respuesta en la detección de cáncer de mama que ofrecen las funcionalidades necesarias para llevar a cabo proyectos de Análisis Predictivo. La herramienta elegida

es Weka, aplicación completa que dispone de interfaces gráficas que permiten abstraer la complejidad matemática del análisis, poniendo a disposición de empresas, instituciones públicas, etc. (p. 51).

1.3.4. Justificación práctica

Esta investigación se realiza porque observamos que existe mucha demora en el diagnóstico de cáncer de mama de las pacientes del Hospital de la Solidaridad, por lo que nuestro modelo predictivo alcanzará los objetivos propuestos para lograr la satisfacción de las mismas pacientes. Por otro lado, consideramos que las mujeres de la actualidad tienen la oportunidad de conocer la enfermedad que quiere invadir su cuerpo.

Banegas (2015) sostiene: “La nueva herramienta de análisis, al estar basada en datos de mujeres, ayudará a predecir el cáncer de mama, puesto que es sumamente importante que la herramienta incluya información de estas mujeres para determinar su puntaje de riesgo” (p. 57).

1.4. Objetivos de la investigación

1.4.1. Objetivo general

Determinar en qué medida la aplicación del algoritmo mejorará el proceso de predecir el riesgo del cáncer de mama de las pacientes del Hospital de la Solidaridad de Villa El Salvador

1.4.2. Objetivos específicos

- Determinar en qué medida la aplicación del algoritmo reducirá el tiempo empleado para verificar los datos del paciente.
- Determinar en qué medida la aplicación del algoritmo disminuirá el tiempo para procesar la data.
- Determinar en qué medida la aplicación del algoritmo reducirá el tiempo para predecir el cáncer de mama.
- Determinar en qué medida la aplicación del algoritmo aumentará el nivel de satisfacción

del paciente.

1.5. Hipótesis

Si se aplica el algoritmo basado en minería de datos, entonces se influye positivamente en el diagnóstico de cáncer de mama de las pacientes del Hospital de la Solidaridad de Villa El Salvador.

1.6. Variable e indicadores

1.6.1. Variable independiente

✓ Indicadores

A. Conceptualización

Variable independiente: Algoritmo

Tabla 3

Cuadro de indicadores de la variable independiente. presencia-ausencia

Indicador: Presencia – Ausencia
Descripción: Este indicador va a permitir dar a conocer su existencia o ausencia. Cuando indique no, es porque no se ha desarrollado el modelo predictivo en el Hospital Solidaridad de Villa El Salvador y aún se encuentra en la situación actual del problema. Cuando indique si, se ha desarrollado el modelo predictivo y se espera tener buenos resultados.

B. Operacionalización

Variable Independiente: Algoritmo

Tabla 4

Cuadro de indicadores de la variable independiente presencia – ausencia

Indicador	Índice
Presencia - Ausencia	No, Sí

1.6.2. Variable dependiente

✓ Indicadores

A. Conceptualización

- **Variable Dependiente:** Riesgo de cáncer de mama de las pacientes del Hospital Solidaridad de Villa El Salvador.

Tabla 5

Cuadro de definición de los indicadores de la variable dependiente

Indicadores	Descripción
Tiempo para procesar el registro de los datos del paciente.	Es el tiempo transcurrido para procesar el registro de los datos del paciente.
Tiempo para procesar la data	Es el tiempo empleado para procesar la data obtenida.
Tiempo para predecir el cáncer de mama	Es el tiempo procesado para predecir el cáncer de mama
Nivel de satisfacción del paciente	Es la satisfacción del paciente, para verificar el cumplimiento de sus atenciones y necesidades.

B. Operacionalización

- **Variable Dependiente:** Riesgo de cáncer de mama de las pacientes del Hospital Solidaridad de Villa El Salvador. (TO-BE)

Tabla 6

Indicador variable dependiente

Dimensión	Indicador	Índice	Unidad de medida	Unidad de observación
Tiempo	Tiempo para procesar el registro de los datos del paciente.	[0.10...0.20]	Segundos	Proceso
	Tiempo para procesar la data.	[60...120]	Segundos	Proceso
	Tiempo para predecir el cáncer de mama.	[12...50]	Segundos	Proceso
Calidad	Nivel de satisfacción del paciente.	No satisfecho, Poco satisfecho, Satisfecho, Muy satisfecho.	Encuesta	Paciente

1.7. Limitaciones de la investigación

- Temporal:** Este presente trabajo de investigación se elaborará en un periodo entre el mes de abril hasta diciembre 2017.
- Espacial:** Esta investigación se enfoca en el área de oncología del Hospital Solidaridad de Villa El Salvador.
- Conceptual:** Utilizaremos la metodología Crisp – DM.

1.8. Diseño de la investigación

Tabla 7

Diseño de la investigación

G_e	O₁	X	O₂
Grupo experimental: Grupo de estudio al que se le aplicará el estímulo (Modelo predictivo).	Datos de la Pre – Prueba para los indicadores de la Variable Dependiente: Mediciones pre – prueba del grupo experimental.	Algoritmo: Estímulo o condición experimental.	Datos de la Post – Prueba para los indicadores de la Variable Dependiente: Mediciones post – prueba del grupo de control.

Interpretación

Se trata de la confrontación de forma aleatoria de un grupo (G_e) conformado por el diagnóstico de un paciente del Hospital de la Solidaridad de Villa El Salvador al que se aplica un estímulo o condición experimental (X), luego se les aplica una prueba posterior a los indicadores de la variable dependiente (O₁). A un segundo grupo (G_e) también conformado de forma aleatoria, grupo al que no se le aplica estímulo alguno, sirviendo sólo como de control; en forma simultánea se les aplica una prueba a los indicadores de la variable dependiente (O₂). Se espera que los valores O₁ sean mejores que los valores O₂.

Los dos grupos están constituidos de forma aleatoria. Tanto en ausencia como en presencia modelo predictivo propuesto.

1.9. Técnicas e instrumentos para recolección de información

1.9.1. Técnicas

Tabla 8

Cuadro de técnicas para la recolección de información

Técnicas para la recolección de información	
Investigación de campo	1. Observación directa <ul style="list-style-type: none">✓ Individual✓ No participante
	2. Realización de entrevistas al personal del área <ul style="list-style-type: none">✓ Dirigida✓ Estructurada
	3. Aplicación de encuestas a los respectivos pacientes y/o involucrados <ul style="list-style-type: none">✓ Abierto
Investigación experimental	1. Seguimiento del proceso administración de historiales clínicos de los pacientes.
	2. Seguimiento del tiempo empleado para realizar una consulta.
	3. Seguimiento del número de consultas realizadas en un tiempo determinado.
	4. Seguimiento de la satisfacción de los pacientes.
Investigación documental	✓ Revistas
	✓ Libros
	✓ Tesis
	✓ Papers
	✓ Monografías físicas y virtuales
	✓ Documentos

1.9.2. Instrumentos

Tabla 9

Cuadro de instrumento para la recolección de información

Instrumento para la recolección de información	
Investigación de campo	✓ Fotografía
	✓ Ficha de permiso de observación (Carta de presentación)
	✓ Formato de entrevista
	✓ Diario de campo
Investigación experimental	✓ Diario de campo
	✓ Fotografías
	✓ Fichas de seguimiento
Investigación documental	✓ Diario de campo
	✓ Libreta de apuntes
	✓ Cámara
	✓ USB
	✓ Computadoras
	✓ Fotocopias
	✓ CD – ROM
	✓ Escáner
✓ Grabadora	
✓ Impresiones	

CAPÍTULO II
MARCO REFERENCIAL

2.1. Antecedentes de la investigación

A) **Autor:** Díaz Pizarro, Gema M.

Título: Detección de lesiones en mamografías y ecografías para el diagnóstico del cáncer de mama.

Tipo: Tesis de pregrado

Año: 2014

Correlación:

Díaz (2014) afirma:

En la presente tesis el autor nos indica que el cáncer de mama es la primera causa de muerte por cáncer entre las mujeres en todo el mundo, es una de las pocas enfermedades cancerosas que tarda en presentar síntomas y que puede ser detectada antes de que lo haga. Así mismo los nuevos avances tecnológicos y la implantación de sistemas informatizados en clínicas y hospitales ofrecen una gran cantidad de posibilidades en diagnóstico automático de enfermedades. El autor tiene como objetivo es la implementación y comparación de métodos computacionales de clasificación automática orientados a la detección de lesiones de mama. Realizó un análisis de imágenes diagnósticas de distintas modalidades, mediante distintas técnicas de inteligencia artificial para la detección de las regiones de interés, de acuerdo al fin, el autor persigue la investigación aplicada con un estudio de tipo analítico. (p. 45).

El autor concluye lo siguiente: La herramienta generada permite la inclusión sencilla de nuevos algoritmos de detección o etapas de procesado previas para continuar mejorando los métodos propuestos. Asimismo, se debe destacar la facilidad de reutilización del código de la capa de dominio para su uso aparte en pruebas por lotes.

Aunque esté orientada a la detección de lesiones en ecografías y mamografías, se podrá utilizar, del mismo modo, para otros tipos de detección de cáncer.

Tomamos esta tesis como referencia porque compartimos la misma problemática ya que hoy en día el cáncer de mama sigue siendo un problema con alto índice de mortalidad

de acuerdo a ello incorporar indicadores a partir de las necesidades del hospital.

B) Autor: Javier Alonso Manrique Loayza

Título: Estudio mamográfico de tamizaje e histológico para el riesgo temprano del cáncer de mama en el centro de prevención y detección del cáncer del Inén.

Tipo: Tesis de pregrado

Año: 2016

Correlación:

El autor nos presenta la problemática: En el Perú, el diagnóstico de cáncer de mama se realiza en estadios avanzados (estadio III y IV), lo cual trae serias limitaciones en el manejo y posible curación de los pacientes. Es por ello que la detección temprana es de suma importancia, dependiendo de la metodología del tamizaje para lograr este resultado. La finalidad de esta tesis busca determinar el valor predictivo positivo de la mamografía de tamizaje para el diagnóstico del Cáncer de mama.

El autor realizó un estudio de enfoque cuantitativo, de diseño no experimental, así mismo consideró a toda la población de pacientes que ingresaron al centro con un estudio mamográfico y diagnóstico mediante anatomía patológica. Los datos obtenidos de la ficha de recolección de datos fueron analizados con el paquete estadístico SPSS 20.2 y para mejorar su rendimiento se utilizó un sistema de lectura e interpretación universal, conocido como BI-RADS (Breast Imaging Reporting and Data System); esto permite clasificar y establecer parámetros en su control, seguimiento y manejo.

El autor concluye que el estudio de las mamografías reportadas como BI-RADS 4 y sub-categorías A, B y C; presentan un valor predictivo positivo que va en aumento dependiendo del hallazgo mamográfico de mayor sospecha de malignidad para la detección del cáncer de mama.

Manrique (2016) sostiene: “El proceso de tamizaje para el diagnóstico temprano

del cáncer de mama puesto que también estamos trabajando con indicadores que le dan similitud a la tesis hallada en cuanto al tiempo de espera o respuesta de esta compleja enfermedad” (p. 54).

C) Autor: Anaís Jackelin, Ascencio Puicón

Título: Resiliencia y estilos de afrontamiento en mujeres con cáncer de un hospital estatal de Chiclayo 2013.

Tipo: Tesis de pregrado

Año: 2013

Correlación:

En la presente tesis las autoras mencionan que el diagnóstico de cáncer supone un importante impacto psíquico en la persona que lo recibe, por ende, las mujeres que lo padecen poseen un evento traumático ya que viven con temor, tristeza, aislamiento o preocupación. Esta investigación tiene como objetivo determinar si existe asociación entre los niveles de resiliencia y los niveles de los estilos de afrontamiento en las pacientes de un hospital estatal de Chiclayo.

Ascencio (2013) afirma:

La población a 70 mujeres diagnosticadas con cáncer de mama de un hospital estatal de Chiclayo, provenientes de distintos lugares de la región Lambayeque, que acudían por consulta externa al Servicio de Oncología para recibir tratamiento. Por otro lado, las autoras tomaron como instrumento la Escala de Resiliencia de Wagnid y Young, dirigida a adolescentes y adultos, esta se aplica de forma individual y colectiva con una duración de 15 minutos aproximadamente. Como conclusión las autoras mostraron que sí existe asociación directa de grado fuerte entre los niveles de resiliencia y los niveles del estilo de afrontamiento centrado a la emoción lo que significa que el paciente oncológico resiliente, conociendo su diagnóstico y mediante la expresión de las emociones, ya sea de forma positiva o negativa, analiza la manera de sobreponerse a ella, lo que le permite sobrellevar la enfermedad. (p. 68).

Tomamos de referencia esta tesis porque es importante saber sobre el estado de una persona y cómo puede sobrellevar a este tipo de situación nos habla del conocimiento y

el tipo de casuística sobre el afrontamiento en mujeres con cáncer de mama y tomarlos en consideración en diversos escenarios. Por otro lado, es importante considerar que la resiliencia puede variar a través del tiempo y las circunstancias, así como de los recursos con los que cuenta para recuperarse y superar la situación.

D) Autor: María Gabriela Camborda Zamudio

Título: Aplicación de árboles de decisión para la predicción del rendimiento académico de los estudiantes de los primeros ciclos de la carrera de ingeniería civil de la universidad continental.

Tipo: Tesis de pregrado

Año: 2014

Correlación:

La presente tesis tiene como objetivo determinar el porcentaje de exactitud de la técnica de árboles de decisión y la medida de concordancia de la técnica de árboles de decisión para predecir el rendimiento académico de los estudiantes de los primeros ciclos de la carrera de Ingeniería Civil de la Universidad Continental.

Camborda (2014) sostiene:

Para definir la población la autora tomó en cuenta todos los Árboles que se pueden formar con la combinación de los atributos (campos) referidos a los registros de los estudiantes I al III Ciclo de la Carrera de Ing. Civil de la Universidad continental. La muestra en su proyecto no fue probabilística-intencionada, ya que la autora seleccionó los datos según su propio criterio, sin ninguna regla matemática o estadística. El investigador procede a seleccionar la muestra en forma intencional, eligiendo aquellos elementos que considera convenientes y cree que son los más representativos.

Los instrumentos que la autora utilizó en su investigación consisten en unos cuestionarios que contienen preguntas cerradas, abiertas y de opción múltiple. Los datos obtenidos mediante la aplicación de las técnicas e instrumentos antes mencionados; fueron incorporados a programas computarizados, tales como el aplicativo de MS Office -Excel y con precisiones porcentuales que son presentados como informaciones en forma de, gráficos y cuadros. (p. 41).

Esta tesis nos sirvió como referencia ya que vimos reflejado uno de nuestros

indicadores en cuenta la satisfacción de los pacientes. Esto nos ayuda a visualizar que la supervivencia general a largo plazo nos ayuda a elegir la mejor estrategia de tratamiento para el cáncer de mama en los pacientes.

E) Autor: Bruno Filipe Aveleira Andrade

Título: Modelo de predicción para mujeres recurrentes del cáncer de mama

Tipo: Tesis de pregrado

Año: 2015

Correlación:

En la presente tesis, el autor señala que, a pesar de los avances considerables en el estudio de Cáncer de mama en las últimas dos décadas, los procesos subyacentes de recurrencia aún no se han entendido completamente ya que indica que el cáncer reaparece aún después del tratamiento y que, en el caso específico de ser cáncer de mama, la recurrencia es muy común y la experimentan aproximadamente un tercio de los pacientes después del diagnóstico inicial. Por ello el autor tiene como finalidad conducir una investigación basada en datos, tratando de construir un modelo de recurrencia para pacientes con esta condición detallando como objetivo estudiar los factores pronósticos para predecir la recurrencia de un particular paciente, en función de sus características personales.

Aveleira (2015) afirma:

Para el desarrollo de la presente tesis el autor a 257 pacientes y usaron los algoritmos de Machine Learning con la capacidad de revelar información oculta en los datos, generalizando de sus patrones subyacentes. A su vez el autor utiliza variables de respuesta binaria en la tarea de clasificación, o tratar de predecir periodos de tiempo. El autor de esta investigación usa algoritmos estadísticos para encontrar las características de los pacientes en cada población de estudio. (p. 39).

El autor concluye que las técnicas estadísticas fueron el enfoque tradicional para descubrir relaciones ocultas entre las variables de datos, pero las técnicas de minería de datos se han adoptado gradualmente, y se han aplicado en varios campos, incluida la

investigación médica.

Tomamos esta tesis como referencia ya que observamos cómo es el cuidado y tratado de esta enfermedad una vez que se haya detectado (post), por otro lado, vemos características y datos similares a que estamos usando en nuestra tesis de acuerdo con nuestros indicadores.

F) Autor: Patricio Miguel Araneda García

Título: Diseño y modelo preliminar de una plataforma de integración de datos clínicos y genómicos: aplicaciones en Alzheimer y cáncer de mama

Tipo: Tesis de maestría

Año: 2016

Correlación:

La presente tesis tiene como objetivo diseñar e implementar un conjunto de herramientas de software para permitir procesos de extracción, transformación y carga (ETL) de información sobre las bases de datos creadas y permitir consultas en línea mediante webservice. Dichos webservice se construyeron utilizando software open source y las mejores prácticas de diseño de interface, fuerte prototipado y técnicas de desarrollo xtreme programming. Por otra parte, el autor implementó la utilización de un repositorio de datos (data warehouse) para datos clínicos y genómicos. El autor indicó que para realizar la tesis tuvo que obtener los datos clínicos mediante recolección por fichas clínicas junto a datos demográficos, incluye mediciones clínicas y menciona también que la información se almacena en un sistema de base de datos híbridos, permitiendo tanto datos clínicos estructurados como datos genómicos de tipo documental.

Araneda (2016) sostiene:

La elección de la estructura de datos seleccionada, específicamente postgresQL (con un modelo NoSQL), resultó ser efectiva en el almacenamiento de información de los distintos dominios, información biomédica que se caracteriza por ser heterogénea según la patología a estudiar. Por otro lado, permitió una mayor simplicidad en la adecuación de los registros y las modificaciones

permanentes que se realizan en una base de datos de este tipo y facilitó el desarrollo del software al unificar las funciones de manejo de los datos. (p. 52).

La tesis nos sirvió como referencia ya que tenemos la misma finalidad en la búsqueda de datos accediendo a mayor información del paciente, lo cual también buscamos definir de mejor manera el fenotipo de su enfermedad y de esa manera decidir el mejor tratamiento.

G) Autor: César Carlos López Portillo

Título: Minería de datos como soporte en el diagnóstico y tratamiento del cáncer de mama.

Tipo: Tesis de maestría

Año: 2013

Correlación:

En la presenta investigación el autor nos indica que el cáncer de mama es una de las enfermedades más comunes que existe y que ocupa el primer lugar en mortalidad. Indica que, en estos casos, el tratamiento de la enfermedad únicamente tiene el propósito de aliviar síntomas, pero no para alcanzar la curación.

Dada esta problemática el autor tiene como objetivo aplicar el estudio de minería de datos en repositorios relacionados con cáncer de mama con la finalidad de descubrir conocimiento útil que apoye al proceso de diagnóstico y tratamiento del cáncer de mama en sus diferentes etapas.

Para el desarrollo de la tesis se aplicó el proceso de minería de datos (Proceso KDD) sobre los repositorios obtenidos en los casos de estudios obtenidos de la clínica, además se realizó un estudio sombra durante la consulta de pacientes diagnosticadas con cáncer de mama, con la finalidad de entender el proceso de toma de decisión que conlleva el diagnóstico y el tratamiento de dicha enfermedad. Así mismo se utilizó el software Weka para detectar con mayor precisión sus valores, en conjunto con el algoritmo Bayes Simple.

López (2013) afirma:

Fue posible clasificar las observaciones de tumores como benignos y malignos en base a los datos observados y medidos en la clínica. Los datos referentes a la descripción morfológica de la observación fueron los más relevantes a la hora de hacer la clasificación e indica que los modelos obtenidos por los algoritmos de árbol de decisión, además de haber obtenido buenos resultados, proveen un diagrama que puede ser interpretado directamente por un profesional de la salud. (p. 64).

Esta tesis nos ayudó como antecedente al desarrollo de nuestra investigación debido a que estamos usando el software para poder realizar nuestros objetivos trazados.

H) Autor: María del Rocío Boone Rojas

Título: Identificación de regiones con altas tasas de incidencia de cáncer mediante la integración y uso de técnicas de la minería de datos: almacenes de datos, agrupamiento y sistema de la información geográficos.

Tipo: Tipo de maestría

Año: 2011

Correlación:

La presente tesis busca desarrollar un proceso mediante la integración, el uso de herramientas y tecnologías especializadas en minería de datos ya que básicamente lo que se quiere es poder simplificar los tiempos de desarrollo y obtener resultados de mejor calidad que contribuya a disminuir los índices de mortalidad por cáncer. Como parte del desarrollo de este trabajo, la autora utilizó el proceso KDD, siendo un proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos. El desarrollo del estudio (uso de herramientas y tecnología de minería de datos) permitirá simplificar los tiempos de desarrollo, obtener resultados de mejor calidad y facilitar las tareas de actualización.

Boone (2011) sostiene:

Se logró una integración y utilización exitosa de las tecnologías especializadas de Minería de Datos, Almacenes de Datos, Agrupamiento y Sistemas de Información Geográficos, para identificar

regiones del país con altas tasas de incidencia de cáncer. Por otro lado, el enfoque adoptado para el desarrollo del estudio realizado, ha permitido obtener resultados de alta calidad simplificando los tiempos de desarrollo y facilitando las actividades de actualización y mantenimiento. Por otra parte, la implementación del almacén de datos, basado en el modelo ROLAP, ha permitido tomar ventaja de facilidades para las Bases de Datos relacionales. (p. 102).

Esta tesis nos sirvió como referencia a nuestra investigación debido a que compartíamos similares objetivos, así como también el uso de herramientas especializadas de la minería de datos dado que se busca la integración y agrupamientos de los mismos a fin de obtener resultados de buena calidad a fin optimizar los tiempos de respuesta.

I) Autor: Verónica Burriel Coll

Título: Diseño y desarrollo de un sistema de información para la gestión de información sobre cáncer de mama.

Tipo: Tesis de doctorado

Año: 2017

Correlación:

En la presente tesis el autor manifiesta que el diagnóstico, tratamiento e investigación sobre enfermedades tan complejas como el cáncer de mama es una tarea cada vez más complicada por la gran cantidad y diversidad de datos implicados. La presente tesis tiene como objetivo diseñar un sistema de gestión de datos clínicos y genómicos para su utilización en el ámbito clínico, manipulación eficiente y fiable de la información sobre el cáncer de mama. Para el desarrollo de la tesis el autor utilizó la metodología desing science ya que esta metodología sigue un patrón estereotípico y su tal uniformidad hace que sea más fácil reconocer y evaluar los resultados de la investigación.

Burriel (2017) manifiesta:

Bajo este planteamiento de problema y objetivos, la tesis ha llevado a cabo el modelado conceptual del dominio del diagnóstico, tratamiento e investigación del cáncer de mama, el diseño de arquetipos

bajo el estándar ISO 13606 para ofrecer interoperabilidad entre sistemas, la integración de datos de distintos orígenes relacionados con el cáncer de mama en una base de datos unificadora y el diseño de un prototipo de herramienta de gestión y análisis de datos clínicos y de expresión génica. La validación del Sistema de Información propuesto, mejora considerablemente la gestión de datos clínicos y biológicos de los pacientes, la realización de análisis y estudios con los mismos y la obtención de resultados. (p. 115).

Esta tesis nos sirvió ya que el sector salud hoy en día es uno de los factores más importantes y en la tesis observamos que el autor está trabajando con la normativa ISO 13606 (Informática de la salud), a su vez vemos cuan productivo es evaluar la precisión de los datos del paciente para desarrollar la eficiencia de los procedimientos y análisis en cuanto a su resultado.

J) Autor: Natalia Pérez Abad

Título: Minería de datos espaciales: Detección de agregaciones espaciales de riesgo de morir por cáncer de mama y próstata en España y su asociación con la contaminación industrial.

Tipo: Tesis de doctorado

Año: 2014

Correlación:

La autora menciona que en la actualidad, el cáncer de mama es el tumor más frecuente en la población de España, lo cual la autora desarrolla la tesis sobre minería de datos teniendo como objetivo evaluar la asociación entre la contaminación industrial y los patrones espaciales de los cánceres de hormonas dependientes más frecuentes en España (mama y próstata), y detectar posibles patrones geográficos de dichas causas en el territorio nacional, con medidas clásicas de estimación de riesgos y con técnicas de minería de datos espaciales. Para la realización de la tesis se dispuso de información sobre la mortalidad municipal por distintos tipos de cáncer de España como de una base de datos muy completa aportada por el ministerio. También utilizaron mapas representando la mortalidad por estos cánceres y análisis de agrupaciones espaciales de riesgo.

Pérez (2014) sostiene:

La autora concluye que las técnicas de minería de datos espaciales pueden ser claves para la comprensión y manejo de datos localizados geográficamente y un buen complemento en estudios epidemiológicos ambientales. Concluye también que, en relación con el análisis de asociación geográfica, los resultados obtenidos muestran un exceso de mortalidad por cáncer de mama. A su vez añadió que las emisiones de determinados grupos industriales pueden suponer un factor de riesgo para el desarrollo del cáncer de mama. Tomamos como referencia esta tesis ya que compartimos el mismo propósito para el diagnóstico de cáncer de mama, por lo cual gracias al desarrollo de los sistemas informáticos existen grandes volúmenes de datos con atributos que proveen información directa asociada a una localización específica que propician la aplicación de técnicas para la extracción de patrones que pueden proveer información nueva, útil y relevante. (p. 76).

K) Autor: Sandro Saúl Camacho Centellas

Título: Método heurístico para el diagnóstico de cáncer de mama basado en Minería de Datos.

Tipo: Tesis de maestría

Año: 2016

Correlación:

En la presente tesis el autor expone un trabajo de investigación documental cuyo propósito es elaborar un método que permita la selección de patrones obtenidos a través de la minería de datos para facilitar el diagnóstico oportuno del cáncer de mama. El cáncer de mama es una de las principales causas de muerte en las mujeres; por esta razón, existe una amplia evidencia de que la detección temprana juega un papel importante en la reducción de la mortalidad de este cáncer, muchos métodos han sido investigados para mejorar una detección temprana.

Camacho (2016) afirma:

El autor nos detalla que es importante evaluar el nivel de uniformidad en las familias, a través de cálculos de los índices de homogeneidad para posteriormente determinar rectángulos de contención de familias y abstraer los niveles de coincidencia, mediante el algoritmo de colonia de hormigas. En este trabajo se propone aumentar la fiabilidad de un diagnóstico con la utilización

de procesamiento de imágenes digitales y análisis de las mismas; mediante un Método Heurístico basado en Minería de Datos para extraer información esencial de las imágenes mamográficas y transformarlas en patrones. “El enfoque propuesto es un método sistemático y consistente para proporcionar recomendaciones y ayudar a la detección de patrones y conformación de familias de patrones” (p. 97).

L) Autor: Eugenio Hernández Martínez y Rodrigo Lorente Sanjuro

Título: Minería de datos aplicada a la detección de cáncer de mama

Tipo: Artículo científico

Año: 2014

Correlación:

El autor nos menciona en este artículo de investigación que el cáncer de mama afecta aproximadamente a un 15% de las mujeres en edad de riesgo de todo el mundo, sin embargo, una detección temprana de la presencia de células cancerosas malignas aumenta la posibilidad de vida de las pacientes.

Dada esta situación, el autor tiene como objetivo, realizar el análisis de los datos correspondientes a un conjunto de mujeres para poder obtener ciertos patrones a la hora de poder aplicarlos para detectar posibles casos de cáncer de mama mediante minería de datos, de manera similar a otros estudios basados en algoritmos genéticos.

Hernández y Lorente (2014) enfatizan:

Para el desarrollo del artículo, el autor trabajó con el software Weka ya que contiene una colección de herramientas de visualización y algoritmos para el análisis de datos y modelos de predicción. El autor concluye que, si bien todos los atributos pueden afectar en mayor o menor medida a la clasificación de un tumor como benigno o maligno. Este artículo nos ayudó a expandir y detallar más de nuestro trabajo de investigación ya que presenta diferentes fases de aplicación de Minería de Datos con el fin de cumplir el mismo objetivo que es obtener ciertos patrones para detectar posibles casos de cáncer de mama. (p. 11).

2.2. Marco teórico

2.2.1. Algoritmo J48 aplicando minería de datos

Cardona et al. (2017) define al algoritmo J48 como:

Una implementación libre en java del algoritmo C4.5, que utiliza el concepto de entropía de la información para la selección de variables que mejor clasifiquen a la variable (clase) estudiada.

La implementación del algoritmo J48 con el software libre Weka 3.8.0 ejecutado desde un clúster en la nube de AWS, el cual fue desarrollado con Starcluster 0.91. Este sistema es utilizado sobre una base de datos que contiene información de clasificación de vidrios a través del algoritmo, junto con datos de entrenamiento y validación cruzada, se logra crear un árbol de clasificación que permitirá predecir a qué clase de material pertenece el vidrio ingresado. (p. 51).

2.2.2. El proceso de la minería de datos

- **Recopilación:** Desde BDMS, datamarts, texto, imágenes, video, sonido, etc.
- **Selección de Datos:** Selección de atributos relevantes.
- **Preparación de Datos (Pre – procesamiento):** Mejora de la calidad de los datos, eliminación de atributos irrelevantes, eliminación de datos extremos, datos faltantes.
- **Minería de Datos:** Generación de modelos desde los datos recopilados y seleccionados. Uso de modelos.
- **Interpretación y evaluación:** Evaluación y uso de los resultados obtenidos, reformulación del modelo.

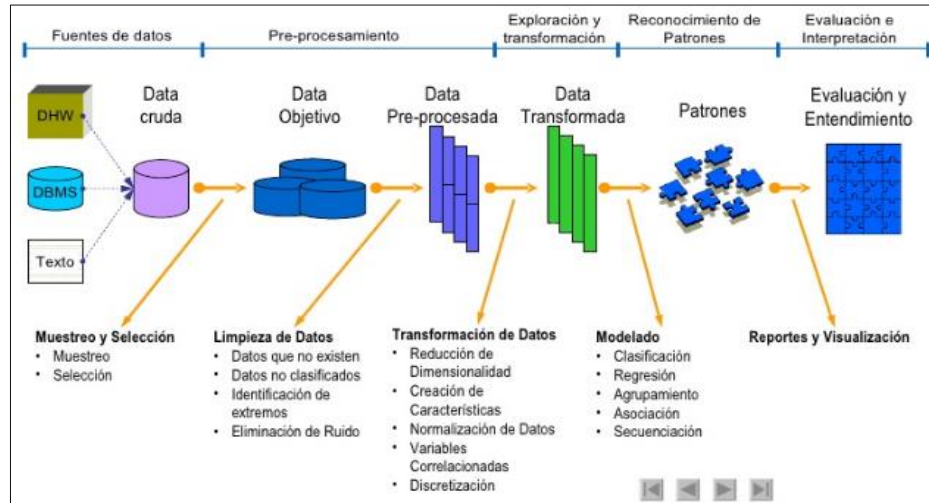


Figura 4. Proceso de minería de datos. Fuente: Universidad Autónoma del Estado de Hidalgo (2016).

El objetivo de los modelos predictivos basados en minería de datos es descubrir una o más de las variables en relación con todas las demás, son conocidas como métodos asimétricos, supervisados o directos. Se llevan a cabo mediante la búsqueda de normas de clasificación o de predicción basada en los datos, estas normas nos ayudan a predecir o clasificar el resultado futuro de una o más variables de respuesta o de destino en relación a lo que ocurre en la práctica con los motivos que la causan o bien en relación con las variables de entrada.

2.2.3. Estadísticas vs minería de datos

Tabla 10

Estadísticas vs minería de datos

	Estadística	Minería de datos
Construcción de modelos	Ceñido a premisas y teoremas	Mayor libertad en la construcción, interpretable
Búsqueda	Test de la razón de la verosimilitud	Metaheurísticos
Transparencia	Más complicados	Más claros y sencillos
Validación	No	Si

Además, en minería de datos se tiene:

- ✓ Mayor volumen de datos (diferencia fundamental)
- ✓ Mayor complejidad de los datos (miles de atributos o dimensiones)
- ✓ Su énfasis está en buscar más modelos que verificar un modelo fijo.

2.2.4. OLAP vs minería de datos

Tabla 11

Olap vs minería de datos

OLAP (ejemplos)	Minería de datos
¿Cuál es la proporción media de enfermedades de cáncer de mama?	Mayor libertad en la construcción, interpretable
¿Cuál es la factura telefónica media de mis clientes y de los que han dejado la compañía?	¿Dejará X la compañía? ¿Qué factores afectan a los abandonados?
¿Cuánto es la compra media diaria de tarjetas robadas y legítimas?	¿Cuáles son los patrones de compra asociados con el fraude de tarjetas?

2.2.5. Técnicas usadas en minería de datos

- **Técnicas no supervisadas:** Todo proceso de modelado se lleva a cabo sobre un conjunto de ejemplos formando tan sólo por entradas al sistema.

- Detección de desviaciones
- Segmentación
- Agrupamiento (clustering)
- Reglas de asociación
- Patrones secuenciales
- Método particional
- Método jerárquico

- **Técnicas supervisadas**

El proceso de modelado se realiza sobre un conjunto de ejemplos formados por entradas al sistema y la respuesta que debería dar para cada entrada.

Regresión lineal global, regresión no lineal ponderada localmente, árbol de decisión, estimador de núcleos, tabla de decisión, inducción de reglas, bayesiana, máquinas de soporte vectorial, basado en ejemplares, redes neuronales, lógica borrosa, técnicas genéticas, algoritmos genéticos, series temporales.

2.2.6. Árbol de decisión

Un árbol de decisión es un modelo de predicción cuyo objetivo principal es el aprendizaje inductivo a partir de observaciones y construcciones lógicas. Son muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva para la solución de un problema. Constituyen probablemente el modelo de clasificación más utilizado y popular.

León (2018) afirma:

Los árboles de decisión permiten representar una serie de reglas de forma gráfica con el fin de tomar una decisión sobre la asignación de un valor de salida. Se encuentran compuestos por nodos, ramas y hojas o nodos hoja, cuyas características son: Los nodos son los datos de entrada, las ramas son grupos de registros en las variables de entrada y los nodos hojas o las hojas son los valores de la variable de salida. (p. 17).

Los nodos finales o nodos hoja corresponden a una decisión, la cual coincide con una de las variables clase del problema a resolver. Este modelo se construye a partir de la descripción narrativa de un problema, ya que provee una visión gráfica de la toma de decisión, especificando las variables que son evaluadas, las acciones que deben ser tomadas y el orden en el que la toma de decisión será efectuada. Cada vez que se ejecuta este tipo de modelo, sólo un camino será seguido dependiendo del valor actual de la

variable evaluada. Los valores que pueden tomar las variables para este tipo de modelos pueden ser discretos o continuos. Esta técnica se encuentra dentro de una metodología de aprendizaje supervisado. Su representación es en forma de árbol en donde cada nodo es una decisión, los cuales a su vez generan reglas para la clasificación de un conjunto de datos.

Ventajas:

- Plantean el problema para que todas las opciones sean analizadas.
- Permiten analizar totalmente las posibles consecuencias de tomar una decisión.
- Proveen un esquema para cuantificar el costo de un resultado y la probabilidad de que suceda.
- Ayuda a realizar las mejores decisiones sobre la base de la información existente y de las mejores suposiciones.

Desventajas:

- Sólo es recomendable para cuando el número de acciones es pequeño y no son posibles todas las combinaciones.

Dorian (2013) afirma:

El primer sistema que construía árboles de decisión fue CLS de Hunt, desarrollado en 1959 y depurado a lo largo de los años sesenta. CLS es un sistema desarrollado por psicólogos como un modelo del proceso cognitivo de formación de conceptos sencillos. Su contribución fundamental fue la propia metodología, pero no resultaba computacionalmente eficiente debido al método que empleaba en la extensión de los nodos. Se guiaba por una estrategia similar al minimax con una función que integraba diferentes costes. En 1979 Quinlan desarrolla el sistema ID3, que él denominaría simplemente herramienta porque la consideraba experimental. (p. 114).

Conceptualmente es fiel a la metodología de CLS pero le aventaja en el método de expansión de los nodos, basado en una función que utiliza la medida de la información de Shannon. La versión definitiva, presentada por su autor Quinlan como un sistema de

aprendizaje, es el sistema C4.5 que expone con cierto detalle en la obra: Programs for machine learning.

2.2.7. Metodología para minería de datos

- CRISP – DM (Cross industry standard process for data mining), es un modelo de proceso de minería de datos que describe una manera en la que los expertos en esta materia abordan el problema.

Galán (2015) sostiene:

En el caso de los proyectos de implementación de Minería de Datos una de las metodologías que ha tenido más apoyo de las empresas privadas y organismos públicos es CRISP-DM, como se puede observar en la siguiente gráfica (figura 3), publicada en kdnuggets.com, y que representa el grado de utilización de las principales pautas de la aplicación de proyectos de minería de datos según las encuestas realizadas. Como se puede observar CRISP-DM ha experimentado un ligero descenso en los últimos años, pero sigue siendo la más empleada de las distintas metodologías. (p. 21).

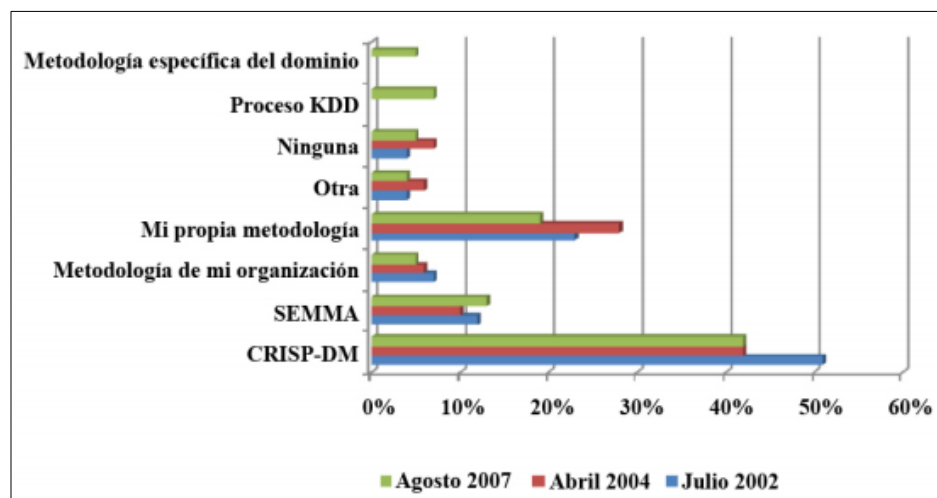


Figura 5. Estadística de metodología de minería de datos. Fuente: *Metodología para desarrollo de proyectos [CRISP-DM]*, 2000

CRISP-DM incluye un modelo y una guía, estructurados en seis fases, algunas de las cuales son bidireccionales, es decir que de una fase en concreto se puede volver a una fase anterior para poder revisarla, por lo que la sucesión de fases no tiene por qué ser

ordenada desde la primera hasta la última. En la figura 6 se puede observar las fases en las que se divide CRISP-DM y las posibles secuencias a seguir entre ellas.

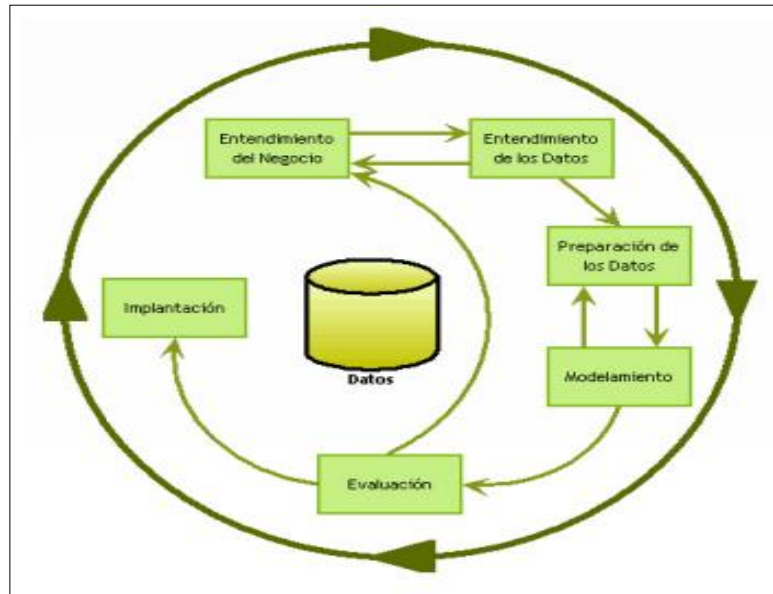


Figura 6. Representación de la metodología CRISP-DM. Fuente: Metodología para desarrollo de proyectos [CRISP-DM], 2000

2.2.8. Fases de la metodología

Jose Alberto (2009) afirma:

La metodología está estructurada en un proceso jerárquico, compuesto por tareas descritas en cuatro niveles diferentes de abstracción, que van desde lo general a lo específico. CRISP-DM, propone en el nivel más alto seis fases para el proceso de minería de datos: entendimiento del negocio, entendimiento de los datos, preparación de los datos, modelado, evaluación e implementación. La sucesión de fases, no es necesariamente rígida. Analizando el nivel más alto de abstracción del modelo, las seis fases que componen el proceso de minería de datos son:

Comprensión del negocio: en esta fase se determinan los objetivos y requerimientos del proyecto desde una perspectiva del negocio, definiendo el problema de minería y el plan de trabajo.

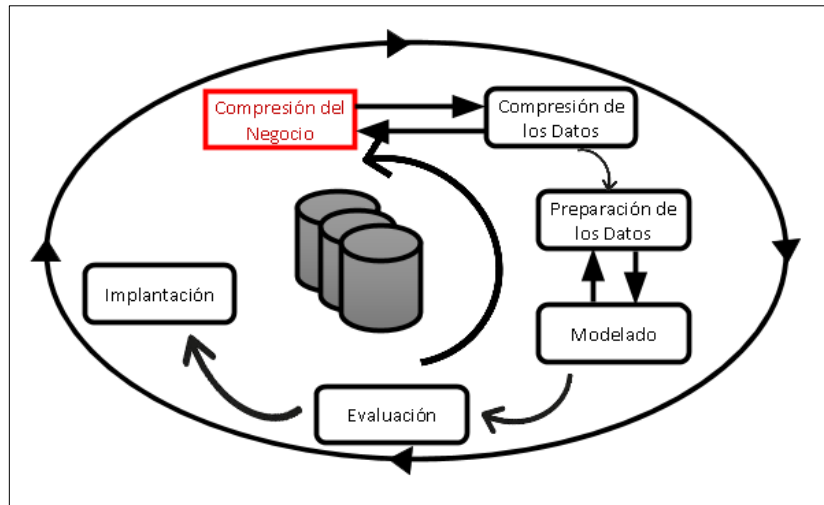


Figura 7. Fase 1 - Comprensión del negocio. Fuente: Metodología para desarrollo de proyectos [CRISP-DM], 2000.

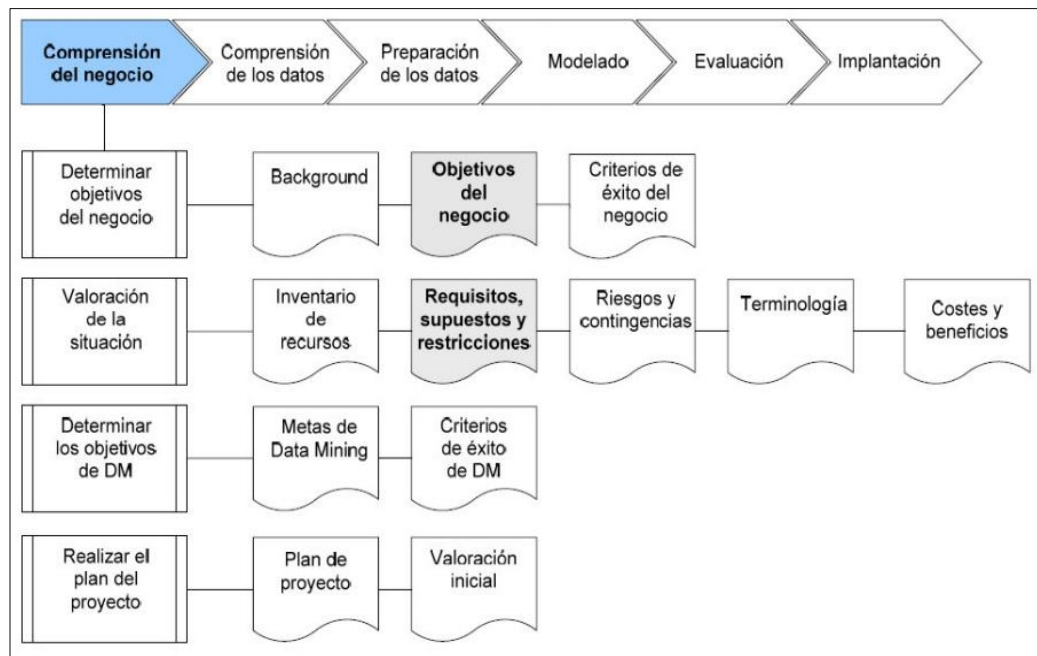


Figura 8. Esquema de la primera fase. Fuente: Metodología para desarrollo de proyectos [CRISP-DM], 2000.

Comprensión de los datos: fase que consiste en la recolección de datos que se utilizarán en el proyecto y la familiarización con los mismos. En esta etapa es posible el surgimiento de las primeras hipótesis acerca de la información que podría estar oculta.

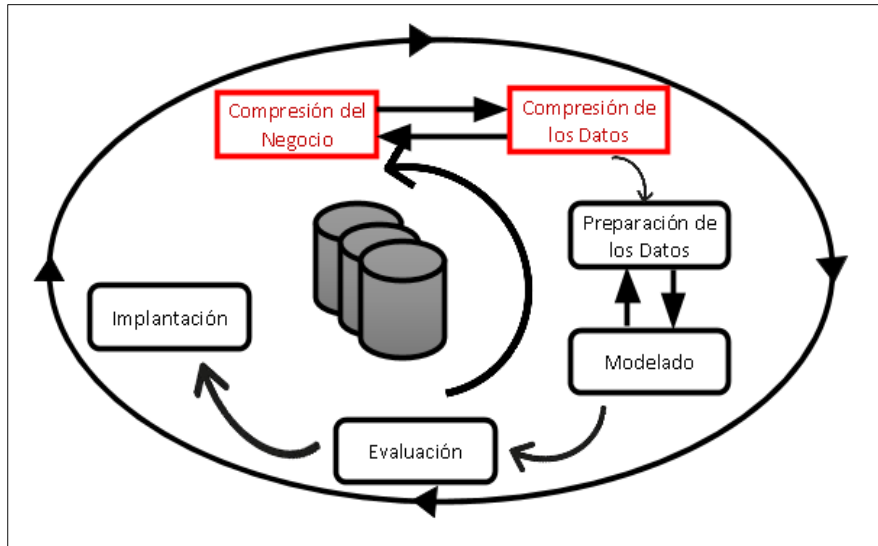


Figura 9. Fase 2 - Compresión de los datos. Fuente: Metodología para desarrollo de proyectos [CRISP-DM], 2000.

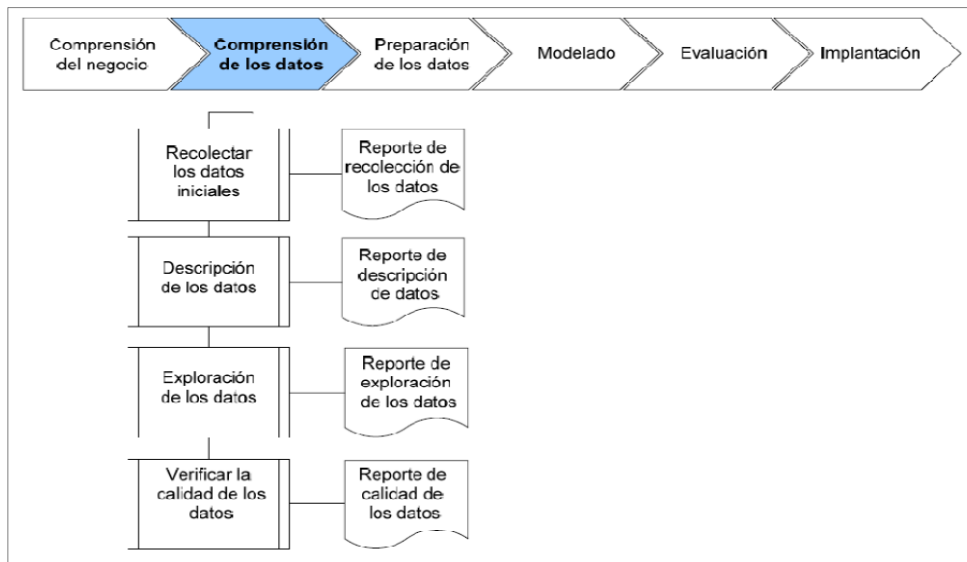


Figura 10. Esquema de la segunda fase. Fuente: Metodología para desarrollo de proyectos [CRISP-DM], 2000.

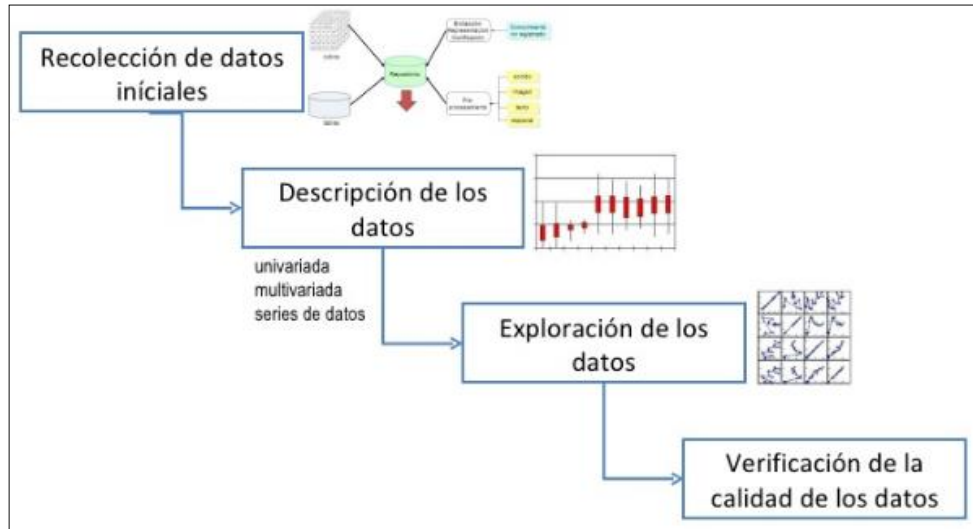


Figura 11. Secuencia de la segunda fase - comprensión de los datos. Fuente: *Metodología para desarrollo de proyectos [CRISP-DM], 2000.*

Preparación de los datos: Comprende las actividades de tratamiento de los datos para construir el conjunto de datos finales sobre el cual se aplicarán las técnicas de minería.

