



**Universidad
Pontificia
Bolivariana**

**MINERÍA DE DATOS DE LA SALUD: ANÁLISIS DE LOS FACTORES QUE
INFLUYEN EN LA REALIZACIÓN DE CIRUGÍAS ESTÉTICAS**

SANDRA MILENA SÁNCHEZ SÁNCHEZ

UNIVERSIDAD PONTIFICIA BOLIVARIANA

ESCUELA INGENIERÍAS

FACULTAD DE INGENIERÍA EN TECNOLOGÍAS DE INFORMACIÓN Y
COMUNICACIÓN

MAESTRÍA EN TECNOLOGÍAS DE INFORMACIÓN Y COMUNICACIÓN

MEDELLÍN - 2017

**MINERÍA DE DATOS DE LA SALUD: ANÁLISIS DE LOS FACTORES QUE
INFLUYEN EN LA REALIZACIÓN DE CIRUGÍAS ESTÉTICAS**

SANDRA MILENA SÁNCHEZ SÁNCHEZ

Trabajo de grado para optar al título de Maestría en Tecnologías de la Información
y la Comunicación

Asesor

ANA ISABEL OVIEDO CARRASCAL

PhD. Ingeniería Electrónica

UNIVERSIDAD PONTIFICIA BOLIVARIANA

ESCUELA INGENIERÍAS

FACULTAD DE INGENIERÍA EN TECNOLOGÍAS DE INFORMACIÓN Y
COMUNICACIÓN

MAESTRÍA EN TECNOLOGÍAS DE INFORMACIÓN Y COMUNICACIÓN

MEDELLÍN - 2017

DECLARACIÓN ORIGINALIDAD

“Declaro que esta tesis (o trabajo de grado) no ha sido presentada para optar a un título, ya sea en igual forma o con variaciones, en esta o cualquier otra universidad”. Art. 82 Régimen Discente de Formación Avanzada, Universidad Pontificia Bolivariana.



FIRMA AUTOR

SANDRA MILENA SÁNCHEZ SÁNCHEZ

C.C. 1.039.449.960

Medellín, 13 de Marzo de 2017

RESUMEN

A nivel mundial, las cirugías estéticas se han convertido en procedimientos de constante realización y año tras año han venido en aumento. Según la Sociedad Internacional de Cirugía Plástica Estética - ISAPS, Colombia se ha mantenido a nivel mundial entre los ocho países con mayor número de prácticas de cirugías estéticas anuales. La cirugía estética requiere, como paso indispensable, la asistencia a una cita de valoración con el especialista, y sólo el 40% de éstas, culminan en la realización de algún procedimiento quirúrgico. Por tal razón, los especialistas en cirugía plástica requieren identificar los factores que influyen en la toma de decisión de los pacientes respecto a practicarse un procedimiento estético. Se pretende entonces, mediante minería de datos, descubrir el perfil de los pacientes con mayor probabilidad de realizarse una cirugía plástica y así, definir los factores que influyen en la decisión de someterse a cirugías estéticas como estrategia para la minimización de tiempo en espera en la asignación de consultas de valoración.

PALABRAS CLAVE: Minería de Datos, perfilamiento de pacientes, priorización de citas en cirugía plástica.

ABSTRACT

Globally, cosmetic surgery procedures have