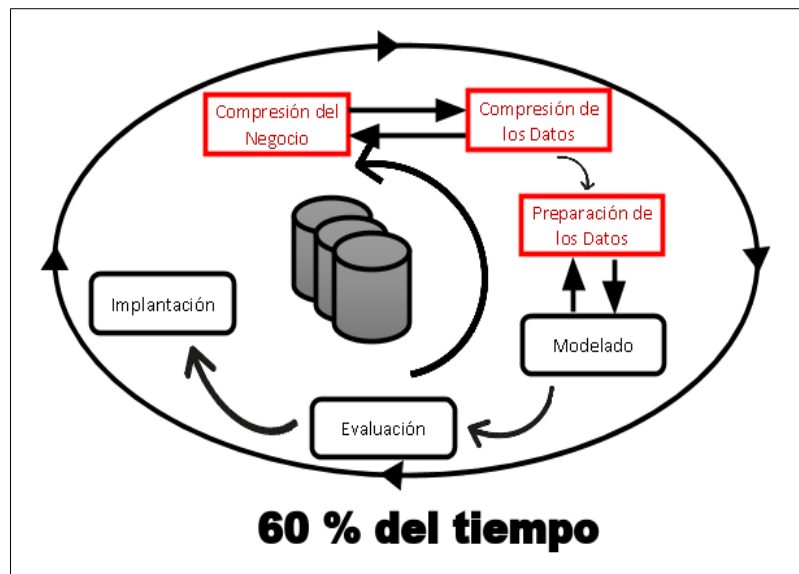


Figura 12. Fase 3 - Preparación de los datos. Fuente: *Metodología para desarrollo de proyectos [CRISP-DM], 2000.*

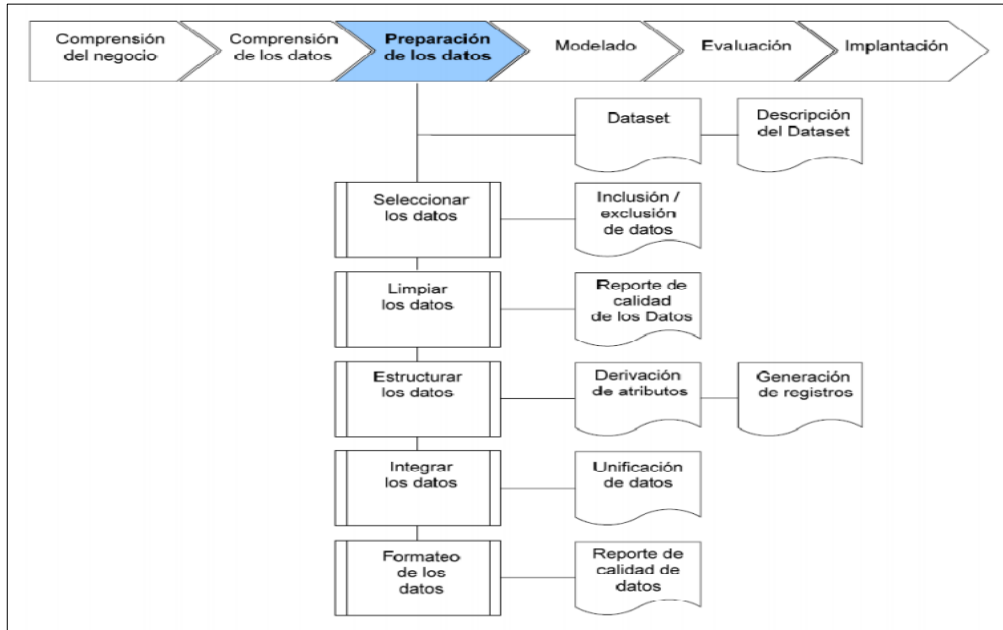


Figura 13. Esquema de la tercera fase. Fuente: *Metodología para desarrollo de proyectos [CRISP-DM], 2000.*

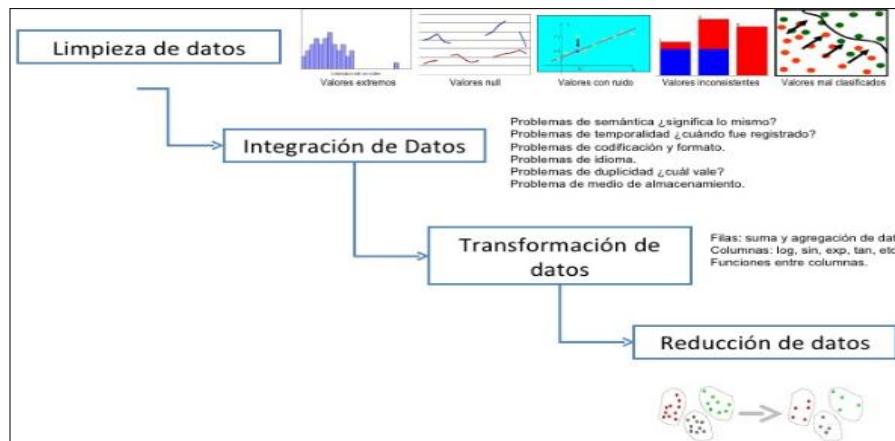


Figura 14. Secuencia de la tercera fase - preparación de los datos. Fuente: *Metodología para desarrollo de proyectos [CRISP-DM], 2000.*

Modelado: en esta etapa se aplican las diferentes técnicas y algoritmos de minería sobre el conjunto de datos para obtener la información oculta y los patrones implicados en ellos.

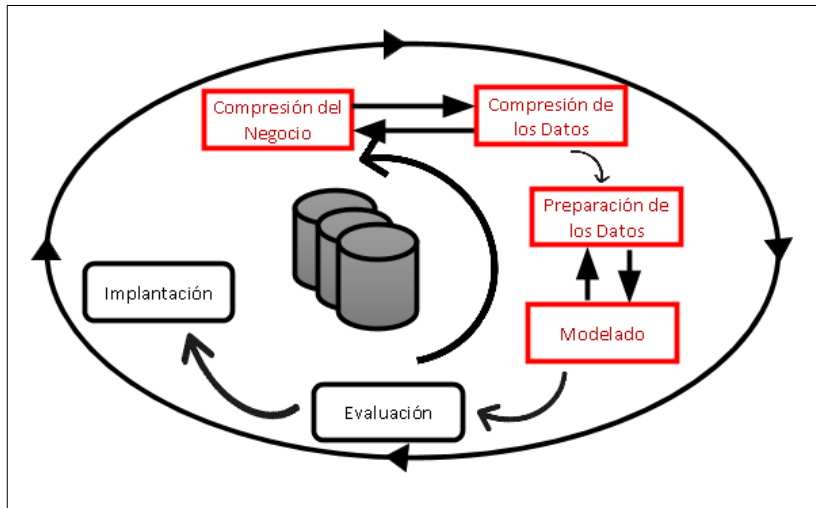


Figura 15. Fase 4 – modelado. Fuente: Metodología para desarrollo de proyectos [CRISP-DM], 2000.

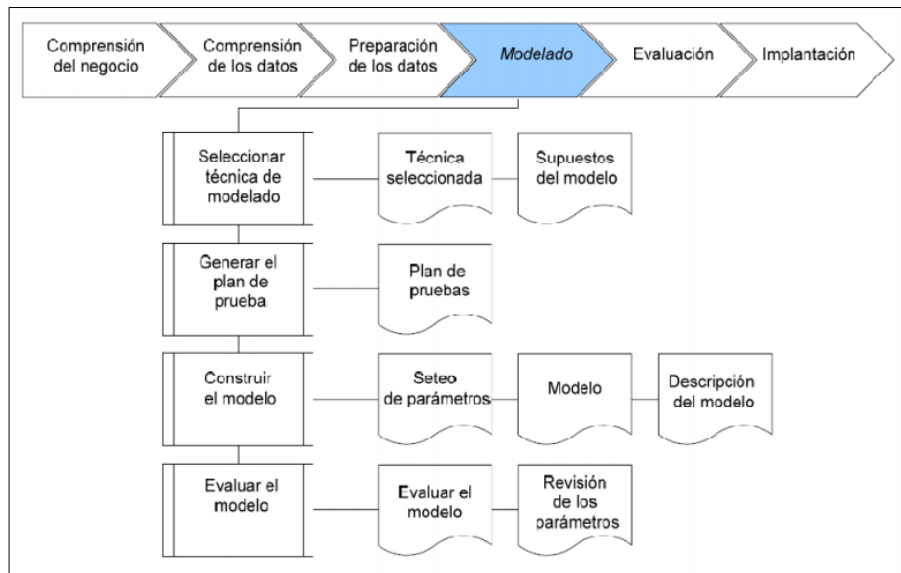


Figura 16. Esquema de la cuarta fase. Fuente: Metodología para desarrollo de proyectos [CRISP-DM], 2000.

Evaluación: La evaluación es la fase donde se analizan los patrones obtenidos en función de los objetivos organizacionales. En esta etapa se debería determinar si se ha omitido algún objetivo importante del negocio y si el nuevo conocimiento será implementado, es decir, si se pasará a la próxima fase.

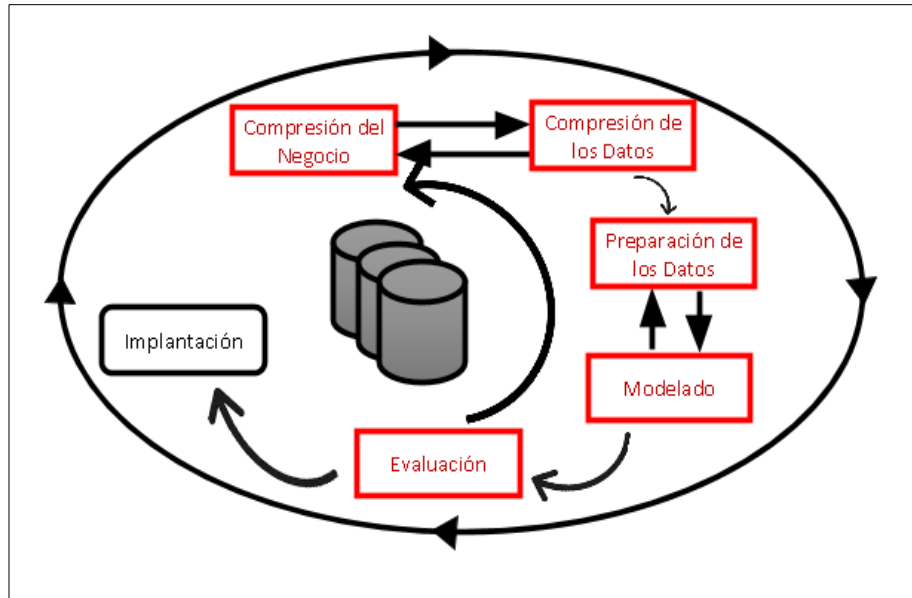


Figura 17. Fase 5 – Evaluación. Fuente: Metodología para desarrollo de proyectos [CRISP-DM], 2000.

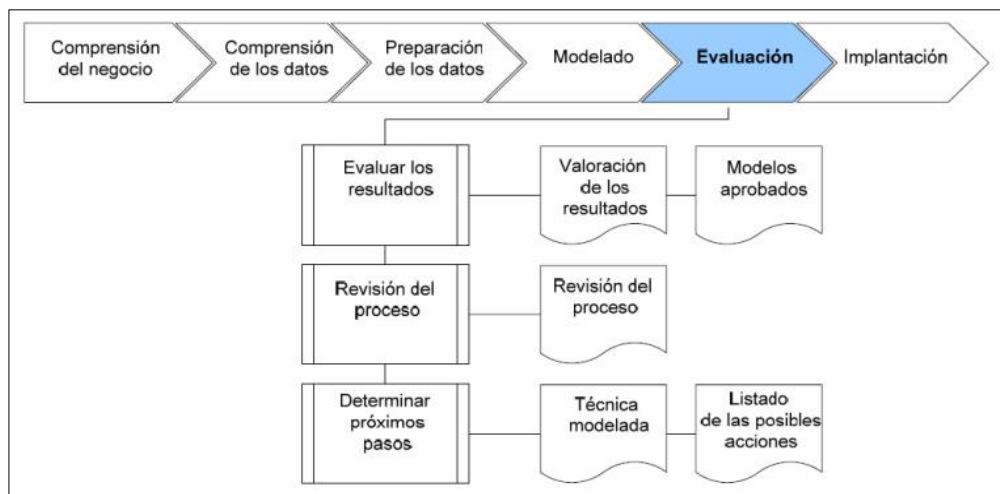


Figura 18. Esquema de la quinta fase. Fuente: Metodología para desarrollo de proyectos [CRISP-DM], 2000.

Implementación: consiste en la comunicación e implementación del nuevo conocimiento, el cual debe ser representado de forma entendible para el usuario. (p. 12).

Galán (2015) afirma:

Cada fase se descompone en una serie de conjunto de actividades genéricas (o generales) de segundo nivel. Estas tareas son genéricas ya que tratan de abarcar la mayoría de las situaciones posibles en minería de datos. A partir del tercer nivel de abstracción, se realiza un “mapeo” de las actividades genéricas definidas en el modelo a situaciones específicas. Por ello, las tareas genéricas se traducen en tareas específicas para casos y proyectos concretos. En el cuarto nivel, hallamos las instancias de estudio, donde se describen las acciones, decisiones y resultados de un proyecto particular de minería de datos. (p. 168).

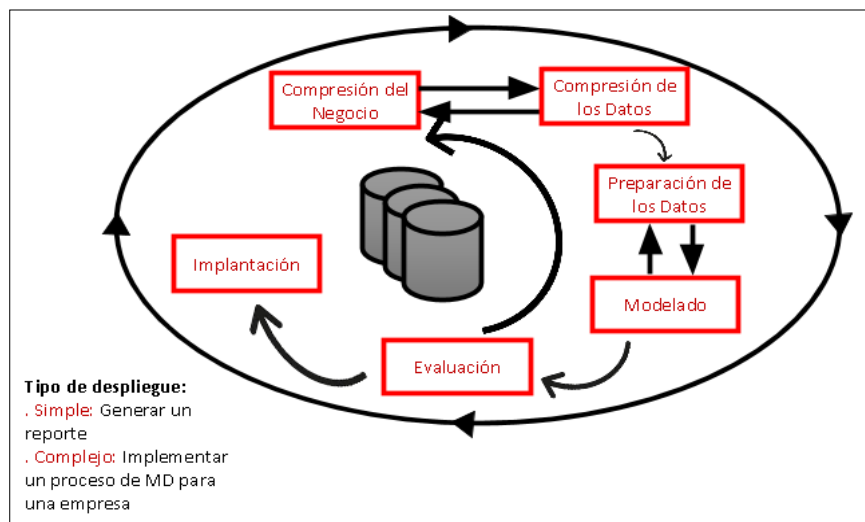


Figura 19. Fase 6 – Implantación. Fuente: Metodología para desarrollo de proyectos [CRISP-DM], 2000.

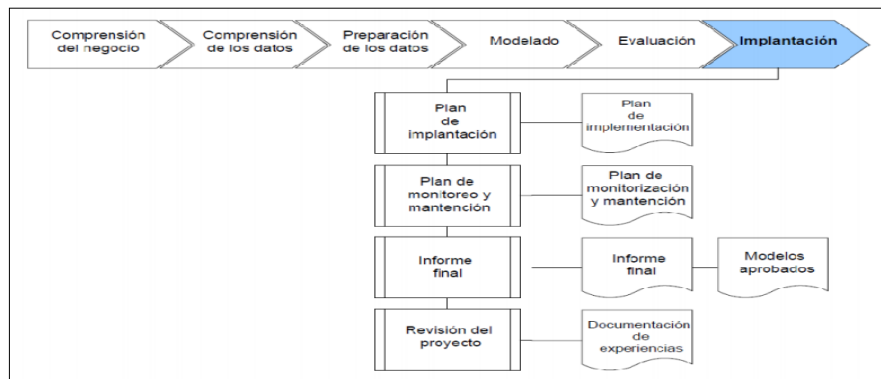


Figura 20. Esquema de la sexta fase. Fuente: Metodología para desarrollo de proyectos [CRISP-DM], 2000.

2.2.9. Tareas y salidas de las fases de la metodología Crisp-Dm

Fase I: Entendimiento del negocio

Ochoa (2016) menciona:

a) Tarea 1. Determinar los objetivos de negocio

Descripción: Entender y establecer cuáles son los objetivos que el cliente quiere alcanzar, desde una perspectiva del negocio.

Salidas:

- Background, con la información que se conoce sobre la situación actual de la organización, incluyendo una descripción general del problema y la solución actual para el mismo.
- Objetivos del negocio, identificando los objetivos principales del cliente.
- Criterios de éxito, describiendo los resultados esperados desde una perspectiva de negocio.

b) Tarea 2. Evaluar la situación de minería de datos

Descripción: Profundizar en la evaluación de la situación actual del negocio. Analizar con mayor profundidad las restricciones y factores que se deben tener en cuenta para el proyecto.

Salidas:

- Inventario de recursos, donde deberán incluirse los recursos disponibles (como los recursos humanos, fuentes de datos, hardware y software).
- Lista de requerimientos del proyecto, supuestos y restricciones que se han detectado.
- Riesgos y planes de contingencia. Consiste en la identificación de los potenciales riesgos para el proyecto y la planificación de las acciones reactivas que se llevarán a cabo (planes de contingencia).
- Glosario con terminología relevante para el proyecto. En el mismo deberá incluirse un glosario de terminología del negocio y otro de minería de datos.
- Análisis costo-beneficio del proyecto.

c) Tarea 3. Determinar los objetivos de la minería de datos

Descripción: Los objetivos del negocio se describen en términos organizacionales, sin embargo, los objetivos de minería de datos describen los objetivos del proyecto en “términos técnicos”, lo cual significa que, si un objetivo de negocio busca aumentar el volumen de ventas, el objetivo de minería de datos podría ser el “agrupamiento” de los clientes para la promoción de nuevas campañas publicitarias.

Salidas:

- Objetivos de minería de datos, describiendo los resultados del proyecto que permiten el logro de los objetivos de negocio.
- Definir un criterio de éxito para el proyecto de minería.

Tarea 4. Crear un plan para el proyecto de minería de datos.

Descripción: Crear una planificación para el proyecto de minería, el cual debe ser consistente con los objetivos planteados.

Salidas:

- Plan de proyecto: Listar las tareas que deben ser realizadas, duraciones y recursos necesarios, así como sus entradas y salidas. El plan del proyecto es un documento dinámico, que debe ser revisado y ajustado al final de cada fase.
- Evaluación de técnicas y herramientas de minería de datos que podrían ser utilizadas en el proyecto. (p. 3).

Fase 2: Entendimiento de los datos

Según Ochoa (2016) menciona:

a) Tarea 1. Recolectar los datos iniciales

Descripción: Recolectar los datos necesarios y específicos en la lista de recursos del proyecto a realizar.

Salidas:

- Reporte de recolección inicial de datos, donde se detalla la forma en la que han sido obtenidos los conjuntos de datos y los problemas que han surgido en el proceso.

b) Tarea 2. Describir los Datos

Descripción: Describir los datos recolectados.

Salidas:

- Descripción de los datos, incluyendo el formato de los mismos y su tamaño (como cantidad de registros y variables).

c) Tarea 3. Explorar los datos

Descripción: Realizar una explotación de los datos, observando la distribución de las variables con mayor relevancia. En esta fase es conveniente el uso de técnicas simples de análisis estadístico.

Salidas:

- Reporte de la exploración de los datos, donde se expongan los resultados del análisis y las hipótesis iniciales con su impacto y alcance en el proyecto.

d) Tarea 4. Verificar la calidad de los datos

Descripción: Examinar la calidad de los datos, incluyendo un análisis de su completitud, de potenciales errores en los mismos y de los datos ausentes.

Salidas:

- Reporte de calidad de los datos, donde se documente el análisis de calidad realizado y se propongan potenciales soluciones a los problemas hallados. (p. 5).

Fase 3: Preparación de los datos

Ochoa (2016) menciona:

a) Tarea 1. Seleccionar los datos

Descripción: Seleccionar los datos que serán utilizados para el análisis. En esta etapa se debe seleccionar con qué atributos (columnas) y con qué observaciones (filas o registros) se trabajará en el proyecto.

Salidas:

- Justificación de la selección. Documento donde se justifiquen las causas por las cuales se incluyeron y excluyeron los datos a trabajar.

b) Tarea 2. Limpieza de datos

Descripción: Es una etapa se busca mejorar la calidad de los por lo cual se deberán tomar decisiones sobre los problemas de calidad hallados en los mismos, como datos ausentes.

Salidas:

- Reporte de limpieza de datos, donde se incluyan las decisiones tomadas sobre los diferentes problemas de calidad de los datos (reportados en la fase “2.4 Verificar la calidad de los datos”).

c) Tarea 3. Construcción de los datos

Descripción: En esta fase se realiza la construcción de nuevos datos, derivados de los disponibles, que son relevantes para el análisis. Estos datos pueden ser atributos calculados o atributos transformados.

Salidas:

- Atributos derivados. Se calculan a partir de otros atributos del mismo registro. Por ejemplo: edad del paciente.

- Registros creados. Estos registros se crean cuando son necesarios en la fase posterior de modelado.

d) Tarea 4. Integrar los datos

Descripción: Consiste en la integración de los datos de diferentes tablas o registros.

Salidas:

- Datos combinados. Resultan de integrar la información de dos o más tablas que tienen diferente información de las mismas observaciones. Por ejemplo, la integración de los datos personales y los datos de las atenciones efectuadas a un paciente en un centro de salud. En esta fase se incluye el cálculo de agregaciones, en la cual se calculan los nuevos datos resumiendo la información de diferentes tablas y registros. Siguiendo con el ejemplo del centro de salud, podríamos integrar en un solo registro los datos personales del paciente, el total de atenciones efectuadas, y el promedio anual de consultas médicas realizadas.

e) Tarea 5. Formatear los datos

Descripción: Esta etapa se refiere al cambio que debe realizarse en el formato de los datos por los requisitos de las técnicas de modelado elegidas. Por ejemplo, el formato de fechas o el orden de set de datos.

- Salidas: Conjunto de datos reformateados. (p. 6).

Fase 4: Modelado

Según Ochoa (2016) menciona:

a) Tarea 1. Seleccionar la técnica del modelado

Descripción: Consiste en seleccionar cual será la técnica de minería de datos que será utilizada. Por ejemplo, en un caso donde se ha definido un problema de agrupamiento (clustering), se puede utilizar el algoritmo k-medias. Si se ha optado por el uso de múltiples técnicas, se debería repetir esta tarea para cada una.

Salidas:

- Técnica de modelado. Realizar en la documentación la técnica de modelado con la que se trabajará.
- Supuesto del modelo. Algunas técnicas asumen supuestos sobre el conjunto de datos, como la distribución normal de una variable.
- Realizar la documentación de todos los supuestos realizados.

b) Tarea 2. Diseñar las pruebas del modelado

Descripción: Una vez que los modelos han sido construidos, se necesitará un mecanismo para determinar su calidad y validez. Por ejemplo, en problemas de agrupamiento y selección se puede utilizar el coeficiente de silueta para evaluar la robustez de los grupos hallados y en problemas de clasificación la tasa de error para estimar la capacidad del clasificador.

En esta fase se dividirá el conjunto de datos en un grupo para entrenar el modelo (training) y otro para probarlo (test).

Salidas:

- Diseño de las pruebas. Determinar y documentar de qué forma se estrenarán y evaluarán los modelos generados. Incluir las decisiones tomadas sobre los datos que se utilizarán para entrenamiento y prueba.

c) Tarea 3. Construir el modelo

Descripción: Aplicar la técnica seleccionada sobre el conjunto de los datos para generar más modelos. En la actual fase, el modelo será evaluado con diferentes valores de parámetros. Por ejemplo, en un algoritmo de agrupamiento k-medias, se podrían generar distintos modelos para diferentes valores de “k” o grupos.

Salidas:

- Parámetros seleccionados. Listar los parámetros que se han proporcionado al modelo.
- Modelos producidos por las herramientas de minería.
- Descripción de los modelos.

d) Tarea 4. Evaluar el modelo

Descripción: En la actual fase el equipo del proyecto interpreta y evalúa el modelo en función de su conocimiento del dominio, los criterios de éxito definidos para el proyecto y las pruebas diseñadas para el modelo.

Salidas:

- Evaluación de modelos. Realizar un reporte de evaluación de los modelos obtenidos, describiendo sus características para los mismos.
- Evaluación de parámetro. (p. 8).

Fase 5: Evaluación

Ochoa (2016) menciona:

a) Tarea 1. Evaluar los resultados

Descripción: En esta etapa se evalúa el modelo en función de los objetivos del negocio, determinando su validez de acuerdo con los intereses organizacionales. Además del modelo, puede haber surgido como parte del proceso nueva información importante y futuras fuentes de investigación.

Salidas:

- Evaluación de los resultados de la minería de datos con respecto a diferentes criterios de éxito y objetivos de negocios.

- Modelos evaluados y aprobados.

b) Tarea 2. Revisión del proceso

Descripción: Realizar una revisión completa del proceso realizado en la búsqueda de posibles errores u omisiones.

Salidas:

- Revisión del proceso, documentando un resumen de este.

c) Tarea 3. Determinar las siguientes etapas

Descripción: En función de la evaluación de resultados y la revisión del proceso, se debe decidir cómo continua el proyecto: si se pasa a la próxima fase (implementación) o si se retorna a una fase anterior.

Salidas:

- Lista de posibles acciones y descripción de la decisión tomada. (p. 9).

Fase 6: Implementación

Ochoa (2016) menciona:

a) Tarea 1. Planificar la implementación

Descripción: En esta etapa se genera el plan de implementación de los resultados obtenidos de la minería de datos.

Salidas:

- Plan de implementación, incluyendo las etapas y cómo llevarlas a cabo.

b) Tarea 2. Planificar el monitoreo y mantenimiento

Descripción: El monitoreo y mantenimiento es de mucha importancia si los resultados de la minería de datos tomaran parte del trabajo diario del negocio y su entorno.

Salidas:

- Plan de mantenimiento

c) Tarea 3. Crear un reporte final

Descripción: Generar un reporte final, que podría resumir el desarrollo del proyecto para mostrar un análisis comprensivo de los resultados obtenidos en el proceso de minería.

Salidas:

- Reporte final del proyecto
- Presentación final al cierre, incluyendo resultados y conclusiones.

d) Tarea 4. Revisión del proyecto

Descripción: Consiste en identificar y analizar los puntos que fueron bien realizados, los que fueron mal realizados, y los que podrían mejorarse.

Salidas:

- Documentación de la experiencia adquirida durante el desarrollo del proyecto. (p. 10).

2.2.10. Estudio comparativo de metodologías para minería de datos

Tabla 12

Cuadro comparativo de diversas metodologías de minería de datos

Metodología	Enfoque
CRISP – DM (Cross Industry Standard Process for Data Mining)	La metodología está estructurada en un proceso jerárquico, compuesto por tareas descriptas en cuatro niveles diferentes de abstracción, que van desde el lado general hasta el lado específico.
SEMMA (Sample, Explore, Modify, Model and Assess)	Se encuentra enfocada especialmente en aspectos técnicos, excluyendo actividades de análisis y comprensión del problema que se está abordando. Fue propuesta especialmente para trabajar con el software de minería de datos de la compañía SAS. Este producto organiza sus herramientas (llamadas “nodos”) en base a las distintas fases que componen la metodología.
KDD (Knowledge Discovery in Database)	El KDD involucra un proceso iterativo e interactivo de búsqueda de modelos, patrones o parámetros. Sus metas son procesar grandes cantidades de datos, identificar los patrones más significativos y relevantes, y presentarlos como conocimiento para satisfacer los objetivos del usuario.
CATALYST (Product, Place, Price, Time, Quantiry)	Las relaciones ente estas variables buscan mantener el producto correcto, en el lugar adecuado, en el momento adecuado, en la cantidad correcta y con el precio correcto. Esta metodología plantea la formulación de dos modelos: el modelo de negocios y el modelo de minería de datos.

2.2.11. Estudio comparativo de metodologías para minería de datos

Tabla 13

Análisis comparativo de cada enfoque

Fases	Kdd	Crisp – Dm	Semma	Catalyst
Análisis y comprensión del negocio	Comprensión del dominio de aplicación	Comprensión del negocio		
	Crear el conjunto de datos	Entendimiento de los datos	Muestreo Comprensión	
Selección y preparación de los datos	Limpieza y procesamiento de los datos		n	Preparación de los datos
	Reducción y proyección de los datos	Preparación de los datos	Modificación	
	Determinar la tarea de minería		n	
Modelado	Determinar el algoritmo de minería	Modelado	Modelado	Selección de herramientas y modelado inicial
	Minería de datos			Refinamiento del modelado
Evaluación	Interpretación	Evaluación	Vibración	
Implementación	Utilización del nuevo conocimiento	Despliegue		Comunicación

2.2.12. Software weka

Fallas (2013) afirma:

Weka es una herramienta de tipo software para el aprendizaje automático y minería de datos diseñado a base de Java y desarrollado en la universidad de Waikato en Nueva Zelanda en el año 1993, esta herramienta por su nombre en inglés (**W**aikato **E**nvironment for **K**nowledge **A**nalysis). Además es una herramienta de distribución de licencia GNU-GLP o software libre y está orientado a la extracción de conocimientos desde bases de datos con grandes cantidades de información. (p. 65).

Moreno (2016) sostiene: “Weka tiene una colección de algoritmos para realizar el análisis de datos y el modelado predictivo, también tiene herramientas para la visualización de estos datos, además provee una interfaz gráfica que unifica las herramientas para que estén a una mejor disposición” (p. 18)

Por otro lado, Weka tiene herramientas necesarias para realizar transformaciones sobre los datos, tareas de clasificación, regresión, clustering, asociación y visualización y está diseñado como una herramienta orientada a la extensibilidad por lo que una de las propiedades más interesantes de este software, es su facilidad para añadir extensiones, modificar métodos etc.

Características de Weka

Weka es una herramienta muy versátil que soporta muchas tareas estándar de la minería de datos en especial tareas de procesamiento de datos, regresión, clasificación, clusterin entre otras, así mismo permite la visualización y la selección de los datos.

Todas las técnicas en WEKA están basadas en la función de datos que están disponibles en un fichero plano o una relación, en donde cada registro de datos esta descrito por un número fijo de atributos nominales o numéricos.

Permite el acceso a otras instancias de bases de datos por medio de SQL, gracias al JDBC, además puede procesar un resultado generado a base de una consulta hecha a una base de datos.

Moreno (2016) afirma:

Weka soporta varias tareas estándar de minería de datos, especialmente, preprocesamiento de datos, clustering, clasicación, regresión, visualización, y selección. Todas las técnicas de Weka se fundamentan en la presunción de que los datos están disponibles en un archivo plano (flat, file, archivo de texto sin formato) o una relación, en la que cada registro de datos está descrito por un número fijo de atributos (normalmente numéricos o nominales, aunque también se soportan otros tipos). (p. 31).

Ventajas de la herramienta Weka

- Es un software de fácil acceso por los usuarios ya que está disponible libremente bajo licencia pública GNU.
- Contiene una gran gama de técnicas para modelado y procesamiento de datos.
- Posee una interfaz de usuario muy sencilla que permite su uso hasta por usuarios sin experiencia.
- Funciona en la mayoría de las plataformas actuales ya que es una implementación en java.

Desventajas de la herramienta Weka

- No incluye algoritmos para el modelado de secuencias.
- Al utilizar métodos de combinación de modelos, los resultados tienden a complicarse, perdiendo comprensibilidad.

2.2.13. Cáncer de mama

El cáncer de mama es el crecimiento descontrolado que se forma en las células mamarias. Estas células se dividen más rápidamente que las células sanas y continúan acumulándose, y de este modo forman un bulto o una masa. Estas células pueden esparcirse (hacer metástasis) por toda la mama hasta los ganglios linfáticos u otras partes del cuerpo.

El Hospital Mayo Clinic (2018) señala:

El cáncer de mama más a menudo comienza en las células de los conductos que producen leche. El cáncer de mama también puede comenzar en el tejido glandular llamado lobulillos, en otras células o tejidos dentro de la mama.

Síntomas

Nódulo mamario o engrosamiento que se siente diferente de los tejidos circundantes.

Cambio de tamaño, forma o apariencia de la mama.

Cambios en la piel de la mama, como hoyuelos.

La inversión reciente del pezón.

Excoriación, descamación, formación de costras o desprendimiento del área de piel pigmentada que rodea el pezón (areola) o de la piel de la mama

Enrojecimiento o depresiones en la piel de la mama, como en la cáscara de una naranja. (p. 53).

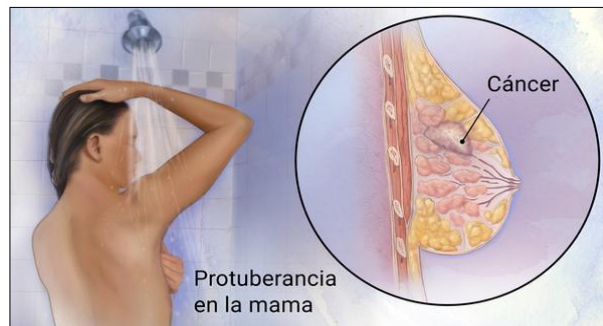


Figura 21. Ilustración del cáncer de mama. Fuente: Hospital Mayo Clinic., 2018.

2.2.14. Factores de riesgo

El Hospital Mayo (2018) afirma:

Un factor de riesgo de cáncer de mama es cualquier cosa que te hace más propensa a tener cáncer de mama. Pero el hecho de tener uno o incluso varios factores de riesgo de cáncer de mama no necesariamente significa que contraerás cáncer de mama. Muchas mujeres que contraen cáncer de mama no presentan otros factores de riesgo que el simple hecho de ser mujer. Los factores asociados a un mayor riesgo de contraer cáncer de mama son los siguientes:

- Ser mujer: Las mujeres son mucho más propensas que los hombres a contraer cáncer de mama.
- Edad avanzada: El riesgo de contraer cáncer de mama aumenta a medida que envejeces.
- Antecedentes personales de afecciones mamarias: Si te has sometido a una biopsia de mama en la cual se detectó un carcinoma lobulillar in situ o una hiperplasia atípica de mama, tienes alto riesgo de contraer cáncer de mama.
- Antecedentes personales de cáncer de mama: Si tuviste cáncer en una mama, tu riesgo de contraer cáncer en la otra mama es mayor.
- Antecedentes familiares de cáncer de mama: Si a tu madre, hermana o hija se le diagnosticó cáncer de mama, particularmente a temprana edad, tu riesgo de contraer cáncer de mama es mayor. De todas maneras, la mayoría de las personas a las que se les diagnostica cáncer de mama no tienen antecedentes familiares de la enfermedad.
- Genes hereditarios que aumentan el riesgo de contraer cáncer: Ciertas mutaciones genéticas que aumentan el riesgo de contraer cáncer de mama pueden transmitirse de padres a hijos. Las mutaciones genéticas más conocidas se conocen como «BRCA1» y «BRCA2». Estos genes pueden aumentar en gran medida tu riesgo de contraer cáncer de mama y otros tipos de cáncer, pero no logran que el cáncer sea inevitable.
- Exposición a la radiación: Si recibiste tratamientos con radiación en el tórax en la niñez o en las primeras etapas de la adultez, tus probabilidades de contraer cáncer de mama son mayores.
- Obesidad: Ser obeso aumenta el riesgo de contraer cáncer de mama.
- Tener tu primera menstruación a una edad temprana: Tener tu primera menstruación antes de los 12 años aumenta el riesgo de contraer cáncer de mama.
- Comenzar la menopausia a una edad mayor: Si en tu caso la menopausia comenzó a una edad mayor, tienes más probabilidades de contraer cáncer de mama.

- Tener tu primer hijo a una edad mayor: Las mujeres que tienen su primer hijo después de los 30 años pueden correr un riesgo mayor de contraer cáncer de mama.
- Nunca has estado embarazada: Las mujeres que nunca estuvieron embarazadas tienen un mayor riesgo de contraer cáncer de mama que las mujeres que tuvieron uno o más embarazos.
- Terapia hormonal posmenopáusica: Las mujeres que toman medicamentos de una terapia hormonal que combinan estrógeno y progesterona para tratar los signos y síntomas de la menopausia tienen un mayor riesgo de contraer cáncer de mama. El riesgo de contraer cáncer de mama disminuye cuando las mujeres dejan de tomar estos medicamentos.
- Consumir alcohol: El consumo de alcohol aumenta el riesgo de contraer cáncer de mama. (p. 117).

2.2.15. Estadios del cáncer de mama

BreastCancer (2015) afirma:

El estadio suele expresarse con un número entre 0 y IV. El estadio 0 corresponde al cáncer no invasivo que permanece en su ubicación original y el estadio IV al cáncer invasivo que se propaga fuera de la mama hacia otras partes del cuerpo.

- Estadio 0: El estadio 0 se utiliza para describir los casos de cáncer de mama que no se encuentran invasivos, como el carcinoma ductal in situ. En el estadio 0, no hay indicios de células cancerígenas o células anómalas no cancerosas que salen de la zona de la mama en la que se originaron o que toman o invaden tejidos normales circundantes.

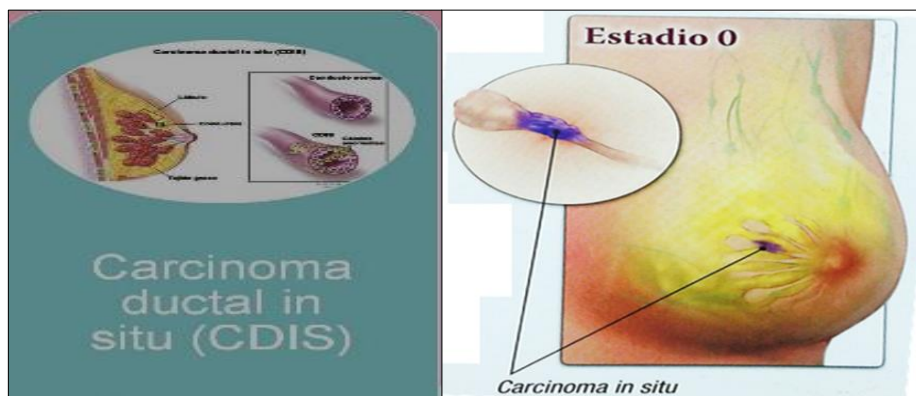


Figura 22. Estadio 0 del cáncer de mama. Fuente: *Hospital Mayo Clinic.*, 2018

- Estadio I: El estadio I describe el cáncer de mama invasivo (las células cancerígenas toman o invaden el tejido mamario normal que las rodea). El estadio I está dividido en subcategorías, conocidas como IA e IB.

El estadio IA describe el cáncer de mama invasivo en el cual el tumor mide hasta 2 cm. y el cáncer no se ha extendido más allá de la mama; no hay ganglios linfáticos afectados.

El estadio IB describe el cáncer de mama invasivo en el que:

No hay tumor en la mama; en cambio se observan en los ganglios linfáticos pequeños grupos de células cancerígenas superiores a 0,2 mm, pero inferiores a 2 mm. o

Se observa un tumor en la mama inferior a 2 cm y pequeños grupos de células cancerígenas superiores a 0,2 mm, pero inferiores a 2 mm en los ganglios linfáticos.

En el cáncer de mama de estadio I, es posible la invasión microscópica. Esto significa que las células cancerígenas recién han comenzado a invadir el tejido ubicado en las paredes del conducto o el lobulillo, pero estas células cancerígenas invasoras no miden más de 1 mm.

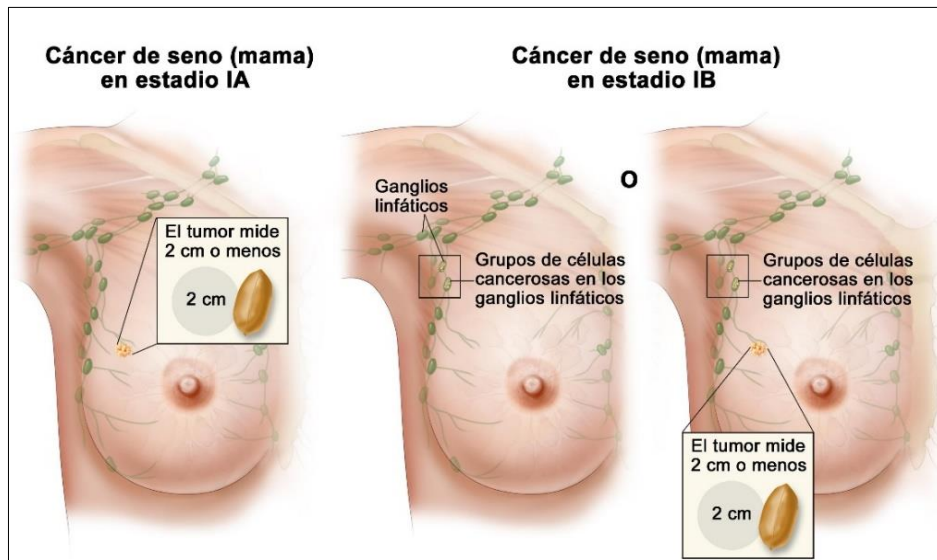


Figura 23. Estadio 1 del cáncer de mama. Fuente: *Hospital Mayo Clinic.*, 2018

- Estadio II: El estadio II se divide en las subcategorías IIA y IIB.

El estadio IIA describe el cáncer de mama invasivo en la cual:

El tumor mide 2 cm o menos, y se ha propagado hacia los ganglios linfáticos axilares o

El tumor mide entre 2 y 5 cm, y no se ha propagado hacia los ganglios linfáticos axilares.

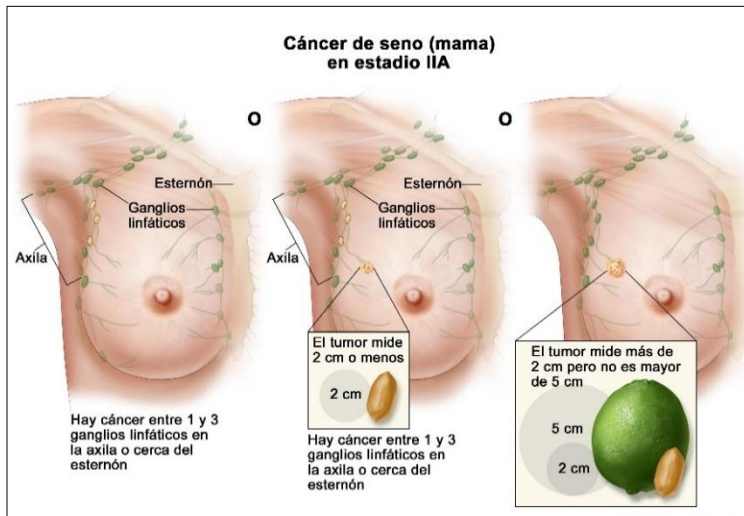


Figura 24. Estadio IIA del cáncer de mama. Fuente: *Hospital Mayo Clinic*, 2018

El estadio IIB describe el cáncer de mama invasivo en la cual:

El tumor mide entre 2 y 5 cm, y se observan en los ganglios linfáticos pequeños grupos de células cancerígenas superiores a 0,2 mm, pero inferiores a 2 mm o

El tumor mide entre 2 y 5 cm, y el cáncer se ha propagado en 1-3 ganglios linfáticos axilares o en los ganglios linfáticos cercanos al esternón (encontrado durante una biopsia del ganglio centinela) o

El tumor mide más de 5 cm, pero no se ha propagado hacia los ganglios linfáticos axilares.

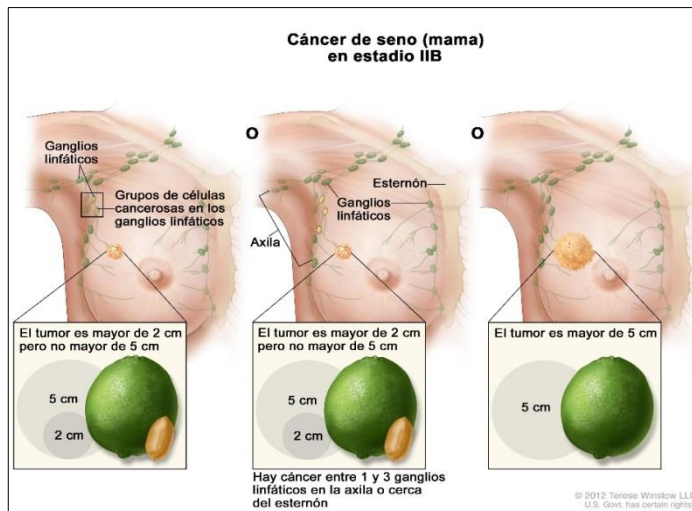


Figura 25. Estadio IIB del cáncer de mama. Fuente: *Hospital Mayo Clinic*, 2018

Estadio III: El estadio III se divide en las subcategorías IIIA, IIIB y IIIC.

El estadio IIIA describe el cáncer de mama invasivo en el que:

No hay tumor en la mama o el tumor puede tener cualquier tamaño, y se detectó la presencia de cáncer en 4-9 ganglios linfáticos axilares o en los ganglios linfáticos cercanos al esternón (encontrado durante estudios de imágenes o una exploración física) o

El tumor mide más de 5 cm, y se observan en los ganglios linfáticos pequeños grupos de células cancerígenas superiores a 0,2 mm, pero inferiores a 2 mm o

El tumor mide más de 5 cm, y el cáncer se ha propagado en 1-3 ganglios linfáticos axilares o en los ganglios linfáticos cercanos al esternón (encontrado durante una biopsia del ganglio centinela).

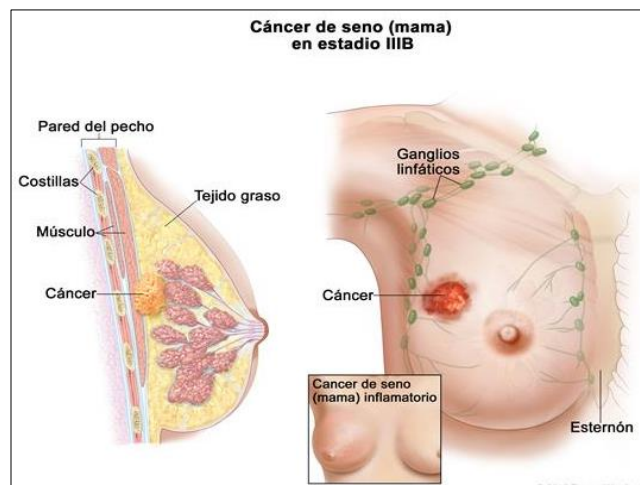


Figura 26. Estadio IIIA del cáncer de mama. Fuente: *Hospital Mayo Clinic.*, 2018

El estadio IIIB describe el cáncer de mama invasivo en el que:

El tumor tiene un tamaño indefinido y se ha propagado hacia la pared torácica o la piel de la mama, y ha provocado inflamación o una úlcera y pudo haberse propagado hacia 9 ganglios linfáticos axilares o Pudo haberse propagado hacia los ganglios linfáticos cercanos al esternón.

Nota: El cáncer de mama inflamatorio pertenece por lo menos al estadio IIIB. Las características típicas del cáncer de mama inflamatorio son:

Enrojecimiento de un gran sector de la piel de la mama.

La mama se siente caliente y puede estar inflamada.

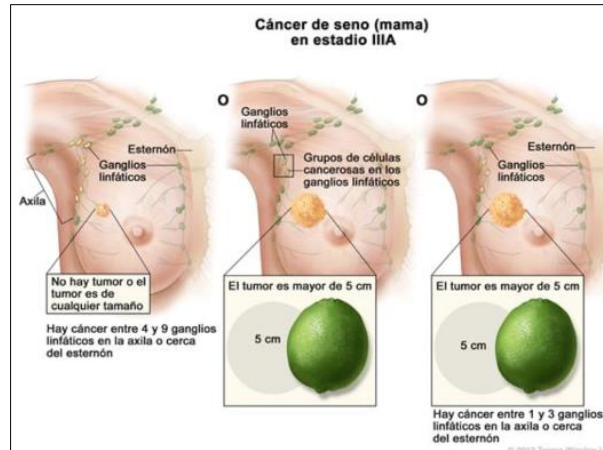


Figura 27. Estadio IIIB del cáncer de mama. Fuente: Hospital Mayo Clinic, 2018

El estadio IIIC describe el cáncer de mama invasivo en el que:

Puede que no haya indicios de la enfermedad en la mama o, si hay un tumor, puede tener cualquier tamaño y haberse propagado hacia la pared torácica o a la piel de la mama y el cáncer se ha propagado en 10 o más ganglios linfáticos axilares o el cáncer se ha propagado hacia los ganglios linfáticos ubicados sobre o debajo de la clavícula.

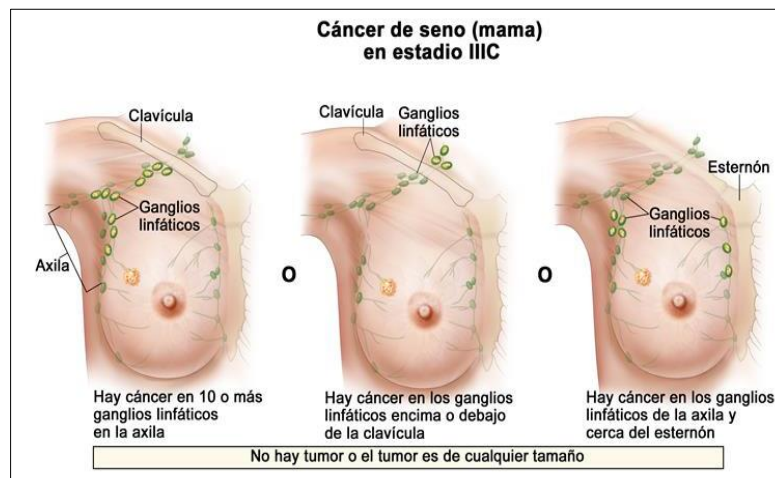


Figura 28. Estadio IIIC del cáncer de mama. Fuente: Hospital Mayo Clinic, 2018

- Estadio IV: El estadio IV describe el cáncer de mama invasivo que se ha propagado más allá de la mama y los ganglios linfáticos circundantes hacia otros órganos del cuerpo, como los pulmones, ganglios linfáticos distantes, la piel, los huesos, el hígado y el cerebro.

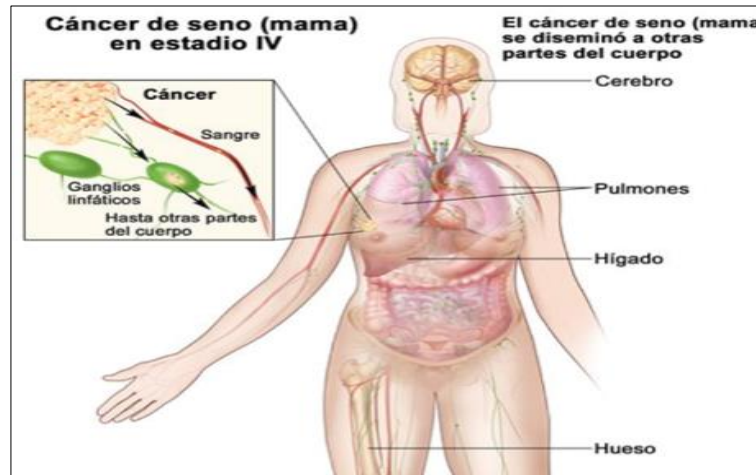


Figura 29. Estadio IV del cáncer de mama. Fuente: *Hospital Mayo Clinic*, 2018

2.2.16. Supervivencia frente al cáncer de mama

Sánchez (2017) señala:

Después de recibir un diagnóstico de cáncer, las prioridades de una persona en lo que respecta a las relaciones, la carrera o el estilo de vida pueden cambiar. Los sobrevivientes del cáncer, las personas con antecedentes de cáncer, a menudo dicen que valoran más la vida. También afirma que han logrado una mayor aceptación de sí mismas. Al mismo tiempo, algunos sobrevivientes también se vuelven ansiosos respecto de su salud. (p. 65).

Fernández (2017) afirma: “Algunos sobrevivientes no saben cómo sobrellevar su vida después del tratamiento, especialmente cuando ya no tienen que asistir a más visitas regulares con sus médicos” (p. 32).

En el Hospital de la Mujer (2014) señala:

Los médicos suelen utilizar las tasas de supervivencia para exponer en forma estándar el pronóstico de una persona. Es posible que algunas mujeres con cáncer de seno quieran conocer las estadísticas de supervivencia para mujeres en situaciones similares, mientras que para otras las cifras pueden no ser útiles e incluso pueden no querer conocerlas (p. 67).

A veces, las personas que han sobrevivido al cáncer consideran a sus amigos y a sus familiares “cosobrevivientes” debido a las experiencias que han tenido al cuidar de la persona con cáncer. Otras

personas con cáncer metastásico no creen que la etiqueta de “sobreviviente” es adecuada para ellos, ya que siguen viviendo con cáncer todos los días (p. 85).

El pronóstico para las mujeres con cáncer de mama varía según la etapa (extensión) del cáncer. En general, las tasas de supervivencia son mayores para las mujeres con cánceres en etapas más tempranas.

- a. La tasa relativa de supervivencia a 5 años para las mujeres con cáncer de mama en etapa 0 o etapa I es casi 100%.
- b. Para las mujeres con cáncer de mama en etapa II, la tasa relativa de supervivencia a 5 años es alrededor de 93%.

La tasa relativa de supervivencia a 5 años para los cánceres de mama en etapa III es aproximadamente 72%. A menudo, sin embargo, las mujeres con estos cánceres de mama pueden tratarse con éxito.

En Diario.es (2016) señala: “El cáncer de mama metaestásico o en etapa IV, tienen una tasa relativa de supervivencia a 5 años de aproximadamente 22%. Aun así, a menudo hay muchas opciones de tratamiento para las mujeres con esta etapa del cáncer de seno” (p. 2).



Figura 30. Gráfico de la tasa de supervivencia del cáncer de mama. Fuente: *Diario.es*, 2016.

2.2.17. Bases teóricas – científicas

León (2018) sostiene que:

El algoritmo J48 que es un clásico entre los árboles de decisión y que es muy fácil de interpretar. Además indica que presenta ventajas importantes sobre otras metodologías debido a su: mayor precisión, permisión para clasificar diversas características de la población y capacidad de identificar diversidad de fuentes que pueden ser tratados. (p. 22).

Loshin (2019) afirma que:

Un modelo predictivo se basa esencialmente en un conjunto de variables, cuyos valores se espera que influyan en las actividades futuras. Los modelos predictivos también se utilizan en muchas aplicaciones empresariales diferentes. Los bancos dependen de modelos que incluyen edad, estado civil, lugar de residencia, historial de crédito y otras variables para evaluar los riesgos asociados con las solicitudes de hipoteca por parte de los clientes potenciales. (p. 29).

CAPÍTULO III
DESARROLLO DE LA SOLUCIÓN

3.1. Estudio de factibilidad

3.1.1. Factibilidad técnica

Este Proyecto es viable ya que el Hospital de la Solidaridad de Villa El Salvador cuenta con los equipos necesarios lo cual facilitarán la implementación de un modelo predictivo para diagnosticar el cáncer de mama.

Hoy en día en el hospital no existen modelos predictivos para el diagnóstico de cáncer de mama lo que a nosotras nos permite poder crear cierto software para que las pacientes puedan detectar si se encuentran propensas a tener cáncer de mama basada a los síntomas que cada paciente presenta.

El Hospital de la Solidaridad de Villa El Salvador avanza con una eficiente y efectiva atención a sus respectivos pacientes. Afortunadamente esta entidad cuenta con equipos altamente sofisticados, donde permitirán la implementación del modelo predictivo a modo escritorio para que cada paciente sea atendida de manera segura y efectiva.

3.1.2. Factibilidad operativa

En el desarrollo de toda la investigación realizada a partir de la investigación preliminar, hemos encontrado apoyo en la Municipalidad de Villa El Salvador, el apoyo de nuestros padres y de los mismos pacientes dispuestos a someterse a dichas pruebas que obtendrán como beneficio una respuesta preliminar descartando a tiempo el cáncer de mama.

El doctor del Hospital de la Solidaridad de Villa El Salvador está de acuerdo de que se lleve a cabo la implementación del modelo predictivo, puesto que aseguran que es una gran herramienta para predecir el cáncer de mama disminuyendo así el error médico en las pacientes.

Este proyecto es viable, ya que se cuenta con los conocimientos en Minería de Datos, habilidades para el manejo de software con el que se pretende trabajar (WEKA) donde cada dato estará ubicado de acuerdo a su atributo, por ejemplo: paciente, edad, género, etc.

Por otro lado, permite al Hospital de la Solidaridad de Villa El Salvador la posibilidad de conseguir poner en marcha la nueva tecnología de Ingeniería, aprovechando los beneficios que ofrece, de igual manera, el correcto funcionamiento y uso del equipo. Así mismo estará sub-editado a la capacidad de los miembros encargados del área de oncología, razón por la cual se debe garantizar el pleno entendimiento y manejo del software en cuestión.

Una parte importante del proyecto lo constituye la planeación detallada a los procedimientos necesarios para implantar exitosamente la solución propuesta. A continuación, se describen las actividades que deben contemplarse para llevar a cabo el desarrollo del modelo predictivo para el área de oncología del Hospital de la Solidaridad de Villa El Salvador.

1. Firma del contrato.
2. Formación del personal
3. Adecuación de su instalación
4. Recopilación de datos
5. Recepción del equipo
6. Instalación del software
7. Pruebas de funcionalidad
8. Entrega del software al área de oncología en modo funcional
9. Puesta en marcha del software
10. Período de seguimiento

3.1.3. Factibilidad económica

Tabla 14

Presupuesto mensual

Presupuesto Mensual						
Concepto	Total de Meses	Precio Unitario (S/.)	Cantidad	Total por mes (S/.)	Total (S/.)	
1. Humanos						
Gamarra Condo, Cyntia	10 meses	1000 c/m	1	1000.00	10 000	
Santos Melo Yadira	10 meses	1000 c/m	1	1000.00	10 000	20 000
2. Materiales y Servicios						
Internet	10 meses	40 c/m	1	40	400	
Impresiones Cantidad Promedio	10 meses	0.25 por impresión	250 impresiones c/m	62.5	625	
Fotocopias Cantidad Promedio, libro, separatas, revista	10 meses	0.1 por fotocopia	300 fotocopias c/m	30	300	1725
Paquete de Hojas Bond tamaño A4	10 meses	15 por paquete	2 paquetes c/m	30	300	
Tinta de impresora (Negro y colores)	10 meses	50 por paquete tinta	2 paquetes por 10 meses	-	100	
3. Técnicos						
Movilidad (visitas al Hospital, bibliotecas y otras visitas que se requiera)	10 meses	120 c/cm	2 tesistas	240	2,400	2,400
4. Alimentos						
Almuerzos y refrigerios por 5 días	10 meses	300 c/m	2 tesistas	600	6000	6000
					MONTO TOTAL	30,125

Tabla 15

Presupuesto de costo único – total del proyecto

Presupuesto de costo único				
Concepto	Precio Unitario (S/.)	Cantidad	Total (S/.)	
Materiales				
USB 8GB	32.00	1	32.00	164.5
USB 4GB	24.00	1	24.00	
Útiles de escritorio (cantidad promedio)	108.5	-	108.5	
4 lápiz				
3 lapiceros c/negro (azul, negro, rojo)				
4 borradores				
2 reglas				
4 cuadernos				
2 cajas plumones gruesos y delgados				
2 cajas colores				
4 tajadores				
2 2 gommas grandes				
4 correctores				
5 paquetes manilas				
6 papelografos cuadriculados				
Equipos				
Impresora Canon MX 360	450	1	450	1724
Laptop	1.099	1	1,099	
Cámara	175	1	175	
Software				
Bizagi Modeler	-	1	Software Libre	
Weka	-	1	Software Libre	
			MONTO TOTAL	1,888.50
Sumatoria total del presupuesto				15,014

3.2. Fase 1 - comprensión del negocio

3.2.1. Determinar los objetivos del negocio

El objetivo de la minería de datos que se va a aplicar en esta presente tesis es el de hacer predicciones lo más fiables posible a partir de los datos de los que ya se disponen de los pacientes en el Hospital de la Solidaridad de Villa El Salvador.

En referencia a la situación de negocio del Hospital de la Solidaridad de Villa El Salvador se puede decir que se cuenta con una base de datos de los pacientes de dicho hospital donde hayan sido atendidos y diagnosticado el cáncer de mama mediante ciertos datos generales, características de características del paciente (síntomas) y medidas de clínica y diagnóstico.

Sin embargo, no existe ningún estudio en profundidad sobre el comportamiento de los estudiantes de los que se puedan sacar conclusiones o patrones para hacer predicciones sobre los futuros estudiantes.

✓ Target organization assessment

Hospital de la Solidaridad de Villa El Salvador

El Hospital de la Solidaridad de Villa El Salvador se encuentra ubicado en Av. Pastor Sevilla S/N Ovalo Cocharcas.

El Hospital de la Solidaridad de Villa El Salvador es una institución cuyo objetivo es brindar servicios integrales de salud a la población tanto varones, mujeres y niños(as), así como otros servicios y acciones que permitan mejorar la calidad de vida de las personas.

El diagnóstico clínico que brindan los médicos del Hospital Solidaridad de Villa El Salvador, es un proceso complejo que manifiesta experiencia, capacitación, reconocimiento de patrones y cálculo de probabilidad condicional.

Tabla 16

Hospital de la Solidaridad de Villa el Salvador

Hospital de la Solidaridad de Villa el Salvador	
Establecimiento	Hospital de la Solidaridad de Villa El Salvador
Clasificación	Policlínicos
Tipo	Sin Internamiento
Dirección	Av. Pastor Sevilla S/N Óvalo Cocharcas
Departamento	Lima
Provincia	Lima
Distrito	Villa El Salvador
Ubigeo	150142
Teléfono	01 292 – 3504

✓ **Visión del producto**

El desarrollo de este modelo predictivo está orientado a predecir o diagnosticar el cáncer de mama mediante la minería de datos, lo cual extrae patrones de los datos históricos y transaccionales para identificar riesgos y oportunidades. Por otro lado, nuestro modelo predictivo identifica relaciones entre diferentes factores que permiten valorar riesgos o probabilidades asociadas sobre la base de un conjunto de condiciones.

Nuestro modelo predictivo se crea usando datos de los pacientes del Hospital de la Solidaridad de Villa El Salvador (comportamiento observado) en un proceso denominado entrenamiento.

Con los datos del paciente (datos observados actualizados) se puede predecir el comportamiento esperado del paciente. Este proceso se denomina scoring.

Primero, se debe entrenar el sistema usando datos de nuestros sistemas de información que contengan el comportamiento observado de los pacientes del Hospital de la Solidaridad de Villa El Salvador.

Después, una vez al mes se realiza un proceso de estimación del comportamiento previsto del paciente usando patrones de uso del servicio actualizados a este momento.

Se presentará una tecnología organizada, a su vez dependerá de un hardware para su ejecución; será una herramienta que apoyará a los procesos de servicios que el Hospital de la Solidaridad brinda a los pacientes, y por ende mejorará la eficiencia en su servicio.

✓ **Lista To Do**

Tabla 17

Lista To Do

IN SCOPE	OUT SCOPE
Datos del paciente	Lugar de tratamiento
Síntomas del paciente	
Resultados	
Recomendaciones de alimentación	

✓ **Equipo**

El equipo de trabajo para llevar a cabo el desarrollo de software de nuestra tesis estará conformado por los siguientes miembros:

Tabla 18

Roles de equipo

Rol	Persona	Área
Director Médico	Orlando Merino Cisneros	Hospital de la Solidaridad de Villa El Salvador
Médico Oncológico	Eduardo Salazar Campos	Oncología
Tesistas	Gamarra Condo, Cyntia Santos Melo, Yadira	Mamografía

✓ **Organigrama estructural**

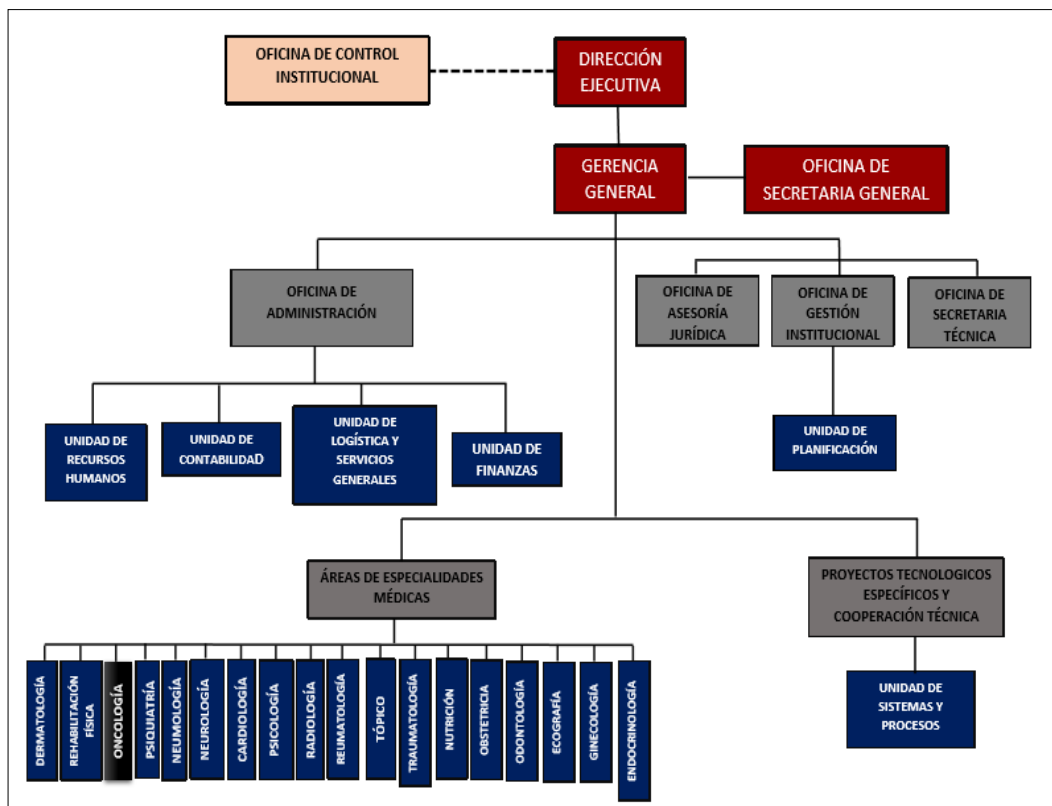


Figura 31. Organigrama del Hospital de la Solidaridad de Villa el Salvador

✓ **Servicios**

El Hospital de la Solidaridad de Villa El Salvador ofrece los siguientes servicios: Unidades de atención medica servicios de diagnóstico, cobertura médica, chequeos médicos, otros servicios. Por la cual cada uno de estos servicios tiene diferentes áreas que se describen a continuación:

a) Unidades de atención médica

- Hospital del día. - Brinda a usted y a su familia servicios de atención clínica y quirúrgica, de lunes a sábado para aquellos pacientes que no tienen tiempo para hospitalizarse y que regresan a casa el mismo día.
- Emergencia. - Promete atención las 24 horas del día durante los 365 días del año, a población adulta y pediátrica, con diferentes niveles de gravedad, incluyendo la atención a pacientes críticos, quemados, con convulsiones, etc.
- Cuidados Intensivos. - Atiende las 24 horas al día pacientes críticamente enfermos o con patologías que exigen una atención médica especializada y el uso de monitoreo continuo de su condición clínica hacia los pacientes.

b) Unidades de atención médica

- Laboratorios Clínicos. - Cuenta con el aval del Colegio Americano de Patología, cubre los requerimientos de las diferentes áreas de: hematología, química, hormonal, gastroenterología, inmunología, genética, microbiología, fertilidad y 9 marcadores tumorales.

c) Cobertura médica

- Medicina Interna. - Es la atención integral del paciente que presenta problemas clínicos en las diferentes especialidades como: Dermatología, rehabilitación física, oncología, psiquiatría, neurología, cardiología, psicología, tóxico, radiología, traumatología, nutrición, obstetricia, pediatría, ecografía, endocrinología, ginecología, medicina general, neumología.
- Cirugía. - se realizan cirugías programadas y de emergencia de mínima y gran complejidad como son: Otorrinolaringología, urología, oftalmología.

d) Chequeos médicos

Los planes de chequeos médicos se realizan para detectar a tiempo, antes de que se presenten los síntomas, cualquier alteración en el organismo del paciente y actuar en consecuencia para corregirla. Principalmente radica en los chequeos que implican un examen global de la persona con consultas médicas especializadas, rigurosos análisis de laboratorio y de diagnóstico por imágenes. De esta manera los chequeos médicos permiten conocer tempranamente varias alteraciones del organismo del paciente y aumentar las probabilidades de curar cualquier enfermedad 12 con un menor costo.

e) Otros servicios

- Farmacia.- Amplio rango de stock en medicamentos, y soporte profesional a través de profesionales farmacéuticas; Abierta desde las 9H00 hasta las 20H00.
- Kioskos.- Ofrece variados componentes de alimentos saludables para atender las necesidades de los visitantes. Abierto desde las 10H30 hasta las 18H30.

✓ **Clientes**

El Hospital de la Solidaridad de Villa El Salvador tiene segmentado el mercado de sus clientes externos siendo esta una Oportunidad, ya que conoce el mercado de sus clientes, con el fin de atender de manera eficiente y de acuerdo a las necesidades que se les presenta a cada tipo de paciente.

A parte de los clientes externos existen los internos que son todos aquellos que forman parte del Hospital de la Solidaridad en sus diferentes áreas como son: Director general, personal médico, enfermeros (as), personal administrativo, técnica enfermera, técnicos farmacéuticos, auxiliar de enfermería, técnicos de laboratorio, personal de limpieza, documentalistas, personal de seguridad, personal asistencial.

✓ **Stakeholders internos**

- ❖ **Director General:** El director de un hospital es aquel que combina un amplio e impecable currículo médico, así como una cualificada formación administrativa, que le avalan como máxima autoridad del centro sanitario.
- ❖ **Enfermeros(as):** Enfermera es aquella mujer que se dedica al cuidado personal e intensivo de un paciente, ya sea en el centro hospitalario o de salud como en su domicilio particular.
- ❖ **Técnica Enfermera:** Un técnico en enfermería es el encargado de la prevención y promoción de la salud mediante la preservación de las condiciones de higiene y seguridad bajo la supervisión del médico y/o enfermera.
- ❖ **Técnicos Farmacéuticos:** El técnico en farmacia es el profesional de la salud que posee formación académica Técnico - Científica y humanista que le permite estar capacitado para prestar apoyo eficiente al químico farmacéutico, en todas las

actividades propias del ámbito de la farmacia.

- ❖ **Auxiliar de enfermería:** El auxiliar de enfermería es el profesional sanitario encargado de proporcionar cuidados auxiliares al paciente y actuar sobre las condiciones sanitarias de su entorno bajo la supervisión del diplomado en enfermería o el facultativo médico.

- ❖ **Técnicos de laboratorio:** Los técnicos de laboratorio médico realizan pruebas de laboratorio rutinarias para el diagnóstico, tratamiento y prevención de la enfermedad. Pueden trabajar bajo la supervisión de un tecnólogo médico.

- ❖ **Doctor Oncológico:** Médico que se encarga de diagnosticar y tratar el cáncer en adultos mediante quimioterapia, terapia con hormonas, terapia biológica y terapia dirigida. A menudo, es el proveedor principal de atención de la salud de alguien que padece de cáncer, brinda cuidados médicos de apoyo y puede coordinar el tratamiento administrado por otros especialistas.

- ❖ **Personal de limpieza:** Las tareas que realiza el personal de limpieza en un centro de atención a personas en situación de dependencia son las propias de la limpieza como barrer, fregar, limpiar el polvo y sacar la basura, en habitaciones y zonas comunes (camas, baños, ventanales, balcones, mobiliario, etc.)

- ❖ **Documentalistas:** Los documentalistas son aquellas personas que se dedican a recoger, preparar y organizar toda clase de datos bibliográficos, informaciones y noticias sobre una determinada materia.

- ❖ **Personal de Seguridad:** Vela por la seguridad, primordialmente en relación a las personas del Hospital de la Solidaridad de Villa El Salvador.

- ❖ **Personal asistencial:** Ayudan a los individuos con discapacidades físicas,

enfermedades mentales, y otras necesidades de atención médica con sus actividades de vida diaria y les proporcionan cuidados de cabecera — incluso procedimientos básicos de enfermería.

✓ **Stakeholders externos**

❖ **Competidores:**

La competencia del Hospital de la Solidaridad radica en la categoría de los Hospitales de nivel III existentes en el distrito de Villa El Salvador y sus alrededores. Por otra parte el Hospital de la Solidaridad de Villa El Salvador brinda servicios médicos principalmente a los asegurados y a personas particulares que requieran de los servicios. Tomando en cuenta la información anterior y la existencia de entidades de salud tanto públicas como privadas que brindan los mismos servicios es importante reflejar la participación en el mercado del hospital del distrito de Villa El Salvador:

En cuanto al sector público existen hospitales que por sus capacidades pueden competir con el Hospital de la Solidaridad de Villa El Salvador, a continuación, se presenta una lista:

- Hospital Villa Salud
- Hospital EsSalud
- Hospital Juan Pablo II
- Nuevo Hospital de Emergencia de Villa El Salvador.

Con estos antecedentes y debido a la diversidad de clientes y el creciente demanda del mercado, es poco probable que los pacientes muestren fidelidad a una marca, más aun con la estandarización de la oferta y la importancia del factor precio (tarifario médico) al momento de definir la contratación del servicio, es por eso que

se convierte en un factor crucial la calidad del servicio a fin de conseguir de esta forma una ventaja competitiva que permita ser líderes en el mercado, y así disminuir en algo la Amenaza que el precio y la falta de fidelidad representa.

❖ Proveedores

En el Hospital de la Solidaridad de Villa El Salvador los proveedores internos son la bodega general de suministros y la farmacia que proveen de suministros y fármacos a todas las áreas; y el hospital como ente principal provee de equipos y área física. A continuación, se describe las funciones de cada uno de los proveedores internos:

- Bodega General de Insumos: Área operativa del hospital encargada de solventar todas las necesidades de equipo médico, instrumental, insumos, materiales y equipo de oficina, muebles y enseres.
- Farmacia: Área operativa del hospital encargada de solventar las necesidades de medicamentos para las recetas de los tratamientos clínicos ambulatorios y de hospitalización.
- Municipalidad de Villa El Salvador: Entidad encargada de solventar todas las necesidades referentes al personal administrativo y médico requerido; además es el facultado del proceso de selección de proveedores de equipo médico, insumos, materiales y fármacos a través del departamento de materiales.

Los proveedores externos son todas aquellas empresas, sociedades, fundaciones y personas naturales que ofertan productos y servicios al hospital.

Tabla 19

Lista de principales proveedores de suministro

Nro	Proveedor	Tipo de producto
1	Alem	Insumos hospitalarios
2	Aventis pharma	Medicamentos
3	Braun	Equipo médico e insumos hospitalarios
4	Farmaventas	Medicamentos
5	Labqualyti	Reactivos y equipos de laboratorios
6	Life	Insumos hospitalarios y medicamentos
7	Roche	Reactivos de laboratorio, medicamentos y equipos
8	Viennatone c.a.	Equipo médico e insumos hospitalarios

Tabla 20

Lista de principales proveedores de servicio

Nro	Proveedor	Tipo de servicio
1	Arquitecto Núñez	Construcción
2	Datapro	Copiado y anillado
3	E.o.a	Mantenimientos de Equipos
4	Emseon	Seguridad Privada
5	Fundación natura	Manejo de desechos Contaminados
6	Imecanic	Mantenimiento de instalaciones
7	Telconet	Internet
8	Truly nolen	Control de Plagas

Todos los productos y servicios brindados por los proveedores son constantemente monitoreados por un proceso de control de calidad (Oportunidad), el cual es muy estricto no sólo por la imagen que brindamos, sino por el compromiso que tiene el Hospital de la Solidaridad de Villa El Salvador.

❖ Organismos gubernamentales

Las instituciones que son organismos de gubernamentales para el Hospital de la Solidaridad de Villa El Salvador son:

- La superintendencia de compañías
- El servicio de rentas internas
- El ministerio de salud
- Superintendencia de bancos
- Ministerio de trabajo

La empresa debe regirse de acuerdo a las leyes, estatutos y ordenanzas que estas instituciones dictan y emiten permanentemente. Estas instituciones controlan y establecen los parámetros para el funcionamiento de los hospitales.

✓ Cadena de valor

Actividades de Soporte	Infraestructura Organizacional	Contabilidad	Libros contables			Evaluación de balance	Control de ingreso y egresos
		Finanzas	Administración de los fondos del Hospital	Administración de presupuesto y tesorería	Elaboración de asignación financiera del estado	Evaluación de estado financiero	Balance financiero
		Administración	Gestión de los Recursos		Planificación Estratégica	Comprobar cumplimiento de pagos	
		Logística	Recepción y control de materiales		Conformidad de Stock	Almacenamiento de materiales	
	Administración de Recursos Humanos	Contratación del Personal	Remuneración para el personal			Capacitación	Elaboración de planilla
	Desarrollo Tecnológico	Sistemas informáticos	Mejora de infraestructura de equipos		Innovación y desarrollo	Administración de sistemas estadísticos	Gestión de requerimientos
	Administración del Abastecimiento	Adquisición de materiales tecnológicas (compras)	Planificación de compras		Adquisición de medicamentos	Transporte	Almacenamiento de equipamiento
MARGEN							
Actividades Primarias	Logística de Entrada (Bodega interna)	Operación: ATENCIÓN CLÍNICO	Logística de Salida (Despacho y entrega)	VENTAS	Aspectos comerciales (publicidades)	Servicios (calidad de atención)	
	Requerimientos de información y conocimientos de los pacientes	Atención de informaión hacia los pacientes	Pacientes informados		Página Web, Área presencial	Trámite de quejas y reclamos	
						Implementación de mejora	
	Requerimientos de los pacientes	Atención de salud	Paciente atendido o tratado	Venta del servicio	Promoción de las campañas ofrecidos por el hospital	Generación de espacios de participación de pacientes.	
				Registro del servicio		Seguimiento al paciente tratado	Apoyo en trámite sociales
				Emisión de boletas o facturas			
Verificación de materiales	Transformación del producto o insumo	Verificación de materiales del producto final	Venta de productos	Servicio completo (SIEMPRE HAY STOCK)	Seguimiento del producto vendido		
Almacenamiento o gestiona de devoluciones		Almacenamiento o gestiona de devoluciones del producto final	Registro del producto				
Control de inventario		Control de inventario del producto final	Emisión de boletas o facturas				

Figura 32. Cadena de valor del área de oncología del Hospital de la Solidaridad.

✓ **Procesos de negocios**

Tabla 21

Procesos de negocio

Inputs	Procesos	Outputs
Recursos	Elaboración de asignación financiera del estado.	
Computadora	Control de ingreso y egresos	
Manual de usuario		Servicio
Software	Remuneración para el personal	
Memoria USB	Capacitación	Paciente atendido
Documentos		Solución a la necesidad que el Hospital presenta
	Innovación y desarrollo	
Boleta de venta		Reportes
Factura	Administración de sistemas estadísticos	
Inventario		Personal capacitado
Formato de atención	Planificación de compras	
		Análisis de información
Información	Adquisición de materiales tecnológicas (compras)	
		Orden de compra
Base de Datos	Adquisición de medicamentos (compras)	
Información del paciente	Atención de salud	
		Paciente atendido o tratado

✓ **Matriz de selección de proceso**

Tabla 22

Matriz de selección de proceso

Matriz de selección de proceso					
Rango permitido	Calidad			Sostenibilidad financiera	Total de priorización (multiplicación de criterios) 1*2*3*4
	Satisfacción del cliente interno	Satisfacción del cliente externo	Calidad de atención		
Elaboración de asignación financiera del estado	3	3	1	3	27
Control de ingresos egresos	1	1	3	1	3
Remuneración para el personal	3	1	1	3	9
Capacitación	1	5	1	3	15
Innovación y desarrollo	3	1	3	3	27
ADMINISTRACIÓN DE SISTEMAS ESTADÍSTICOS	5	1	3	1	15
Planificación de compras	1	3	3	3	27
ADQUISICIÓN DE MATERIALES TECNOLÓGICAS	1	3	5	3	45
ADQUISICIÓN DE MEDICAMENTOS	1	3	5	3	45
ATENCIÓN DE SALUD	5	3	5	1	75
Paciente atendido o tratado	1	1	5	5	15

✓ **Modelado del proceso**

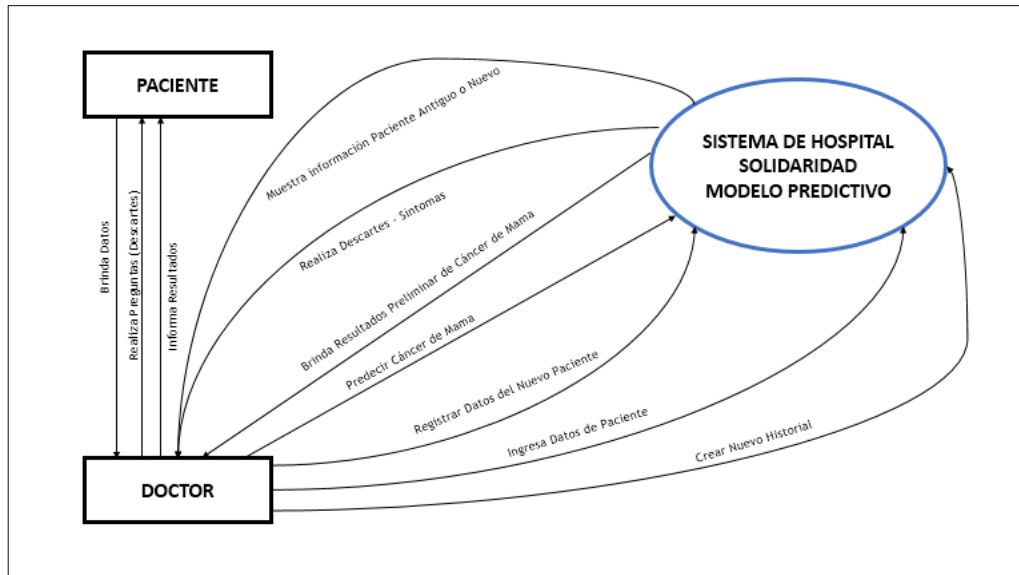


Figura 33. Modelo de contexto.

✓ **Diagrama de descomposición de funciones**

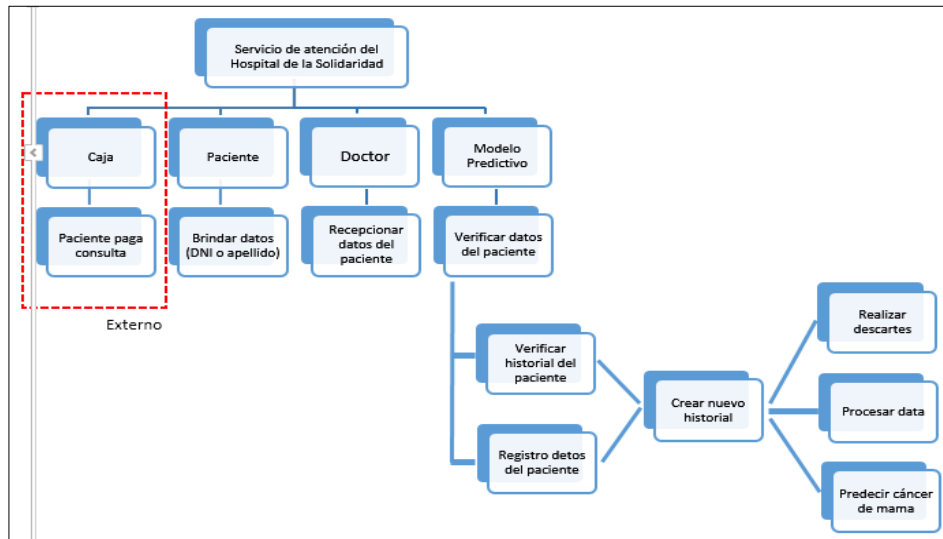


Figura 34. Diagrama de descomposición de funciones.

✓ **Flujograma de actividades**

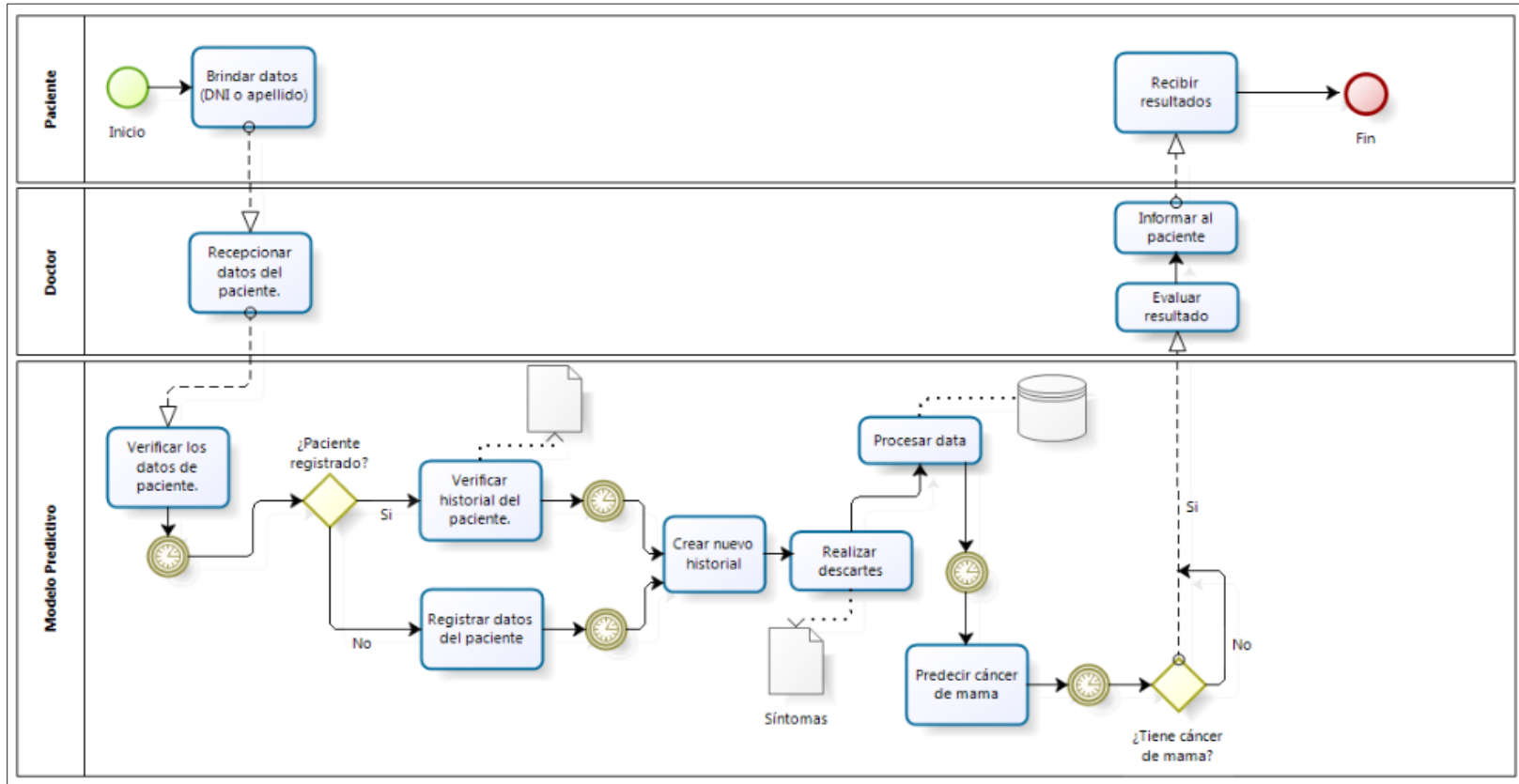


Figura 35. Flujograma de actividades del Hospital de la Solidaridad de Villa el Salvador.

✓ **Objetivos del negocio**

El Hospital de la Solidaridad de Villa El Salvador requiere diagnosticar el cáncer de mama en las pacientes de dicho hospital, a su vez desea reducir el error médico para tener una mayor eficiencia; para ello se crea el modelo predictivo usando minería de datos, el cual será trabajado en la herramienta Weka (Software libre), la cual permite la experimentación de análisis de datos mediante la aplicación, análisis y evaluación.

Se aplicará la metodología CRISP – DM, lo cual permitirá obtener los datos reales de la paciente, es decir, explorar grandes bases de datos de manera automática con el objetivo de encontrar patrones de acuerdo a los síntomas de cada paciente, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto.

Los objetivos del negocio como ya se ha mencionado son la predicción de datos para las pacientes ya existentes del Hospital de la Solidaridad de Villa El Salvador y a su vez para los nuevos pacientes que van a ser atendidos en dicho hospital.

Se han definido el mayor objetivo:

- Hacer predicciones acerca de los síntomas que presentan las pacientes del hospital para comprobar si son o no propensas a tener cáncer de mama.

Se ha considera la herramienta Weka por las siguientes razones:

- Es una herramienta que es fácil de utilizar, ya que dispone de una interfaz gráfica de usuario.
- Weka tiene implementados varios algoritmos, con los cuales se pueden realizar tareas de minería de datos, permitiendo el pre procesamiento de datos.
- Permite realizar manipulaciones en los datos aplicando filtros ya sea a los atributos o a las instancias de los mismos, entre algunos de los filtros que nos permite aplicar Weka están los filtros remove y discretize, que eliminan atributos y discretizan atributos numéricos, respectivamente.
- Weka además, proporciona información adicional, que ayudan a evaluar el nivel de

efectividad del modelo predictivo obtenido.

✓ **Criterios de éxito del negocio**

Se establece los siguientes criterios de éxito:

- Realizar predicciones de nuevos pacientes que serán atendidos en el Hospital de la Solidaridad de Villa El Salvador (Sistema).
- Disminuir un gran porcentaje de grado cáncer de mama que puede tener el paciente.
- Mejorar el servicio de atención de cada paciente del hospital.

3.2.2. Valoración de la situación

Se cuenta con una base de datos .XLS con información detallada de las pacientes que han sido atendidas en el Hospital de la Solidaridad de Villa El Salvador, por la cual se puede disponer con dicha información relevante para dar solución a nuestro problema planteado. Por otro lado nuestra información incluye los datos personales, el historial clínico y los síntomas del paciente para poder realizar nuestra minería de datos.

Tabla 23
Inventario de recursos

		Descripción
SW	WEKA	<p>Contiene una colección de algoritmos para realizar un análisis de datos y un modelado predictivo, además tiene herramientas para la visualización de estos datos. Por otro lado provee una interfaz gráfica que unifica las herramientas para que estén a una mejor disposición.</p>
	Laptop 1	<p>Windows 8.1 Pro</p> <ul style="list-style-type: none"> - Procesador: Intel® Corel TM i3-2350M CPU @ 2.30 GHz 2.30 GHz. - Memoria instalada (RAM): 4.00 GB. - Tipo de sistema: Sistema operativo de 64 bits, procesador x64.
HW	Laptop 2	<p>Windows 10 Pro</p> <ul style="list-style-type: none"> - Procesador: Intel® Corel TM i3-2350M CPU @ 2.30 GHz 2.30 GHz. - Memoria instalada (RAM): 4.00 GB. - Tipo de sistema: Sistema operativo de 64 bits, procesador x64.
	Computadora	<p>Windows 7 Ultimate</p> <ul style="list-style-type: none"> - Procesador: Intel® Corel TM i3-2350M CPU @ 2.30 GHz 2.30 GHz. - Memoria instalada (RAM): 4.00 GB. - Tipo de sistema: Sistema operativo de 64 bits, procesador x64.

✓ **Requisitos, supuestos y restricciones**

- Requisitos: base de datos de los datos personales de las pacientes del Hospital de la Solidaridad De Villa El Salvador. síntomas sobre la enfermedad del cáncer de mama.
- Supuestos: Datos característicos de las pacientes de la solidaridad de Villa El Salvador.
- Restricciones: Al no poder utilizar los datos personales de las pacientes de dicho hospital debido a un tema legal y de confidencialidad, para ello se ha tenido que utilizar la base de datos ficticia de las características de las pacientes, es decir, datos no reales.

✓ **Terminología**

- Análisis prospectivo de datos: El análisis de datos predice futuras tendencias, comportamientos o eventos basado en datos históricos.
- Análisis exploratorio de datos: El uso de técnicas estadísticas tanto gráficas como descriptivas para aprender acerca de la estructura de un conjunto de datos.
- Árbol de decisión: Es la estructura en forma de árbol que representa un conjunto de decisiones. Estas decisiones generan reglas para clasificar un conjunto de datos.
- Clasificación: Proceso de dividir un conjunto de datos en grupos mutuamente excluyentes de tal manera que cada miembro de un grupo esté lo "más cercano" posible a otro, y grupos diferentes estén lo "más lejos" posible uno del otro, donde la distancia está medida con respecto a variable(s) específica(s) las cuales se están tratando de predecir.
- Análisis estadístico: es un componente del análisis de datos. En el contexto de la inteligencia de negocios (BI), el análisis estadístico requiere recoger y escudriñar cada muestra de datos individual en una serie de artículos desde los cuales se puede extraer las muestras.

- Análisis retrospectivo de datos: Análisis de datos que provee una visión de las tendencias, comportamientos o eventos basado en datos históricos.
- Clustering (agrupamiento): Proceso de dividir un conjunto de datos en grupos mutuamente excluyentes de tal manera que cada miembro de un grupo esté lo "más cercano" posible a otro, y grupos diferentes estén lo "más lejos" posible uno del otro, donde la distancia está medida con respecto a todas las variables disponibles.
- Data Mining: La extracción de información predecible escondida en grandes bases de datos.
- Dimensión: En una base de datos relacional o plana, cada campo en un registro representa una dimensión. En una base de datos multidimensional, una dimensión es un conjunto de entidades similares; por ejemplo una base de datos multidimensional de ventas podría incluir las dimensiones producto, tiempo y ciudad.
- Modelo analítico: Una estructura y proceso para analizar un conjunto de datos. Por ejemplo, un árbol de decisión es un modelo para la clasificación de un conjunto de datos.
- Modelo lineal: Un modelo analítico que asume relaciones lineales entre una variable seleccionada (dependiente) y sus predictores (variables independientes).
- Modelo no lineal: Un modelo analítico que no asume una relación lineal en los coeficientes de las variables que son estudiadas.
- Modelo predictivo: Estructura y proceso para predecir valores de variables especificadas en un conjunto de datos.
- Navegación de datos: Proceso de visualizar diferentes dimensiones, "fetas" y niveles de una base de datos multidimensional.

✓ **Costes y beneficios**

- Costes: Algunos datos de nuestras presentes tesis no requieren ningún coste adicional al hospital ya que estos datos pertenecen al propio hospital.
- Beneficios: En cuanto a beneficios, no se puede decir que este proyecto genere algún beneficio económico para el hospital directamente, pero sí puede suponerlo indirectamente ya que el objetivo de este proyecto es mejorar la calidad de los servicios ofrecidos a los pacientes por parte del hospital, y por tanto la satisfacción de los pacientes, y esto se traduce en prestigio para el hospital, lo cual hará que más pacientes consideren atenderse en dicho hospital.

3.2.3. Determinar los objetivos de la minería de datos

✓ **Metas de Data Mining**

- Determinar el perfil de los pacientes respecto a las características y síntomas que presenta.
- Predecir si la paciente tiene cáncer de mama mediante los síntomas o signos que presenta en la actualidad.
- Describir de clases lo cual provee una clasificación concisa y resumida de un conjunto de datos para poder distinguirlos con otros.
- Extraer información interesante/útil a partir de datos. En muchos casos esta información tiene la forma de patrones no evidentes en los datos y, dependiendo de su complejidad, puede llegarse a hablar de conocimiento.
- Correlacionar los datos adquiridos para posteriormente realizar una limpieza masiva en la cual sea de gran utilidad para poder predecir.

✓ Criterios de éxito

Realizar predicciones de los pacientes tanto antiguos como nuevos del Hospital de la Solidaridad de Villa El Salvador con un alto valor de fiabilidad.

3.2.4. Realizar el plan de trabajo

Tabla 24

Plan de trabajo

Etapas	Descripción	Tiempo estimado
Etapa 1	Análisis de la estructura de los datos y la información de la base de datos.	2 semanas
Etapa 2	Ejecución de consultas para tener muestras representativas de los datos.	1 semana.
Etapa 3	Preparación de los datos (selección, limpieza, conversión y formateo, si fuera necesario) para facilitar la minería de datos sobre ellos.	3 semanas.
Etapa 4	Elección de las técnicas de modelado y ejecución de las mismas sobre los datos.	1 semana.
Etapa 5	Análisis de los resultados obtenidos en la etapa anterior, si fuera necesario repetir la etapa	1 semana.
Etapa 6	Producción de informes con los resultados obtenidos en función de los objetivos de negocio y los criterios de éxito establecidos.	1 semana.
Etapa 7	Presentación de los resultados finales.	1 semana.

3.3. Fase 2- Comprensión de los datos

En esta segunda etapa de la metodología CRISP-DM se realizará la recolección inicial de los datos recolectados para poder establecer un primer contacto con el problema detectado, familiarizarse con los datos recolectados y averiguar la calidad de información, así como identificar las relaciones más evidentes para formular las primeras hipótesis.

3.3.1. Recolectar los datos iniciales

✓ Reporte de recolección de datos

Los datos utilizados en nuestra presente tesis son datos personales referentes a las pacientes del Hospital de la Solidaridad de Villa El Salvador, entre ellos tenemos: apellido paterno, apellido materno, primer nombre segundo nombre (Opcional), dni, sexo, fecha de nacimiento, dirección, teléfono lo cual posee cierta información relevante que se ajustan a las características que desde un origen se necesita.

Aclarando que la base de datos de la información personal de los pacientes, se obtuvo del mismo hospital, mientras que sus características y síntomas se obtuvieron de datos ficticios, puesto que por un tema de confidencialidad e ilegalidad el hospital no nos pudo facilitar cierta información.

Las tablas a considerar de acuerdo a la información obtenida es la siguiente:

Especialidad, paciente, síntoma, triaje, usuarios

A continuación, un breve detalle de las descripciones de cada tabla:

Tabla 25

Reporte de datos

	Descripción
Pacientes	Cada paciente está identificado por su número de DNI o carnet de extranjería.
Síntomas	Cada síntoma está identificado por alfanumérico.
Especialidades	Son áreas específicas que maneja el Hospital de la Solidaridad de Villa El Salvador.
Usuarios (Doctores)	Profesional capacitado que practica la medicina e intenta mantener y recuperar la salud humana mediante el estudio, el diagnóstico y el tratamiento de la enfermedad o lesión del paciente, quienes realizaran el uso del Sistema para sus funciones laborables.
Triaje	Proceso que se encarga de tomar datos principales de los pacientes (talla, peso, presión arterial) para luego clasificarlos de acuerdo a la urgencia de la atención.

3.3.2. Descripción de los datos

Las características de la base de datos fueron creadas por los autores de la presente tesis puesto que por un tema de confidencialidad e ilegalidad el hospital no nos pudo facilitar cierta información. Cabe mencionar que los siguientes datos a detallar se están tomando en cuenta para llevar a cabo el desarrollo de nuestro sistema de modelo predictivo para el diagnóstico cáncer de mama.

Tabla 26

Descripción de los atributos

Tablas	Descripción de atributos	Tipos de datos
Especialidad	<p>Se podrá verificar, crear las distintas áreas que existen en el Hospital Solidaridad de Villa El Salvador, las cuales se consideró 3 atributos:</p> <ul style="list-style-type: none"> - Id_Especialidad. - Nombre de especialidad. - Descripción de especialidad. 	<p>Id: Int Nombre: Varchar Descripción: Varchar Ar</p>
Paciente	<p>En esta tabla se podrá crear, modificar, eliminar, guardar pacientes, las cuales se consideró 10 atributos:</p> <ul style="list-style-type: none"> -Id_Paciente -Nombre -Apellido Paterno -Apellido Materno -Dni -Fecha de Nacimiento -Estado Civil -Sexo -Teléfono -Dirección 	<p>Id_Paciente: Int Nombre: Varchar Apellido Paterno: Varchar Apellido Materno: Varchar Dni: Varchar Fecha de Nacimiento: Date Estado Civil: Char Sexo: Char Teléfono: Varchar Dirección: Varchar</p>
Síntomas	<p>Se consideró todos los posibles síntomas relacionados al cáncer de mama, la cual estarán como preguntas cerradas en el sistema. Así mismo se consideró 17 atributos:</p> <ul style="list-style-type: none"> -Id -IdTriage -Diagnostico -Grado -FechaHoraRegistro -S001 al S012 (Refiere la cantidad de síntomas consideradas, el detalle lo podrá verificar en la parte 3.2.4 del documento.) 	<p>Id: Int IdTriage: Int Diagnóstico: Int Grado: Int FechaHoraRegistro: DateTime S001 al S012: Int</p>
Triage	<p>Se podrá realizar nuevo triaje, verificar los antiguos, al igual que se podrá eliminar; la cual se consideró 9 atributos:</p> <ul style="list-style-type: none"> -Id -IdPaciente -IdEspecialidad -Peso -Talla -Temperatura -Presión -Descripción -FechaHoraRegistro 	<p>Id: Int IdPaciente: Int IdEspecialidad: Int Peso: Varchar Talla: Varchar Temperatura: Varchar</p> <p>Presión: Varchar Descripción: Varchar FechaHoraRegistro: DateTime</p>

Usuario	<p>En esta tabla se podrá crear, editar, eliminar los usuarios quienes harán uso del sistema. Se consideró 6 atributos:</p> <ul style="list-style-type: none"> -Id -Nombre -ApellidoPaterno -ApellidoMaterno -Usuario -Password 	<p>Id: Int Nombre: Varchar ApellidoPaterno: Varchar ApellidoMaterno: Varchar Usuario: Varchar Password: Varchar</p>
---------	---	--

3.3.3. Exploración de los datos

En esta fase de la metodología Crisp-Dm se podrá apreciar las primeras conclusiones de nuestra tesis con el Modelo Predictivo para el diagnóstico de cáncer de mama de las pacientes del hospital de Villa El Salvador, obteniendo como primeros resultados positivos porcentajes de asertividad del pronóstico de cáncer de mama de los datos almacenados de las pacientes; utilizando la herramienta Weka como se muestra en las siguientes figuras.

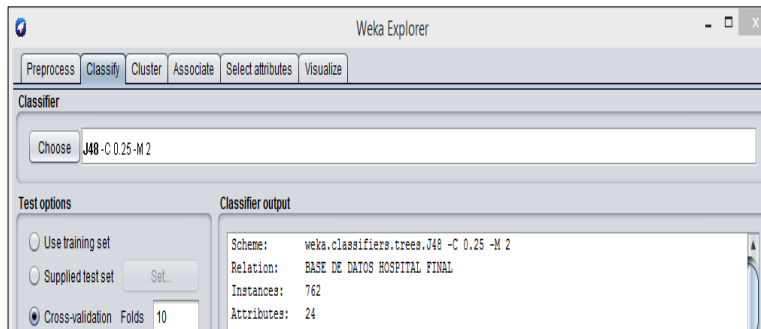


Figura 36. Exploración de los datos parte I

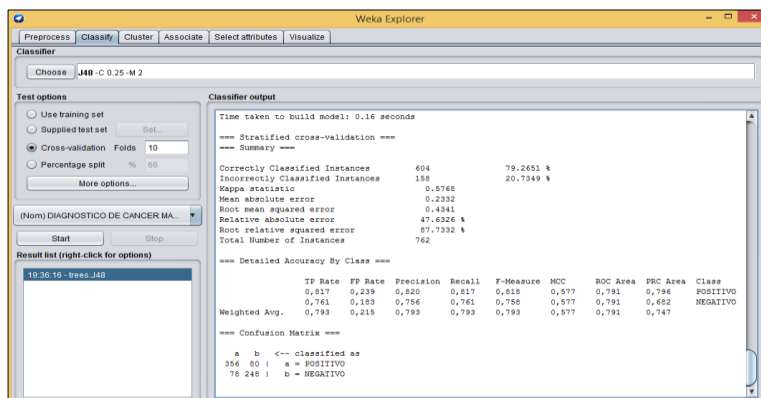


Figura 37. Exploración de los datos parte II

Después de realizar un análisis profundo llegamos a nuestra primera conclusión de que el software Weka nos muestra porcentajes de asertividad de pacientes almacenados en la base de datos, más no cuando se requiere la detección temprana de un nuevo paciente.

Para ello es necesaria la implementación de un sistema web lo cual nos permite una mayor facilidad de uso para el usuario

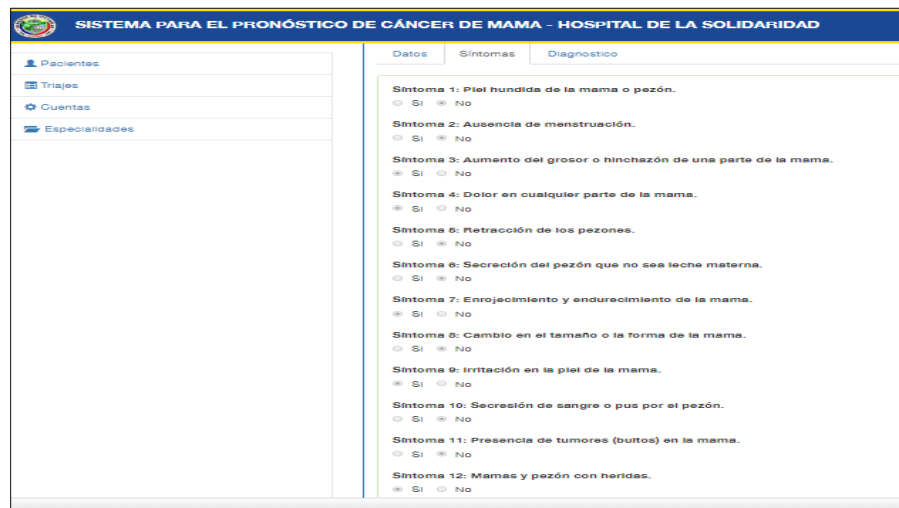


Figura 38. Muestra de selección de síntomas del paciente

3.3.4. Verificar la calidad de los datos

Clínica y Diagnóstico - Síntomas:

S001: Piel hundida de la mama o pezón

S002: Ausencia de menstruación

S003: Aumento del grosor o hinchazón de una parte de la mama.

S004: Dolor en cualquier parte de la mama.

S005: Retracción de los pezones

S006: Secreción del pezón que no sea leche materna

S007: Enrojecimiento y endurecimiento de la mama.

S008: Cambio en el tamaño o la forma de la mama.

S009: Irritación en la piel de la mama.

S010: Secreción de sangre o pus por el pezón.

S011: Presencia de tumores (bultos) en la mama.

S012: Mamas y pezón con heridas.

- Campos a predecir
 - Diagnóstico de cáncer de mama
 - Grado de posición a detectar el cáncer de mama
 - Síntomas del paciente.

✓ **Reporte de la calidad de los datos**

Cuentas		Especialidades		Show 10 entries							Buscar:
DNI	Nombre	Fecha Nacimiento	Estado Civil	Sexo	Teléfono	Acción					
06337054	VIZARRETA CARDENAS JESSICA	1990-04-17	S	F	123456	Triajes 0	Editar	Eliminar			
06771677	AGREDA MINAYA CARLA ANDREA	1094-03-12	S	F	567890	Triajes 0	Editar	Eliminar			
07270721	AGUAYO DAVILA DAVILA MARITZAJULIETA	1990-09-07	V	F	123456	Triajes 0	Editar	Eliminar			
07472678	ACEVEDO SOTO LUISA	1990-09-10	S	F	684245	Triajes 0	Editar	Eliminar			
07472678	ACEVEDO SOTO MANUELA	0000-00-00	V	F	23423	Triajes 1	Editar	Eliminar			
07977448	ABARCA LOPEZ ROCIO	1994-01-25	S	F	123123	Triajes 1	Editar	Eliminar			
08275540	AGUAYO CALMET RAPHAELA ROSA	1995-03-14	S	F	123456	Triajes 0	Editar	Eliminar			
08523409	AGREDA GUZMANIO JUANA GREGORIA	1997-06-24	S	F	456789	Triajes 0	Editar	Eliminar			
08808616	ADRIANZEN RONCEROS ALONDRA	1996-02-15	S	F	678432	Triajes 0	Editar	Eliminar			

© Sistema de Gestión de Registros de Pacientes del Hospital de la Solidaridad 2018

Figura 39. Datos generales de las pacientes

CO DE CÁNCER DE MAMA - HOSPITAL DE LA SOLIDARIDAD

LISTA DE TRIAJES - Oncología - VOLKERT SCHUSTER EDITH EILENN #11

Datos | Sintomas | Diagnostico

Paciente: VOLKERT SCHUSTER EDITH EILENN | Especialidad: Oncología | Edad: 20

Peso (Kg): 20 | Talla (M): 2 | Presión: 20

Situación Física: Situación Física | Edad Menarquia: 2 | Cantidad Embarazos: 2

Edad Primer Embarazo: 2 | Cantidad Abortos: 2 | Cantidad Hijos: 2

Antecedentes Familiares: 2 | Descripción: 2

Guardar

Figura 40. Características del paciente

LISTA DE TRIAJES - Oncología - ACEVEDO SOTO MANUELA #16

Datos | Sintomas | Diagnostico

Se diagnostica que el paciente ACEVEDO SOTO MANUELA SI presenta cancer de mama con un grado de 2.

Figura 41. Clínica y diagnóstico del paciente

3.4. Fase 3 – Preparación de los datos

3.4.1. Seleccionar los datos

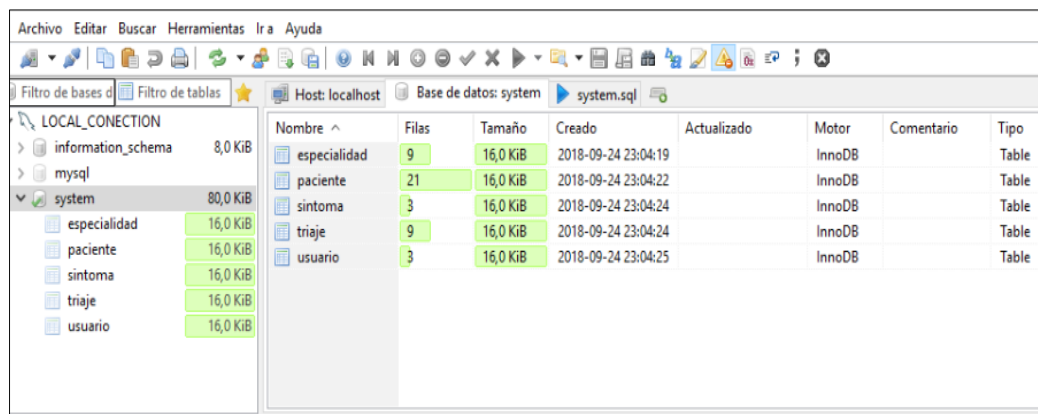
En términos de registros, se van a utilizar todos los registros dentro de cada tabla que compone la base de datos, ya que al ser ésta una base de datos específicamente creada para este proyecto, el número de registros que se han insertado ha sido elegido a propósito. Sin embargo, hay campos dentro de estos registros que no son necesarios para nuestros objetivos de minería de datos, por lo que se puede prescindir de algunos de ellos.

3.4.2. Limpiar los datos

La base de datos con la que contamos para nuestra tesis tiene toda la información necesaria para poder cumplir con los objetivos de la minería de datos, puesto que son datos limpios y por lo tanto no hay necesidad de hacer una limpieza más profunda sobre ellos.

Tampoco tenemos campos en los que falten valores, por lo tanto no se consideran como datos faltantes, por lo que no es necesario realizar ningún tipo de estimación de valores faltantes.

Para generar el modelo relacionado con el primer objetivo de la minería de datos, es decir, el de la predicción del diagnóstico de cáncer de mama de las pacientes, es necesario utilizar un filtro que nos proporciona la herramienta Weka y de este modo sólo seleccionar aquellas filas que serán muy relevantes para nuestra tesis.



The screenshot shows a MySQL database interface. On the left, a tree view shows the database structure: LOCAL_CONNECTION, information_schema (8,0 KiB), mysql, and system (80,0 KiB). The 'system' database is expanded, showing tables: especialidad (16,0 KiB), paciente (16,0 KiB), sintoma (16,0 KiB), triaje (16,0 KiB), and usuario (16,0 KiB). On the right, a table structure view for 'system' is displayed with the following data:

Nombre	Filas	Tamaño	Creado	Actualizado	Motor	Comentario	Tipo
especialidad	9	16,0 KiB	2018-09-24 23:04:19		InnoDB		Table
paciente	21	16,0 KiB	2018-09-24 23:04:22		InnoDB		Table
sintoma	3	16,0 KiB	2018-09-24 23:04:24		InnoDB		Table
traje	9	16,0 KiB	2018-09-24 23:04:24		InnoDB		Table
usuario	3	16,0 KiB	2018-09-24 23:04:25		InnoDB		Table

Figura 42. Base de datos

3.4.3. Estructurar los datos

En nuestra presente tesis no estructuramos los campos, debido a que nosotras mismas creamos los datos de la BD con el único fin de darle un mejor uso de investigación y beneficio a las pacientes para una detección temprana de cáncer de mama.

3.4.4. Integrar los datos

En esta parte del proceso se verifica que dos tablas tuvieron que integrar los atributos de otra tabla para obtener resultados positivos sobre el modelo predictivo para la detección del cáncer de mama.

Siendo que la tabla triaje requiera del idpaciente, idespecialidad a fin de que pueda obtener los datos del paciente y de la especialidad que se atendió anteriormente, lográndose visualizar los historiales de las pacientes de atenciones pasadas.

En base a la tabla síntoma se necesita del idtriaje, para que haga una combinación entre los historiales y la nueva información que se recibe por parte de la paciente para una mejor predicción del cáncer de mama.

#	Nombre	Tipo de datos	Longitud/Conjunto	Sin signo	Permitir NULL	Rellenar con c...	Predeterminado
1	Id	INT	11	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	AUTO_INCREME
2	IdTriaje	INT	11	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	NULL
3	S001	INT	11	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	NULL
4	S002	INT	11	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	NULL
5	S003	INT	11	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	NULL
6	S004	INT	11	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	NULL
7	S005	INT	11	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	NULL
8	S006	INT	11	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	NULL
9	S007	INT	11	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	NULL
10	S008	INT	11	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	NULL
11	S009	INT	11	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	NULL
12	S010	INT	11	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	NULL
13	S011	INT	11	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	NULL
14	S012	INT	11	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	NULL
15	Diagnostico	INT	11	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	NULL

Figura 43. Tabla síntoma

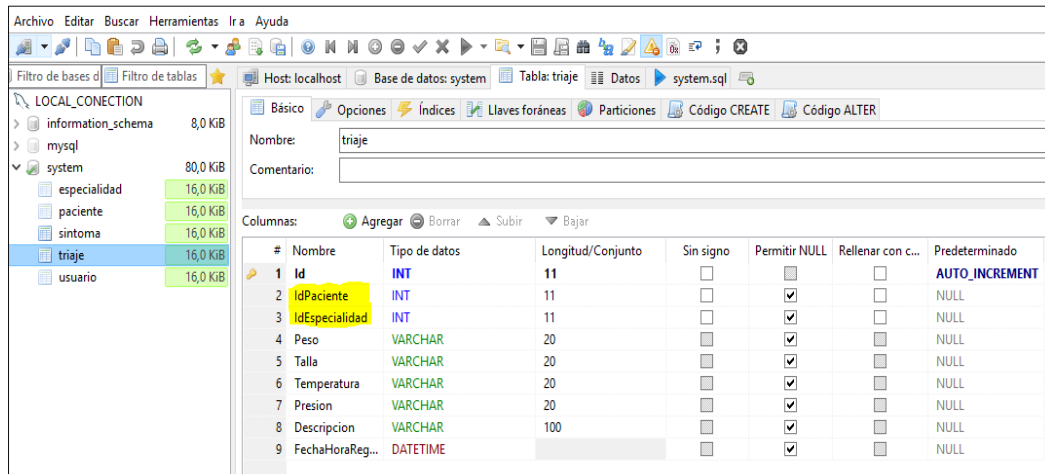


Figura 44. Tabla triaje

3.4.5. Formateo de los datos

El campo con la información referente a los diversos síntomas que presenta el paciente ha sido codificado con valores numérico ya que la herramienta de minería de datos exige que los datos a estudiar sean numéricos. Inicialmente este campo contenía el nombre de cada síntoma escrito con caracteres alfanuméricos, pero después se optó por asignar un número a cada centro.

Relation: BASE DE DATOS HOSPITAL FINAL												
No	1 EDAD	2 TALLA	3 PESO	4 FRECUENCIA CARDIACA	5 SITUACION FISICA	6 EDAD MENARQUIA	7 CANTIDAD EMBARAZOS	8 EDAD PRIMER EMBARAZO	9 CANTIDAD ABORTOS	10 CANTIDAD HIJOS		
	Numerico	Numerico	Numerico	Numerico	Nominal	Numerico	Numerico	Numerico	Numerico	Numerico		
1	48.0	1.68	69.0	80.0	NORMAL	10.0	2.0	19.0	2.0	0.0		
2	44.0	1.48	59.0	70.0	NORMAL	15.0	3.0	17.0	1.0	2.0		
3	47.0	1.65	52.0	68.0	SOBREPESO	14.0	5.0	18.0	3.0	2.0		
4	48.0	1.53	55.0	77.0	SOBREPESO	12.0	3.0	20.0	0.0	3.0		
5	52.0	1.6	66.0	72.0	NORMAL	14.0	4.0	17.0	3.0	1.0		
6	53.0	1.67	66.0	82.0	NORMAL	14.0	6.0	20.0	2.0	4.0		
7	53.0	1.67	80.4	70.0	SOBREPESO	11.0	3.0	16.0	3.0	0.0		
8	51.0	1.67	66.0	80.0	NORMAL	12.0	0.0	17.0	0.0	0.0		
9	61.0	1.68	69.0	64.0	SOBREPESO	14.0	0.0	15.0	0.0	0.0		
10	52.0	1.48	52.0	90.0	SOBREPESO	14.0	3.0	16.0	3.0	0.0		
11	49.0	1.52	55.0	74.0	SOBREPESO	14.0	3.0	20.0	2.0	1.0		
12	50.0	1.65	66.0	71.0	NORMAL	11.0	4.0	20.0	0.0	4.0		
13	55.0	1.53	59.0	75.0	SOBREPESO	10.0	5.0	21.0	0.0	5.0		
14	53.0	1.52	62.0	68.0	SOBREPESO	11.0	6.0	15.0	2.0	4.0		
15	43.0	1.67	66.0	74.0	SOBREPESO	14.0	6.0	22.0	2.0	4.0		
16	57.0	1.48	49.0	70.0	SOBREPESO	14.0	6.0	15.0	2.0	4.0		
17	61.0	1.55	60.0	81.0	NORMAL	10.0	1.0	17.0	0.0	1.0		
18	65.0	1.67	71.0	81.0	SOBREPESO	10.0	2.0	24.0	0.0	2.0		
19	51.0	1.57	57.0	63.0	NORMAL	11.0	6.0	20.0	2.0	4.0		
20	56.0	1.6	65.4	88.0	SOBREPESO	11.0	6.0	20.0	2.0	4.0		
21	49.0	1.75	78.2	90.0	NORMAL	12.0	5.0	18.0	2.0	3.0		
22	63.0	1.56	60.3	80.0	SOBREPESO	11.0	5.0	19.0	0.0	5.0		
23	50.0	1.75	80.4	90.0	SOBREPESO	15.0	3.0	16.0	0.0	3.0		
24	52.0	1.8	87.6	68.0	SOBREPESO	15.0	0.0	19.0	0.0	0.0		
25	45.0	1.5	50.0	70.0	SOBREPESO	11.0	7.0	15.0	2.0	5.0		
26	65.0	1.52	56.3	75.0	NORMAL	15.0	3.0	23.0	3.0	0.0		
27	47.0	1.6	54.3	90.0	SOBREPESO	13.0	7.0	18.0	3.0	4.0		
28	44.0	1.8	85.8	83.0	SOBREPESO	11.0	2.0	21.0	2.0	0.0		
29	46.0	1.66	70.3	68.0	SOBREPESO	13.0	4.0	21.0	3.0	1.0		
30	61.0	1.58	65.2	64.0	NORMAL	12.0	8.0	24.0	3.0	5.0		
31	51.0	1.62	60.4	67.0	SOBREPESO	10.0	3.0	24.0	1.0	2.0		
32	60.0	1.65	67.8	75.0	SOBREPESO	15.0	4.0	22.0	0.0	4.0		
33	54.0	1.72	77.9	62.0	SOBREPESO	12.0	2.0	23.0	0.0	2.0		
34	56.0	1.75	75.3	86.0	NORMAL	14.0	5.0	23.0	2.0	3.0		
35	46.0	1.5	57.8	70.0	SOBREPESO	13.0	4.0	20.0	3.0	1.0		
36	46.0	1.79	84.2	84.0	NORMAL	14.0	5.0	16.0	0.0	5.0		
37	43.0	1.63	66.7	80.0	NORMAL	13.0	2.0	17.0	2.0	0.0		

Figura 45. Reporte de datos del paciente: edad, talla, peso, frecuencia cardíaca, edad menarquía, cantidad embarazos, cantidad primer embarazo, cantidad abortos, cantidad hijos

11: ANTECEDENTES FAMILIARES 12: PIEL HUNDIDA DE LA MAMA O PEZÓN 13: AUSENCIA DE MENSTRUACIÓN 14: AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA 15: DOLOR EN CUALQUIER PARTE DE LA MAMA							
Numario	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal
0.0	NO	SI	SI	SI	SI	SI	SI
2.0	SI	SI	SI	NO	NO	NO	NO
0.0	NO	NO	NO	NO	NO	SI	SI
2.0	SI	SI	SI	SI	SI	NO	NO
2.0	SI	SI	SI	SI	SI	NO	NO
3.0	NO	NO	NO	NO	NO	SI	SI
2.0	SI	NO	NO	NO	NO	NO	NO
4.0	SI	SI	SI	NO	NO	NO	NO
2.0	SI	NO	NO	NO	NO	SI	SI
1.0	SI	SI	SI	NO	NO	SI	SI
1.0	NO	SI	SI	NO	NO	NO	NO
3.0	NO	SI	SI	SI	SI	SI	SI
4.0	SI	NO	NO	NO	NO	SI	SI
4.0	NO	SI	SI	NO	NO	NO	NO
0.0	NO	NO	NO	SI	SI	SI	SI
3.0	SI	SI	SI	NO	NO	SI	SI
2.0	SI	NO	NO	NO	NO	SI	SI
1.0	SI	SI	SI	NO	NO	SI	SI
3.0	SI	SI	SI	SI	SI	SI	SI
0.0	SI	SI	SI	SI	SI	SI	SI
0.0	SI	NO	NO	SI	SI	NO	NO
2.0	SI	SI	SI	SI	SI	SI	SI
4.0	NO	NO	NO	SI	SI	NO	NO
4.0	NO	NO	NO	SI	SI	SI	SI
1.0	NO	SI	SI	SI	SI	SI	SI
4.0	SI	NO	NO	NO	NO	SI	SI
4.0	NO	SI	SI	NO	NO	NO	NO
4.0	SI	NO	NO	NO	NO	NO	NO
4.0	SI	SI	SI	SI	SI	SI	SI
2.0	NO	SI	SI	NO	NO	NO	NO
3.0	SI	NO	NO	NO	NO	NO	NO
3.0	NO	SI	SI	NO	NO	NO	NO
3.0	NO	SI	SI	NO	NO	SI	SI
1.0	SI	SI	SI	SI	SI	NO	NO
1.0	NO	SI	SI	SI	SI	SI	SI
2.0	NO	SI	SI	NO	NO	NO	NO
0.0	SI	NO	NO	SI	SI	SI	SI

Figura 46. Reporte de datos del paciente: ausencia de menstruación y dolor de la mama.

16: RETRACCIÓN DE LOS PEZONES 17: SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA 18: ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA 19: CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA 20: IRRITACIÓN EN LA PIEL DE LA MAMA							
Nominal		Nominal		Nominal		Nominal	
NO	SI	NO	SI	NO	SI	NO	SI
SI	SI	NO	SI	SI	SI	NO	NO
SI	NO	NO	NO	NO	SI	NO	NO
NO	NO	NO	NO	NO	NO	SI	NO
NO	SI	NO	NO	NO	SI	SI	NO
NO	NO	NO	NO	NO	SI	SI	NO
NO	SI	NO	NO	NO	SI	SI	NO
SI	NO	NO	NO	NO	SI	SI	NO
NO	NO	SI	NO	NO	SI	SI	NO
SI	SI	NO	NO	NO	SI	SI	NO
NO	NO	NO	NO	NO	NO	SI	NO
SI	SI	NO	NO	NO	SI	SI	NO
SI	NO	NO	NO	NO	SI	SI	NO
SI	NO	NO	NO	NO	SI	SI	NO
SI	NO	NO	NO	NO	SI	SI	NO
SI	SI	NO	NO	NO	SI	SI	NO
NO	SI	NO	NO	NO	SI	SI	NO
SI	NO	NO	NO	NO	NO	SI	NO
SI	SI	NO	NO	NO	NO	SI	NO
NO	SI	NO	NO	NO	SI	SI	NO
NO	NO	SI	NO	NO	NO	SI	NO
SI	SI	NO	NO	NO	SI	SI	NO
NO	NO	NO	NO	NO	NO	SI	NO
NO	NO	NO	NO	NO	NO	SI	NO
NO	NO	NO	NO	NO	NO	SI	NO
SI	NO	NO	NO	NO	SI	SI	NO
SI	SI	NO	NO	NO	NO	SI	NO
SI	SI	SI	SI	SI	SI	SI	NO

Figura 47. Características generales del paciente: retracción de los pezones, secreción, endurecimiento de la mama, cambio en el tamaño de la mama e irritación.

21: SECRESIÓN DE SANGRE O PUS POR EL PEZÓN		22: PRESENCIA DE TUMORES (BULTOS) EN LA MAMA		23: MAMAS Y PEZÓN CON HERIDAS		24: DIAGNOSTICO DE CANCER MAMA	
Nominal		Nominal		Nominal		Nominal	
NO		SI		SI		POSITIVO	
NO		NO		SI		NEGATIVO	
NO		NO		SI		POSITIVO	
NO		NO		NO		POSITIVO	
NO		NO		SI		POSITIVO	
SI		NO		SI		NEGATIVO	
NO		SI		NO		NEGATIVO	
SI		SI		SI		POSITIVO	
NO		NO		NO		POSITIVO	
NO		SI		NO		POSITIVO	
NO		NO		NO		POSITIVO	
NO		SI		SI		POSITIVO	
NO		SI		NO		POSITIVO	
SI		SI		NO		POSITIVO	
SI		NO		SI		POSITIVO	
SI		NO		SI		POSITIVO	
SI		NO		SI		POSITIVO	
SI		NO		SI		POSITIVO	
NO		NO		NO		NEGATIVO	
NO		NO		SI		POSITIVO	
NO		NO		SI		POSITIVO	
SI		NO		SI		POSITIVO	
NO		NO		SI		POSITIVO	
SI		NO		SI		NEGATIVO	
SI		SI		NO		POSITIVO	
NO		NO		NO		POSITIVO	
NO		SI		NO		NEGATIVO	
NO		NO		SI		NEGATIVO	
SI		SI		NO		NEGATIVO	
SI		SI		NO		POSITIVO	
SI		SI		SI		NEGATIVO	
SI		NO		NO		NEGATIVO	
SI		SI		SI		POSITIVO	
SI		SI		NO		NEGATIVO	
NO		SI		SI		POSITIVO	
NO		NO		NO		POSITIVO	
NO		SI		NO		NEGATIVO	
SI		NO		NO		POSITIVO	

Figura 48. Características generales del paciente: secreción de sangre, mamas con heridas.

3.5. Fase 4 - Modelado

3.5.1. Seleccionar la técnica del modelado

✓ Técnica seleccionada

Para la realización de nuestra presente tesis se va utilizar la técnica del árbol debido a que va realizar un modelo de predicción que se utiliza en el campo de la inteligencia artificial a partir de una base de datos en la cual se construyen los diagramas de construcciones lógicas

Es un sistema similar a la predicción que está basado en normas que sirven para la representación de una serie de condiciones que se presentan de manera sucesiva en la resolución de problemas.

Así mismo decidimos utilizar la técnica del árbol puesto nos permite plantear claramente el problema de tal manera que todas las opciones sean analizadas, hacer un análisis rápido de todas las consecuencias de las posibles decisiones, ya que utiliza un esquema que cuantifica los resultados y las probabilidades de que los diferentes resultados (valga la redundancia) aparezcan, nos ayuda a tomar decisiones adecuadamente.

Los árboles de decisión, son representaciones gráficas y analíticas de datos introducidos por medio de una base de datos ya establecida. Estos ayudan en la toma de decisiones, desde un punto de vista probabilístico con el fin de elegirla opción más conveniente.

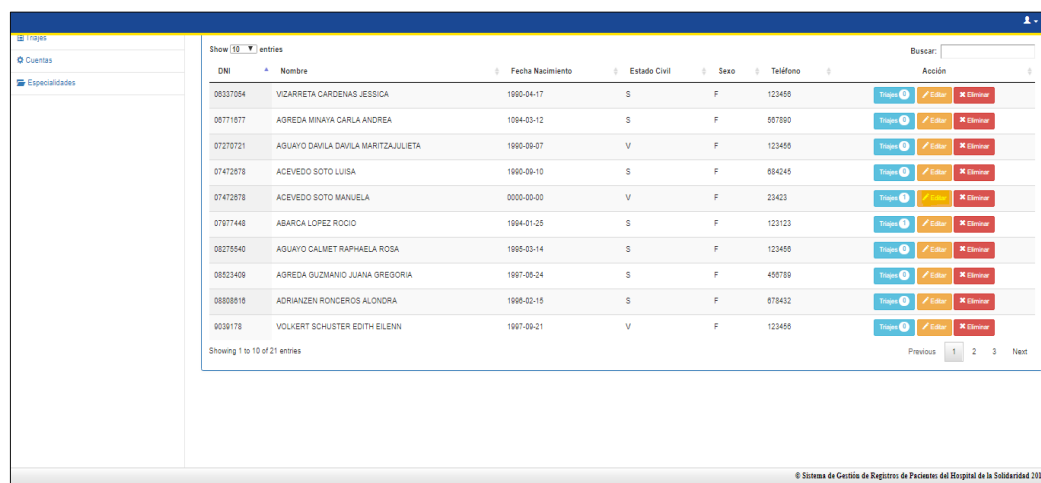
Los árboles de decisión se caracterizan por ser una técnica utilizada en minería, en otros se busca solucionar problemas de predicción, clasificación y segmentación, una vez analizada todos los datos.

3.5.2. Generar el plan de prueba

- Se tiene la base de datos seleccionada con los síntomas de los pacientes para proceder con el modelo predictivo:

Para ello tenemos dos bases de datos:

- a) Una base de datos que representa pacientes antiguos que han sido detectados cáncer de mama en la herramienta Weka que representa una cantidad de: 599 pacientes antiguos.
- b) Una base de datos que representa pacientes nuevos que han sido detectados cáncer de mama con el Sistema Web implementado que representa una cantidad de: 15 pacientes nuevos.



DNI	Nombre	Fecha Nacimiento	Estado Civil	Sexo	Teléfono	Acción
06337054	VIZARRETA CARDENAS JESSICA	1990-04-17	S	F	123456	Trasferir Editar Eliminar
06771877	AGREDA MINAYA CARLA ANDREA	1094-03-12	S	F	667890	Trasferir Editar Eliminar
07270721	AGUAYO DAVILA DAVILA MARITZAJULIETA	1990-09-07	V	F	123456	Trasferir Editar Eliminar
07472978	ACEVEDO SOTO LUISA	1990-09-10	S	F	684245	Trasferir Editar Eliminar
07472978	ACEVEDO SOTO MANUELA	0000-00-00	V	F	23423	Trasferir Editar Eliminar
07977448	ABARCA LOPEZ RODIO	1994-01-25	S	F	123123	Trasferir Editar Eliminar
08278540	AGUAYO CALMET RAPHAELA ROSA	1995-03-14	S	F	123456	Trasferir Editar Eliminar
08823409	AGREDA GUZMANIO JUANA GREGORIA	1997-06-24	S	F	456789	Trasferir Editar Eliminar
08808918	ADRIANZEN RONCEROS ALONDRA	1998-02-15	S	F	678432	Trasferir Editar Eliminar
9030178	VOLKERT SCHUSTER EDITH EILENN	1997-09-21	V	F	123456	Trasferir Editar Eliminar

Figura 49. Base de datos de nuevos pacientes

2.- Una vez teniendo la base de datos en excel se procedió con el convertidor de la base de datos a archivos de Weka:

- CSV
- ARFF



Figura 50. Conversión de archivos.

3.-Luego se procede convertir de la extensión CSV a ARFF para ello se realiza en el mismo Weka como se muestra en la siguiente imagen:



Figura 51. Interfaz del software weka

Explorer: Preprocesamiento (Preprocess)

El explorer permite visualizar y aplicar distintos algoritmos de aprendizaje a un conjunto de datos. Cada una de las tareas de minería de datos viene representada por una pestaña en la parte superior. Estas son:

- **Preprocess:** visualización y preprocesador de los datos (Aplicación de filtros).
- **Classify:** Aplicación de algoritmos de clasificación y regresión.
- **Cluster:** Agrupación
- **Associate:** Asociación
- **Select Attributes:** Selección de atributos
- **Visualize:** Visualización de los datos por parejas de atributos.

En este procedimiento agregamos el archivo convertido CSV guardado en uno del disco de unidades.

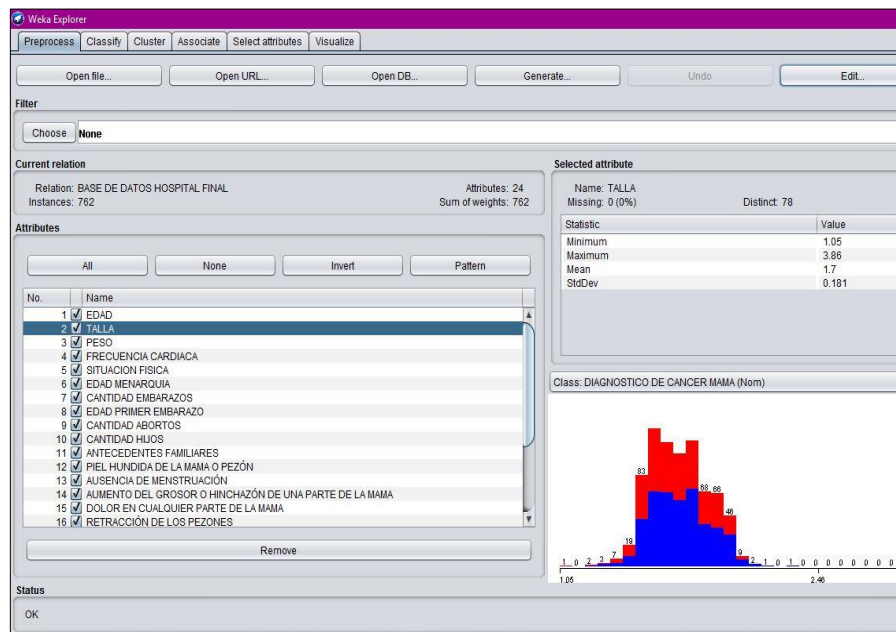


Figura 52. Síntomas de los pacientes.

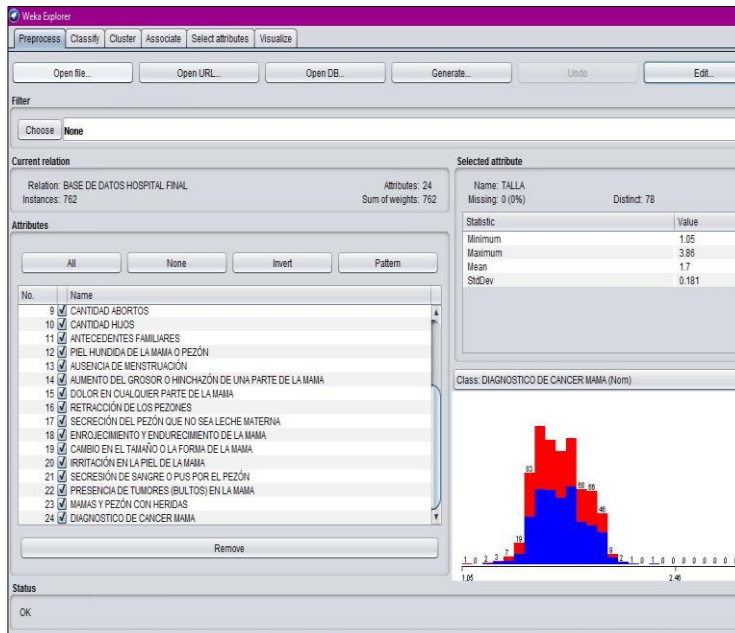


Figura 53. Características generales de los pacientes

También podemos observar en la pantalla principal que el grado_posicion_cáncer de mama se encuentra clasificado como se ha definido en la base de datos que se realizó.

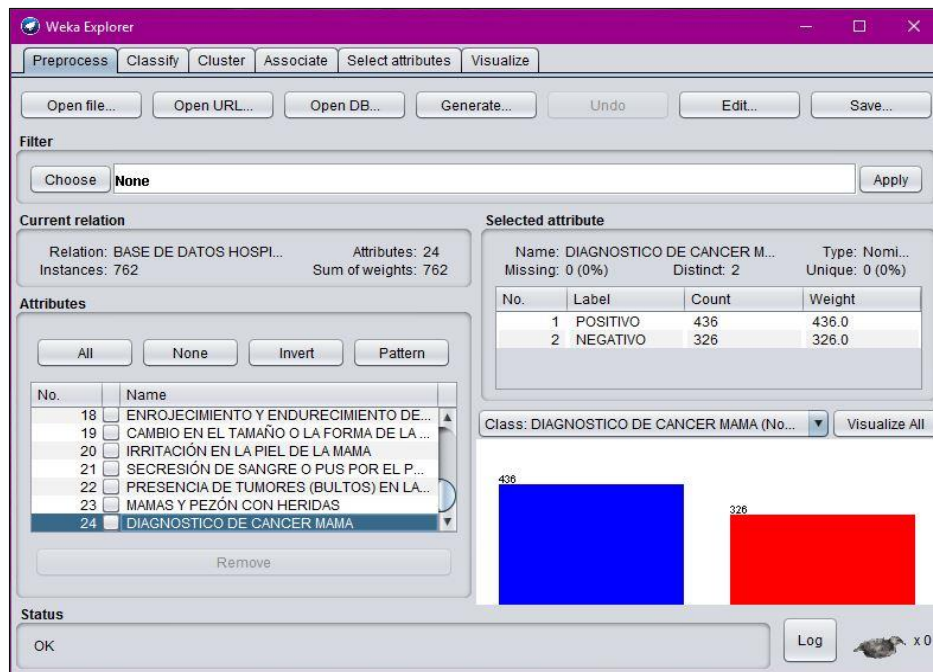


Figura 54. Número de grado del paciente.

Número de grados:

- **Grado 1:** Describe el cáncer de mama invasivo (las células cancerígenas toman o invaden el tejido mamario normal que las rodea), es posible la invasión microscópica. Esto significa que las células cancerígenas recién han comenzado a invadir el tejido ubicado en las paredes del conducto o el lobulillo, pero estas células cancerígenas no miden más de 1 mm.
- **Grado 2:** No se visualiza algún tumor en la mama, pero se detectan células cancerígenas (que superan los 2 mm) en 1-3 ganglios linfáticos axilares (debajo del brazo) o en los ganglios linfáticos cerca del esternón (encontrado durante una biopsia del ganglio centinela) O; el tumor mide 2 cm o menos, y se ha propagado hacia los ganglios linfáticos axilares O; el tumor mide entre 2 y 5 cm, y no se ha propagado hacia los ganglios linfáticos axilares.
- **Grado 3:** El tumor puede tener cualquier tamaño, y se detectó la presencia de cáncer en 4-9 ganglios linfáticos axilares o en los ganglios linfáticos cercanos al esternón; enrojecimiento del sector de la piel de la mama, la mama se siente caliente y puede estar inflamada.
- **Grado 4:** Describe el cáncer de mama que se ha propagado más allá de la mama y los ganglios linfáticos circundantes hacia otros órganos del cuerpo, como los pulmones, ganglios linfáticos distantes, la piel, los huesos, el cerebro y el hígado.

Clasificación (Classify)

Vamos ahora a construir un primer clasificado para los datos. Pulsemos la pestaña classify, que está arriba del todo a la izquierda en la ventana del Explorer. Primero tendremos que elegir el clasificador, en choose, arriba a la izquierda (actualmente está seleccionado zeror). Debajo de choose, podemos elegir las opciones de test, es decir, la manera de computar el porcentaje esperado de aciertos (en clasificación), o el error cuadrático medio (entre otros, en regresión). Estas opciones son:

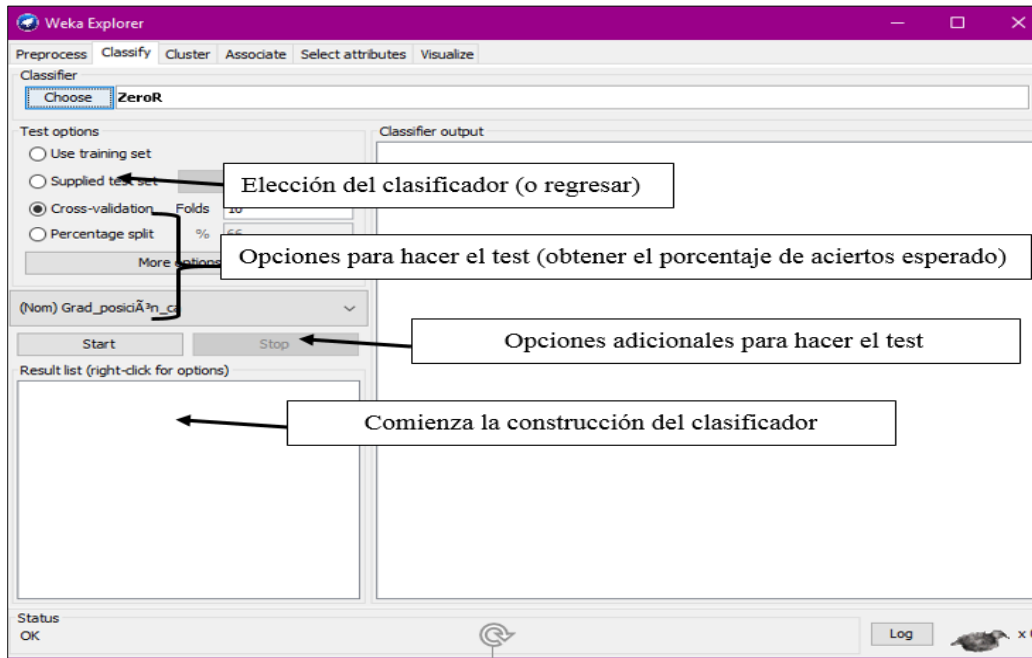


Figura 55. Clasificador de atributos de las pacientes.

También se puede aplicar algoritmos de clasificación estadística y análisis de regresión a los conjuntos de datos resultantes. Tomando en cuenta la base de datos de pacientes antiguos.

Modelo gen

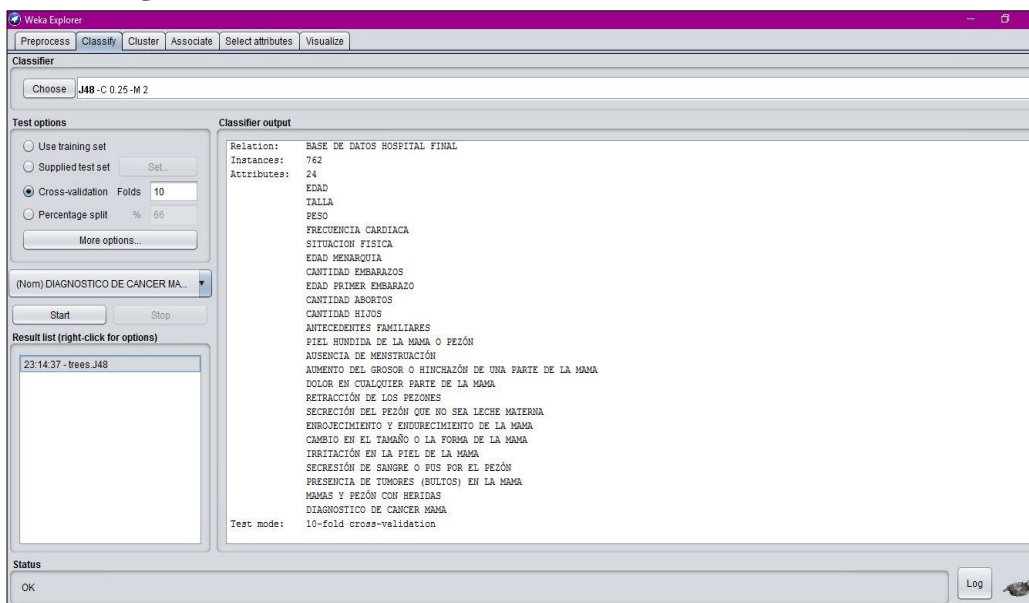


Figura 56. Formación del árbol con atributos de las pacientes

Descripción y evaluación del modelo predictivo

- Modelo 1:

```
Classifier output
=== Evaluation on training set ===
Time taken to test model on training data: 0.01 seconds

=== Summary ===
Correctly Classified Instances      700          91.8635 %
Incorrectly Classified Instances    62           8.1365 %
Kappa statistic                    0.8373
Mean absolute error                0.1331
Root mean squared error            0.258
Relative absolute error            26.6315 %
Root relative squared error        51.6057 %
Total Number of Instances          762

=== Detailed Accuracy By Class ===
                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0,921  0,083  0,916  0,921  0,918  0,837  0,966  0,963  NORMAL
                0,917  0,079  0,921  0,917  0,919  0,837  0,966  0,966  SOBREPESO
Weighted Avg.   0,919  0,081  0,919  0,919  0,919  0,837  0,966  0,964

=== Confusion Matrix ===
  a  b  <-- classified as
348 30 | a = NORMAL
 32 352 | b = SOBREPESO
```

Figura 57. Formación del árbol con atributos de las pacientes del modelo 1.

- Modelo 2:

```
Classifier output
Time taken to test model on training data: 0.01 seconds

=== Summary ===
Correctly Classified Instances      703          92.2572 %
Incorrectly Classified Instances    59           7.7428 %
Kappa statistic                    0.8446
Mean absolute error                0.1219
Root mean squared error            0.2469
Relative absolute error            24.4251 %
Root relative squared error        49.4219 %
Total Number of Instances          762

=== Detailed Accuracy By Class ===
                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0,947  0,104  0,908  0,947  0,927  0,845  0,973  0,972  SI
                0,896  0,053  0,940  0,896  0,917  0,845  0,973  0,971  NO
Weighted Avg.   0,923  0,080  0,923  0,923  0,922  0,845  0,973  0,972

=== Confusion Matrix ===
  a  b  <-- classified as
376 21 | a = SI
 38 327 | b = NO

=== Source code ===
```

Figura 58. Formación del árbol con atributos de las pacientes del modelo 2.

- Modelo 3:

```

Classifier output

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances      696          91.3386 %
Incorrectly Classified Instances     66           8.6614 %
Kappa statistic                     0.8262
Mean absolute error                  0.1373
Root mean squared error              0.262
Relative absolute error              27.5904 %
Root relative squared error          52.5269 %
Total Number of Instances           762

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                -----  -----  -
                0,921    0,093    0,896    0,921    0,908    0,827    0,967    0,959    SI
                0,907    0,079    0,930    0,907    0,918    0,827    0,967    0,969    NO
Weighted Avg.   0,913    0,086    0,914    0,913    0,913    0,827    0,967    0,965

=== Confusion Matrix ===

  a  b  <-- Classified as
326 28 | a = SI
 38 370 | b = NO

=== Source code ===

```

Figura 59. Formación del árbol con atributos de las pacientes del modelo 3.

- Modelo 4:

```

Classifier output

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances      706          92.6509 %
Incorrectly Classified Instances     56           7.3491 %
Kappa statistic                     0.853
Mean absolute error                  0.1213
Root mean squared error              0.2463
Relative absolute error              24.2727 %
Root relative squared error          49.2673 %
Total Number of Instances           762

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                -----  -----  -
                0,928    0,075    0,923    0,928    0,926    0,853    0,971    0,968    SI
                0,925    0,072    0,930    0,925    0,927    0,853    0,971    0,968    NO
Weighted Avg.   0,927    0,073    0,927    0,927    0,927    0,853    0,971    0,968

=== Confusion Matrix ===

  a  b  <-- Classified as
349 27 | a = SI
 29 357 | b = NO

=== Source code ===

```

Figura 60. Formación del árbol con atributos de las pacientes del modelo 4.

- Modelo 5:

```

Classifier output
=====
Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances      705          92.5197 %
Incorrectly Classified Instances     57           7.4803 %
Kappa statistic                     0.8499
Mean absolute error                  0.126
Root mean squared error              0.251
Relative absolute error              25.2471 %
Root relative squared error          50.2466 %
Total Number of Instances           762

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                -----  -----  -
                0,942    0,094    0,917     0,942    0,930     0,850    0,966    0,963    SI
                0,906    0,058    0,935     0,906    0,920     0,850    0,966    0,964    NO
Weighted Avg.   0,925    0,077    0,925     0,925    0,925     0,850    0,966    0,963

=== Confusion Matrix ===
  a  b  <-- classified as
376 23 | a = SI
 34 329 | b = NO
=====
Source code

```

Figura 61. Formación del árbol con atributos de las pacientes del modelo 5.

- Modelo 6:

```

Classifier output
=====
=== Evaluation on training set ===
Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances      700          91.8635 %
Incorrectly Classified Instances     62           8.1365 %
Kappa statistic                     0.8373
Mean absolute error                  0.136
Root mean squared error              0.2607
Relative absolute error              27.1967 %
Root relative squared error          52.1505 %
Total Number of Instances           762

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                -----  -----  -
                0,907    0,069    0,931     0,907    0,919     0,838    0,962    0,961    SI
                0,931    0,093    0,907     0,931    0,919     0,838    0,962    0,957    NO
Weighted Avg.   0,919    0,081    0,919     0,919    0,919     0,838    0,962    0,959

=== Confusion Matrix ===
  a  b  <-- classified as
350 36 | a = SI
 26 350 | b = NO
=====
Source code

```

Figura 62. Formación del árbol con atributos de las pacientes del modelo 6.

3.5.3. Construir el modelo

- Indicador número 1:

Tiempo para procesar el registro de los datos del paciente.



The screenshot shows a web application interface for a breast cancer prognosis system. The main content area is titled 'LISTA DE PACIENTES' and contains a table with one entry. A 'Cronómet...' window is overlaid on the bottom left, showing a timer at 00:00:06 and a power button icon.

DNI	Nombre	Fecha Nacimiento	Estado Civil	Sexo	Teléfono
89064326	ESCOBAR VILCHEZ ANTONELA	1996-10-04	S	F	234667

Showing 1 to 1 of 1 entries (filtered from 22 total entries)

Figura 63. Tiempo de demora del KPI1.

- Indicador número 2:

Tiempo para procesar la data.



The screenshot shows the same web application interface as Figure 63, but with a table containing four entries. The 'Cronómet...' window is overlaid on the bottom left, showing a timer at 00:00:09 and a power button icon.

DNI	Nombre	Fecha Nacimiento	Estado Civil	Sexo	Teléfono
34967693	sanchez sanchez juia	1967-01-22	S	F	940986706
60320166	Garza Vitez Shelya	1983-08-23	S	F	940986706
70897018	Garza Soto Marcelo	1978-02-26	S	F	940986799
89064326	ESCOBAR VILCHEZ ANTONELA	1996-10-04	S	F	234667

Showing 21 to 24 of 24 entries

Figura 64. Tiempo de demora del KPI2.

- Indicador número 3:

Tiempo para predecir el cáncer de mama.



Figura 65. Tiempo de demora de KPI3.

3.6. Fase 5 - Evaluación

3.6.1. Evaluar los resultados

Desde un inicio se estableció como criterio de éxito principal el poder realizar predicciones con un porcentaje de fiabilidad “aceptable”, este criterio puede ser algo subjetivo, por lo que es inevitable apoyarse principalmente en los criterios de éxito desde el punto de vista de la minería de datos son mucho más que específicos y precisos. Además, para poder calificar como aceptable o no las predicciones que se van a realizar es necesario tener una base objetiva, como lo son los indicadores estadísticos que se han obtenido al ejecutar los modelos. También sería conveniente la evaluación de los resultados por parte de un grupo de expertos en la minería de datos, si se contara con ellos.

3.7. Fase 6 - Implantación

3.7.1. Plan de implantación

- En primer lugar

Para poder implantar este proyecto en el negocio real sería necesario en primer lugar tener acceso a la base de datos real del negocio, es decir la base de datos que contiene toda la información relativa de los pacientes del hospital. A partir de ahí, los pasos a seguir serían los mismos que se han seguido en este documento desde la comprensión del negocio hasta la implantación. Si bien, cabe decir que habrá algunas fases, como la de comprensión y preparación de los datos, que en el negocio real probablemente sean más complejas y llevarán más tiempo que en este proyecto ya que se puede esperar que en la base de datos real se tengan muchos más registros y estos mismos contengan más ruido que en nuestra base de datos ficticia creada específicamente para este uso.

- En segundo lugar

En segundo lugar, sería necesario que en el negocio (el hospital) se use una base de datos Oracle, de no ser así se tendrían dos opciones, la primera sería exportar la base de datos actual a una base de datos SQL, y la segunda sería utilizar otro software de minería de datos distinto al utilizado en este proyecto, para este propósito se podría utilizar alguna de las herramientas que la minería de datos nos brinda.

3.7.2. Plan de monitoreo y mantención

- Plan de monitoreo y mantención

El mantenimiento y la supervisión de la implementación del presente proyecto es una fase importante del mismo debido a que los datos que se procesan con mucha frecuencia pueden ser modificados por el personal del hospital. Los datos pueden ser modificados por diferentes motivos como haber realizado una codificación incorrecta. El volumen de estos datos en movimiento es grande motivo por el cual la extracción de las

muestras debe ser realizada cuidadosamente y realizando siempre backups de los datos explotados en cada proceso.

La minería de datos debería ser realizada en un periodo de seis meses ya que esta es la medida de tiempo utilizada en el hospital para realizar los exámenes y asignar las notas finales a los alumnos, sin embargo, esta medida podría variar en cualquier momento en función del plan de estudios que esté vigente en cada momento.

De acuerdo con el plan de supervisión y mantenimiento se podría establecer los siguientes procesos:

- ✓ La extracción y almacenamiento de los datos guardando la información obtenida en formato de hoja de cálculo.
- ✓ Distribución de los datos en función de los modelos de software de minería de datos a trabajar.
- ✓ Los archivos de la explotación de datos deberán ser guardados en el área de Soporte o Administración en el propio hospital, almacenándolos por ejemplo en carpetas ordenadas por procesos según lo requiera el hospital.
- ✓ Los resultados obtenidos en cada explotación de datos deberán ser llevados a formato de hoja de cálculo y generar gráficas de distintos tipos para una mejor visualización e interpretación de los resultados obtenidos en cada periodo.

3.7.3. Informe final

- Modelos aprobados

El uso de la metodología CRISP-DM en este proyecto ha permitido hallar un comportamiento predictivo al momento de diagnosticar el cáncer de mama de las pacientes del Hospital de la Solidaridad de Villa El Salvador.

De los tres objetivos de minería de datos iniciales que se habían fijado se han podido alcanzar dos de ellos (objetivos 1 y 2). Además, al margen de estos objetivos, se han sacado otras conclusiones a partir de los datos analizados y estudiados.

Por otro lado, dentro de las etapas que hemos seguido para la realización de nuestra tesis podemos decir que:

La primera etapa ha sido una de las más laboriosas por no tener una base de datos de la cual poder partir. Esto ha supuesto que tengamos que generar nosotros mismos un conjunto de datos sobre el que trabajar. Para poder hacer una simulación lo más real posible, no valía con generar datos aleatorios, si no que se ha tenido que desarrollar un pequeño programa en Java que generase estos datos de manera automática, debido a la gran cantidad de datos que necesitábamos manejar para hacer una estimación lo más precisa posible.

Cuando ya disponíamos de la base de datos sobre la que ejercer la minería de datos, se hizo un análisis de la estructura de los datos y la información contenida.

Se realizaron consultas significativas en MySQL para tener muestras representativas de los datos, y sacar más conclusiones al margen de los objetivos iniciales de la minería.

(Realización del plan de proyecto).

A continuación, se realizó la elección de las técnicas del modelado y la ejecución de dichas técnicas sobre los datos empleando la herramienta escogida para ello (Weka). Esta herramienta nos facilitó por completo la aplicación de los modelos ya que nos permitió ver de manera muy visual e intuitiva para analizar cuáles eran las técnicas más adecuadas para nuestra base de datos. Por último, una vez obtenidos los modelos, se analizaron para determinar la adecuación o no adecuación de los mismos.

3.7.4. Revisión del proyecto

- Documentación de experiencias

En primer lugar, y como ya se ha comentado anteriormente en el informe final de la última fase, el mayor problema que hemos tenido a lo largo de este proyecto es el de no disponer de una base de datos real sobre la que actuar ya que esto condiciona en gran medida los resultados obtenidos.

A pesar de haber intentado generar unos datos lo más veraces posible, no cabe duda que existen multitud de factores que no podemos manejar y que disponer de los datos reales y demás características de los pacientes incrementaría aún más la fiabilidad de los modelos de minería de datos elegidos en este proyecto.

Esto se puede interpretar como algo positivo ya que, si hemos dado por válidos dos de los tres modelos empleados, y sin disponer de la cantidad y veracidad de los datos que se manejan en la base de datos del hospital, esto quiere decir que si el proyecto saliera en real los resultados mejorarían aún más.

CAPÍTULO IV
ANÁLISIS DE RESULTADOS Y CONTRASTACIÓN
DE LA HIPÓTESIS

4.1. Población y muestra

4.1.1. Población

Está constituida por todos los pacientes registrados y nuevos del Hospital de la Solidaridad de Villa El Salvador.

N= Indeterminado

4.1.2. Muestra

Actualmente existen varios procedimientos estadísticos de forma aleatoria para calcular el tamaño de la muestra debido a que tanto los elementos de la población como de la muestra en estudio tienen similares características. Por lo cual para esta investigación se tomó una muestra valor 30, ya que es un valor adecuado, estándar y se utiliza en varios procesos de investigación, según lo informa el autor Peter Pande.

Por lo tanto: Tamaño de muestra es $n = 30$

4.2. Nivel de confianza y grado de significancia

El nivel de confianza será de 95% dada la inexperiencia de los investigadores.

4.3. Resultados genéricos

a) Modelado del negocio

- Estructura Organizacional
- Productos y clientes
- Stakeholders internos y externos
- Cadena de valor
- Procesos de negocios
- Matriz de selección de proceso

b) Modelado del proceso

- Modelado de contexto
- Diagrama de descomposición de funciones
- Flujograma de actividades

c) Incepción del proyecto

- Modelado de contexto
- Visión del producto

d) Planificación del proyecto

- Comprensión del negocio
- Comprensión de los datos
- Preparación de los datos
- Modelado
- Evaluación
- Implantación

e) Ejecución del proyecto

- Diseño de la base de datos
- Selección de la técnica del modelado
- Construcción de modelos predictivos según el número de objetivos

4.4. Resultados específicos

A continuación, se muestran las medidas de los KPIs para la Pre-Prueba y Post-Prueba.

Tabla 27

Resultados de pre prueba y post prueba para los KPI1, KPI2, KPI3 y KPI4

Número	KPI 1: Tiempo para procesar el registro de los datos del paciente		KPI 2: Tiempo para procesar la data		KPI 3: Tiempo para predecir el cáncer de mama		KPI 4: Nivel de Satisfacción del paciente	
	Pre-prueba (seg.)	Post-prueba (seg.)	Pre-prueba (seg.)	Post-prueba (seg.)	Pre-prueba (seg.)	Post-prueba (seg.)	Pre-prueba	Post-prueba
1	120.0	0.20	420.00	120.00	172800	0.35	No satisfecho	Satisfecho
2	180.0	0.15	480.00	120.00	172800	0.23	No satisfecho	Satisfecho
3	120.0	0.23	420.00	120.00	345600	0.19	Satisfecho	Muy satisfecho
4	180.0	0.14	480.00	60.00	259200	0.27	Satisfecho	Muy satisfecho
5	240.0	0.10	540.00	120.00	172800	0.18	Poco satisfecho	Satisfecho
6	240.0	0.17	540.00	60.00	172800	0.25	No satisfecho	Satisfecho
7	180.0	0.19	600.00	120.00	345600	0.18	Poco satisfecho	Satisfecho
8	240.0	0.12	480.00	60.00	259200	0.26	Poco satisfecho	Muy satisfecho
9	180.0	0.11	540.00	60.00	172800	0.16	No satisfecho	Satisfecho
10	240.0	0.13	540.00	60.00	345600	0.34	Satisfecho	Muy satisfecho
11	180.0	0.10	600.00	120.00	259200	0.36	Satisfecho	Muy satisfecho

12	120.0	0.13	360.00	60.00	345600	0.27	Poco satisfecho	Satisfecho
13	120.0	0.15	540.00	120.00	172800	0.38	No satisfecho	Muy satisfecho
14	180.0	0.11	420.00	60.00	259200	0.17	Satisfecho	Muy satisfecho
15	240.0	0.16	420.00	120.00	259200	0.19	No satisfecho	Muy satisfecho
16	180.0	0.12	540.00	60.00	172800	0.27	No satisfecho	Muy satisfecho
17	240.0	0.13	420.00	120.00	345600	0.41	Satisfecho	Muy satisfecho
18	180.0	0.11	540.00	120.00	259200	0.30	Poco satisfecho	Muy satisfecho
19	180.0	0.14	600.00	120.00	172800	0.20	Poco satisfecho	Muy satisfecho
20	120.0	0.12	540.00	120.00	259200	0.18	Poco satisfecho	Muy satisfecho
21	240.0	0.14	480.00	120.00	172800	0.26	Poco satisfecho	Satisfecho
22	240.0	0.15	480.00	60.00	172800	0.15	No satisfecho	Satisfecho
23	180.0	0.14	420.00	120.00	172800	0.37	Poco satisfecho	Satisfecho
24	120.0	0.13	540.00	60.00	259200	0.50	Poco satisfecho	Muy satisfecho
25	180.0	0.12	420.00	60.00	345600	0.16	Satisfecho	Muy satisfecho
26	120.0	0.14	480.00	120.00	172800	0.27	No satisfecho	Satisfecho
27	240.0	0.11	600.00	60.00	259200	0.46	Poco satisfecho	Satisfecho
28	180.0	0.12	540.00	120.00	345600	0.31	Poco satisfecho	Muy satisfecho
29	120.0	0.10	480.00	60.00	172800	0.12	Satisfecho	Muy satisfecho
30	180.0	0.11	600.00	60.00	345600	0.17	Satisfecho	Muy satisfecho

4.5. Análisis e interpretación de resultados

A. Indicador: Tiempo para procesar el registro de los datos del paciente. KPI₁

Tabla 28

Resultados de pre-prueba y post-prueba para el KPI1

	Pre prueba	Post prueba		
	120.0	0.20	0.20	0.20
	180.0	0.15	0.15	0.15
	120.0	0.23	0.23	0.23
	180.0	0.14	0.14	0.14
	240.0	0.10	0.10	0.10
	240.0	0.17	0.17	0.17
	180.0	0.19	0.19	0.19
	240.0	0.12	0.12	0.12
	180.0	0.11	0.11	0.11
	240.0	0.13	0.13	0.13
	180.0	0.10	0.10	0.10
	120.0	0.13	0.13	0.13
	120.0	0.15	0.15	0.15
	180.0	0.11	0.11	0.11
	240.0	0.16	0.16	0.16
	180.0	0.12	0.12	0.12
	240.0	0.13	0.13	0.13
	180.0	0.11	0.11	0.11
	180.0	0.14	0.14	0.14
	120.0	0.12	0.12	0.12
	240.0	0.14	0.14	0.14
	240.0	0.15	0.15	0.15
	180.0	0.14	0.14	0.14
	120.0	0.13	0.13	0.13
	180.0	0.12	0.12	0.12
	120.0	0.14	0.14	0.14
	240.0	0.11	0.11	0.11
	180.0	0.12	0.12	0.12
	120.0	0.10	0.10	0.10
	180.0	0.11	0.11	0.11
Promedio	182.0		0.14	
Meta planteada			0.15	
N° menor al promedio		13	13	30
% menor al promedio		43.3 %	43.3 %	100

- El 43.3 % del tiempo para procesar el registro de los datos del paciente en la Post-Prueba fueron menores que su tiempo promedio.
- El 43.3 % del tiempo para procesar el registro de los datos del paciente en la Post-Prueba fueron menores que la meta planteada.
- El 100.0 % del tiempo para procesar el registro de los datos del paciente en la Post-Prueba fueron menores que el tiempo promedio en la Pre-Prueba.

- **Con estadística descriptiva – Pre Prueba**

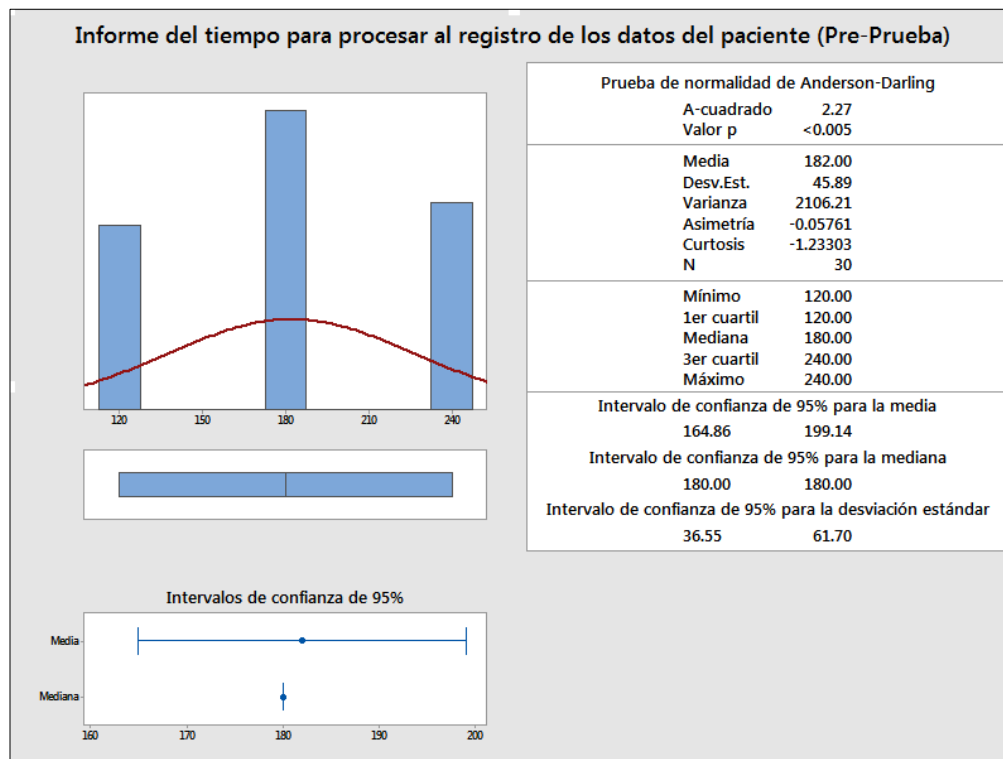


Figura 66. Informe del resumen KPI1 pre prueba.

- Con estadística descriptiva – Post Prueba

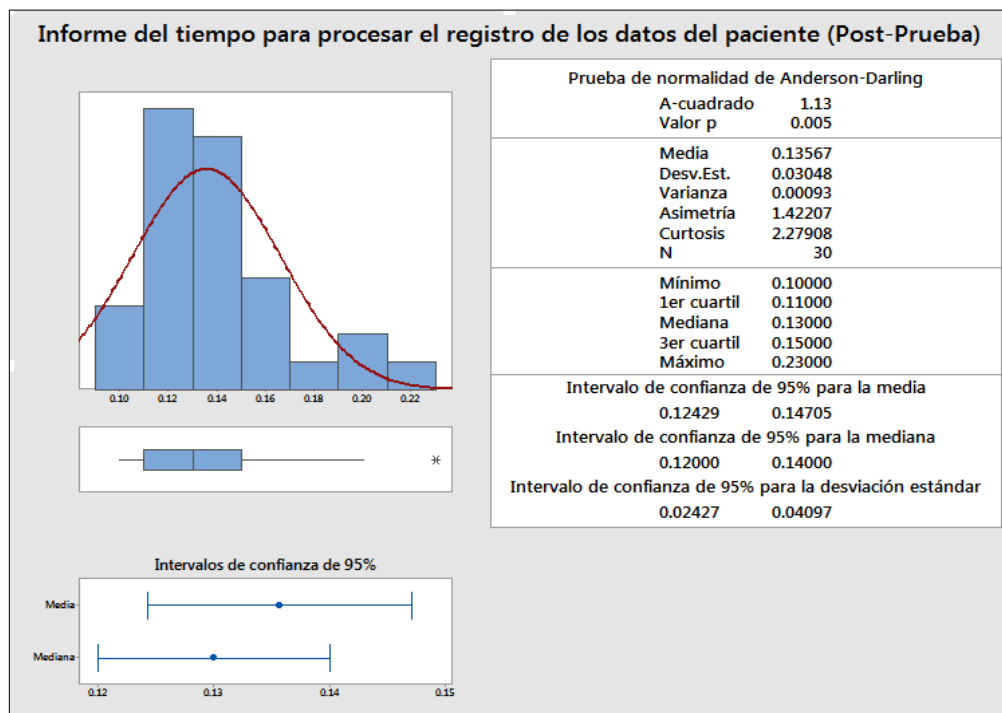


Figura 67. Informe del resumen KPI1 post prueba.

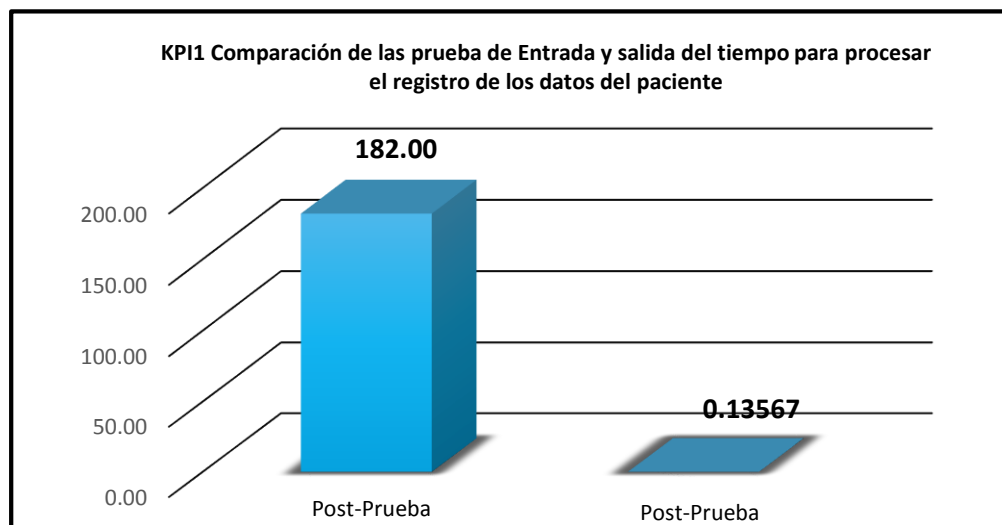


Figura 68. Comparación de pruebas.

- Se obtuvo como media del tiempo para procesar el registro de los datos del paciente, en la Pre – Prueba muestra el valor de 182 seg., mientras que en la Post – Prueba el valor fue de 0.13 segundos; lo cual indica una gran reducción del tiempo con la solución.
- Monitoreado la varianza, la dispersión de los datos con respecto al tiempo para procesar el registro de los datos del paciente, en la Pre-prueba (2106.21) frente a la Post-prueba (0.00093) esto demuestra que tuvo un impacto favorable ya que disminuyó los tiempos del proceso en los registros de las pacientes.
- Con respecto a los cuartiles Q1 y Q3 para la Pre-prueba:
El 25% de datos es menor que o igual a los 120.00 segundos.
El 75% de los datos es menor que o igual a los 240.00 segundos.
- Con respecto a los cuartiles Q1 y Q3 para la Post-prueba:
El 25 % de datos es menor que o igual a los 0.1100 segundos.
El 75% de los datos es menor que o igual a los 0.1500 segundos.
- La asimetría en la Pre-Prueba fue de -0.05761 esto nos indica que existen demoras en el proceso para el registro de las pacientes, lo que es adverso a la simetría de la Post-prueba con 1.42207 ello nos acerca a un menor tiempo en procesar el registro de los datos de las pacientes, mejorando en la atención del hospital solidaridad de Villa El Salvador.

B. Indicador: Tiempo para procesar la data: KPI₂

Tabla 29

Resultados de pre-prueba y post-prueba para el KPI₂

	Pre prueba	Post prueba		
	420.00	120.00	120.00	120.00
	480.00	120.00	120.00	120.00
	420.00	120.00	120.00	120.00
	480.00	60.00	60.00	60.00
	540.00	120.00	120.00	120.00
	540.00	60.00	60.00	60.00
	600.00	120.00	120.00	120.00
	480.00	60.00	60.00	60.00
	540.00	60.00	60.00	60.00
	540.00	60.00	60.00	60.00
	600.00	120.00	120.00	120.00
	360.00	60.00	60.00	60.00
	540.00	120.00	120.00	120.00
	420.00	60.00	60.00	60.00
	420.00	120.00	120.00	120.00
	540.00	60.00	60.00	60.00
	420.00	120.00	120.00	120.00
	540.00	120.00	120.00	120.00
	600.00	120.00	120.00	120.00
	540.00	120.00	120.00	120.00
	480.00	120.00	120.00	120.00
	480.00	60.00	60.00	60.00
	420.00	120.00	120.00	120.00
	540.00	60.00	60.00	60.00
	420.00	60.00	60.00	60.00
	480.00	120.00	120.00	120.00
	600.00	60.00	60.00	60.00
	540.00	120.00	120.00	120.00
	480.00	60.00	60.00	60.00
	600.00	60.00	60.00	60.00
Promedio	502.00		92	
Meta planteada			2.00	
N° menor al promedio		14	14	30
% menor al promedio		46.6	46.6	100

- El 46.6 % del tiempo para procesar la data en la Post-Prueba fueron menores que su tiempo promedio.
- El 46.6 % del tiempo para procesar la data en la Post-Prueba fueron menores que la meta planteada.
- El 100.0 % del tiempo para procesar la data en la Post-Prueba fueron mayores que el tiempo promedio en la Pre-Prueba.

- Con estadística descriptiva – Pre prueba

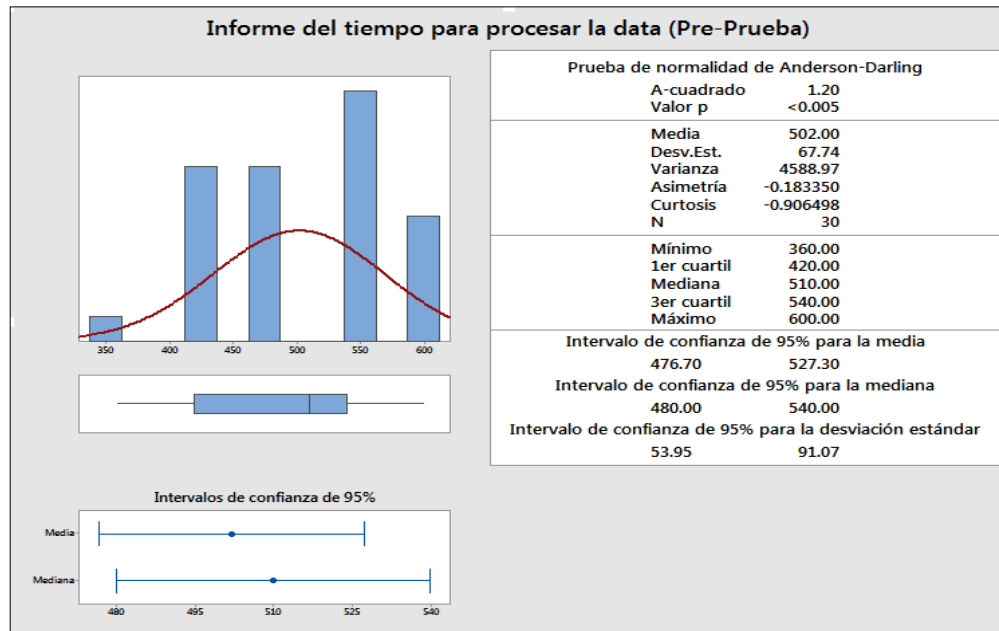


Figura 69. Informe del resumen kpi 2 pre prueba

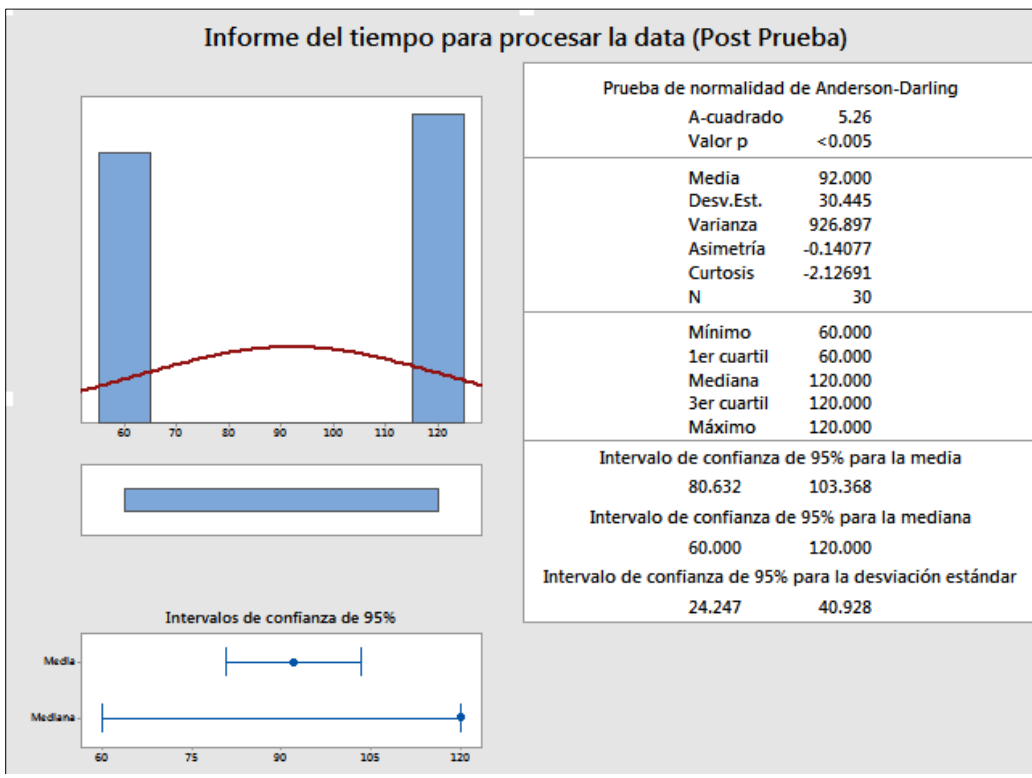


Figura 70. Informe del resumen KPI2 post prueba.

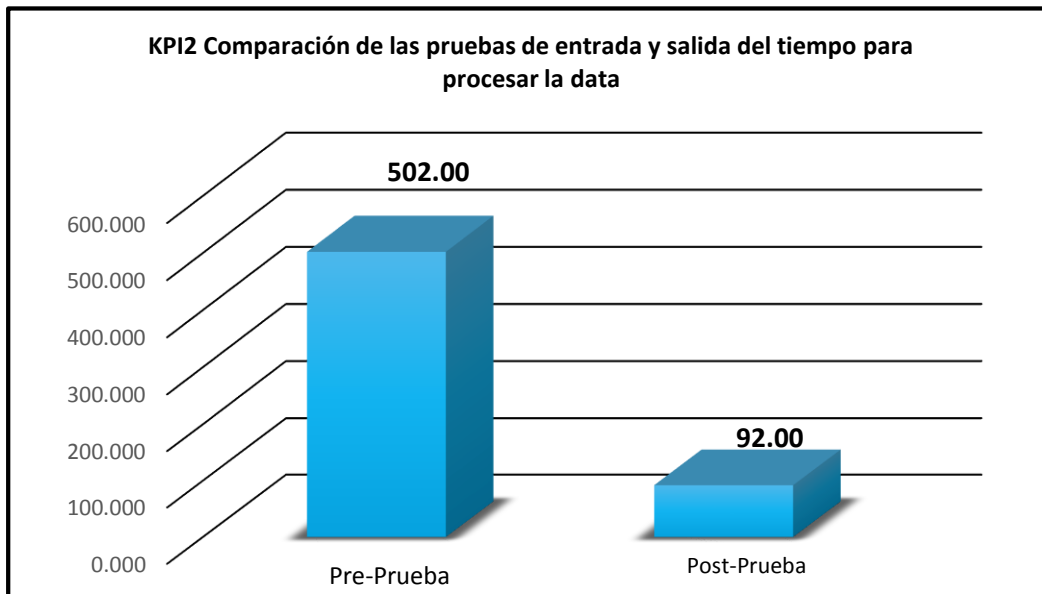


Figura 71. Comparación de pruebas.

- Se obtuvo como media del tiempo para procesar la data, en la Pre – Prueba muestra el valor de 502.00 segundos mientras que en la Post – Prueba el valor fue de 92.000 segundos; lo cual indica una gran reducción del tiempo con la solución. También podemos indicar que los valores mínimos del tiempo para procesar la data fueron 360.00 segundos antes y 60.000 segundos después.
- Analizando la varianza de dispersión de los datos con respecto al tiempo para procesar la data, en la Pre-prueba (4588.97) frente a la Post-Prueba (926.897) demuestra ampliamente la variación en los tiempos de procesamiento lo cual resulta un impacto favorable.
- Con respecto a los cuartiles Q1 y Q3 para la Pre-prueba:
 El 25% de los datos es menor o igual a los 420.00 segundos.
 El 75% de los datos es menor que o igual a los 540.00 segundos.
- Con respecto a los cuartiles Q1 y Q3 para la Post-prueba:
 El 25% de los datos es menor que o igual a los 60.000 segundos.
 El 75% de los datos es menor que o igual a los 120.000 segundos.

- La asimetría en la Pre-prueba fue de -0.183350, esto nos indica que existen tiempos de procesamiento de la data muy elevados, lo que es adverso a la asimetría de la de la Post-prueba con -0.14077, ellos nos acerca a los tiempos de procesamiento de la data mucho más bajos y aceptables para las pacientes del Hospital de la Solidaridad.

C. Indicador: Tiempo para predecir el cáncer de mama: KPI₃

Tabla 30

Resultados de pre-prueba y post-prueba para el KPI₃

	Pre prueba	Post prueba		
	172800	0.35	0.35	0.35
	172800	0.23	0.23	0.23
	345600	0.19	0.19	0.19
	259200	0.27	0.27	0.27
	172800	0.18	0.18	0.18
	172800	0.25	0.25	0.25
	345600	0.18	0.18	0.18
	259200	0.26	0.26	0.26
	172800	0.16	0.16	0.16
	345600	0.34	0.34	0.34
	259200	0.36	0.36	0.36
	345600	0.27	0.27	0.27
	172800	0.38	0.38	0.38
	259200	0.17	0.17	0.17
	259200	0.19	0.19	0.19
	172800	0.27	0.27	0.27
	345600	0.41	0.41	0.41
	259200	0.30	0.30	0.30
	172800	0.20	0.20	0.20
	259200	0.18	0.18	0.18
	172800	0.26	0.26	0.26
	172800	0.15	0.15	0.15
	172800	0.37	0.37	0.37
	259200	0.50	0.50	0.50
	345600	0.16	0.16	0.16
	172800	0.27	0.27	0.27
	259200	0.46	0.46	0.46
	345600	0.31	0.31	0.31
	172800	0.12	0.12	0.12
	345600	0.17	0.17	0.17
Promedio	244800		0.2636	
Meta planteada			0.46	
N° menor al promedio		17	17	30
% menor al promedio		56.6	56.6	100

- El 56.6 % del tiempo para predecir el cáncer de mama data en la Post-Prueba fueron

menores que su tiempo promedio.

- El 56.6 % del tiempo para predecir el cáncer de mama en la Post-Prueba fueron menores que la meta planeada.
- El 100.0 % tiempo para predecir el cáncer de mama en la Post Prueba fueron mayores que el tiempo promedio en la Pre-Prueba.

- **Con estadística descriptiva – Pre prueba**

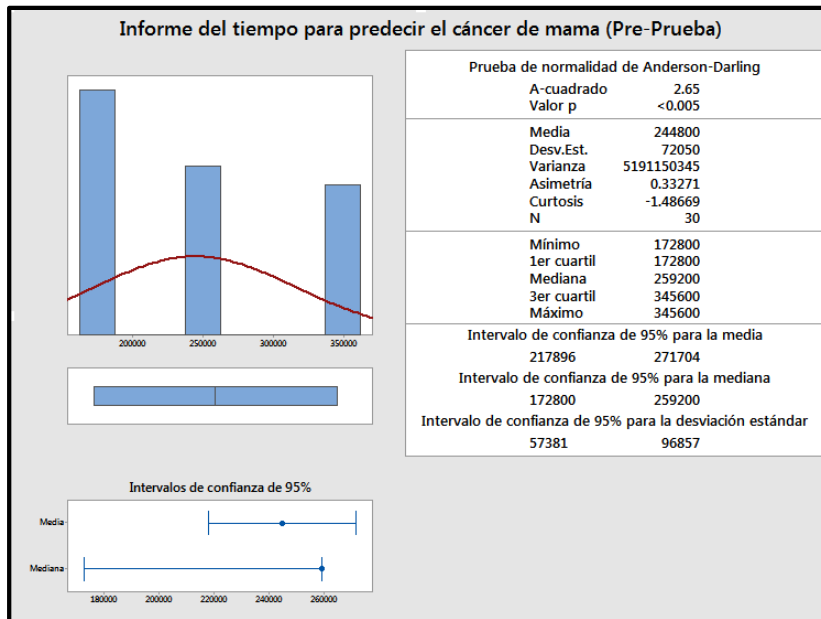


Figura 72. Informe del resumen kpi3 pre prueba.

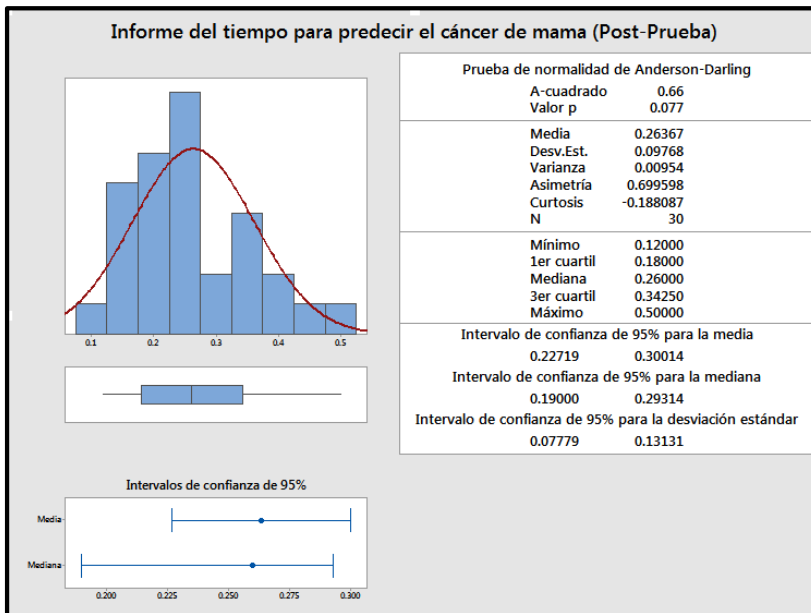


Figura 73. Informe del resumen kpi3 post prueba

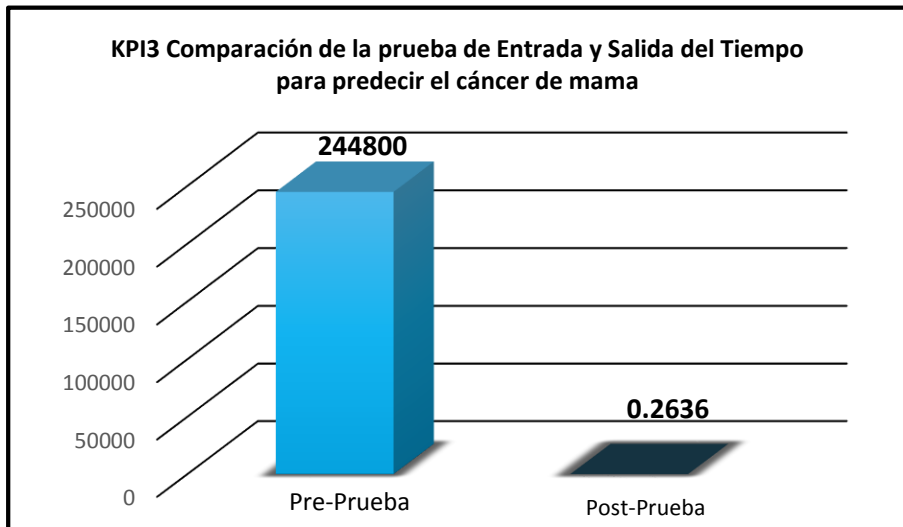


Figura 74. Comparación de las pruebas

- Se obtuvo como media del tiempo para predecir el cáncer de mama, en la Pre – Prueba muestra el valor de 244800 segundos, mientras que en la Post – Prueba el valor fue de 0.26367 segundos; lo cual indica una gran reducción del tiempo con la solución.
- Analizando la varianza de dispersión de los datos con respecto al tiempo para procesar la data, en la Pre-prueba (5191150345) frente a la Post-prueba (0.00954) demuestra ampliamente la variación en el tiempo para predecir el cáncer de mama, lo cual resulta un impacto favorable.
- Con respecto a los cuartiles Q1 YQ3 para la Pre-prueba:
 - El 25% de los datos es menor que o igual a los 172800 segundos.
 - El 75% de los datos es menor que o igual a los a los 345600 segundos.
- Con respecto a los cuartiles Q1 y Q3 para la Post-prueba:
 - El 25% de los datos es menor que o igual a los 0.18000 segundos.
 - El 75% de los datos es menor que o igual a los 0.34250 segundos.
- La asimetría en la Pre-prueba fue de 0.33271, esto nos indica que existen tiempos de procesamiento de la data muy elevados, lo que es adverso a la asimetría de la Post-prueba con 0.699598, ello nos acerca a los tiempos de procesamiento de la data mucho más bajos y aceptables para las pacientes del Hospital de la Solidaridad.

D. Indicador: Nivel de satisfacción del paciente: KPI 4

Valores de Pre-prueba

Tabla 31

Valores de pre-prueba

Pre Prueba					
1	No satisfecho	11	Satisfecho	21	Poco satisfecho
2	No satisfecho	12	Poco satisfecho	22	No satisfecho
3	Satisfecho	13	No satisfecho	23	Poco satisfecho
4	Satisfecho	14	Satisfecho	24	Satisfecho
5	Poco satisfecho	15	No satisfecho	25	Satisfecho
6	No satisfecho	16	No satisfecho	26	No satisfecho
7	Poco satisfecho	17	Satisfecho	27	Poco satisfecho
8	Poco satisfecho	18	Poco satisfecho	28	Poco satisfecho
9	No satisfecho	19	Poco satisfecho	29	Satisfecho
10	Satisfecho	20	Poco satisfecho	30	Satisfecho

Tabla 32

Valores de estado y frecuencia de pre-prueba

Estado	Frecuencia
Muy satisfecho	0
Satisfecho	9
Poco satisfecho	12
No satisfecho	9
Total	30

Tabla 33

Valores de estado y frecuencia de pre-prueba

Estado	Frecuencia
Satisfecho	9
No satisfecho	21

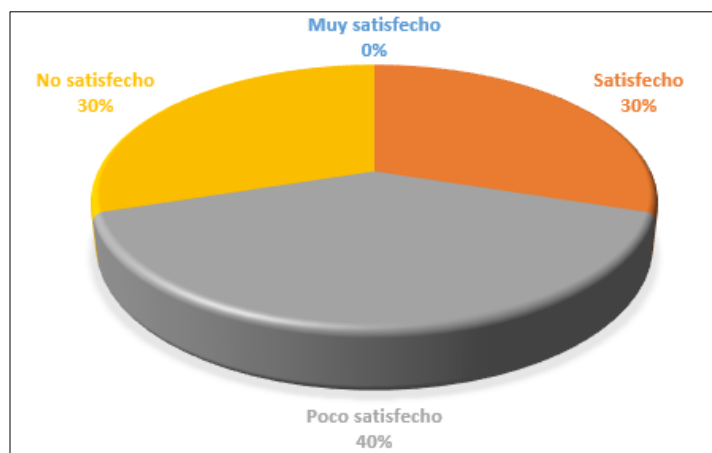


Figura 75. Porcentaje de pre prueba del KPI4

- 30 % de las veces el nivel de satisfacción del paciente fue catalogado como no satisfecho por los pacientes atendidos.
- El 30 % de las veces el nivel de satisfacción del paciente fue catalogado como satisfecho por los pacientes atendidos.
- Se determina que el 30 % de las veces el nivel de satisfacción es satisfecho.
- Se determina que el 70 % de las veces el nivel de satisfacción es no satisfecho.

Valores de Post-prueba

Tabla 34

Valores de post-prueba

Post Prueba					
1	Satisfecho	11	Muy satisfecho	21	Satisfecho
2	Satisfecho	12	Satisfecho	22	Satisfecho
3	Muy satisfecho	13	Muy satisfecho	23	Satisfecho
4	Muy satisfecho	14	Muy satisfecho	24	Muy satisfecho
5	Satisfecho	15	Muy satisfecho	25	Muy satisfecho
6	Satisfecho	16	Muy satisfecho	26	Satisfecho
7	Satisfecho	17	Muy satisfecho	27	Satisfecho
8	Muy satisfecho	18	Muy satisfecho	28	Muy satisfecho
9	Satisfecho	19	Muy satisfecho	29	Muy satisfecho
10	Muy satisfecho	20	Muy satisfecho	30	Muy satisfecho

Tabla 35

Valores de estado y frecuencia post-prueba

Estado	Frecuencia
Muy satisfecho	18
No satisfecho	12

Tabla 36

Valores de estado y frecuencia post-prueba

Estado	Frecuencia
Muy satisfecho	18
Satisfecho	12
Poco satisfecho	0
No satisfecho	0
Total	30

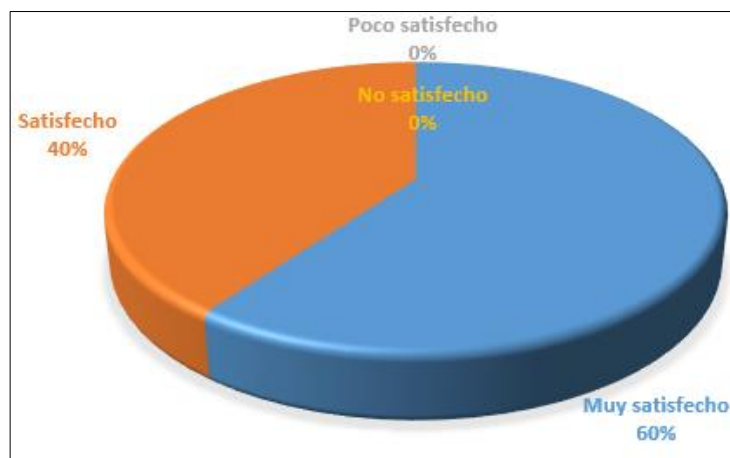


Figura 76. Porcentaje de post prueba del KPI4.

- 0% de las veces el nivel de satisfacción del paciente fue catalogado como no satisfecho y poco satisfecho por los pacientes atendidos.
- El 40 % de las veces el nivel de satisfacción del paciente fue catalogado como satisfecho por los pacientes atendidos.
- El 60 % de las veces el nivel de satisfacción del paciente fue catalogado como muy satisfecho por los pacientes atendidos.
- Se determina que el 100% de las veces el nivel de satisfacción es satisfecho.

4.6. Constrastación de hipótesis

En este caso, se presentan las medias de los KPIs para la Pre-Prueba y Post-Prueba. Resultados numéricos.

Tabla 37

Recurso de indicadores

Indicador	Pre – Prueba	Post - Prueba	Comentario
Tiempo para procesar el registro de los datos del paciente.	182 seg.	0.14 seg.	----
Tiempo para procesar la data	502.00 seg.	92 seg.	----
Tiempo para predecir el cáncer de mama.	244800 seg.	0.2336 seg.	----
Nivel de satisfacción del paciente.	Poco satisfeho	Satisfecho, Muy satisfecho	No contrastado. Indicador cualitativo

A. Contrastación para el indicador: Tiempo para procesar el registro de los datos de los pacientes. KPI₁

Se debe validar el impacto que tiene el uso del modelo predictivo en el tiempo para procesar el registro de los datos de los pacientes, llevado a cabo en la muestra. Se realiza una medición antes de utilizar el modelo predictivo (Pre-Prueba) y otra después de utilizar el modelo predictivo (Post-Prueba). La tabla contiene los tiempos para tomar los datos del paciente para las dos muestras.

Tabla 38

Valores de pre-prueba. KPII

Pre									
120.0	120.0	120.0	120.0	120.0	120.0	120.0	120.0	120.0	120.0
180.0	180.0	180.0	180.0	180.0	180.0	180.0	180.0	180.0	180.0
120.0	120.0	120.0	120.0	120.0	120.0	120.0	120.0	120.0	120.0

Tabla 39

Valores de post-prueba. KPII

Post									
0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23

- **Hipótesis alternativa (H_a):**

El uso del modelo predictivo determinará el tiempo para procesar el registro de los datos del paciente (Post Prueba) con respecto al porcentaje que no se aplicó (Pre Prueba).

- **Hipótesis nula (H₀):**

El uso del modelo predictivo no determina el tiempo para procesar el registro de los datos del paciente (Post prueba) con respecto al porcentaje que no se aplicó (Pre Prueba).

a) Planteamiento de la hipótesis:

μ_1 = Media del tiempo para procesar el registro de los datos del paciente. Pre Prueba.

μ_2 = Media del tiempo para procesar el registro de los datos del paciente Post Prueba.

H₀: $\mu_1 = \mu_2$ H_a: $\mu_1 < \mu_2$

b) Criterios de decisión:

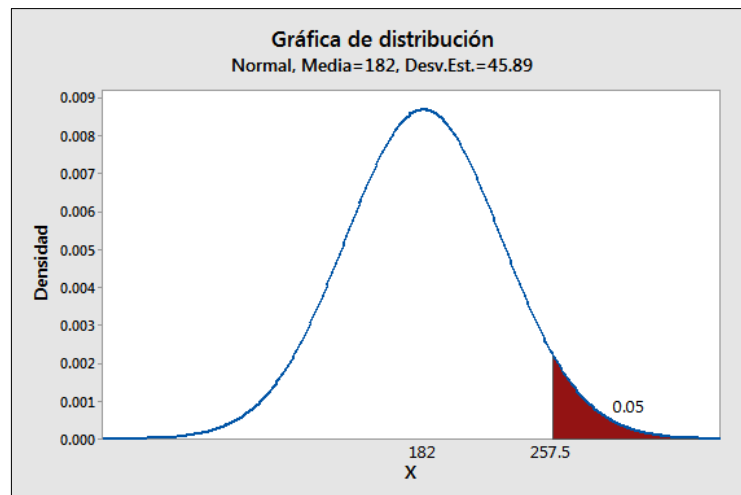


Figura 77. Gráfica de distribución KPI1.

c) Cálculo:

Para las pruebas se utilizó la técnica de Mann-Whitney por ser dos datos independientes no paramétricos.

Tabla 40

Estadísticas descriptivas

	Pre – Prueba	Post - Prueba
Media (\bar{x})	182.00	0.13567
Desviación Estándar (S)	45.89	0.03048
Observaciones	30	30
Estimación de la diferencia		179.87
Confianza lograda		94.9
IC de 95% para la diferencia		(179.86, 179.89)
W		1365.0
Prueba significativa (ajustado por empate)		0.000

d) Decisión Estadística:

Puesto que el valor- $p=0.000 < \alpha=0.05$, los resultados proporcionan suficiente evidencia para rechazar la hipótesis nula (H_0), y aceptar la hipótesis alterna (H_a) es cierta. La prueba resultó ser significativa con un nivel de confianza de 95%.

A. Contrastación para el indicador: Tiempo para procesar la data. KPI₂

Se debe validar el impacto que tiene el uso del modelo predictivo en el tiempo para procesar la data, llevado a cabo en la muestra. Se realiza una medición antes de utilizar el Modelo Predictivo (Pre-Prueba) y otra después de utilizar el Modelo Predictivo (Post-Prueba). La tabla contiene los Tiempos para tomar los datos del paciente para las dos muestras:

Tabla 41

Valores de pre-prueba. KPI2

Pre									
420.00	420.00	420.00	420.00	420.00	420.00	420.00	420.00	420.00	420.00
480.00	480.00	480.00	480.00	480.00	480.00	480.00	480.00	480.00	480.00
420.00	420.00	420.00	420.00	420.00	420.00	420.00	420.00	420.00	420.00

Tabla 42

Valores de post-prueba. KPI2

Post									
120.00	120.00	120.00	120.00	120.00	120.00	120.00	120.00	120.00	120.00
120.00	120.00	120.00	120.00	120.00	120.00	120.00	120.00	120.00	120.00
120.00	120.00	120.00	120.00	120.00	120.00	120.00	120.00	120.00	120.00

- **Hipótesis alternativa (Ha):**

El uso del modelo predictivo determinará el tiempo para procesar la data (Post-Prueba) con respecto a la cantidad que no se aplicó (Pre- Prueba).

- **Hipótesis nula (Ho):**

El uso del modelo predictivo no determina el tiempo para procesar la data (Post-prueba) con respecto a la cantidad que no se aplicó (Pre Prueba).

Solución:

a) Planteamiento de la hipótesis:

μ_1 = Media del para procesar la data. Pre-Prueba.

μ_2 = Media del tiempo para procesar la data. Post-Prueba.

Ho: $\mu_1 = \mu_2$ Ha: $\mu_1 < \mu_2$

b) Criterios de decisión:

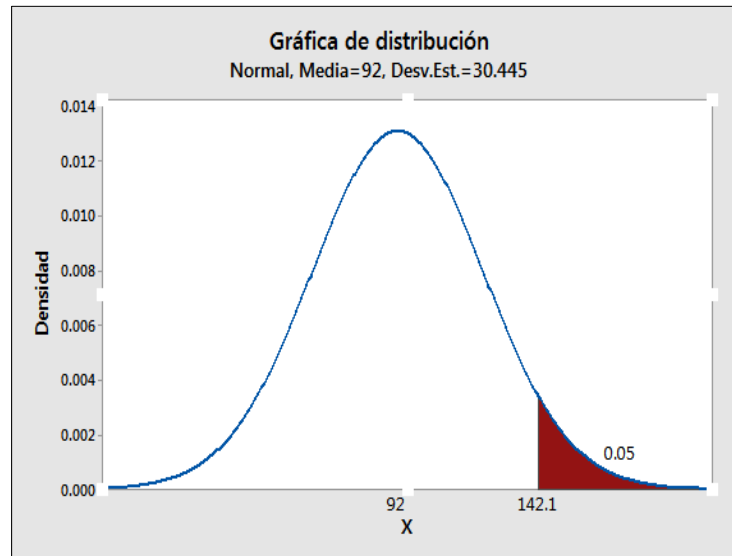


Figura 78. Gráfica de distribución KPI2.

c) Cálculo:

Para las pruebas se utilizó la técnica de Mann-Whitney por ser dos datos independientes no paramétricos.

Tabla 43

Estadísticas descriptivas

	Pre – Prueba	Post - Prueba
Media (x)	502.00	92.000
Desviación Estándar (S)	67.74	30.445
Observaciones	30	30
Estimación de la diferencia		420.00
Confianza lograda		94.9
IC de 95% para la diferencia		(360.00, 419.98)
W		1365.0
Prueba significativa (ajustado por empate)		0.000

d) Decisión estadística:

Puesto que el valor- $p=0.000 < \alpha=0.05$, los resultados proporcionan suficiente evidencia para rechazar la hipótesis nula (H_0), y aceptar la hipótesis alterna (H_a) es cierta. La prueba resultó ser significativa con un nivel de confianza de 95%.

B. Contrastación para el indicador: Tiempo para predecir el cáncer de mama.

KPI₃

Se debe validar el impacto que tiene el uso del modelo predictivo en el tiempo para predecir el cáncer de mama, llevado a cabo en la muestra. Se realiza una medición antes de utilizar el modelo predictivo (Pre-Prueba) y otra después de utilizar el modelo predictivo (Post-Prueba). La tabla contiene los tiempos para tomar los datos del paciente para las dos muestras:

Tabla 44

Valores de pre-prueba. KPI₃

Pre									
172800	172800	172800	172800	172800	172800	172800	172800	172800	172800
172800	172800	172800	172800	172800	172800	172800	172800	172800	172800
345600	345600	345600	345600	345600	345600	345600	345600	345600	345600

Tabla 45

Valores de post-prueba. KPI₃

Post									
0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23
0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19

• **Hipótesis alternativa (H_a):**

El uso del modelo predictivo determinará el tiempo para predecir el cáncer de mama (Post-Prueba) con respecto a la cantidad que no se aplicó (Pre-Prueba).

- **Hipótesis Nula (Ho):**

El uso del modelo predictivo no determina el tiempo predecir el cáncer de mama (Post-prueba) con respecto a la cantidad que no se aplicó (Pr- Prueba).

Solución:

a) Planteamiento de la hipótesis:

μ_1 = Media del tiempo para procesar la data. Pre-Prueba.

μ_2 = Media del tiempo para procesar la data. Post-Prueba.

Ho: $\mu_1 = \mu_2$ Ha: $\mu_1 < \mu_2$

b) Criterios de decisión:

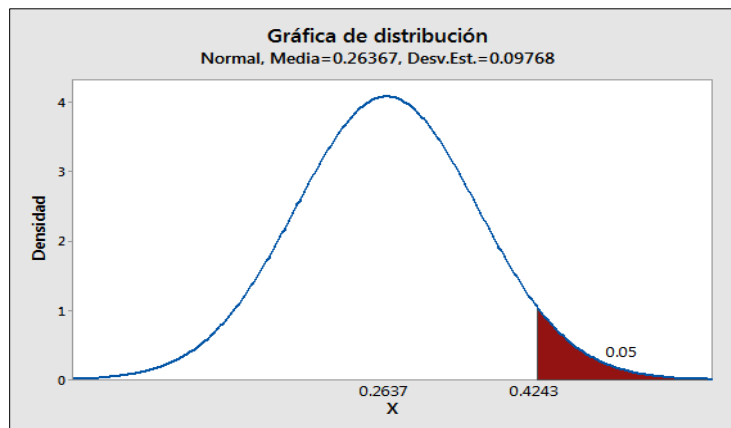


Figura 79. Gráfica de distribución KPI3.

c) Cálculo:

Para las pruebas se utilizó la técnica de Mann-Whitney por ser dos datos independientes no paramétricos.

Tabla 46

Estadísticas descriptivas

	Pre - Prueba	Post - Prueba
Media (\bar{x})	244800	026367
Desviación Estándar (S)	72050	0.09768
Observaciones	30	30
Estimación de la diferencia		259200
Confianza lograda		94.9
IC de 95% para la diferencia		(172800, 259200)
W		1365.0
Prueba significativa (ajustado por empate)		0.000

d) Decisión estadística:

Puesto que el valor- $p=0.000 < \alpha=0.05$, los resultados proporcionan suficiente evidencia para rechazar la hipótesis nula (H_0), y aceptar la hipótesis alterna (H_a) es cierta. La prueba resultó ser significativa con un nivel de confianza de 95%.

CAPÍTULO V
CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

- a) En nuestra primera conclusión, se determina que el tiempo promedio para procesar el registro de los datos de las pacientes del Hospital de la Solidaridad de Villa El Salvador, era de 182 segundos, lo cual con la implementación se llega a reducir en un tiempo promedio de 0.14 segundos que equivale a un 75%. Sustento según la tabla 26 de la estadística descriptiva del KPI 1.

- b) Como segunda conclusión se verifica que al inicio el tiempo para procesar la data era de 502 segundos, logrando reducir el tiempo en 92 segundos en procesar la data para cualquier tipo de operación permitido dentro del sistema lo cual equivale a un 85%. Sustento según tabla 27 de la estadística descriptiva del KPI 2.

- c) Como tercera conclusión, en el desarrollo del sistema del modelo predictivo, dio como resultado significativo en un 95% de reducción del tiempo para predecir el cáncer de mama, pasando de horas de espera (68 horas) y en segundos, la cantidad de 244800 a sólo 0.2637 segundos en promedio. Sustento según tabla 28 de la estadística descriptiva del KPI 3.

- d) Como cuarta conclusión, verificamos que al inicio de las evaluaciones se contaba con un 70% de satisfacción de las pacientes, sin embargo, con el desarrollo del modelo predictivo se logró demostrar el aumento del nivel de satisfacción a un 100% de las pacientes del Hospital de la Solidaridad de Villa El Salvador. Sustento según las tablas 29 y 30 de la estadística descriptiva del KPI 4.

5.2. Recomendaciones

- a) Se sugiere implementar y desarrollar esta iniciativa de proyecto con fines de mejorar parte de nuestro sector de salud de bajos recursos.
- b) Se aconseja investigar en el mercado sobre las nuevas tendencias en cuanto a las herramientas y/o metodologías que se usan en correlación con la minería de datos.
- c) Se propone realizar planes de contingencia post – implementación, para darle mayor continuidad y calidad al modelo predictivo.
- d) Se recomienda capacitar a los usuarios del sistema del modelo predictivo para beneficiar su funcionalidad y mejor uso para las pacientes del Hospital de la Solidaridad de Villa El Salvador.

REFERENCIAS BIBLIOGRÁFICAS

Artículos

- Becerra, F. (2015). Leptina y lactancia materna: Beneficios fisiológicos. *ResearchGate*, 63(1), 119-126. Recuperado de: https://www.researchgate.net/publication/276444842_Leptina_y_Lactancia_materna_Beneficios_fisiologicos
- Cardona, C., Gelvez, N. y Palacios, J. (2016). Análisis de datos mediante el algoritmo de clasificación J48, sobre un cluster en la nube de AWS. *Redes de Ingeniería*, 3-15. Recuperado de <https://revistas.udistrital.edu.co/index.php/REDES/article/view/11985/12615>
- Hernández, E. y Lorente, R. (2014). *Minería de datos aplicada a la detección de cáncer de mama*. Universidad Carlos III de Madrid. Recuperado de <http://www.it.uc3m.es/jvillena/irc/practicass/08-09/14.pdf>
- Martínez, J. (2017). Análisis de minería de datos distribuida con Weka Parallel en computadoras con múltiples procesadores físicos y lógicos. *Economía y Administración (E&A)*, 6(2), 155-166. Recuperado de: <https://www.lamjol.info/index.php/EyA/article/view/4307>
- Organización Mundial de la Salud. (2013). El enigma del cáncer de mama. *Boletín de la Organización Mundial de la Salud*, 91(9), 621-715. Recuperado de: <http://www.who.int/bulletin/volumes/91/9/13-020913/es/>
- Vargas, Z. (2009). La Investigación Aplicada: una forma de conocer las realidades con evidencia científica. *Revista Educación*, 33(1), 155-165. Recuperado de: <https://www.redalyc.org/pdf/440/44015082010.pdf>

Blog

Córdoba, L. (16 de junio del 2011). *Minería de datos*. [Mensaje en un blog]. Recuperado de: <http://cor-mineriadedatos.blogspot.com/2011/06/weka.html>

Periódicos

El Diario.es (19 de octubre de 2016). La tasa de supervivencia del cáncer de mama, del 100% si se diagnostica pronto. *El Diario. es*. Recuperado de: https://www.eldiario.es/sociedad/tasa-supervivencia-cancermama_0_571143004.html

El Comercio. (25 de abril de 2016). Tratamiento contra cáncer de mama más agresivo llega al Perú. *El Comercio*. Recuperado de: <https://elcomercio.pe/tecnologia/ciencias/tratamiento-cancer-mama-agresivo-llega-peru-193785-noticia/>

Sausa, M. (19 de octubre de 2017). Cáncer de mama: en el Perú los casos se detectan desde los 25 años. *Perú 21*. Recuperado de: <https://peru21.pe/vida/salud/cancer-mama-peru-casos-detectan-25-anos-380782-noticia/>

Wong, S. (12 de mayo de 2014). Casos de cáncer de mama en aumento en el Perú. *La Prensa*. Recuperado de <https://laprensa.peru.com/actualidad/noticia-cancer-mama-crecimiento-economico-peru-25462>

Sitios Web

American Cancer Society. (10 de diciembre de 2019). *Tasas de supervivencia del cáncer de seno*. Atlanta: Cancer.org. Recuperado de: <https://www.cancer.org/es/cancer/cancer-de-seno/compreension-deundiagnostico-de-cancer-de-seno/tasas-de-supervivencia-del-cancer-de-seno.html>

American Cancer Society. (setiembre, 2018). *Cáncer de seno*. Atlanta: Cancer.org. Recuperado de: <https://www.cancer.org/es/cancer/cancer-de-seno/acerca/que-tan-comun-es-el-cancer-de-seno.html>

Asociación Gastro-Onco Latinoamericana. (2016). *Definición de sobrevivencia*. Mérida; Oncología Araujo. Recuperado de <http://gastrooncologo.com/sobrevivencia/>

Breastcancer. Org. (2 de diciembre de 2015). *Estadios del cáncer de mama*. Filadelfia: Breastcancer. Org. Recuperado de: https://www.breastcancer.org/es/sintomas/diagnostico/estadios?gclid=EAIaIQobChMIp83F79zG6gIVGI7ICh2Z_wIeEAAYASAAEgJwAfD_BwE

Cancer.net (octubre, 2018). *Cáncer de mama: Diagnóstico*. Washington D.C.: American Society of Clinical Oncology. Recuperado de <https://www.cancer.net/es/tipos-de-cancer/cancer-de-mama/diagnostico>

Mayo Clinic. (22 de noviembre de 2019). *Cáncer de mama*. Minnesota: Mayo Clinic. Recuperado de: <https://www.mayoclinic.org/es-es/diseases-conditions/breast-cancer/symptoms-causes/syc-20352470>

Loshin, D. (7 de agosto de 2017). *Modelado predictivo exitoso requiere conjuntos correctos de datos para analizar*. Atlanta: TechTarget. Recuperado de: <https://searchdatacenter.techtarget.com/es/consejo/Modelado-predictivo-exitoso-requiere-conjuntos-correctos-de-datos-para-analizar>

Organización Mundial de la Salud. (2020). *Cáncer de mama: prevención y control*. Ginebra: OMS. Recuperado de: <http://www.who.int/topics/cancer/breastcancer/es/>

Organización Panamericana de la Salud – OPS. (2013). *Cáncer de mama*. Washington D.C.: Paho.org. Recuperado de: https://www.paho.org/hq/index.php?option=com_content&view=article&id=50

41:2011-breast-cancer&Itemid=3639&lang=es

Sisol Salud. (2020). *Sisol Salud de Villa el Salvador*. Lima: SISOL. Recuperado de:
<http://www.sisol.gob.pe/lima/villa-el-salvador/>

Tesis

Álvarez, O. (2015). *Particularidades del afrontamiento en mujeres con cáncer de mama no metastásico en intervalo libre de enfermedad* (Tesis de maestría). Recuperado de
<https://dspace.uclv.edu.cu/bitstream/handle/123456789/3819/Tesis%20Maestria%20Olga%20Lidia%20Alvarez.pdf?sequence=1&isAllowed=y>

Araneda, P. (2016). *Diseño y modelo preliminar de una plataforma de integración de datos clínicos y genómicos: aplicaciones en Alzheimer y cáncer de mama* (Tesis de maestría). Recuperado de
<https://pdfs.semanticscholar.org/37f3/694b00d4be70654a679b9be46b7528ac56ae.pdf>

Ascencio, A. (2013). *Resiliencia y estilos de afrontamiento en mujeres con cáncer de un hospital estatal de Chiclayo 2013* (Tesis de pregrado). Recuperado de
http://tesis.usat.edu.pe/bitstream/20.500.12423/336/1/TL_Ascencio_Puicon_AnisJackelin.pdf

Batallanos, M. (2018). *Factores asociados a cáncer de mama en mujeres posmenopáusicas del Hospital Nacional Dos De Mayo 2016-2017* (Tesis de pregrado). Recuperado de
<http://repositorio.urp.edu.pe/bitstream/handle/URP/1205/16-TEISIS-MBatallanos.pdf?sequence=1&isAllowed=y>

Boone, M. (2011). *Identificación de regiones con altas tasas de incidencia de cáncer mediante la integración y uso de técnicas de la minería de datos: almacenes de datos, agrupamiento y sistema de la información geográficos*. (Tesis de

maestría). Recuperado de
<http://www.cenidet.edu.mx/subplan/biblio/seleccion/Tesis/MC%20Maria%20del%20Rocio%20Bonne%20Rojas%202011.pdf>

Burriel, V. (2017). *Diseño y Desarrollo de un Sistema de Información para la gestión de información sobre cáncer de mama* (Tesis doctoral). Recuperado de <https://cutt.ly/0rt0TZG>

Camacho, S. (2016). *Método Heurístico para el Diagnóstico de Cáncer de Mama basado en Minería de Datos* (Tesis de maestría). Recuperado de <https://repositorio.umsa.bo/bitstream/handle/123456789/10786/T.3254.pdf?sequence=1&isAllowed=y>

Camborda, M. (2014). *Aplicación de árboles de decisión para la predicción del rendimiento académico de los estudiantes de los primeros ciclos de la carrera de ingeniería civil de la universidad continental* (Tesis de pregrado). Recuperado de <http://repositorio.uncp.edu.pe/bitstream/handle/UNCP/1477/Tesis%20Maria%20Gabriela%20Camborda%20Zamudio.pdf?sequence=1&isAllowed=y>

Bocanegra, L. (2016). *Hallazgos radiológicos en estudios de screening, cribado para detectar cáncer de mama en el Hospital I Octavio Mongrut Muñoz en el periodo de enero – diciembre del 2014* (Tesis de pregrado). Recuperado de http://repositorio.urp.edu.pe/bitstream/handle/urp/486/Bocanegra_1.pdf?sequence=1&isAllowed=y

Galán, V. (2015). *Aplicación de la metodología CRISP-DM a un proyecto de minería de datos en el entorno universo* (Tesis de pregrado). Recuperado de: https://e-archivo.uc3m.es/bitstream/handle/10016/22198/PFC_Victor_Galan_Cortina.pdf

- Gallardo, J. (2009). *Metodología para la definición de requisitos en proyectos de data mining* (Tesis doctoral). Recuperado de: http://oa.upm.es/1946/1/JOSE_ALBERTO_GALLARDO_ARANCIBIA.pdf
- León, H. (2018). *Desarrollo de un modelo algorítmico basado en árboles de decisión para la predicción de la permanencia de un paciente en un proceso psicoterapéutico* (Tesis de pregrado). Recuperado de: http://tesis.pucp.edu.pe/repositorio/bitstream/handle/20.500.12404/11868/leon_heli_modelo_algoritmico_arboles_decision_prediccion.pdf?sequence=1&isallowed=y
- Manrique, J. (2016). *Estudio mamográfico de tamizaje e histológico para el riesgo temprano del cáncer de mama en el centro de prevención y detección del cáncer del Inén* (Tesis de pregrado). Recuperado de http://repositorio.urp.edu.pe/bitstream/handle/urp/548/Manrique_J.pdf?sequence=1&isAllowed=y
- Moreno, F. (junio, 2016). *Introducción de Minería de Datos con Weka: Aplicación a un problema económico* (Tesis de pregrado). Recuperado de: <http://tauja.ujaen.es/bitstream/10953.1/6984/1/TFG%20-%20navas%20moreno%2C%20%20Francisco.pdf>
- Ochoa, M. (2016). *Uso de técnicas de educación para el entendimiento del negocio* (Tesis de maestría). Recuperado de: <http://laboratorios.fi.uba.ar/lfi/rgm/tesis/tesisdemagister.pdf>
- Saura, C. (2017). *Tratamiento del cáncer de mama diagnosticado durante el embarazo* (Tesis doctoral). Recuperado de <https://www.tdx.cat/bitstream/handle/10803/454826/csm1de1.pdf?sequence=1&isAllowed=y>

ANEXOS

Anexo 1: Matriz de consistencia

Título: “Aplicación del algoritmo para predecir el riesgo de cáncer de mama en las pacientes del Hospital de la Solidaridad de villa el salvador”

PROBLEMA	OBJETIVOS	HIPÓTESIS	VARIABLES	INDICADORES	ÍNDICE	UNIDAD	
¿En qué medida la aplicación del algoritmo mejorará el proceso de predicción del cáncer de mama en las pacientes del Hospital Solidaridad de Villa El Salvador?	Determinar en qué la aplicación del algoritmo mejorará el proceso de predicción del cáncer de mama en las pacientes del Hospital Solidaridad de Villa El Salvador	Si se implementa un algoritmo basado en minería de datos, entonces se influye positivamente en la predicción de cáncer de mama de las pacientes del Hospital de la Solidaridad de Villa El Salvador.	Variable Independiente	Presencia-Ausencia	-----	-----	Tipo de Investigación
			Algoritmo				-Aplicada
			Variable Dependiente	Tiempo para procesar el registro de los datos del paciente	[0.10...0.20]	Proceso	Nivel de investigación
			Predicción del cáncer de mama basado en minería de datos de las pacientes del Hospital de la Solidaridad de Villa El Salvador.	Tiempo para procesar la data	[60...120]	Proceso	-Explicativa - Correlacional
				Tiempo para predecir el cáncer de mama	[12...50]	Proceso	Diseño de investigación
				Nivel de Satisfacción del paciente		Encuesta	G _e O ₁ X O ₂
							Población Pacientes

					<p>No satisfecho, Poco satisfecho, Satisfecho, Muy satisfecho.</p>		<p>registrados y nuevos del Hospital de la Solidaridad de Villa El Salvador. N = Indeterminad o</p> <p>Muestra Se toma una muestra con un valor de 30 historiales de las pacientes del Hospital de la Solidaridad de Villa El Salvador.</p>
--	--	--	--	--	--	--	---

Anexo 2: Formato de encuesta

ENCUESTA

Fecha: Día Mes Año

Nombre del Establecimiento de Salud:

La presente encuesta pretende valorar cual es la relación entre el nivel de conocimiento sobre el cáncer de mama y el autoexamen mamario, así como las actitudes con la práctica del autoexamen en las mujeres derechohabientes del Hospital de la Solidaridad de Villa El Salvador.

Aspectos Generales

¿Cuál es su edad? Años Sexo M F

¿En dónde vive?

Cuál es su Grado de Instrucción:

1 = Analfabeto

4 = Super

1 = Analfabeto

4 = Superior técnica

2 = Primaria

5 = Superior Universitaria

3 = Secundaria

6 = No sabe/ no opina

¿Por qué escogió venir a este establecimiento de salud?	Está más cerca <input type="checkbox"/>	No tengo seguro <input type="checkbox"/>
	Es barato <input type="checkbox"/>	Hay buena atención <input type="checkbox"/>
	No hay otro Establecimiento adonde acudir <input type="checkbox"/>	Otro: <input type="checkbox"/>
	Cuál?	

Parte I: Percepción del paciente respecto a la atención recibida.

1. ¿Durante su permanencia en el establecimiento cómo lo trataron? Bien Mal

2. ¿El personal de informes le orientó y explicó de manera clara y adecuada sobre los pasos o trámites para la atención en consulta externa? Sí No

3. ¿Su atención se realiza respetando la programación y el orden de llegada? Sí No

3.1 Si es No, ¿Por que?.....

4. El tiempo que usted esperó para ser atendido fue: Mucho Adecuado Poco

5. En términos generales ¿Ud. se siente satisfecho con la atención recibida? Sí No

5.1.¿Porque?.....

Parte II. Conocimiento sobre el cáncer de mama

1.- ¿Qué conoce sobre el cáncer de mama? Responda al enunciado correcto

El cáncer de mama es la presencia de tumores malignos en la mama. ()

El cáncer de mama afecta solo a mujeres. ()

Una mujer gestante (embarazada) puede sufrir cáncer de mama. ()

Los tumores malignos de la mama pueden invadir otros órganos. ()

Los tumores benignos son característicos del cáncer de mama. ()

El cáncer de mama es curable si se detecta tempranamente. ()

2.- ¿Cuáles son los signos y síntomas que Usted conoce acerca del cáncer de mama?

Coloque una (X) según crea usted conveniente:

Piel hundida de la mama o pezón ()

Diferencias de tamaño y forma de ambas mamas. ()

Enrojecimiento y endurecimiento de la mama. ()

Mamas y pezón con heridas. ()

Secreción de sangre o pus por el pezón ()

Presencia de tumores (bultos) en la mama, axila y/o alrededor del cuello ()

3.-¿Cuál de las siguientes situaciones aumenta el riesgo de cáncer de mama? Responda

con una (X) al enunciado correcto

Tener familiares por parte de madre que han sufrido cáncer de mama. ()

Beber alcohol y/o fumar en forma habitual. ()

Ser obesas y realizar poca actividad física. ()

Tener muchos hijos. ()

Presentar la primera menstruación antes de los 12 años. ()

Ser una mujer menos de 30 años. ()

Usar hormonas frecuentemente. ()

4.- Con respecto a las técnicas de detección del cáncer de mama coloque una (X) en los recuadros según usted crea conveniente.

El autoexamen de mamas es realizado por la misma mujer. ()

El ultrasonido es útil para el estudio de la mama en mujeres menores de 35 años. ()

El autoexamen de las mamas se realiza mensualmente a partir de la aparición de la primera menstruación. ()

La mastografía debe realizarse anualmente en toda mujer mayor de 40 años. ()

El autoexamen de mamas puede realizarse en cualquier momento que la mujer lo desee. ()

La mujer mayor de 30 años debe acudir al médico o enfermera para su examen de mama de manera anual. ()

La tomografía es un método para detectar el cáncer de mama ()

Las mujeres que usan hormonas deben realizarse mamografía ()

5.- ¿Conoce cómo se realiza el autoexamen de mama? Existen ciertos pasos a seguir. De acuerdo a estos, coloque una (X) en los recuadros según corresponda.

Observar frente al espejo las características de la mama y el pezón ()

Las mamas se pueden observar con los brazos caídos, manos en la cintura o detrás de la cabeza. ()

Palpar sus manos con la yema de los dedos en forma circular para buscar la presencia de tumores (bultos). ()

Examinar y palpar las mamas por encima de la ropa interior ()

Colocar un brazo detrás de la nuca y con el otro revisar la mama opuesta, axila y luego cuello. ()

6.- ¿Cree usted que el doctor se demora demasiado al realizarle la entrevista de sus síntomas?

SI () Cuanto tiempo: _____

NO ()

7.- Cree usted que el médico se demora mucho tiempo en el diagnóstico?

SI () Cuanto tiempo: _____

NO ()

8.- Cree usted que los resultados del tipo de tratamiento que va recibir se toma demasiado tiempo?

SI () Cuanto tiempo: _____

NO ()

Parte III: Opiniones y sugerencias respecto de la atención recibida.

.....
.....
.....
.....
.....

GRACIAS!!

Anexo 4: Manual de usuario

MANUAL DE USUARIO

INTRODUCCIÓN

El presente manual, es de uso exclusivo para los trabajadores del sector de salud en el área de oncología del Hospital de la Solidaridad de Villa El Salvador.

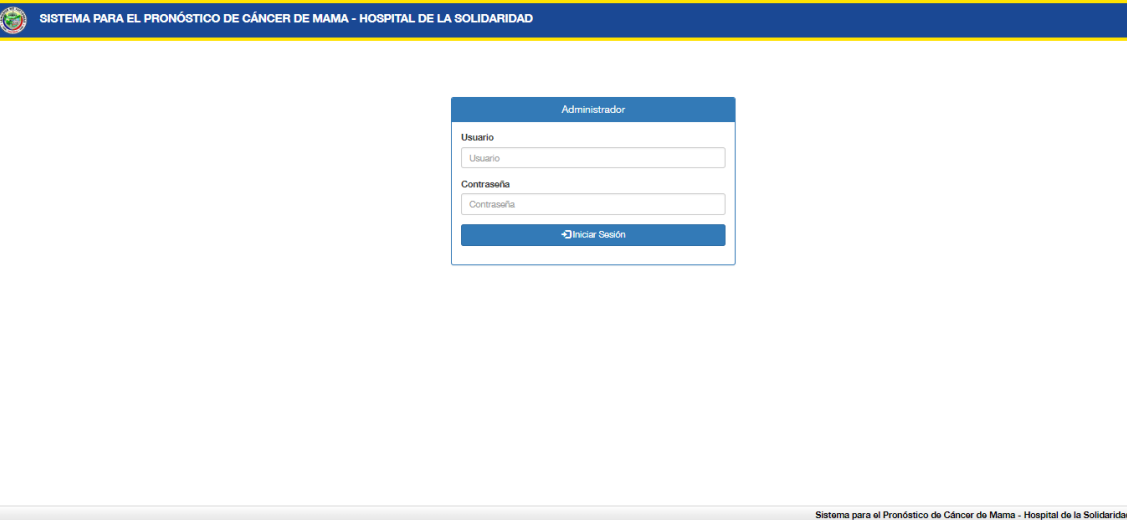
La finalidad del presente manual es gestionar el proceso y parte del diagnóstico en la detección de cáncer de mama en el Hospital de la Solidaridad de Villa El Salvador. Así los usuarios del área de oncología podrán trabajar de una forma más óptima y precisa, rápida obteniendo una gran satisfacción por las pacientes.

	MANUAL DE USUARIO	Código: HSVE - 01
		Versión 1
		Fecha: Febrero 2019

DESARROLLO DEL MANUAL DE USUARIO

2.1. Ingreso:

Esta es la vista de Login, deberá colocar su usuario y debajo su contraseña respectiva, luego presionar el botón Ingresar.



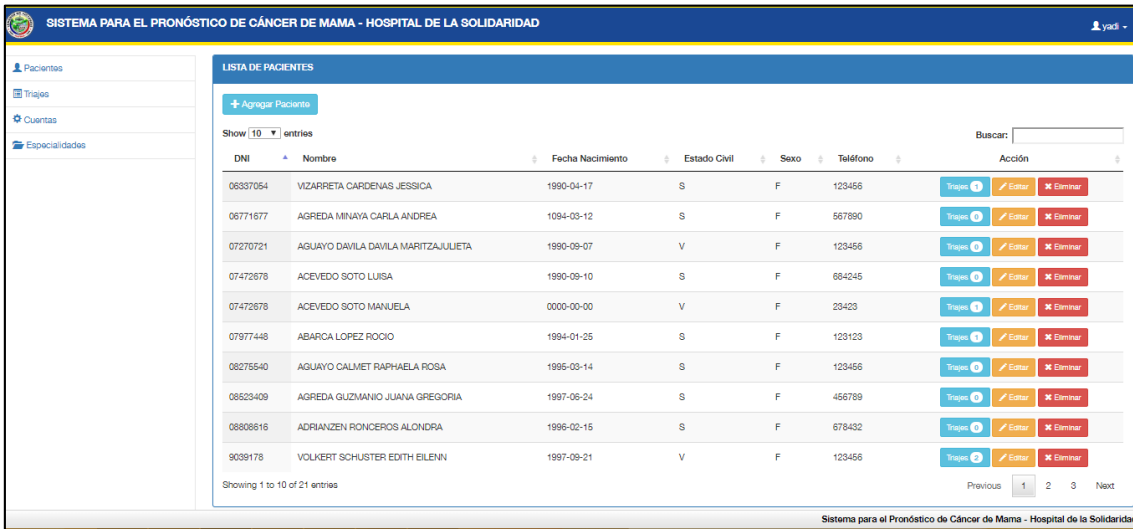
Ejemplo:



	MANUAL DE USUARIO	Código: HSVE - 01
		Versión 1
		Fecha: Febrero 2019

2.2. Inicio de Sesión

Una vez ingresada a la sesión, se mostrará el siguiente entorno de la Figura N° 3 donde podrá crear un nuevo paciente, ver los triajes de dicho paciente, editar y eliminar.



SISTEMA PARA EL PRONÓSTICO DE CÁNCER DE MAMA - HOSPITAL DE LA SOLIDARIDAD

LISTA DE PACIENTES

+ Agregar Paciente

Show 10 entries

Buscar:

DNI	Nombre	Fecha Nacimiento	Estado Civil	Sexo	Teléfono	Acción
06337054	VIZARRETA CARDENAS JESSICA	1990-04-17	S	F	123456	Triage Edit Eliminar
06771677	AGREDA MINAYA CARLA ANDREA	1094-03-12	S	F	567890	Triage Edit Eliminar
07270721	AGUAYO DAVILA DAVILA MARITZA JULIETA	1990-09-07	V	F	123456	Triage Edit Eliminar
07472678	ACEVEDO SOTO LUISA	1990-09-10	S	F	984245	Triage Edit Eliminar
07472678	ACEVEDO SOTO MANUELA	0000-00-00	V	F	23423	Triage Edit Eliminar
07977448	ABARCA LOPEZ ROCIO	1994-01-25	S	F	123123	Triage Edit Eliminar
08275540	AGUAYO CALMET RAPHAELA ROSA	1996-03-14	S	F	123456	Triage Edit Eliminar
08523409	AGREDA GUZMANIO JUANA GREGORIA	1997-06-24	S	F	456789	Triage Edit Eliminar
08808616	ADRIANZEN RONCEROS ALONDRA	1996-02-15	S	F	678432	Triage Edit Eliminar
9039178	VOLKERT SCHUSTER EDITH EILENN	1997-09-21	V	F	123456	Triage Edit Eliminar

Showing 1 to 10 of 21 entries

Previous 1 2 3 Next

Sistema para el Pronóstico de Cáncer de Mama - Hospital de la Solidaridad

2.3. Búsqueda de Paciente

Ingresaremos a la opción Pacientes, para realizar la búsqueda de un paciente que anteriormente ha sido atendido en el Hospital, buscaremos por su nombre o Número de Dni como se muestra en las siguientes capturas de pantallas.



SISTEMA PARA EL PRONÓSTICO DE CÁNCER DE MAMA - HOSPITAL DE LA SOLIDARIDAD

LISTA DE PACIENTES

+ Agregar Paciente

Show 10 entries

Buscar: abensur

DNI	Nombre	Fecha Nacimiento	Estado Civil	Sexo	Teléfono	Acción
10315202	ABENSUR MARTINEZ MILAGROS	1990-04-30	S	F	678097	Triage Edit Eliminar
10315202	ABENSUR MARTINEZ MAYRA	2016-08-21	S	F	11113	Triage Edit Eliminar

Showing 1 to 2 of 2 entries (filtered from 21 total entries)

Previous 1 Next

	MANUAL DE USUARIO	Código: HSVE - 01
		Versión 1
		Fecha: Febrero 2019

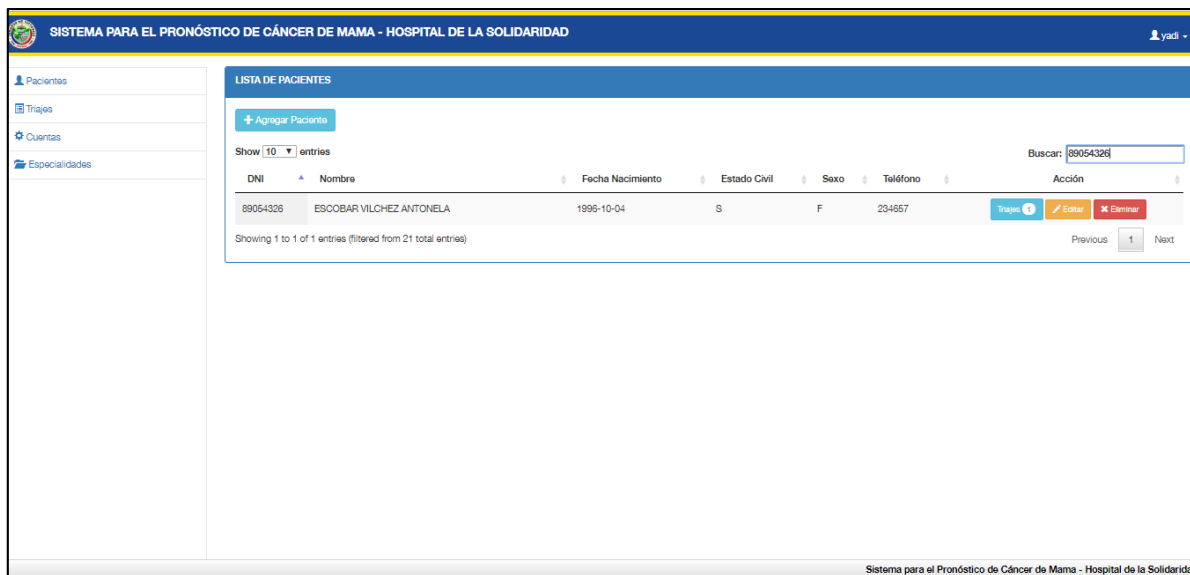
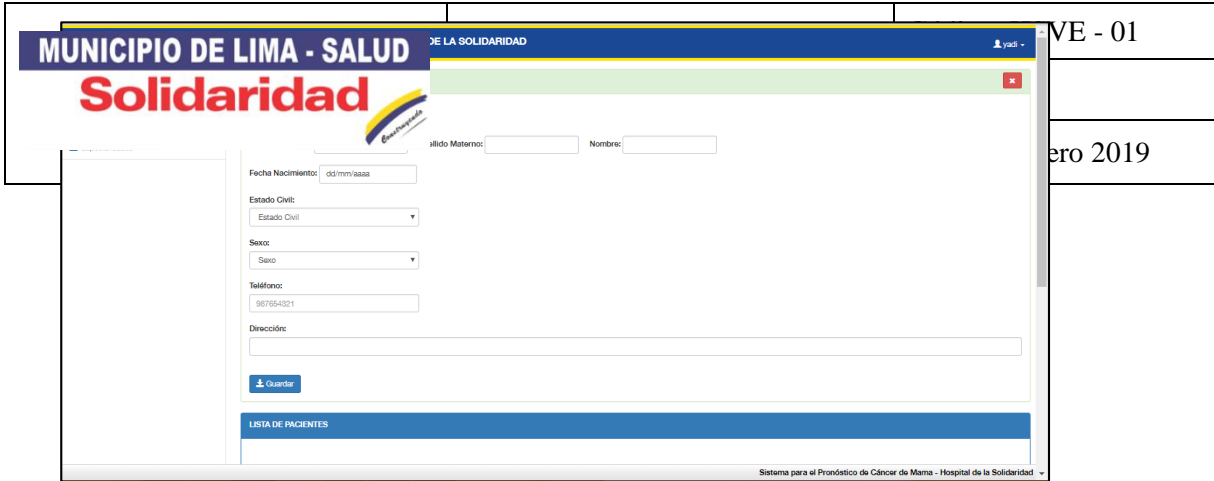


Figura N° 5

Como se ha podido observar en la imagen N° 5 se ha realizado la búsqueda en automático al colocar el número de Dni del paciente. Cabe mencionar que se visualiza botones como Triage, Editar, Eliminar.

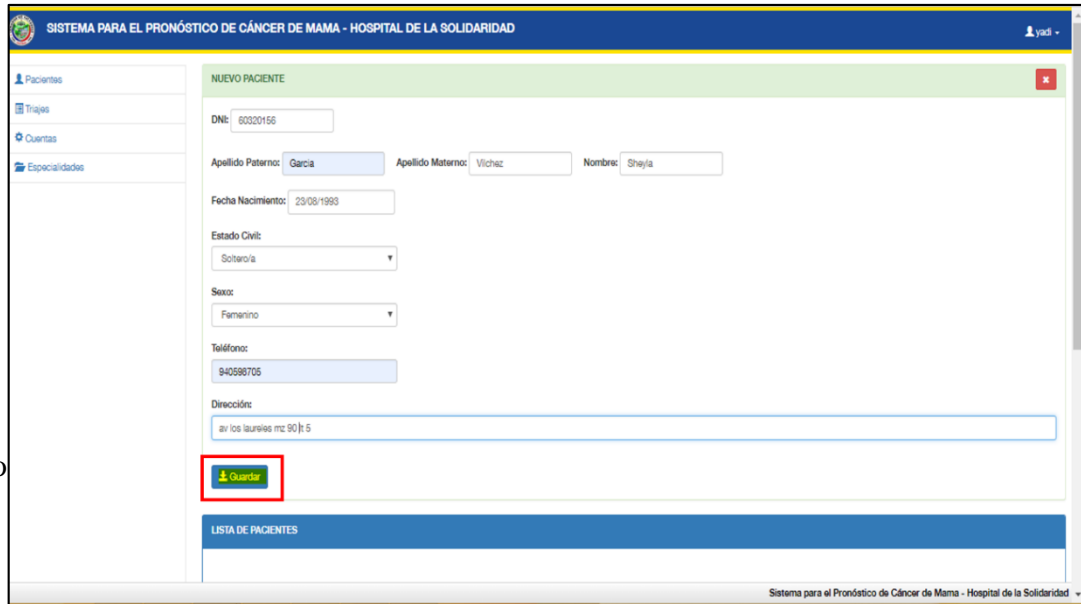
2.4. Crear Pacientes

En el entorno paciente visualizará el botón “Agregar Paciente” y deberá colocar los datos requeridos y al final botón Guardar como se muestra en las siguientes figuras de ejemplo.



VE - 01

ero 2019



Como

SISTEMA PARA EL PRONÓSTICO DE CÁNCER DE MAMA - HOSPITAL DE LA SOLIDARIDAD yadi

- [Pacientes](#)
- [Triajes](#)
- [Cuentas](#)
- [Especialidades](#)

LISTA DE PACIENTES

[+ Agregar Paciente](#)

Show entries Buscar: vil

DNI	Nombre	Fecha Nacimiento	Estado Civil	Sexo	Teléfono	Acción
07270721	AGUAYO DAVILA DAVILA MARITZAJULIETA	1990-09-07	V	F	123456	Triage Editar Eliminar
10083219	VILA CASTILLO ROXANA	1990-02-24	S	F	---	Triage Editar Eliminar
10319012	AGUAYO DAVILA JULIA ROSA	1985-03-16	S	F	435678	Triage Editar Eliminar
80320198	García Vilchez Shelya	1993-08-23	S	F	940698705	Triage Editar Eliminar
88054326	ESDOBAR VILCHEZ ANTONELA	1996-10-04	S	F	234667	Triage Editar Eliminar

Showing 1 to 5 of 5 entries (filtered from 22 total entries) Previous Next

Sistema para el Pronóstico de Cáncer de Mama - Hospital de la Solidaridad

2.5. Visualizar y crear triajes de Paciente

Ingresaremos a la opción de Triage para visualizar el historial del paciente ya sea en cualquier área que haya sido atendido en el Hospital.

SISTEMA PARA EL PRONÓSTICO DE CÁNCER DE MAMA - HOSPITAL DE LA SOLIDARIDAD yadi

- [Pacientes](#)
- [Triajes](#)
- [Cuentas](#)
- [Especialidades](#)

LISTA DE TRIAJES

[+ Agregar Triage](#)

Show entries Buscar:

DNI	Nombre	Especialidad	Fecha Hora Registro	Descripción	Acción
9039178	VOLKERT SCHUSTER EDITH EILENN	Oncología	2018-12-08 18:27:38	2	Crear Eliminar
9039178	VOLKERT SCHUSTER EDITH EILENN	Oncología	2018-12-08 18:15:03	ok	Detalles Eliminar

Showing 1 to 2 of 2 entries Previous Next

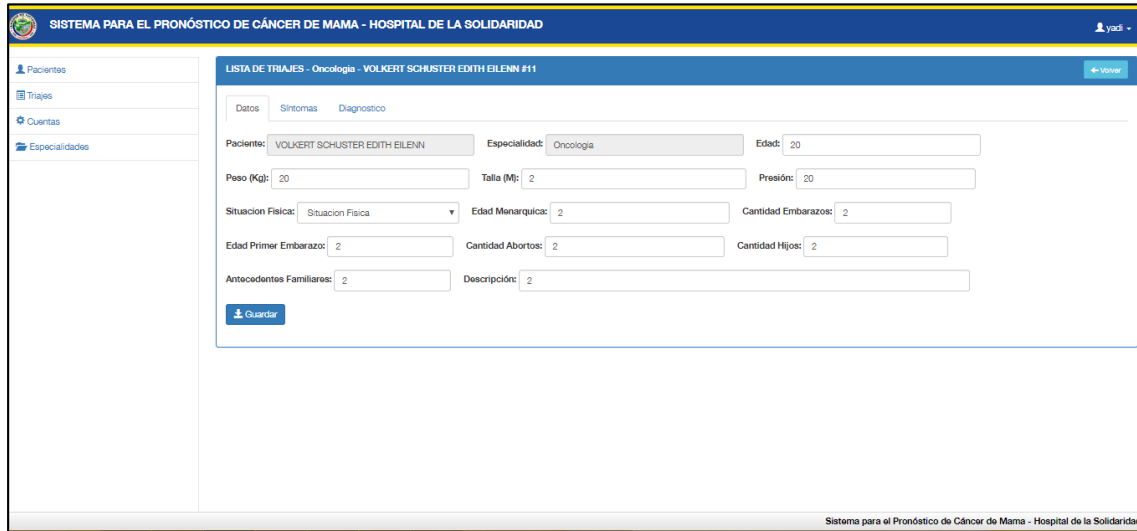
Podremos apreciar el botón detalles y crear triajes.

Detalle:

Visualizaremos tres opciones: Datos, Síntomas y Diagnostico para predecir el cáncer de mama

	MANUAL DE USUARIO	Código: HSVE - 01
		Versión 1
		Fecha: Febrero 2019

Datos: Se podrá apreciar el detalle del paciente actualmente ingresado.



SISTEMA PARA EL PRONÓSTICO DE CÁNCER DE MAMA - HOSPITAL DE LA SOLIDARIDAD

LISTA DE TRIAJES - Oncología - VOLKERT SCHUSTER EDITH ELENIN #11

Datos Síntomas Diagnostico

Paciente: VOLKERT SCHUSTER EDITH ELENIN Especialidad: Oncología Edad: 20

Peso (Kg): 20 Talla (M): 2 Presión: 20

Situación Física: Situación Física Edad Menarquia: 2 Cantidad Embarazos: 2

Edad Primer Embarazo: 2 Cantidad Abortos: 2 Cantidad Hijos: 2

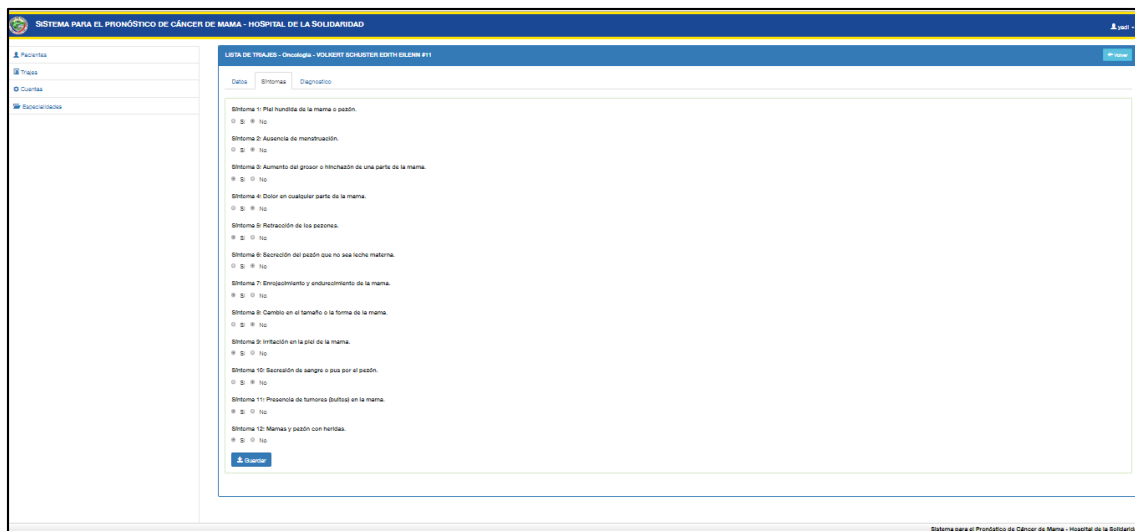
Antecedentes Familiares: 2 Descripción: 2

[Guardar](#)

Sistema para el Pronóstico de Cáncer de Mama - Hospital de la Solidaridad

Síntomas:

Síntomas de cáncer de mama: Presionar Si/No de acuerdo a las respuestas obtenidas por el paciente.



SISTEMA PARA EL PRONÓSTICO DE CÁNCER DE MAMA - HOSPITAL DE LA SOLIDARIDAD

LISTA DE TRIAJES - Oncología - VOLKERT SCHUSTER EDITH ELENIN #11

Datos Síntomas Diagnostico

Síntoma 1: Mal humor de la mama a pecho.
 Si No

Síntoma 2: Ausencia de menstruación.
 Si No

Síntoma 3: Aumento del grosor o hinchazón de una parte de la mama.
 Si No

Síntoma 4: Dolor en cualquier parte de la mama.
 Si No

Síntoma 5: Formación de los pecanos.
 Si No

Síntoma 6: Secreción del pezón que no sea leche materna.
 Si No

Síntoma 7: Enrojecimiento y endurecimiento de la mama.
 Si No

Síntoma 8: Cambio en el tamaño o la forma de la mama.
 Si No

Síntoma 9: Irritación en la piel de la mama.
 Si No

Síntoma 10: Secreción de sangre o pus por el pezón.
 Si No

Síntoma 11: Presencia de tumores (bultos) en la mama.
 Si No

Síntoma 12: Mamas y pezón con heridas.
 Si No

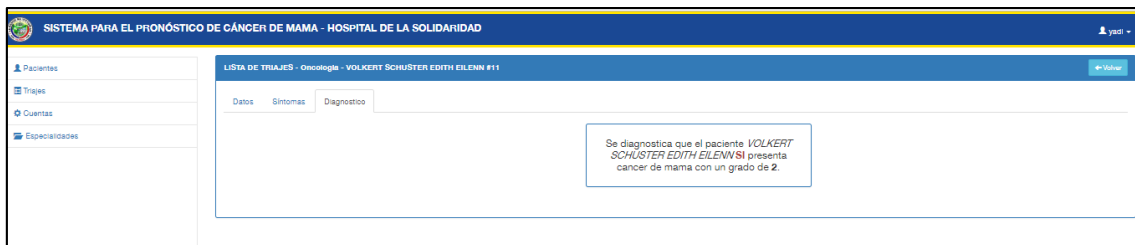
[Guardar](#)

Sistema para el Pronóstico de Cáncer de Mama - Hospital de la Solidaridad

	MANUAL DE USUARIO	Código: HSVE - 01
		Versión 1
		Fecha: Febrero 2019

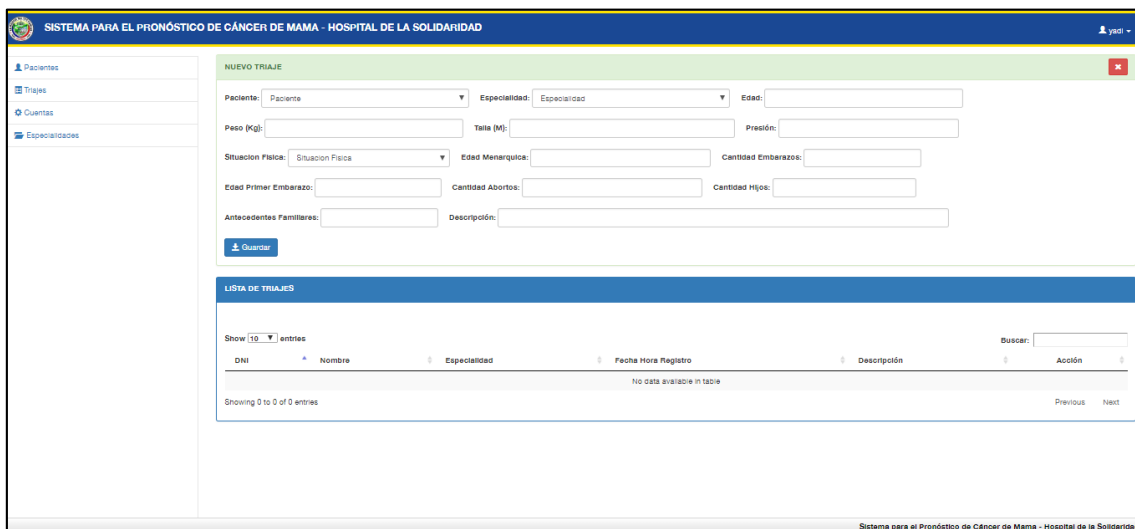
Una vez ingresado todas las encuestas realizadas al paciente con los descartes respectivos sobre el cáncer de mama, previamente figurara el Botón Predecir para obtener un diagnostico si el paciente tenga posiblemente Cáncer de Mama.

Diagnóstico



The screenshot shows a web application interface. At the top, it says 'SISTEMA PARA EL PRONÓSTICO DE CÁNCER DE MAMA - HOSPITAL DE LA SOLIDARIDAD'. On the left is a navigation menu with 'Pacientes', 'Triajes', 'Cuentas', and 'Especialidades'. The main content area is titled 'LISTA DE TRIAJES - Oncología - VOLKERT SCHUSTER EDITH ELENNSI #11'. There are tabs for 'Datos', 'Síntomas', and 'Diagnostico'. A central message box states: 'Se diagnostica que el paciente VOLKERT SCHUSTER EDITH ELENNSI presenta cáncer de mama con un grado de 2.'

Agregar triaje



The screenshot shows the 'NUEVO TRIAJE' (New Triage) form. It includes fields for:

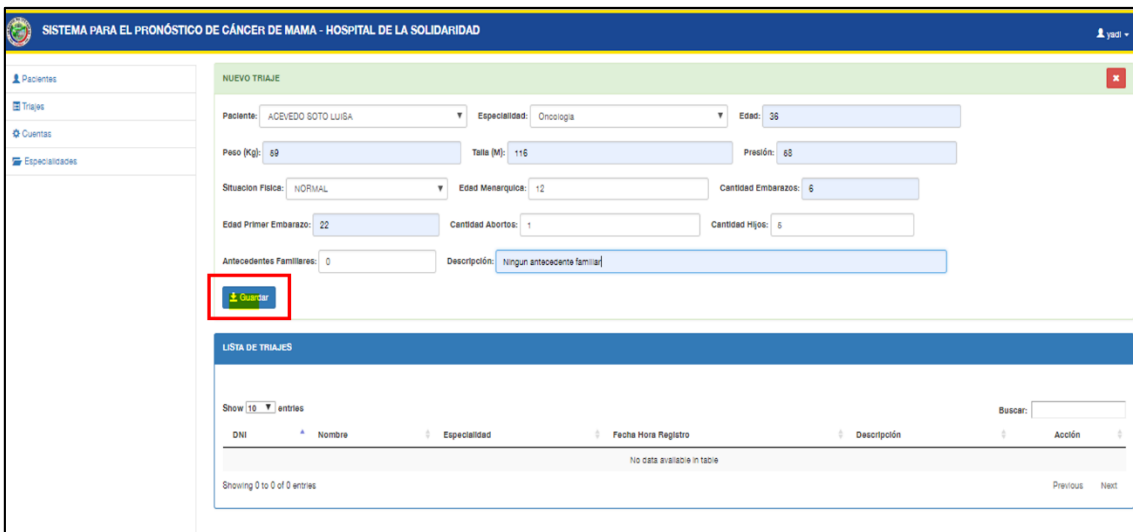
- Paciente:
- Especialidad:
- Edad:
- Peso (Kg):
- Talla (M):
- Presión:
- Situación Física:
- Edad Menarquia:
- Cantidad Embarazos:
- Edad Primer Embarazo:
- Cantidad Abortos:
- Cantidad Hijos:
- Antecedentes Familiares:
- Descripción:

 Below the form is a 'Guardar' button. Underneath is a 'LISTA DE TRIAJES' section with a search bar and a table. The table has columns: DNI, Nombre, Especialidad, Fecha Hora Registro, Descripción, and Acción. The table is currently empty, showing 'No data available in table'.

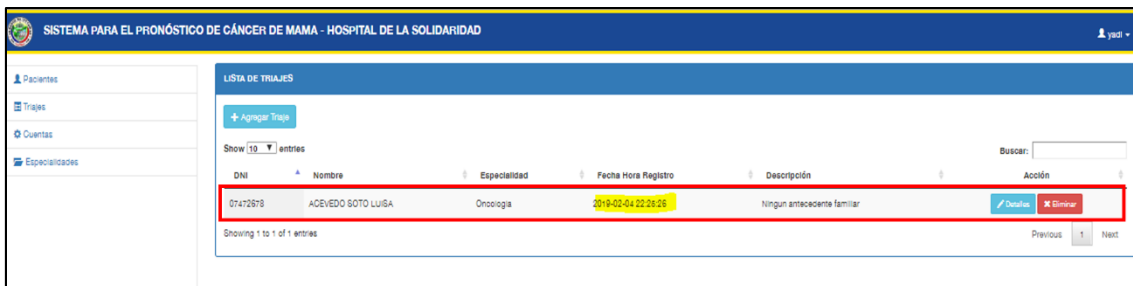
	MANUAL DE USUARIO	Código: HSVE - 01
		Versión 1
		Fecha: Febrero 2019

Ejemplo

Como observamos en la Figura N° 15 se ingresa los datos solicitados y luego guardar.



Finalmente podremos visualizar el nuevo triaje agregados, la cual podremos visualizar en detalle, como también eliminarlos.



DNI	Nombre	Especialidad	Fecha Hora Registro	Descripción	Acción
07472878	ACEVEDO SOTO LUISA	Oncología	2019-02-04 22:28:26	Ningun antecedente familiar	Detalle Eliminar

2.6. Otras Opciones

Adicionalmente podremos visualizar más opciones tales como Triage, Cuentas, Especialidades.

SISTEMA PARA EL PRONÓSTICO DE CÁNCER DE MAMA - HOSPITAL DE LA SOLIDARIDAD

VE - 01

enero 2019

Pacientes
Trajes
Cuentas
Especialidades

LISTA DE PACIENTES

Agregar Paciente

Show 10 entries

Busca:

DNI	Nombre	Fecha Nacimiento	Estado Civil	Sexo	Teléfono	Acción
06331064	VIZARRIETA CARDENAS JESSICA	1990-04-17	S	F	123456	Trasladar Editar Eliminar
06717877	AGREDA MINAYA CARLA ANDREA	1984-03-12	S	F	667890	Trasladar Editar Eliminar
01210121	AGUIVIO DAVILA MARITZA JULIETA	1990-09-07	V	F	123456	Trasladar Editar Eliminar
07472978	ADEVEDO SOTO LUISA	1990-09-10	S	F	664246	Trasladar Editar Eliminar
07472978	ADEVEDO SOTO MANUELA	0000-00-00	V	F	23423	Trasladar Editar Eliminar
01977448	ABARCA LOPEZ RODIO	1964-01-26	S	F	123123	Trasladar Editar Eliminar
08278640	AGUIVIO CALMET RAPHAELA ROSA	1994-03-14	S	F	123456	Trasladar Editar Eliminar
08623409	AGREDA GUZMAN JOANA GREGORIA	1971-06-24	S	F	466789	Trasladar Editar Eliminar
08088916	ADRIANZEN RONCEROS ALONDRA	1996-01-18	S	F	678432	Trasladar Editar Eliminar
9039178	VOLKERT SCHUSTER EDITH ELENN	1971-09-21	V	F	123456	Trasladar Editar Eliminar

Showing 1 to 10 of 22 entries

Previous 1 2 3 Next

Sistema para el Pronóstico de Cáncer de Mama - Hospital de la Solidaridad

2.7. Especialidades

En esta opción se puede Agregar, Editar y Eliminar

SISTEMA PARA EL PRONÓSTICO DE CÁNCER DE MAMA - HOSPITAL DE LA SOLIDARIDAD

NUEVA ESPECIALIDAD

Nombre:

Descripción:

Guardar

LISTA DE ESPECIALIDADES

Show 10 entries

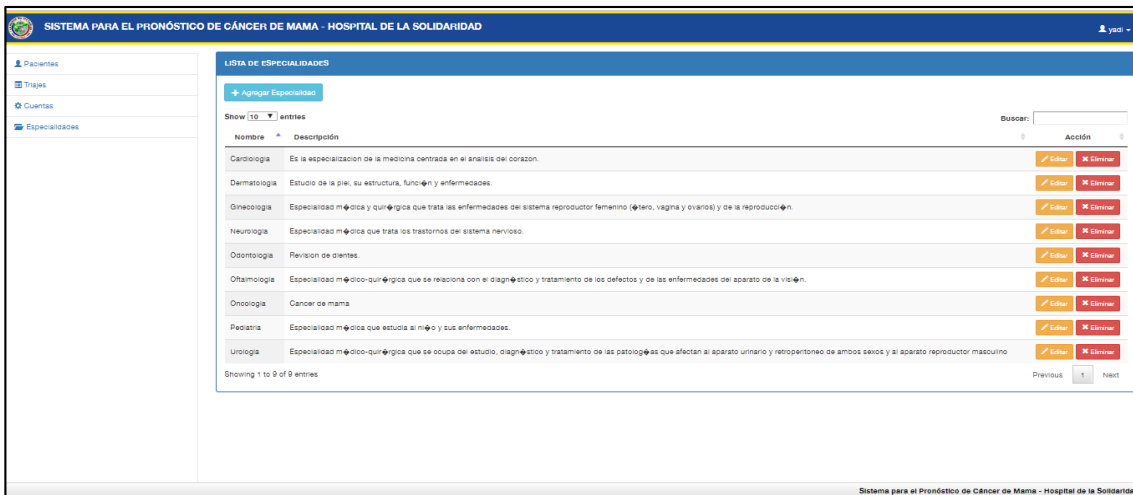
Busca:

Nombre	Descripción	Acción
Cardiología	Es la especialización de la medicina centrada en el análisis del corazón.	Edita Eliminar
Dermatología	Estudio de la piel, su estructura, función y enfermedades.	Edita Eliminar
Ginecología	Especialidad médica y quirúrgica que trata las enfermedades del sistema reproductor femenino (útero, vagina y ovarios) y de la reproducción.	Edita Eliminar
Neurología	Especialidad médica que trata los trastornos del sistema nervioso.	Edita Eliminar
Odonatología	Revisión de dientes.	Edita Eliminar
Oftalmología	Especialidad médico-quirúrgica que se relaciona con el diagnóstico y tratamiento de los defectos y de las enfermedades del aparato de la visión.	Edita Eliminar
Oncología	Cáncer de mama	Edita Eliminar

Sistema para el Pronóstico de Cáncer de Mama - Hospital de la Solidaridad

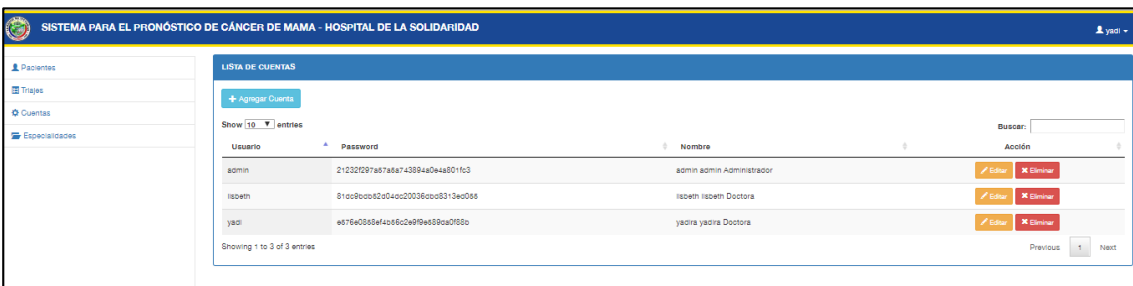
	MANUAL DE USUARIO	Código: HSVE - 01
		Versión 1
		Fecha: Febrero 2019

En la Figura N.º 19 podemos observar que se puede agregar una nueva especialidad y luego presionar el botón Guardar.



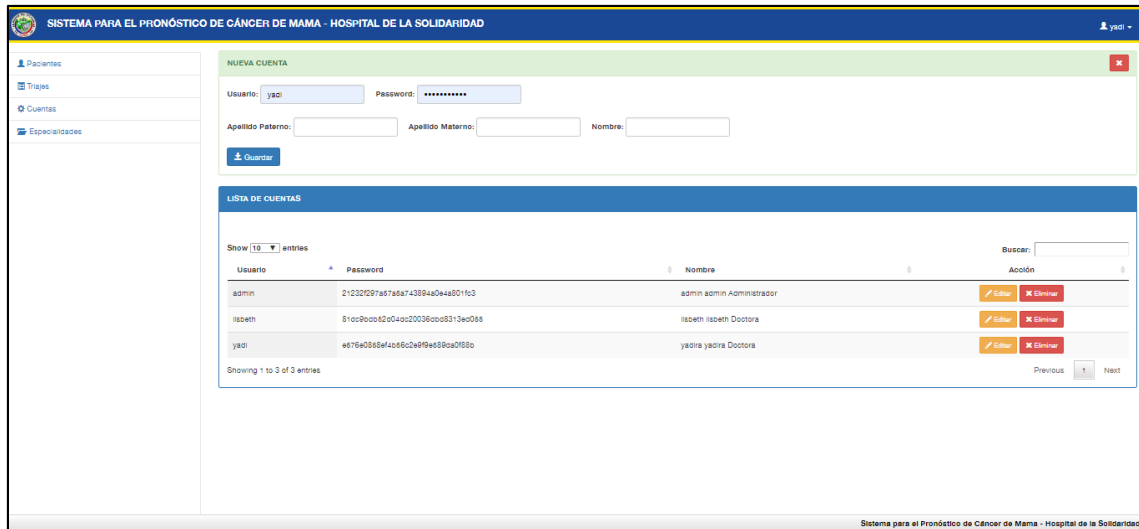
2.8. Cuentas

Finalmente podemos observar en la opción de Cuentas los perfiles de los Doctores



	MANUAL DE USUARIO	Código: HSVE - 01
		Versión 1
		Fecha: Febrero 2019

Lo cual podemos agregar una nueva cuenta, como se muestra en la siguiente captura de pantalla:



2.9. Cerrar Sesión

Podrá cerrar su sesión haciendo click a su nombre de Usuario una vez dándole click se abrirá un menú llamado Cerrar Sesión le da click y podrá cerrar la sesión.

SISTEMA PARA EL PRONÓSTICO DE CÁNCER DE MAMA - HOSPITAL DE LA SOLIDARIDAD

Clear Sesión

Pacientes
Triajes
Cuentas
Especialidades

LISTA DE PACIENTES

+ Agregar Paciente

Show 10 entries

Buscar:

DNI	Nombre	Fecha Nacimiento	Estado Civil	Sexo	Teléfono	Acción
06337064	VIZARRA CARDENAS JESSICA	1990-04-17	S	F	123486	Triaje / Editar / Eliminar
06716177	AGREDA MINAYA CARLA ANDREA	1094-03-12	S	F	667090	Triaje / Editar / Eliminar
07270721	AGUIYO DAVILA DAVILA MARITZAJULIETA	1990-09-07	V	F	123486	Triaje / Editar / Eliminar
07472878	AGUEDO SOTO LUISA	1990-09-10	S	F	664246	Triaje / Editar / Eliminar
07472878	AGUEDO SOTO MANUELA	0000-00-00	V	F	23423	Triaje / Editar / Eliminar
07977448	ABARCA LOPEZ ROCIO	1994-01-26	S	F	123123	Triaje / Editar / Eliminar
08278640	AGUIYO CALMET RAPHAELA ROSA	1996-03-14	S	F	123486	Triaje / Editar / Eliminar
08623409	AGREDA GUZMANIO JUANA GREGORIA	1997-06-24	S	F	468789	Triaje / Editar / Eliminar
08808916	ADRIANZEN RONCEROS ALONDRA	1996-02-16	S	F	678432	Triaje / Editar / Eliminar
9039178	VOLKERT SCHUSTER EDITH EILENN	1997-09-21	V	F	123486	Triaje / Editar / Eliminar

Showing 1 to 10 of 22 entries

Previous 1 2 3 Next

Sistema para el Pronóstico de Cáncer de Mama - Hospital de la Solidaridad

Anexo 5: Desarrollo de programación desarrollado en el Software Weka

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: BASE DE DATOS HOSPITAL FINAL

Instances: 762

Attributes: 24

EDAD

TALLA

PESO

FRECUENCIA CARDIACA

SITUACION FISICA
EDAD MENARQUIA
CANTIDAD EMBARAZOS
EDAD PRIMER EMBARAZO
CANTIDAD ABORTOS
CANTIDAD HIJOS
ANTECEDENTES FAMILIARES
PIEL HUNDIDA DE LA MAMA O PEZÓN
AUSENCIA DE MENSTRUACIÓN
AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA
DOLOR EN CUALQUIER PARTE DE LA MAMA
RETRACCIÓN DE LOS PEZONES
SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA
ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA
CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA
IRRITACIÓN EN LA PIEL DE LA MAMA
SECRECIÓN DE SANGRE O PUS POR EL PEZÓN
PRESENCIA DE TUMORES (BULTOS) EN LA MAMA
MAMAS Y PEZÓN CON HERIDAS
DIAGNOSTICO DE CANCER MAMA

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

DOLOR EN CUALQUIER PARTE DE LA MAMA = SI

| SITUACION FISICA = NORMAL

| | RETRACCIÓN DE LOS PEZONES = SI

| | | AUSENCIA DE MENSTRUACIÓN = SI: POSITIVO (43.0/3.0)

| | | AUSENCIA DE MENSTRUACIÓN = NO

| | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = SI

| | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI: POSITIVO (16.0)

| | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO

| | | | | MAMAS Y PEZÓN CON HERIDAS = SI

| | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI: NEGATIVO (3.0)

| | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO: POSITIVO (5.0/1.0)

| | | | | MAMAS Y PEZÓN CON HERIDAS = NO

| | | | | | EDAD <= 62: POSITIVO (7.0)

| | | | | | EDAD > 62: NEGATIVO (2.0)

| | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = NO

| | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI

| | | | | TALLA <= 1.68: NEGATIVO (8.0)

| | | | | TALLA > 1.68: POSITIVO (4.0)

| | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO

| | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI

| | | | | | TALLA <= 1.73: POSITIVO (5.0/1.0)

| | | | | | TALLA > 1.73: NEGATIVO (3.0)

| | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO: POSITIVO (6.0)

| | RETRACCIÓN DE LOS PEZONES = NO

| | | AUSENCIA DE MENSTRUACIÓN = SI

| | | MAMAS Y PEZÓN CON HERIDAS = SI

| | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI

| | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI: NEGATIVO (5.0/1.0)

| | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO

| | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI:
POSITIVO (6.0/1.0)

| | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO:
NEGATIVO (3.0/1.0)

| | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO: POSITIVO
(11.0)

| | | | MAMAS Y PEZÓN CON HERIDAS = NO

| | | | PESO <= 67.8: POSITIVO (5.0/1.0)

| | | | PESO > 67.8: NEGATIVO (8.0)

| | | AUSENCIA DE MENSTRUACIÓN = NO

| | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI

| | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA
= SI

| | | | | CANTIDAD HIJOS <= 0: NEGATIVO (4.0/1.0)

| | | | | CANTIDAD HIJOS > 0: POSITIVO (4.0)

| | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA
= NO

| | | | | EDAD PRIMER EMBARAZO <= 21: NEGATIVO (9.0)

| | | | | EDAD PRIMER EMBARAZO > 21

| | | | | ANTECEDENTES FAMILIARES <= 1: NEGATIVO (4.0/1.0)

| | | | | ANTECEDENTES FAMILIARES > 1: POSITIVO (2.0)

| | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO: NEGATIVO (29.0/1.0)

| SITUACION FISICA = SOBREPESO

| | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = SI:
POSITIVO (95.0)

| | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = NO

| | | RETRACCIÓN DE LOS PEZONES = SI: POSITIVO (60.0/3.0)

| | | RETRACCIÓN DE LOS PEZONES = NO

| | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI: POSITIVO (32.0/5.0)
| | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO
| | | | AUSENCIA DE MENSTRUACIÓN = SI
| | | | MAMAS Y PEZÓN CON HERIDAS = SI: POSITIVO (6.0)
| | | | MAMAS Y PEZÓN CON HERIDAS = NO: NEGATIVO (5.0/1.0)
| | | | AUSENCIA DE MENSTRUACIÓN = NO: NEGATIVO (9.0/1.0)
DOLOR EN CUALQUIER PARTE DE LA MAMA = NO
| SITUACION FISICA = NORMAL
| | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI
| | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = SI
| | | AUSENCIA DE MENSTRUACIÓN = SI: POSITIVO (23.0/3.0)
| | | AUSENCIA DE MENSTRUACIÓN = NO
| | | | ANTECEDENTES FAMILIARES ≤ 0 : POSITIVO (2.0)
| | | | ANTECEDENTES FAMILIARES > 0
| | | | CANTIDAD ABORTOS ≤ 1 : NEGATIVO (6.0)
| | | | CANTIDAD ABORTOS > 1
| | | | | ANTECEDENTES FAMILIARES ≤ 1 : POSITIVO (2.0)
| | | | | ANTECEDENTES FAMILIARES > 1 : NEGATIVO (6.0/1.0)
| | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = NO
| | | | RETRACCIÓN DE LOS PEZONES = SI
| | | | PESO ≤ 66.7
| | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI: POSITIVO (6.0)
| | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO
| | | | | EDAD PRIMER EMBARAZO ≤ 18 : NEGATIVO (3.0)
| | | | | EDAD PRIMER EMBARAZO > 18 : POSITIVO (6.0/1.0)
| | | | PESO > 66.7 : NEGATIVO (11.0)
| | | | RETRACCIÓN DE LOS PEZONES = NO

| | | | | ANTECEDENTES FAMILIARES <= 3: NEGATIVO (20.0)

| | | | | ANTECEDENTES FAMILIARES > 3

| | | | | TALLA <= 1.76: NEGATIVO (5.0)

| | | | | TALLA > 1.76: POSITIVO (2.0)

| | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO: NEGATIVO (94.0/4.0)

| SITUACION FISICA = SOBREPESO

| | RETRACCIÓN DE LOS PEZONES = SI

| | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = SI: POSITIVO (42.0/3.0)

| | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = NO

| | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI

| | | | | AUSENCIA DE MENSTRUACIÓN = SI: POSITIVO (8.0)

| | | | | AUSENCIA DE MENSTRUACIÓN = NO

| | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI: POSITIVO (4.0)

| | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO

| | | | | | CANTIDAD ABORTOS <= 0: POSITIVO (2.0)

| | | | | | CANTIDAD ABORTOS > 0: NEGATIVO (3.0)

| | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO

| | | | | FRECUENCIA CARDIACA <= 81

| | | | | | AUSENCIA DE MENSTRUACIÓN = SI: POSITIVO (9.0/1.0)

| | | | | | AUSENCIA DE MENSTRUACIÓN = NO: NEGATIVO (4.0/1.0)

| | | | | FRECUENCIA CARDIACA > 81: NEGATIVO (8.0)

| | RETRACCIÓN DE LOS PEZONES = NO

| | | AUSENCIA DE MENSTRUACIÓN = SI

| | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = SI

| | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI: POSITIVO (14.0)

| | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO

| | | | | EDAD MENARQUIA <= 12: NEGATIVO (3.0)
 | | | | | EDAD MENARQUIA > 12: POSITIVO (8.0/2.0)
 | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA
 = NO
 | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI: NEGATIVO
 (8.0/1.0)
 | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO
 | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI: NEGATIVO (5.0/1.0)
 | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO: POSITIVO (3.0)
 | | | AUSENCIA DE MENSTRUACIÓN = NO
 | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA
 = SI
 | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI
 | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI: NEGATIVO
 (10.0)
 | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO
 | | | | | EDAD <= 54: NEGATIVO (7.0)
 | | | | | EDAD > 54: POSITIVO (2.0)
 | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO
 | | | | | EDAD <= 46: POSITIVO (4.0)
 | | | | | EDAD > 46
 | | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI
 | | | | | PESO <= 77.41: POSITIVO (2.0)
 | | | | | PESO > 77.41: NEGATIVO (2.0)
 | | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO: NEGATIVO (4.0)
 | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA
 = NO: NEGATIVO (25.0)

Number of Leaves : 64

Size of the tree : 127

Time taken to build model: 0.37 seconds

==== Stratified cross-validation ====

==== Summary ====

Correctly Classified Instances	604	79.2651 %
Incorrectly Classified Instances	158	20.7349 %
Kappa statistic	0.5768	
Mean absolute error	0.2332	
Root mean squared error	0.4341	
Relative absolute error	47.6326 %	
Root relative squared error	87.7332 %	
Total Number of Instances	762	

==== Detailed Accuracy By Class ====

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	
	0,817	0,239	0,820	0,817	0,818	0,577	0,791	0,796	POSITIVO
	0,761	0,183	0,756	0,761	0,758	0,577	0,791	0,682	NEGATIVO
Weighted Avg.	0,793	0,215	0,793	0,793	0,793	0,577	0,791	0,747	

==== Confusion Matrix ====

a b <-- classified as

356 80 | a = POSITIVO

78 248 | b = NEGATIVO

Modelo 1: Situación física

J48 pruned tree

DIAGNOSTICO DE CANCER MAMA = POSITIVO

| ANTECEDENTES FAMILIARES <= 3

| | RETRACCIÓN DE LOS PEZONES = SI

| | | CANTIDAD ABORTOS <= 0

| | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI

| | | | | EDAD <= 48: NORMAL (6.0/1.0)

| | | | | EDAD > 48: SOBREPESO (26.0/5.0)

| | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO: SOBREPESO (25.0/3.0)

| | | CANTIDAD ABORTOS > 0

| | | | CANTIDAD HIJOS <= 4

| | | | | AUSENCIA DE MENSTRUACIÓN = SI

| | | | | MAMAS Y PEZÓN CON HERIDAS = SI

| | | | | | FRECUENCIA CARDIACA <= 85

| | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI

| | | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI:
NORMAL (5.0)

| | | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO

| | | | | | | | | CANTIDAD HIJOS <= 0: NORMAL (3.0)

| | | | | | | | | CANTIDAD HIJOS > 0

| | | | | | | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = SI

| | | | | | | | | | EDAD MENARQUIA <= 13: NORMAL (3.0)

| | | | | | | | | | EDAD MENARQUIA > 13: SOBREPESO (3.0)

| | | | | | | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = NO:
SOBREPESO (2.0)

| | | | | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO

| | | | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI:
SOBREPESO (6.0)

| | | | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO

| | | | | | | | | | CANTIDAD ABORTOS <= 1: SOBREPESO (2.0)

| | | | | | | | | | CANTIDAD ABORTOS > 1: NORMAL (6.0/1.0)

| | | | | | | | | | FRECUENCIA CARDIACA > 85: SOBREPESO (7.0)

| | | | | | | | | | MAMAS Y PEZÓN CON HERIDAS = NO

| | | | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI

| | | | | | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI: NORMAL (12.0/2.0)

| | | | | | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO

| | | | | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI

| | | | | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI:
SOBREPESO (3.0/1.0)

| | | | | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO:
NORMAL (4.0)

| | | | | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO:
SOBREPESO (3.0)

| | | | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO

| | | | | | | | | | CANTIDAD HIJOS <= 1: SOBREPESO (4.0)

| | | | | | | | | | CANTIDAD HIJOS > 1

| | | | | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI

| | | | | | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE
LA MAMA = SI: NORMAL (4.0)

| | | | | | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE
LA MAMA = NO

| | | | | | | | | | EDAD PRIMER EMBARAZO <= 19: NORMAL (2.0)

| | | | | | | | | | EDAD PRIMER EMBARAZO > 19: SOBREPESO (4.0/1.0)

| | | | | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO

| | | | | | | | | CANTIDAD HIJOS <= 3

| | | | | | | | | FRECUENCIA CARDIACA <= 72: NORMAL (3.0)

| | | | | | | | | FRECUENCIA CARDIACA > 72: SOBREPESO (4.0/1.0)

| | | | | | | | | CANTIDAD HIJOS > 3: SOBREPESO (2.0)

| | | | | AUSENCIA DE MENSTRUACIÓN = NO

| | | | | EDAD <= 58

| | | | | EDAD <= 53

| | | | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI:
SOBREPESO (18.0/3.0)

| | | | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO

| | | | | | | | | EDAD PRIMER EMBARAZO <= 18: SOBREPESO (7.0)

| | | | | | | | | EDAD PRIMER EMBARAZO > 18

| | | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA =
SI: NORMAL (5.0)

| | | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA =
NO

| | | | | | | | | FRECUENCIA CARDIACA <= 75: SOBREPESO (5.0/1.0)

| | | | | | | | | FRECUENCIA CARDIACA > 75: NORMAL (3.0)

| | | | | | | EDAD > 53: NORMAL (9.0/1.0)

| | | | | | EDAD > 58: SOBREPESO (11.0)

| | | | CANTIDAD HIJOS > 4

| | | | | AUSENCIA DE MENSTRUACIÓN = SI: SOBREPESO (9.0)

| | | | | AUSENCIA DE MENSTRUACIÓN = NO

| | | | | TALLA <= 1.7: NORMAL (2.0)

| | | | | TALLA > 1.7: SOBREPESO (4.0)

| | | RETRACCIÓN DE LOS PEZONES = NO

| | | EDAD MENARQUIA <= 14

| | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA
= SI

| | | | | AUSENCIA DE MENSTRUACIÓN = SI

| | | | | MAMAS Y PEZÓN CON HERIDAS = SI
| | | | | PESO <= 67.9
| | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI:
SOBREPESO (7.0)
| | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO
| | | | | EDAD <= 49: SOBREPESO (2.0)
| | | | | EDAD > 49: NORMAL (4.0)
| | | | | PESO > 67.9: NORMAL (11.0/1.0)
| | | | | MAMAS Y PEZÓN CON HERIDAS = NO
| | | | | EDAD MENARQUIA <= 12: SOBREPESO (8.0)
| | | | | EDAD MENARQUIA > 12
| | | | | PESO <= 56.49: NORMAL (4.0)
| | | | | PESO > 56.49
| | | | | FRECUENCIA CARDIACA <= 78: SOBREPESO (7.0)
| | | | | FRECUENCIA CARDIACA > 78: NORMAL (3.0/1.0)
| | | | | AUSENCIA DE MENSTRUACIÓN = NO
| | | | | CANTIDAD HIJOS <= 4: SOBREPESO (18.0/2.0)
| | | | | CANTIDAD HIJOS > 4: NORMAL (3.0/1.0)
| | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA
= NO
| | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI
| | | | | EDAD <= 63: SOBREPESO (24.0/3.0)
| | | | | EDAD > 63: NORMAL (3.0)
| | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO: SOBREPESO (9.0)
| | | | EDAD MENARQUIA > 14
| | | | EDAD PRIMER EMBARAZO <= 22: SOBREPESO (21.0)
| | | | EDAD PRIMER EMBARAZO > 22
| | | | | EDAD <= 51: NORMAL (2.0)
| | | | | EDAD > 51: SOBREPESO (2.0)

| ANTECEDENTES FAMILIARES > 3
| | EDAD PRIMER EMBARAZO <= 15: SOBREPESO (11.0/1.0)
| | EDAD PRIMER EMBARAZO > 15
| | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI
| | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI
| | | | | PESO <= 45.3: NORMAL (2.0)
| | | | | PESO > 45.3: SOBREPESO (15.0/1.0)
| | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO
| | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = SI
| | | | | TALLA <= 1.72: NORMAL (7.0/1.0)
| | | | | TALLA > 1.72: SOBREPESO (3.0)
| | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = NO: SOBREPESO (4.0/1.0)
| | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO
| | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI
| | | | | EDAD MENARQUIA <= 10: SOBREPESO (7.0/2.0)
| | | | | EDAD MENARQUIA > 10
| | | | | MAMAS Y PEZÓN CON HERIDAS = SI: NORMAL (10.0)
| | | | | MAMAS Y PEZÓN CON HERIDAS = NO
| | | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = SI
| | | | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = SI
| | | | | | | | PESO <= 72
| | | | | | | | PESO <= 56.49: NORMAL (2.0)
| | | | | | | | PESO > 56.49: SOBREPESO (3.0)
| | | | | | | | PESO > 72: NORMAL (5.0)
| | | | | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = NO: SOBREPESO
(2.0)
| | | | | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = NO: NORMAL (4.0)
| | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO
| | | | | PESO <= 54: NORMAL (5.0)

| | | | | PESO > 54

| | | | | CANTIDAD HIJOS <= 2

| | | | | CANTIDAD HIJOS <= 0: SOBREPESO (2.0)

| | | | | CANTIDAD HIJOS > 0: NORMAL (5.0)

| | | | | CANTIDAD HIJOS > 2: SOBREPESO (9.0/1.0)

DIAGNOSTICO DE CANCER MAMA = NEGATIVO

| DOLOR EN CUALQUIER PARTE DE LA MAMA = SI

| | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = SI:
NORMAL (31.0)

| | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = NO

| | | MAMAS Y PEZÓN CON HERIDAS = SI

| | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI: NORMAL (11.0)

| | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO

| | | | | EDAD MENARQUIA <= 10: SOBREPESO (3.0)

| | | | | EDAD MENARQUIA > 10: NORMAL (19.0/3.0)

| | | MAMAS Y PEZÓN CON HERIDAS = NO

| | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI

| | | | | EDAD MENARQUIA <= 14: NORMAL (11.0)

| | | | | EDAD MENARQUIA > 14

| | | | | EDAD PRIMER EMBARAZO <= 18: NORMAL (2.0)

| | | | | EDAD PRIMER EMBARAZO > 18: SOBREPESO (3.0)

| | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO

| | | | | RETRACCIÓN DE LOS PEZONES = SI: NORMAL (5.0/1.0)

| | | | | RETRACCIÓN DE LOS PEZONES = NO

| | | | | ANTECEDENTES FAMILIARES <= 1: NORMAL (4.0/1.0)

| | | | | ANTECEDENTES FAMILIARES > 1

| | | | | ANTECEDENTES FAMILIARES <= 3: SOBREPESO (7.0)

| | | | | ANTECEDENTES FAMILIARES > 3

| | | | | EDAD MENARQUIA <= 12: NORMAL (2.0)

| | | | | | | | EDAD MENARQUIA > 12: SOBREPESO (2.0)
| DOLOR EN CUALQUIER PARTE DE LA MAMA = NO
| | RETRACCIÓN DE LOS PEZONES = SI
| | | CANTIDAD HIJOS <= 3: NORMAL (55.0/9.0)
| | | CANTIDAD HIJOS > 3
| | | | MAMAS Y PEZÓN CON HERIDAS = SI
| | | | ANTECEDENTES FAMILIARES <= 3: NORMAL (11.0/1.0)
| | | | ANTECEDENTES FAMILIARES > 3: SOBREPESO (2.0)
| | | | MAMAS Y PEZÓN CON HERIDAS = NO
| | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = SI: SOBREPESO (7.0/2.0)
| | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = NO: NORMAL (4.0/1.0)
| | RETRACCIÓN DE LOS PEZONES = NO
| | | AUSENCIA DE MENSTRUACIÓN = SI
| | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI
| | | | | CANTIDAD EMBARAZOS <= 4: NORMAL (13.0)
| | | | | CANTIDAD EMBARAZOS > 4
| | | | | FRECUENCIA CARDIACA <= 72: SOBREPESO (3.0)
| | | | | FRECUENCIA CARDIACA > 72: NORMAL (4.0)
| | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO
| | | | | FRECUENCIA CARDIACA <= 70: NORMAL (15.0/2.0)
| | | | | FRECUENCIA CARDIACA > 70
| | | | | FRECUENCIA CARDIACA <= 85
| | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI:
SOBREPESO (6.0)
| | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO
| | | | | | FRECUENCIA CARDIACA <= 76: NORMAL (2.0)
| | | | | | FRECUENCIA CARDIACA > 76: SOBREPESO (4.0)
| | | | | | FRECUENCIA CARDIACA > 85: NORMAL (4.0/1.0)
| | | AUSENCIA DE MENSTRUACIÓN = NO

| | | | EDAD <= 61

| | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI

| | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI

| | | | | | | CANTIDAD EMBARAZOS <= 6

| | | | | | | PESO <= 51.2: NORMAL (3.0/1.0)

| | | | | | | PESO > 51.2: SOBREPESO (17.0)

| | | | | | | CANTIDAD EMBARAZOS > 6: NORMAL (2.0)

| | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO

| | | | | | | PESO <= 53.5: NORMAL (4.0)

| | | | | | | PESO > 53.5

| | | | | | | SECRECIÓN DE SANGRE O PUS POR EL PEZÓN = SI

| | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI

| | | | | | | | CANTIDAD HIJOS <= 2: SOBREPESO (3.0)

| | | | | | | | CANTIDAD HIJOS > 2: NORMAL (2.0)

| | | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO:
SOBREPESO (4.0)

| | | | | | | | SECRECIÓN DE SANGRE O PUS POR EL PEZÓN = NO

| | | | | | | | CANTIDAD EMBARAZOS <= 4: NORMAL (5.0/1.0)

| | | | | | | | CANTIDAD EMBARAZOS > 4: SOBREPESO (2.0)

| | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO

| | | | | | EDAD <= 43: NORMAL (7.0)

| | | | | | EDAD > 43

| | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI

| | | | | | | ANTECEDENTES FAMILIARES <= 1

| | | | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI: NORMAL (4.0)

| | | | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO: SOBREPESO
(3.0/1.0)

| | | | | | | | ANTECEDENTES FAMILIARES > 1: SOBREPESO (11.0/1.0)

| | | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO

| | | | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = SI

| | | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI

| | | | | | | | CANTIDAD EMBARAZOS \leq 5: NORMAL (6.0/1.0)

| | | | | | | | CANTIDAD EMBARAZOS $>$ 5: SOBREPESO (2.0)

| | | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO: SOBREPESO (2.0)

| | | | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = NO

| | | | | | | | TALLA \leq 1.49: SOBREPESO (2.0)

| | | | | | | | TALLA $>$ 1.49: NORMAL (10.0/1.0)

| | | | EDAD $>$ 61: NORMAL (7.0)

Modelo 2: Piel hundida de la mama o pezón

J48 pruned tree

DIAGNOSTICO DE CANCER MAMA = POSITIVO

| SITUACION FISICA = NORMAL

| | DOLOR EN CUALQUIER PARTE DE LA MAMA = SI

| | | EDAD \leq 62

| | | | FRECUENCIA CARDIACA \leq 61: SI (7.0)

| | | | FRECUENCIA CARDIACA $>$ 61

| | | | | AUSENCIA DE MENSTRUACIÓN = SI

| | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI

| | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI: NO (9.0/1.0)

| | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO

| | | | | | ANTECEDENTES FAMILIARES \leq 1: NO (5.0)

| | | | | | ANTECEDENTES FAMILIARES $>$ 1

| | | | | | CANTIDAD ABORTOS \leq 0: NO (3.0)

| | | | | | CANTIDAD ABORTOS $>$ 0: SI (5.0)

| | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO

| | | | | | | MAMAS Y PEZÓN CON HERIDAS = SI

| | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI: SI (7.0/1.0)

| | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO

| | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI

| | | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI

| | | | | | | | | CANTIDAD EMBARAZOS <= 5: NO (2.0)

| | | | | | | | | CANTIDAD EMBARAZOS > 5: SI (3.0/1.0)

| | | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO: SI (3.0)

| | | | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO: NO (3.0)

| | | | | | | MAMAS Y PEZÓN CON HERIDAS = NO

| | | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI: SI (6.0)

| | | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO

| | | | | | | | | EDAD <= 52: NO (3.0)

| | | | | | | | | EDAD > 52: SI (2.0)

| | | | | AUSENCIA DE MENSTRUACIÓN = NO

| | | | | | FRECUENCIA CARDIACA <= 66: NO (5.0/1.0)

| | | | | | FRECUENCIA CARDIACA > 66: SI (39.0/11.0)

| | | EDAD > 62: SI (10.0)

| | DOLOR EN CUALQUIER PARTE DE LA MAMA = NO: SI (42.0/4.0)

| SITUACION FISICA = SOBREPESO

| | TALLA <= 1.49: SI (16.0/1.0)

| | TALLA > 1.49

| | | RETRACCIÓN DE LOS PEZONES = SI

| | | | EDAD MENARQUIA <= 13

| | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI

| | | | | AUSENCIA DE MENSTRUACIÓN = SI

| | | | | | | | EDAD MENARQUIA <= 12

| | | | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = SI

| | | | | | | | MAMAS Y PEZÓN CON HERIDAS = SI: NO (8.0/1.0)

| | | | | | | | MAMAS Y PEZÓN CON HERIDAS = NO

| | | | | | | | PESO <= 63.5: SI (2.0)

| | | | | | | | PESO > 63.5: NO (2.0)

| | | | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = NO

| | | | | | | | ANTECEDENTES FAMILIARES <= 0: NO (2.0)

| | | | | | | | ANTECEDENTES FAMILIARES > 0

| | | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI: SI (6.0)

| | | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO

| | | | | | | | FRECUENCIA CARDIACA <= 67: SI (3.0)

| | | | | | | | FRECUENCIA CARDIACA > 67: NO (3.0)

| | | | | | | | EDAD MENARQUIA > 12: SI (4.0)

| | | | | | | | AUSENCIA DE MENSTRUACIÓN = NO: SI (17.0)

| | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO

| | | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI

| | | | | | | | ANTECEDENTES FAMILIARES <= 3

| | | | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = SI

| | | | | | | | AUSENCIA DE MENSTRUACIÓN = SI

| | | | | | | | EDAD MENARQUIA <= 12: SI (3.0)

| | | | | | | | EDAD MENARQUIA > 12: NO (2.0)

| | | | | | | | AUSENCIA DE MENSTRUACIÓN = NO: NO (3.0)

| | | | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = NO

| | | | | | | | EDAD PRIMER EMBARAZO <= 15: NO (2.0)

| | | | | | | | | EDAD PRIMER EMBARAZO > 15: SI (6.0)
| | | | | | | | | ANTECEDENTES FAMILIARES > 3: NO (7.0)
| | | | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO
| | | | | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI
| | | | | | | | | EDAD PRIMER EMBARAZO <= 15: NO (2.0)
| | | | | | | | | EDAD PRIMER EMBARAZO > 15: SI (12.0/1.0)
| | | | | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO
| | | | | | | | | CANTIDAD ABORTOS <= 1
| | | | | | | | | MAMAS Y PEZÓN CON HERIDAS = SI: SI (5.0)
| | | | | | | | | MAMAS Y PEZÓN CON HERIDAS = NO: NO (3.0/1.0)
| | | | | | | | | CANTIDAD ABORTOS > 1: NO (4.0)
| | | | | EDAD MENARQUIA > 13
| | | | | | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = SI
| | | | | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI: NO (11.0)
| | | | | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO
| | | | | | | | | EDAD PRIMER EMBARAZO <= 21
| | | | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI: SI (7.0/1.0)
| | | | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO
| | | | | | | | | AUSENCIA DE MENSTRUACIÓN = SI: NO (3.0)
| | | | | | | | | AUSENCIA DE MENSTRUACIÓN = NO
| | | | | | | | | EDAD PRIMER EMBARAZO <= 15: NO (3.0)
| | | | | | | | | EDAD PRIMER EMBARAZO > 15: SI (3.0)
| | | | | | | | | EDAD PRIMER EMBARAZO > 21: NO (4.0)
| | | | | | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = NO
| | | | | | | | | EDAD MENARQUIA <= 14
| | | | | | | | | EDAD PRIMER EMBARAZO <= 23
| | | | | | | | | ANTECEDENTES FAMILIARES <= 1: NO (8.0)
| | | | | | | | | ANTECEDENTES FAMILIARES > 1

| | | | | | | | ANTECEDENTES FAMILIARES <= 3: SI (4.0)

| | | | | | | | ANTECEDENTES FAMILIARES > 3: NO (4.0)

| | | | | | | | EDAD PRIMER EMBARAZO > 23: SI (3.0)

| | | | | | | | EDAD MENARQUIA > 14

| | | | | | | | CANTIDAD EMBARAZOS <= 6

| | | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI

| | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI:
NO (2.0)

| | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO:
SI (5.0/1.0)

| | | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO: SI
(10.0/1.0)

| | | | | | | | CANTIDAD EMBARAZOS > 6: NO (3.0)

| | | | | | | | RETRACCIÓN DE LOS PEZONES = NO

| | | | | | | | CANTIDAD EMBARAZOS <= 5

| | | | | | | | CANTIDAD EMBARAZOS <= 0

| | | | | | | | SECRECIÓN DE SANGRE O PUS POR EL PEZÓN = SI

| | | | | | | | FRECUENCIA CARDIACA <= 72: SI (3.0)

| | | | | | | | FRECUENCIA CARDIACA > 72: NO (2.0)

| | | | | | | | SECRECIÓN DE SANGRE O PUS POR EL PEZÓN = NO: NO (4.0)

| | | | | | | | CANTIDAD EMBARAZOS > 0

| | | | | | | | MAMAS Y PEZÓN CON HERIDAS = SI

| | | | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI: SI (19.0/4.0)

| | | | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO

| | | | | | | | AUSENCIA DE MENSTRUACIÓN = SI

| | | | | | | | CANTIDAD HIJOS <= 3

| | | | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE
LA MAMA = SI: SI (4.0)

| | | | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE
LA MAMA = NO

| | | | | | | | | | EDAD <= 50: SI (2.0)
| | | | | | | | | | EDAD > 50: NO (2.0)
| | | | | | | | | | CANTIDAD HIJOS > 3: NO (2.0)
| | | | | | | | | | AUSENCIA DE MENSTRUACIÓN = NO: NO (4.0/1.0)
| | | | | | | | | | MAMAS Y PEZÓN CON HERIDAS = NO
| | | | | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI
| | | | | | | | | | EDAD PRIMER EMBARAZO <= 20
| | | | | | | | | | PESO <= 65.2: NO (6.0/1.0)
| | | | | | | | | | PESO > 65.2: SI (4.0)
| | | | | | | | | | EDAD PRIMER EMBARAZO > 20: SI (8.0)
| | | | | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO: SI
(21.0/1.0)
| | | | | CANTIDAD EMBARAZOS > 5
| | | | | CANTIDAD ABORTOS <= 1: NO (2.0)
| | | | | CANTIDAD ABORTOS > 1
| | | | | PESO <= 71.15
| | | | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI
| | | | | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI: SI
(2.0)
| | | | | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO: NO
(3.0)
| | | | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO: SI
(7.0/1.0)
| | | | | | | | | | PESO > 71.15: NO (5.0)
DIAGNOSTICO DE CANCER MAMA = NEGATIVO
| | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = SI
| | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = SI
| | | | | CANTIDAD HIJOS <= 0
| | | | | MAMAS Y PEZÓN CON HERIDAS = SI: NO (6.0)
| | | | | MAMAS Y PEZÓN CON HERIDAS = NO: SI (4.0/1.0)

| | | CANTIDAD HIJOS > 0: NO (21.0)

| | DOLOR EN CUALQUIER PARTE DE LA MAMA = NO

| | | MAMAS Y PEZÓN CON HERIDAS = SI

| | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI

| | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI

| | | | | TALLA <= 1.47: SI (2.0)

| | | | | TALLA > 1.47: NO (11.0/1.0)

| | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO

| | | | | CANTIDAD ABORTOS <= 0: SI (3.0)

| | | | | CANTIDAD ABORTOS > 0: NO (12.0/4.0)

| | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO: NO (26.0/1.0)

| | | MAMAS Y PEZÓN CON HERIDAS = NO

| | | | AUSENCIA DE MENSTRUACIÓN = SI: NO (4.0)

| | | | AUSENCIA DE MENSTRUACIÓN = NO

| | | | | EDAD MENARQUIA <= 14

| | | | | RETRACCIÓN DE LOS PEZONES = SI: NO (7.0)

| | | | | RETRACCIÓN DE LOS PEZONES = NO

| | | | | | SITUACION FISICA = NORMAL

| | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI: SI (3.0/1.0)

| | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO: NO (5.0)

| | | | | | | SITUACION FISICA = SOBREPESO

| | | | | | | CANTIDAD EMBARAZOS <= 5: SI (7.0)

| | | | | | | CANTIDAD EMBARAZOS > 5: NO (3.0/1.0)

| | | | | EDAD MENARQUIA > 14: SI (4.0)

| AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = NO

| | CANTIDAD HIJOS <= 3

| | | CANTIDAD EMBARAZOS <= 5

| | | | CANTIDAD EMBARAZOS <= 4

| | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI

| | | | TALLA <= 1.54: NO (6.0)

| | | | TALLA > 1.54

| | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI

| | | | CANTIDAD ABORTOS <= 1

| | | | ANTECEDENTES FAMILIARES <= 1

| | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI: SI (4.0)

| | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO: NO (4.0/1.0)

| | | | ANTECEDENTES FAMILIARES > 1: NO (2.0)

| | | | CANTIDAD ABORTOS > 1: SI (7.0)

| | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO

| | | | MAMAS Y PEZÓN CON HERIDAS = SI

| | | | TALLA <= 1.86

| | | | SITUACION FISICA = NORMAL: SI (6.0)

| | | | SITUACION FISICA = SOBREPESO: NO (3.0/1.0)

| | | | TALLA > 1.86: NO (5.0)

| | | | MAMAS Y PEZÓN CON HERIDAS = NO

| | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI:
SI (3.0/1.0)

| | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO:
NO (10.0/1.0)

| | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO

| | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI

| | | | AUSENCIA DE MENSTRUACIÓN = SI

| | | | SITUACION FISICA = NORMAL

| | | | CANTIDAD ABORTOS <= 0: SI (2.0)

| | | | CANTIDAD ABORTOS > 0

| | | | EDAD PRIMER EMBARAZO <= 18: SI (4.0/1.0)

| | | | | | | | | EDAD PRIMER EMBARAZO > 18: NO (5.0)

| | | | | | | | | SITUACION FISICA = SOBREPESO: NO (2.0)

| | | | | | | | | AUSENCIA DE MENSTRUACIÓN = NO

| | | | | | | | | CANTIDAD EMBARAZOS <= 2

| | | | | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI

| | | | | | | | | ANTECEDENTES FAMILIARES <= 0: NO (2.0)

| | | | | | | | | ANTECEDENTES FAMILIARES > 0: SI (3.0)

| | | | | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO: SI (5.0)

| | | | | | | | | CANTIDAD EMBARAZOS > 2

| | | | | | | | | CANTIDAD ABORTOS <= 1: NO (3.0)

| | | | | | | | | CANTIDAD ABORTOS > 1

| | | | | | | | | FRECUENCIA CARDIACA <= 75

| | | | | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI

| | | | | | | | | EDAD <= 55: SI (2.0)

| | | | | | | | | EDAD > 55: NO (2.0)

| | | | | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO: SI (4.0)

| | | | | | | | | FRECUENCIA CARDIACA > 75: NO (4.0)

| | | | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO

| | | | | | | | | EDAD MENARQUIA <= 10: SI (7.0)

| | | | | | | | | EDAD MENARQUIA > 10

| | | | | | | | | EDAD MENARQUIA <= 14

| | | | | | | | | MAMAS Y PEZÓN CON HERIDAS = SI

| | | | | | | | | RETRACCIÓN DE LOS PEZONES = SI: NO (2.0)

| | | | | | | | | RETRACCIÓN DE LOS PEZONES = NO

| | | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA
= SI: SI (3.0)

| | | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA
= NO

| | | | | | | | | EDAD MENARQUIA <= 12: NO (2.0)

| | | | | | | | | | | EDAD MENARQUIA > 12: SI (3.0)
 | | | | | | | | | | MAMAS Y PEZÓN CON HERIDAS = NO
 | | | | | | | | | | RETRACCIÓN DE LOS PEZONES = SI
 | | | | | | | | | | PESO <= 78.5: NO (2.0)
 | | | | | | | | | | PESO > 78.5: SI (2.0)
 | | | | | | | | | | RETRACCIÓN DE LOS PEZONES = NO: NO (5.0)
 | | | | | | | | | | EDAD MENARQUIA > 14: SI (6.0)
 | | | | CANTIDAD EMBARAZOS > 4
 | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI: NO (12.0/1.0)
 | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO
 | | | | MAMAS Y PEZÓN CON HERIDAS = SI
 | | | | PESO <= 55.11: NO (2.0)
 | | | | PESO > 55.11: SI (4.0)
 | | | | MAMAS Y PEZÓN CON HERIDAS = NO: NO (3.0)
 | | | CANTIDAD EMBARAZOS > 5: SI (8.0/1.0)
 | | CANTIDAD HIJOS > 3
 | | ANTECEDENTES FAMILIARES <= 3
 | | | EDAD <= 49
 | | | | RETRACCIÓN DE LOS PEZONES = SI: NO (6.0/1.0)
 | | | | RETRACCIÓN DE LOS PEZONES = NO: SI (18.0/5.0)
 | | | EDAD > 49: NO (24.0/3.0)
 | | ANTECEDENTES FAMILIARES > 3: NO (11.0)

Modelo 3: Ausencia de menstruación

J48 pruned tree

DIAGNOSTICO DE CANCER MAMA = POSITIVO

| RETRACCIÓN DE LOS PEZONES = SI

| | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = SI

| | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI
| | | | MAMAS Y PEZÓN CON HERIDAS = SI: SI (41.0/12.0)
| | | | MAMAS Y PEZÓN CON HERIDAS = NO
| | | | | FRECUENCIA CARDIACA <= 63: SI (7.0)
| | | | | FRECUENCIA CARDIACA > 63
| | | | | CANTIDAD EMBARAZOS <= 3: NO (9.0)
| | | | | CANTIDAD EMBARAZOS > 3
| | | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = SI
| | | | | | FRECUENCIA CARDIACA <= 69: NO (3.0)
| | | | | | FRECUENCIA CARDIACA > 69
| | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI
| | | | | | | EDAD <= 47: NO (2.0)
| | | | | | | EDAD > 47: SI (4.0)
| | | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO: SI
(6.0)
| | | | | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = NO
| | | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI
| | | | | | | | CANTIDAD HIJOS <= 3: SI (2.0)
| | | | | | | | CANTIDAD HIJOS > 3: NO (5.0/1.0)
| | | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO: NO
(3.0)
| | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO
| | | | SITUACION FISICA = NORMAL
| | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = SI
| | | | | FRECUENCIA CARDIACA <= 64: SI (5.0)
| | | | | FRECUENCIA CARDIACA > 64: NO (19.0/2.0)
| | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = NO: SI (8.0)
| | | | SITUACION FISICA = SOBREPESO
| | | | | TALLA <= 1.57: SI (5.0/1.0)

| | | | | TALLA > 1.57

| | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI: NO
(25.0/3.0)

| | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO

| | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = SI: NO (6.0)

| | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = NO

| | | | | FRECUENCIA CARDIACA <= 78: SI (4.0)

| | | | | FRECUENCIA CARDIACA > 78: NO (2.0)

| | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = NO

| | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI

| | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = SI

| | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = SI

| | | | | EDAD <= 56: SI (9.0/2.0)

| | | | | EDAD > 56: NO (5.0)

| | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = NO

| | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI

| | | | | SITUACION FISICA = NORMAL: NO (3.0/1.0)

| | | | | SITUACION FISICA = SOBREPESO

| | | | | ANTECEDENTES FAMILIARES <= 1: NO (4.0/1.0)

| | | | | ANTECEDENTES FAMILIARES > 1: SI (5.0)

| | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO

| | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI

| | | | | FRECUENCIA CARDIACA <= 75: NO (2.0)

| | | | | FRECUENCIA CARDIACA > 75: SI (4.0)

| | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO: SI (6.0)

| | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = NO

| | | | CANTIDAD HIJOS <= 4: SI (17.0/1.0)

| | | | CANTIDAD HIJOS > 4: NO (3.0/1.0)

| | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO

| | | | TALLA <= 1.45: NO (5.0)
| | | | TALLA > 1.45
| | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = SI
| | | | CANTIDAD ABORTOS <= 0: SI (9.0)
| | | | CANTIDAD ABORTOS > 0
| | | | SITUACION FISICA = NORMAL
| | | | ANTECEDENTES FAMILIARES <= 1: SI (3.0)
| | | | ANTECEDENTES FAMILIARES > 1: NO (7.0)
| | | | SITUACION FISICA = SOBREPESO
| | | | FRECUENCIA CARDIACA <= 65: NO (4.0/1.0)
| | | | FRECUENCIA CARDIACA > 65: SI (8.0)
| | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = NO
| | | | CANTIDAD ABORTOS <= 0
| | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI: NO (5.0)
| | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO
| | | | CANTIDAD EMBARAZOS <= 3: NO (4.0)
| | | | CANTIDAD EMBARAZOS > 3: SI (2.0)
| | | | CANTIDAD ABORTOS > 0
| | | | ANTECEDENTES FAMILIARES <= 1
| | | | MAMAS Y PEZÓN CON HERIDAS = SI: NO (5.0)
| | | | MAMAS Y PEZÓN CON HERIDAS = NO: SI (4.0/1.0)
| | | | ANTECEDENTES FAMILIARES > 1: SI (13.0/2.0)
| RETRACCIÓN DE LOS PEZONES = NO
| | CANTIDAD EMBARAZOS <= 5
| | CANTIDAD EMBARAZOS <= 4
| | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI
| | | | CANTIDAD ABORTOS <= 0: SI (17.0)
| | | | CANTIDAD ABORTOS > 0

| | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI
| | | | | | EDAD <= 51: NO (6.0)
| | | | | | EDAD > 51: SI (4.0)
| | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO
| | | | | | PESO <= 79.56: SI (9.0)
| | | | | | PESO > 79.56: NO (3.0/1.0)
| | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO
| | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = SI
| | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA
MAMA = SI
| | | | | | MAMAS Y PEZÓN CON HERIDAS = SI
| | | | | | SITUACION FISICA = NORMAL: SI (4.0)
| | | | | | SITUACION FISICA = SOBREPESO: NO (2.0)
| | | | | | MAMAS Y PEZÓN CON HERIDAS = NO
| | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI: NO
(10.0/1.0)
| | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO
| | | | | | | CANTIDAD ABORTOS <= 0: NO (3.0/1.0)
| | | | | | | CANTIDAD ABORTOS > 0: SI (3.0)
| | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA
MAMA = NO
| | | | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = SI
| | | | | | | EDAD MENARQUIA <= 13: SI (4.0)
| | | | | | | EDAD MENARQUIA > 13: NO (5.0/1.0)
| | | | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = NO: SI (8.0/1.0)
| | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = NO
| | | | | CANTIDAD EMBARAZOS <= 1: SI (7.0)
| | | | | CANTIDAD EMBARAZOS > 1
| | | | | | SITUACION FISICA = NORMAL: SI (3.0)

| | | | | | | SITUACION FISICA = SOBREPESO: NO (4.0/1.0)

| | | CANTIDAD EMBARAZOS > 4

| | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI

| | | | | SITUACION FISICA = NORMAL

| | | | | TALLA <= 1.64: NO (2.0)

| | | | | TALLA > 1.64: SI (2.0)

| | | | | SITUACION FISICA = SOBREPESO: NO (12.0/2.0)

| | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO

| | | | | EDAD MENARQUIA <= 14: SI (12.0/3.0)

| | | | | EDAD MENARQUIA > 14: NO (3.0)

| | CANTIDAD EMBARAZOS > 5: SI (30.0/3.0)

DIAGNOSTICO DE CANCER MAMA = NEGATIVO

| AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = SI

| | RETRACCIÓN DE LOS PEZONES = SI

| | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI

| | | | CANTIDAD ABORTOS <= 2: NO (9.0/1.0)

| | | | CANTIDAD ABORTOS > 2: SI (2.0)

| | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO: NO (18.0)

| | RETRACCIÓN DE LOS PEZONES = NO

| | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI

| | | | MAMAS Y PEZÓN CON HERIDAS = SI

| | | | | CANTIDAD HIJOS <= 3: SI (4.0/1.0)

| | | | | CANTIDAD HIJOS > 3: NO (5.0)

| | | | MAMAS Y PEZÓN CON HERIDAS = NO: NO (15.0)

| | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO

| | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI

| | | | | CANTIDAD ABORTOS <= 1

| | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI

| | | | | | CANTIDAD ABORTOS <= 0: NO (3.0/1.0)

| | | | | | CANTIDAD ABORTOS > 0: SI (6.0/1.0)

| | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO: NO
(12.0/2.0)

| | | | | | CANTIDAD ABORTOS > 1: NO (18.0/1.0)

| | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO

| | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI

| | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI: NO
(8.0)

| | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO

| | | | | | CANTIDAD EMBARAZOS <= 4: SI (2.0)

| | | | | | CANTIDAD EMBARAZOS > 4: NO (2.0)

| | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO

| | | | | | EDAD <= 43: NO (3.0)

| | | | | | EDAD > 43: SI (11.0/2.0)

| AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = NO

| | DOLOR EN CUALQUIER PARTE DE LA MAMA = SI

| | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI

| | | | TALLA <= 1.84: NO (20.0)

| | | | TALLA > 1.84

| | | | | SITUACION FISICA = NORMAL

| | | | | | CANTIDAD EMBARAZOS <= 5: SI (4.0)

| | | | | | CANTIDAD EMBARAZOS > 5: NO (2.0)

| | | | | SITUACION FISICA = SOBREPESO: NO (2.0)

| | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO

| | | | SECRECIÓN DE SANGRE O PUS POR EL PEZÓN = SI

| | | | | CANTIDAD EMBARAZOS <= 2: SI (4.0)

| | | | | CANTIDAD EMBARAZOS > 2

| | | | | CANTIDAD HIJOS <= 3: NO (9.0)
| | | | | CANTIDAD HIJOS > 3
| | | | | PESO <= 67: NO (3.0)
| | | | | PESO > 67: SI (3.0)
| | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = NO
| | | | CANTIDAD EMBARAZOS <= 4: NO (7.0)
| | | | CANTIDAD EMBARAZOS > 4
| | | | | SITUACION FISICA = NORMAL
| | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI: SI (3.0)
| | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO
| | | | | EDAD <= 51: NO (2.0)
| | | | | EDAD > 51: SI (2.0)
| | | | | SITUACION FISICA = SOBREPESO: NO (8.0/1.0)
| | DOLOR EN CUALQUIER PARTE DE LA MAMA = NO
| | | CANTIDAD ABORTOS <= 2
| | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI
| | | | TALLA <= 1.51: SI (8.0/1.0)
| | | | TALLA > 1.51
| | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI
| | | | | MAMAS Y PEZÓN CON HERIDAS = SI
| | | | | EDAD PRIMER EMBARAZO <= 22: SI (6.0)
| | | | | EDAD PRIMER EMBARAZO > 22: NO (4.0/1.0)
| | | | | MAMAS Y PEZÓN CON HERIDAS = NO
| | | | | TALLA <= 1.8: NO (9.0)
| | | | | TALLA > 1.8: SI (3.0/1.0)
| | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO
| | | | | PESO <= 65.4: NO (17.0/1.0)
| | | | | PESO > 65.4

| | | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI
 | | | | | | | | CANTIDAD EMBARAZOS <= 5: NO (6.0)
 | | | | | | | | CANTIDAD EMBARAZOS > 5: SI (2.0)
 | | | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO: SI
 (5.0)
 | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO
 | | | | FRECUENCIA CARDIACA <= 78: NO (31.0/2.0)
 | | | | FRECUENCIA CARDIACA > 78
 | | | | CANTIDAD EMBARAZOS <= 5
 | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI
 | | | | EDAD MENARQUIA <= 12: NO (8.0)
 | | | | EDAD MENARQUIA > 12: SI (3.0/1.0)
 | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO: SI (5.0/1.0)
 | | | | CANTIDAD EMBARAZOS > 5: SI (5.0)
 | | | CANTIDAD ABORTOS > 2
 | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI: SI (15.0/3.0)
 | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO
 | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI
 | | | | PESO <= 75: SI (6.0/1.0)
 | | | | PESO > 75: NO (3.0/1.0)
 | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO: NO (3.0)

Modelo 4: Aumento del grosor o hinchazon

J48 pruned tree

DIAGNOSTICO DE CANCER MAMA = POSITIVO

| DOLOR EN CUALQUIER PARTE DE LA MAMA = SI
 | | SITUACION FISICA = NORMAL
 | | | RETRACCIÓN DE LOS PEZONES = SI

| | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI
| | | | TALLA <= 1.66
| | | | EDAD MENARQUIA <= 10: NO (3.0/1.0)
| | | | EDAD MENARQUIA > 10: SI (20.0/1.0)
| | | | TALLA > 1.66
| | | | TALLA <= 1.77: NO (10.0/1.0)
| | | | TALLA > 1.77: SI (7.0/1.0)
| | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO
| | | | AUSENCIA DE MENSTRUACIÓN = SI
| | | | CANTIDAD ABORTOS <= 0: NO (5.0)
| | | | CANTIDAD ABORTOS > 0
| | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI: SI
(4.0/1.0)
| | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO
| | | | PESO <= 57.87: SI (5.0/1.0)
| | | | PESO > 57.87: NO (5.0)
| | | | AUSENCIA DE MENSTRUACIÓN = NO
| | | | ANTECEDENTES FAMILIARES <= 1: SI (5.0)
| | | | ANTECEDENTES FAMILIARES > 1
| | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI
| | | | SECRECIÓN DE SANGRE O PUS POR EL PEZÓN = SI: NO (5.0)
| | | | SECRECIÓN DE SANGRE O PUS POR EL PEZÓN = NO
| | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI: NO (2.0)
| | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO: SI (2.0)
| | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO
| | | | CANTIDAD HIJOS <= 1: NO (3.0/1.0)
| | | | CANTIDAD HIJOS > 1: SI (5.0)
| | | | RETRACCIÓN DE LOS PEZONES = NO
| | | | FRECUENCIA CARDIACA <= 85

| | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI
| | | | | PESO <= 61.7
| | | | | SECRECIÓN DE SANGRE O PUS POR EL PEZÓN = SI
| | | | | EDAD <= 49: NO (2.0)
| | | | | EDAD > 49: SI (2.0)
| | | | | SECRECIÓN DE SANGRE O PUS POR EL PEZÓN = NO: NO (3.0)
| | | | | PESO > 61.7: SI (9.0)
| | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO: SI (10.0)
| | | | FRECUENCIA CARDIACA > 85: NO (5.0/1.0)
| | SITUACION FISICA = SOBREPESO
| | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI
| | | CANTIDAD EMBARAZOS <= 0: SI (6.0)
| | | CANTIDAD EMBARAZOS > 0
| | | | SECRECIÓN DE SANGRE O PUS POR EL PEZÓN = SI
| | | | | CANTIDAD HIJOS <= 2
| | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI
| | | | | PESO <= 66.56: SI (5.0)
| | | | | PESO > 66.56
| | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI
| | | | | MAMAS Y PEZÓN CON HERIDAS = SI: SI (2.0)
| | | | | MAMAS Y PEZÓN CON HERIDAS = NO: NO (4.0/1.0)
| | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO: NO
(3.0)
| | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO
| | | | | EDAD <= 50: SI (3.0/1.0)
| | | | | EDAD > 50: NO (5.0)
| | | | | CANTIDAD HIJOS > 2: SI (20.0/3.0)
| | | | | SECRECIÓN DE SANGRE O PUS POR EL PEZÓN = NO
| | | | | EDAD PRIMER EMBARAZO <= 16

| | | | | | | RETRACCIÓN DE LOS PEZONES = SI: NO (8.0)
| | | | | | | RETRACCIÓN DE LOS PEZONES = NO
| | | | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI: NO (3.0)
| | | | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO: SI (2.0)
| | | | | | | EDAD PRIMER EMBARAZO > 16
| | | | | | | CANTIDAD EMBARAZOS <= 5
| | | | | | | PESO <= 76.18
| | | | | | | ANTECEDENTES FAMILIARES <= 1
| | | | | | | MAMAS Y PEZÓN CON HERIDAS = SI
| | | | | | | FRECUENCIA CARDIACA <= 68: NO (2.0)
| | | | | | | FRECUENCIA CARDIACA > 68: SI (4.0)
| | | | | | | MAMAS Y PEZÓN CON HERIDAS = NO: NO (7.0/1.0)
| | | | | | | ANTECEDENTES FAMILIARES > 1
| | | | | | | EDAD <= 49: NO (4.0/1.0)
| | | | | | | EDAD > 49: SI (11.0)
| | | | | | | PESO > 76.18: NO (6.0)
| | | | | | | CANTIDAD EMBARAZOS > 5: SI (9.0/1.0)
| | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO
| | | | CANTIDAD EMBARAZOS <= 1
| | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI
| | | | | TALLA <= 1.56: SI (2.0)
| | | | | TALLA > 1.56: NO (2.0)
| | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO: SI (5.0)
| | | | CANTIDAD EMBARAZOS > 1
| | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI
| | | | | CANTIDAD EMBARAZOS <= 6
| | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI
| | | | | CANTIDAD EMBARAZOS <= 2: SI (3.0/1.0)

| | | | | | | | CANTIDAD EMBARAZOS > 2: NO (15.0/1.0)
| | | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO
| | | | | | | | CANTIDAD EMBARAZOS <= 4
| | | | | | | | FRECUENCIA CARDIACA <= 67: SI (4.0)
| | | | | | | | FRECUENCIA CARDIACA > 67
| | | | | | | | EDAD MENARQUIA <= 11: SI (4.0/1.0)
| | | | | | | | EDAD MENARQUIA > 11: NO (6.0)
| | | | | | | | CANTIDAD EMBARAZOS > 4: NO (4.0)
| | | | | | | | CANTIDAD EMBARAZOS > 6: SI (2.0)
| | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO
| | | | | | | | TALLA <= 1.58: SI (6.0/1.0)
| | | | | | | | TALLA > 1.58
| | | | | | | | EDAD PRIMER EMBARAZO <= 17: SI (6.0/1.0)
| | | | | | | | EDAD PRIMER EMBARAZO > 17
| | | | | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI: NO (13.0/1.0)
| | | | | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO
| | | | | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = SI: SI (5.0/1.0)
| | | | | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = NO: NO (6.0/1.0)
| | | | | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = NO
| | | | | | | | RETRACCIÓN DE LOS PEZONES = SI
| | | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI
| | | | | | | | MAMAS Y PEZÓN CON HERIDAS = SI
| | | | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI: SI (12.0)
| | | | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO
| | | | | | | | EDAD PRIMER EMBARAZO <= 18
| | | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI: SI (5.0)
| | | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO
| | | | | | | | PESO <= 57: SI (2.0)

| | | | | | | PESO > 57: NO (2.0)
| | | | | | EDAD PRIMER EMBARAZO > 18: NO (8.0/1.0)
| | | | MAMAS Y PEZÓN CON HERIDAS = NO
| | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI
| | | | | EDAD MENARQUIA <= 11: NO (5.0)
| | | | | EDAD MENARQUIA > 11
| | | | | FRECUENCIA CARDIACA <= 67: NO (2.0)
| | | | | FRECUENCIA CARDIACA > 67: SI (7.0/1.0)
| | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO: NO (7.0)
| | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO
| | | | CANTIDAD HIJOS <= 2: SI (18.0/1.0)
| | | | CANTIDAD HIJOS > 2
| | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI
| | | | | AUSENCIA DE MENSTRUACIÓN = SI
| | | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI: NO (6.0/2.0)
| | | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO: SI (2.0)
| | | | | | AUSENCIA DE MENSTRUACIÓN = NO: SI (6.0)
| | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO
| | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI: NO (5.0)
| | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO
| | | | | | MAMAS Y PEZÓN CON HERIDAS = SI: NO (2.0)
| | | | | | MAMAS Y PEZÓN CON HERIDAS = NO: SI (2.0)
| | | RETRACCIÓN DE LOS PEZONES = NO: SI (46.0/7.0)
DIAGNOSTICO DE CANCER MAMA = NEGATIVO
| | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI
| | AUSENCIA DE MENSTRUACIÓN = SI: NO (31.0/3.0)
| | AUSENCIA DE MENSTRUACIÓN = NO
| | | DOLOR EN CUALQUIER PARTE DE LA MAMA = SI: NO (27.0/3.0)

| | | DOLOR EN CUALQUIER PARTE DE LA MAMA = NO

| | | | ANTECEDENTES FAMILIARES <= 3

| | | | | EDAD MENARQUIA <= 14

| | | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = SI: NO (17.0/1.0)

| | | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = NO

| | | | | | TALLA <= 1.57: SI (3.0)

| | | | | | TALLA > 1.57

| | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI

| | | | | | | | CANTIDAD HIJOS <= 1

| | | | | | | | | EDAD MENARQUIA <= 11: NO (2.0)

| | | | | | | | | EDAD MENARQUIA > 11: SI (3.0)

| | | | | | | | | CANTIDAD HIJOS > 1: NO (4.0)

| | | | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO: NO (9.0)

| | | | | EDAD MENARQUIA > 14

| | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI: SI (5.0)

| | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO: NO (4.0/1.0)

| | | | ANTECEDENTES FAMILIARES > 3

| | | | | CANTIDAD ABORTOS <= 1: SI (7.0)

| | | | | CANTIDAD ABORTOS > 1

| | | | | | CANTIDAD ABORTOS <= 2: NO (3.0)

| | | | | | CANTIDAD ABORTOS > 2

| | | | | | | EDAD <= 55: SI (2.0)

| | | | | | | EDAD > 55: NO (2.0)

| PIEL HUNDIDA DE LA MAMA O PEZÓN = NO

| | MAMAS Y PEZÓN CON HERIDAS = SI

| | | RETRACCIÓN DE LOS PEZONES = SI

| | | | AUSENCIA DE MENSTRUACIÓN = SI

| | | | | CANTIDAD HIJOS <= 2: SI (3.0/1.0)
| | | | | CANTIDAD HIJOS > 2: NO (10.0)
| | | | | AUSENCIA DE MENSTRUACIÓN = NO
| | | | | CANTIDAD ABORTOS <= 0: SI (4.0)
| | | | | CANTIDAD ABORTOS > 0
| | | | | SITUACION FISICA = NORMAL
| | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI
| | | | | TALLA <= 1.77: SI (9.0/1.0)
| | | | | TALLA > 1.77: NO (4.0)
| | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO: NO
(6.0/1.0)
| | | | | SITUACION FISICA = SOBREPESO: NO (4.0)
| | | | | RETRACCIÓN DE LOS PEZONES = NO
| | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = SI
| | | | | CANTIDAD EMBARAZOS <= 1: SI (5.0)
| | | | | CANTIDAD EMBARAZOS > 1
| | | | | SITUACION FISICA = NORMAL
| | | | | EDAD PRIMER EMBARAZO <= 21
| | | | | EDAD MENARQUIA <= 11
| | | | | CANTIDAD EMBARAZOS <= 2: NO (3.0/1.0)
| | | | | CANTIDAD EMBARAZOS > 2: SI (2.0)
| | | | | EDAD MENARQUIA > 11: NO (8.0)
| | | | | EDAD PRIMER EMBARAZO > 21: SI (3.0)
| | | | | SITUACION FISICA = SOBREPESO: NO (4.0)
| | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = NO
| | | | | FRECUENCIA CARDIACA <= 78
| | | | | CANTIDAD ABORTOS <= 2
| | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = SI
| | | | | TALLA <= 1.56: SI (6.0)

| | | | | | | | TALLA > 1.56

| | | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI

| | | | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA =
SI: SI (3.0)

| | | | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA =
NO

| | | | | | | | | | EDAD MENARQUIA <= 11

| | | | | | | | | | EDAD <= 57: NO (2.0)

| | | | | | | | | | EDAD > 57: SI (2.0)

| | | | | | | | | | EDAD MENARQUIA > 11: NO (2.0)

| | | | | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO:
NO (3.0)

| | | | | | | | SECRECIÓN DE SANGRE O PUS POR EL PEZÓN = NO: SI (6.0/1.0)

| | | | | | CANTIDAD ABORTOS > 2: NO (6.0/1.0)

| | | | | FRECUENCIA CARDIACA > 78: SI (16.0/1.0)

| | MAMAS Y PEZÓN CON HERIDAS = NO

| | | SITUACION FISICA = NORMAL

| | | | CANTIDAD EMBARAZOS <= 6

| | | | | SECRECIÓN DE SANGRE O PUS POR EL PEZÓN = SI

| | | | | EDAD <= 54

| | | | | | FRECUENCIA CARDIACA <= 80: SI (13.0)

| | | | | | FRECUENCIA CARDIACA > 80

| | | | | | CANTIDAD EMBARAZOS <= 4: NO (3.0)

| | | | | | CANTIDAD EMBARAZOS > 4: SI (2.0)

| | | | | | EDAD > 54

| | | | | | ANTECEDENTES FAMILIARES <= 2: NO (4.0)

| | | | | | ANTECEDENTES FAMILIARES > 2: SI (3.0/1.0)

| | | | | SECRECIÓN DE SANGRE O PUS POR EL PEZÓN = NO

| | | | | | PESO <= 85.31

| | | | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = SI
 | | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI: SI
 (2.0)
 | | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO
 | | | | | | | EDAD PRIMER EMBARAZO <= 22: NO (5.0)
 | | | | | | | EDAD PRIMER EMBARAZO > 22: SI (2.0)
 | | | | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = NO
 | | | | | | | TALLA <= 1.88: NO (15.0/1.0)
 | | | | | | | TALLA > 1.88: SI (3.0/1.0)
 | | | | | | PESO > 85.31: SI (4.0)
 | | | | CANTIDAD EMBARAZOS > 6: NO (8.0)
 | | | SITUACION FISICA = SOBREPESO
 | | | | ANTECEDENTES FAMILIARES <= 0: SI (2.0)
 | | | | ANTECEDENTES FAMILIARES > 0: NO (30.0/2.0)

MODELO 5: DOLOR EN CUALQUIER PARTE DE LA MAMA

J48 pruned tree

DIAGNOSTICO DE CANCER MAMA = POSITIVO

| AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = SI
 | | MAMAS Y PEZÓN CON HERIDAS = SI
 | | | SITUACION FISICA = NORMAL
 | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI
 | | | | | CANTIDAD ABORTOS <= 0: SI (9.0/1.0)
 | | | | | CANTIDAD ABORTOS > 0
 | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI
 | | | | | | TALLA <= 1.84: NO (6.0)
 | | | | | | TALLA > 1.84: SI (2.0)
 | | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO
 | | | | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = SI

| | | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI
| | | | | | | | CANTIDAD HIJOS <= 3: NO (3.0)
| | | | | | | | CANTIDAD HIJOS > 3: SI (2.0)
| | | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO: SI
(4.0)
| | | | | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = NO
| | | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI
| | | | | | | | EDAD <= 49: NO (2.0)
| | | | | | | | EDAD > 49: SI (3.0)
| | | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO: NO
(2.0)
| | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO: SI (15.0/1.0)
| | | SITUACION FISICA = SOBREPESO
| | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI
| | | | | EDAD MENARQUIA <= 11
| | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI
| | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI
| | | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI: SI (3.0/1.0)
| | | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO: NO (3.0)
| | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO: SI (2.0)
| | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO: SI (6.0)
| | | | | EDAD MENARQUIA > 11
| | | | | PESO <= 83.23: NO (20.0/2.0)
| | | | | PESO > 83.23: SI (2.0)
| | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO
| | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI
| | | | | RETRACCIÓN DE LOS PEZONES = SI
| | | | | EDAD PRIMER EMBARAZO <= 20
| | | | | EDAD MENARQUIA <= 14

| | | | | | | | CANTIDAD EMBARAZOS <= 5: NO (5.0)
| | | | | | | | CANTIDAD EMBARAZOS > 5: SI (3.0/1.0)
| | | | | | | | EDAD MENARQUIA > 14: SI (2.0)
| | | | | | | | EDAD PRIMER EMBARAZO > 20: SI (5.0)
| | | | | | | | RETRACCIÓN DE LOS PEZONES = NO
| | | | | | | | FRECUENCIA CARDIACA <= 67: SI (2.0)
| | | | | | | | FRECUENCIA CARDIACA > 67: NO (7.0)
| | | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO
| | | | | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = SI
| | | | | | | | TALLA <= 1.72: NO (6.0)
| | | | | | | | TALLA > 1.72: SI (9.0/1.0)
| | | | | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = NO: SI (10.0/1.0)
| | MAMAS Y PEZÓN CON HERIDAS = NO
| | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI
| | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = SI
| | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI
| | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI
| | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI: NO
(5.0/1.0)
| | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO: SI
(5.0/1.0)
| | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO
| | | | | | CANTIDAD HIJOS <= 3
| | | | | | | EDAD PRIMER EMBARAZO <= 15: NO (3.0/1.0)
| | | | | | | EDAD PRIMER EMBARAZO > 15: SI (10.0)
| | | | | | | CANTIDAD HIJOS > 3: NO (2.0)
| | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO: SI (14.0/1.0)
| | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = NO
| | | | | FRECUENCIA CARDIACA <= 83

| | | | | | RETRACCIÓN DE LOS PEZONES = SI

| | | | | | ANTECEDENTES FAMILIARES <= 2

| | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI

| | | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI

| | | | | | | | FRECUENCIA CARDIACA <= 71: NO (2.0)

| | | | | | | | FRECUENCIA CARDIACA > 71: SI (3.0)

| | | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO: SI
(2.0)

| | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO:
NO (5.0/1.0)

| | | | | | | ANTECEDENTES FAMILIARES > 2: SI (4.0)

| | | | | | RETRACCIÓN DE LOS PEZONES = NO

| | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI: SI (8.0)

| | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO

| | | | | | | FRECUENCIA CARDIACA <= 78: NO (2.0)

| | | | | | | FRECUENCIA CARDIACA > 78: SI (2.0)

| | | | | FRECUENCIA CARDIACA > 83

| | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI

| | | | | | CANTIDAD HIJOS <= 4: NO (5.0/1.0)

| | | | | | CANTIDAD HIJOS > 4: SI (2.0)

| | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO: NO (7.0)

| | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO: SI (44.0/7.0)

| AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = NO

| | RETRACCIÓN DE LOS PEZONES = SI

| | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI

| | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI

| | | | | AUSENCIA DE MENSTRUACIÓN = SI

| | | | | CANTIDAD EMBARAZOS <= 5

| | | | | | CANTIDAD HIJOS <= 2

| | | | | | | | CANTIDAD HIJOS <= 0: NO (4.0)
| | | | | | | | CANTIDAD HIJOS > 0
| | | | | | | | FRECUENCIA CARDIACA <= 62: NO (2.0)
| | | | | | | | FRECUENCIA CARDIACA > 62: SI (5.0)
| | | | | | | | CANTIDAD HIJOS > 2: NO (7.0)
| | | | | | CANTIDAD EMBARAZOS > 5: SI (4.0)
| | | | | AUSENCIA DE MENSTRUACIÓN = NO
| | | | | PESO <= 64: NO (2.0)
| | | | | PESO > 64: SI (6.0)
| | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO
| | | | ANTECEDENTES FAMILIARES <= 1
| | | | | CANTIDAD EMBARAZOS <= 4
| | | | | FRECUENCIA CARDIACA <= 86: NO (6.0)
| | | | | FRECUENCIA CARDIACA > 86: SI (2.0)
| | | | | CANTIDAD EMBARAZOS > 4: SI (3.0)
| | | | | ANTECEDENTES FAMILIARES > 1: SI (27.0/3.0)
| | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO
| | | | EDAD PRIMER EMBARAZO <= 15
| | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI: NO (6.0/1.0)
| | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO: SI (2.0)
| | | | EDAD PRIMER EMBARAZO > 15: SI (51.0/7.0)
| | RETRACCIÓN DE LOS PEZONES = NO
| | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI: SI (25.0/1.0)
| | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO
| | | | CANTIDAD EMBARAZOS <= 6
| | | | | PESO <= 48.4: NO (3.0/1.0)
| | | | | PESO > 48.4: SI (21.0/2.0)
| | | | CANTIDAD EMBARAZOS > 6: NO (2.0)

DIAGNOSTICO DE CANCER MAMA = NEGATIVO

| EDAD <= 60

| | SITUACION FISICA = NORMAL

| | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI

| | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI

| | | | | AUSENCIA DE MENSTRUACIÓN = SI: NO (14.0/1.0)

| | | | | AUSENCIA DE MENSTRUACIÓN = NO

| | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI: NO (6.0)

| | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO

| | | | | ANTECEDENTES FAMILIARES <= 1: SI (4.0)

| | | | | ANTECEDENTES FAMILIARES > 1

| | | | | FRECUENCIA CARDIACA <= 74: SI (3.0)

| | | | | FRECUENCIA CARDIACA > 74: NO (4.0)

| | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO

| | | | | EDAD MENARQUIA <= 13

| | | | | FRECUENCIA CARDIACA <= 66: NO (14.0/1.0)

| | | | | FRECUENCIA CARDIACA > 66

| | | | | ANTECEDENTES FAMILIARES <= 1

| | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI: NO (4.0)

| | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO

| | | | | AUSENCIA DE MENSTRUACIÓN = SI: NO (2.0)

| | | | | AUSENCIA DE MENSTRUACIÓN = NO

| | | | | EDAD <= 46: NO (2.0)

| | | | | EDAD > 46: SI (3.0)

| | | | | ANTECEDENTES FAMILIARES > 1: SI (15.0/4.0)

| | | | | EDAD MENARQUIA > 13

| | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI: SI (10.0/1.0)

| | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO

| | | | | | ANTECEDENTES FAMILIARES <= 2
| | | | | | ANTECEDENTES FAMILIARES <= 0: SI (2.0)
| | | | | | ANTECEDENTES FAMILIARES > 0: NO (4.0)
| | | | | | ANTECEDENTES FAMILIARES > 2: SI (4.0)
| | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO
| | | | RETRACCIÓN DE LOS PEZONES = SI: NO (35.0/5.0)
| | | | RETRACCIÓN DE LOS PEZONES = NO
| | | | EDAD <= 55
| | | | | FRECUENCIA CARDIACA <= 83
| | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI
| | | | | CANTIDAD EMBARAZOS <= 2: SI (6.0)
| | | | | CANTIDAD EMBARAZOS > 2
| | | | | TALLA <= 1.82
| | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI
| | | | | | CANTIDAD HIJOS <= 2: SI (2.0)
| | | | | | CANTIDAD HIJOS > 2: NO (4.0)
| | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO:
NO (7.0)
| | | | | | TALLA > 1.82: SI (4.0)
| | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO
| | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA
MAMA = SI: NO (8.0/1.0)
| | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA
MAMA = NO
| | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI: SI (4.0)
| | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO: NO
(4.0/1.0)
| | | | | FRECUENCIA CARDIACA > 83: NO (9.0)
| | | | EDAD > 55: NO (11.0)
| | SITUACION FISICA = SOBREPESO

| | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = SI: NO (31.0)

| | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = NO

| | | | RETRACCIÓN DE LOS PEZONES = SI

| | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI

| | | | | EDAD MENARQUIA <= 11: SI (2.0)

| | | | | EDAD MENARQUIA > 11: NO (6.0)

| | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO: NO (9.0)

| | | | | RETRACCIÓN DE LOS PEZONES = NO

| | | | | TALLA <= 1.87: NO (38.0/6.0)

| | | | | TALLA > 1.87

| | | | | EDAD MENARQUIA <= 11: NO (2.0)

| | | | | EDAD MENARQUIA > 11: SI (6.0)

| EDAD > 60

| | MAMAS Y PEZÓN CON HERIDAS = SI

| | | PESO <= 82.33

| | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI: NO (9.0)

| | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO

| | | | | EDAD <= 61: SI (2.0)

| | | | | EDAD > 61: NO (4.0)

| | | PESO > 82.33: SI (3.0)

| | MAMAS Y PEZÓN CON HERIDAS = NO

| | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI: SI (15.0/1.0)

| | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO

| | | | RETRACCIÓN DE LOS PEZONES = SI: NO (5.0)

| | | | RETRACCIÓN DE LOS PEZONES = NO

| | | | | EDAD <= 64: SI (6.0)

| | | | | EDAD > 64: NO (3.0)

Modelo 6: Retracción de los pezones

J48 pruned tree

DIAGNOSTICO DE CANCER MAMA = POSITIVO

- | AUSENCIA DE MENSTRUACIÓN = SI
- | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = SI
- | | | CANTIDAD EMBARAZOS ≤ 7
- | | | | CANTIDAD ABORTOS ≤ 0
- | | | | | SITUACION FISICA = NORMAL
- | | | | | MAMAS Y PEZÓN CON HERIDAS = SI: NO (8.0)
- | | | | | MAMAS Y PEZÓN CON HERIDAS = NO
- | | | | | EDAD PRIMER EMBARAZO ≤ 17 : SI (2.0)
- | | | | | EDAD PRIMER EMBARAZO > 17 : NO (3.0)
- | | | | | SITUACION FISICA = SOBREPESO
- | | | | | PESO ≤ 76.18
- | | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI
- | | | | | | CANTIDAD EMBARAZOS ≤ 2
- | | | | | | PESO ≤ 49.93 : NO (3.0)
- | | | | | | PESO > 49.93 : SI (3.0)
- | | | | | | CANTIDAD EMBARAZOS > 2 : NO (5.0)
- | | | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO: SI (10.0/1.0)
- | | | | | PESO > 76.18 : NO (6.0)
- | | | | CANTIDAD ABORTOS > 0
- | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI
- | | | | | SITUACION FISICA = NORMAL: SI (19.0/1.0)
- | | | | | SITUACION FISICA = SOBREPESO
- | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI

| | | | | | | | EDAD <= 56

| | | | | | | | CANTIDAD EMBARAZOS <= 2: SI (3.0/1.0)

| | | | | | | | CANTIDAD EMBARAZOS > 2: NO (7.0)

| | | | | | | | EDAD > 56: SI (4.0)

| | | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO

| | | | | | | | CANTIDAD ABORTOS <= 1

| | | | | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI: NO (5.0/1.0)

| | | | | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO: SI (2.0)

| | | | | | | | CANTIDAD ABORTOS > 1: SI (11.0)

| | | | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO

| | | | | | | | FRECUENCIA CARDIACA <= 85

| | | | | | | | ANTECEDENTES FAMILIARES <= 3

| | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI

| | | | | | | | CANTIDAD ABORTOS <= 2

| | | | | | | | EDAD PRIMER EMBARAZO <= 19

| | | | | | | | SITUACION FISICA = NORMAL

| | | | | | | | | | CANTIDAD ABORTOS <= 1: SI (4.0/1.0)

| | | | | | | | | | CANTIDAD ABORTOS > 1: NO (2.0)

| | | | | | | | | | SITUACION FISICA = SOBREPESO: NO (4.0)

| | | | | | | | | | EDAD PRIMER EMBARAZO > 19: SI (3.0)

| | | | | | | | | | CANTIDAD ABORTOS > 2: NO (4.0)

| | | | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO

| | | | | | | | | | SECRECIÓN DE SANGRE O PUS POR EL PEZÓN = SI

| | | | | | | | | | EDAD <= 49: SI (2.0)

| | | | | | | | | | EDAD > 49: NO (3.0)

| | | | | | | | | | SECRECIÓN DE SANGRE O PUS POR EL PEZÓN = NO: SI (6.0/1.0)

| | | | | | | | | | ANTECEDENTES FAMILIARES > 3: SI (10.0/1.0)

| | | | | | | | | | FRECUENCIA CARDIACA > 85: NO (5.0)

| | | CANTIDAD EMBARAZOS > 7: NO (8.0/1.0)

| | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = NO

| | | MAMAS Y PEZÓN CON HERIDAS = SI

| | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = SI

| | | | | EDAD MENARQUIA <= 14

| | | | | CANTIDAD ABORTOS <= 2

| | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI

| | | | | | CANTIDAD HIJOS <= 4

| | | | | | | SITUACION FISICA = NORMAL

| | | | | | | PESO <= 71.29: NO (3.0)

| | | | | | | PESO > 71.29: SI (2.0)

| | | | | | | SITUACION FISICA = SOBREPESO: NO (5.0)

| | | | | | | CANTIDAD HIJOS > 4: SI (2.0)

| | | | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO

| | | | | | | CANTIDAD ABORTOS <= 0: NO (2.0)

| | | | | | | CANTIDAD ABORTOS > 0

| | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI: SI
(6.0)

| | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO:
NO (3.0/1.0)

| | | | | | CANTIDAD ABORTOS > 2: NO (4.0)

| | | | | EDAD MENARQUIA > 14

| | | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = SI

| | | | | | EDAD PRIMER EMBARAZO <= 22: SI (4.0)

| | | | | | EDAD PRIMER EMBARAZO > 22: NO (2.0)

| | | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = NO: SI (6.0)

| | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = NO

| | | | | ANTECEDENTES FAMILIARES <= 0: NO (2.0)

| | | | | ANTECEDENTES FAMILIARES > 0: SI (11.0/1.0)

| | | MAMAS Y PEZÓN CON HERIDAS = NO

| | | | CANTIDAD HIJOS <= 2

| | | | | SITUACION FISICA = NORMAL: SI (9.0/1.0)

| | | | | SITUACION FISICA = SOBREPESO

| | | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = SI

| | | | | | CANTIDAD ABORTOS <= 1

| | | | | | | EDAD PRIMER EMBARAZO <= 20: NO (2.0)

| | | | | | | EDAD PRIMER EMBARAZO > 20: SI (6.0)

| | | | | | | CANTIDAD ABORTOS > 1: NO (10.0/1.0)

| | | | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = NO: SI (6.0/1.0)

| | | | CANTIDAD HIJOS > 2: SI (29.0/2.0)

| AUSENCIA DE MENSTRUACIÓN = NO

| | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI

| | | SITUACION FISICA = NORMAL

| | | | CANTIDAD ABORTOS <= 2

| | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = SI

| | | | | | EDAD <= 48: SI (7.0)

| | | | | | EDAD > 48

| | | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = SI: NO (8.0/1.0)

| | | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = NO: SI (5.0/1.0)

| | | | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = NO

| | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI

| | | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = SI: SI (6.0/1.0)

| | | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = NO: NO (2.0)

| | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO: SI (7.0)

| | | | CANTIDAD ABORTOS > 2: SI (8.0)

| | | SITUACION FISICA = SOBREPESO

| | | | CANTIDAD EMBARAZOS <= 5

| | | | | CANTIDAD ABORTOS <= 1

| | | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = SI

| | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI

| | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI

| | | | | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = SI: NO (4.0)

| | | | | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = NO: SI (3.0/1.0)

| | | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO

| | | | | | | | | CANTIDAD EMBARAZOS <= 4: SI (4.0)

| | | | | | | | | CANTIDAD EMBARAZOS > 4: NO (2.0)

| | | | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO

| | | | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = SI

| | | | | | | | | PESO <= 69.56: SI (4.0)

| | | | | | | | | PESO > 69.56: NO (2.0)

| | | | | | | | | CAMBIO EN EL TAMAÑO O LA FORMA DE LA MAMA = NO: SI (8.0/1.0)

| | | | | | | | | DOLOR EN CUALQUIER PARTE DE LA MAMA = NO: SI (15.0/1.0)

| | | | | CANTIDAD ABORTOS > 1

| | | | | | EDAD MENARQUIA <= 11: NO (7.0)

| | | | | | EDAD MENARQUIA > 11

| | | | | | | EDAD <= 46: NO (4.0)

| | | | | | | EDAD > 46

| | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI: SI (8.0/1.0)

| | | | | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO

| | | | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI: SI (6.0/1.0)

| | | | | | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO: NO
(5.0/1.0)

| | | | CANTIDAD EMBARAZOS > 5: SI (7.0)

| | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO

| | | SITUACION FISICA = NORMAL: SI (16.0/1.0)

| | | SITUACION FISICA = SOBREPESO

| | | | FRECUENCIA CARDIACA <= 80: SI (27.0/2.0)

| | | | FRECUENCIA CARDIACA > 80

| | | | | EDAD <= 57: NO (11.0/3.0)

| | | | | EDAD > 57: SI (4.0)

DIAGNOSTICO DE CANCER MAMA = NEGATIVO

| FRECUENCIA CARDIACA <= 83

| | SITUACION FISICA = NORMAL

| | | DOLOR EN CUALQUIER PARTE DE LA MAMA = SI: NO (63.0/13.0)

| | | DOLOR EN CUALQUIER PARTE DE LA MAMA = NO

| | | | AUSENCIA DE MENSTRUACIÓN = SI

| | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = SI: NO (19.0/2.0)

| | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = NO

| | | | | CANTIDAD HIJOS <= 0: SI (4.0)

| | | | | CANTIDAD HIJOS > 0

| | | | | | EDAD MENARQUIA <= 14: NO (13.0/2.0)

| | | | | | EDAD MENARQUIA > 14: SI (2.0)

| | | | AUSENCIA DE MENSTRUACIÓN = NO

| | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI

| | | | | | EDAD PRIMER EMBARAZO <= 21

| | | | | | PESO <= 79.04: NO (13.0)

| | | | | | PESO > 79.04: SI (2.0)

| | | | | | EDAD PRIMER EMBARAZO > 21

| | | | | | PESO <= 70.3: SI (5.0)

| | | | | PESO > 70.3: NO (2.0)

| | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO

| | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = SI

| | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = SI

| | | | | EDAD MENARQUIA <= 13

| | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = SI

| | | | | | EDAD MENARQUIA <= 11: NO (3.0)

| | | | | | EDAD MENARQUIA > 11: SI (2.0)

| | | | | | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = NO: SI (3.0)

| | | | | | EDAD MENARQUIA > 13: NO (4.0)

| | | | | | SECRESIÓN DE SANGRE O PUS POR EL PEZÓN = NO

| | | | | | MAMAS Y PEZÓN CON HERIDAS = SI: SI (8.0/1.0)

| | | | | | MAMAS Y PEZÓN CON HERIDAS = NO

| | | | | | TALLA <= 1.7: NO (5.0)

| | | | | | TALLA > 1.7: SI (3.0)

| | | | | ENROJECIMIENTO Y ENDURECIMIENTO DE LA MAMA = NO

| | | | | CANTIDAD ABORTOS <= 2

| | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = SI

| | | | | | EDAD MENARQUIA <= 13: NO (2.0)

| | | | | | EDAD MENARQUIA > 13: SI (5.0/1.0)

| | | | | | IRRITACIÓN EN LA PIEL DE LA MAMA = NO: NO (9.0/1.0)

| | | | | | CANTIDAD ABORTOS > 2: SI (3.0)

| | SITUACION FISICA = SOBREPESO

| | | EDAD MENARQUIA <= 10

| | | | EDAD PRIMER EMBARAZO <= 17: SI (6.0/1.0)

| | | | EDAD PRIMER EMBARAZO > 17: NO (11.0/1.0)

| | | EDAD MENARQUIA > 10: NO (60.0/5.0)

| FRECUENCIA CARDIACA > 83

| | MAMAS Y PEZÓN CON HERIDAS = SI

| | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = SI

| | | | EDAD PRIMER EMBARAZO <= 17: NO (11.0)

| | | | EDAD PRIMER EMBARAZO > 17

| | | | EDAD MENARQUIA <= 13

| | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = SI

| | | | | CANTIDAD HIJOS <= 3: SI (5.0/1.0)

| | | | | CANTIDAD HIJOS > 3: NO (2.0)

| | | | | PRESENCIA DE TUMORES (BULTOS) EN LA MAMA = NO: SI (2.0)

| | | | | EDAD MENARQUIA > 13: NO (4.0)

| | | AUMENTO DEL GROSOR O HINCHAZÓN DE UNA PARTE DE LA MAMA = NO

| | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI

| | | | | TALLA <= 1.72: NO (8.0)

| | | | | TALLA > 1.72: SI (2.0)

| | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO

| | | | | CANTIDAD EMBARAZOS <= 2: NO (4.0/1.0)

| | | | | CANTIDAD EMBARAZOS > 2: SI (10.0)

| | MAMAS Y PEZÓN CON HERIDAS = NO

| | | EDAD PRIMER EMBARAZO <= 16: SI (6.0)

| | | EDAD PRIMER EMBARAZO > 16

| | | | TALLA <= 1.84

| | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = SI

| | | | | EDAD MENARQUIA <= 14: NO (5.0/1.0)

| | | | | EDAD MENARQUIA > 14: SI (3.0)

| | | | | SECRECIÓN DEL PEZÓN QUE NO SEA LECHE MATERNA = NO

| | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = SI: NO (3.0/1.0)

| | | | | PIEL HUNDIDA DE LA MAMA O PEZÓN = NO: SI (10.0/1.0)

| | | | TALLA > 1.84: NO (4.0)

GLOSARIO DE TÉRMINOS

A

Alteraciones: Es el cambio que se presenta en la esencia o forma de una cosa.

Análisis estadísticos: Es un componente del análisis de datos. En el contexto de la inteligencia de negocios (BI), el análisis estadístico requiere recoger y escudriñar cada muestra de datos individual en una serie de artículos desde los cuales se puede extraer las muestras.

Análisis retrospectivos: El análisis retrospectivo se utiliza para calcular la precisión futura esperada de un método de pronóstico, lo que resulta útil para evaluar qué modelo de pronóstico debería considerarse más preciso.

Atoexamen: La habilidad para observar y tocar diferentes partes del cuerpo (genitales masculinos y femeninos, incluyendo las mamas) con el fin de detectar alguna anomalía para prevenir y detectar enfermedades y así preservar la salud sexual y reproductiva.

B

Base de datos: Es una herramienta para recopilar y organizar información.

Benigno: No canceroso; no invade los tejidos vecinos ni se disemina a otras partes del cuerpo.

B2-quimerina: Es una proteína con actividad GAP específica de Rac. Rac es una RhoGTPasa que regula múltiples procesos celulares, muchos de los cuales se encuentran alterados en cáncer de mama, como son proliferación, supervivencia, motilidad y desdiferenciación celular.

C

Cáncer de mama: Es un tumor maligno que se origina en las células del seno.

Catalyst: La cadena de valor empresarial, es un modelo teórico popularizado por Michael Porter que define las actividades de la empresa que van añadiendo valor al producto a medida que éste pasa por cada una de ellas.

D

Diagnóstico: Un diagnóstico son el o los resultados que se arrojan luego de un -----, evaluación o análisis sobre determinado ámbito u objeto.

E

Ecografía: Técnica de diagnóstico por la imagen que utiliza ultrasonidos para generar imágenes de distintos órganos.

G

Ganglios linfáticos: Son unas estructuras nodulares que forman parte del sistema linfático y forman agrupaciones en forma de racimos.

Gestión clínica: Es un cambio cultural que aportará nuevos valores al quehacer de los profesionales y al conjunto de la organización.

Glóbulo blanco: Células sanguíneas que intervienen en la destrucción de los microorganismos causantes de las infecciones.

Glóbulo rojo: células sanguíneas cuya función es transportar el oxígeno a los distintos tejidos.

H

Hormona: Sustancia producida por células vivas que circula a través del torrente

sanguíneo y que afecta a las células y a los tejidos que se encuentran lejos del lugar en donde ésta se originó.

I

Indicadores: Muestran los puntos problemáticos del proceso y nos ayudarán a caracterizarlos, comprenderlos y confirmarlos.

K

Kdd: Es un proceso iterativo e interactivo. **Iterativo** ya que la salida de alguna de las fases puede retroceder a pasos anteriores y porque a menudo son necesarias iteraciones para extraer conocimiento de alta calidad. Es **interactivo** por usuario, o más generalmente un experto en el dominio del problema, debe ayudar a la preparación de los datos y validación del conocimiento extraído.

L

Lesiones: Es un cambio anormal en la morfología o estructura de una parte del cuerpo producida por un daño externo o interno. Las heridas en la piel pueden considerarse lesiones producidas por un daño externo como los traumatismos.

M

Mamografía: Una imagen radiográfica de las mamas.

Médico: Es un profesional que practica la medicina que intenta mantener y recuperar la salud humana mediante el estudio, el diagnóstico y el tratamiento de la enfermedad o lesión del paciente.

Medida: Medida o disposición que se toma de manera anticipada para evitar que suceda una cosa considerada negativa.

Metodología: Es la serie de métodos y técnicas de rigor científico que se aplican sistemáticamente durante un proceso de investigación para alcanzar un resultado teóricamente válido.

Microcalcificaciones: Son pequeños depósitos de calcio dentro de los conductos. Se ven en la mamografía como pequeños puntos blancos. Según su tamaño, aspecto o distribución, pueden ser sospechosas de cáncer o no.

Minería de datos: La minería de datos es el proceso de detectar la información procesable de los conjuntos grandes de datos. Utiliza el análisis matemático para deducir los patrones y tendencias que existen en los datos.

Modelo predictivo: Los modelos predictivos son modelos de la relación entre el rendimiento específico de un sujeto en una muestra y uno o más atributos o características del mismo sujeto. El objetivo del modelo es evaluar la probabilidad de que un sujeto similar tenga el mismo rendimiento en una muestra diferente.

Morfología: La morfología humana estudia las estructuras del cuerpo humano desde distintos puntos de vista: se encarga de revisar los aspectos macroscópicos; también forma parte de la morfología humana el estudio microscópico de los tejidos que lo conforman (histología) y también se incluye dentro del área de la morfología humana la forma en que se desarrollan los tejidos desde el momento de la concepción (desarrollo embrionario)

P

Pacientes: Es aquella persona que sufre o padece algún dolor o malestar, es el sujeto

que recibe los servicios de un médico u otro profesional de la salud y se somete a un examen, a un tratamiento o a una intervención.

Patrones: Aquella serie de variables constantes, identificables dentro de un conjunto mayor de datos.

Predicción: Hecho o situación que se anuncia que sucederá en el futuro.

Pruebas de laboratorio: Procedimiento médico en el que se analiza una muestra de sangre, orina u otra sustancia del cuerpo. Las pruebas de laboratorio pueden ayudar a determinar un diagnóstico, planificar el tratamiento, verificar si el tratamiento es eficaz o vigilar la enfermedad con el transcurso del tiempo.

Q

Quimioterapia: El uso de medicamentos para tratar enfermedades; cuando se los utiliza para tratar un cáncer, el objetivo es destruir las células cancerosas.

S

Semna: Establece un conjunto de cinco fases para llevar a cabo el proceso de minería: Sample (Muestreo), Explore (Exploración), Modify (Modificación), Model (Modelado) y Assess (Evaluación). Está especialmente enfocada al desarrollo del modelo de minería, y quedan fuera de su alcance otros aspectos del proyecto como el conocimiento del problema en estudio o la planificación de la implementación.

Sobreviviente del cáncer de mama: Alguien que vive con, a través y más allá de un diagnóstico de cáncer de mama.

T

Tamizaje: El concepto de tamizaje se refiere a la evaluación masiva de sujetos asintomáticos respecto de una patología específica y antes que ellos consulten espontáneamente.

Tasa de mortalidad: Es la proporción de personas que mueren por una enfermedad entre los afectados por la misma en un periodo y área determinados. Es un indicador de la virulencia o de la gravedad de una enfermedad.

Técnica: Conjunto de procedimientos o recursos que se usan en un arte, en una ciencia o en una actividad determinada, en especial cuando se adquieren por medio de su práctica y requieren habilidad.

Tratamiento: Conjunto de medios de cualquier clase (higiénicos, farmacológicos, quirúrgicos o físicos) cuya finalidad es la curación o el alivio (paliación) de las enfermedades o síntomas.

Tumor: Un grupo de células cancerosas.

V

Vigilancia epidemiológico: Es el conjunto de actividades que permiten reunir la información indispensable para conocer en todo momento la conducta o historia natural de la enfermedad (o los problemas o eventos de salud), detectar o prever cualquier cambio que pueda ocurrir, con el fin de recomendar oportunamente las medidas indicadas que lleven a la prevención y el control de la enfermedad.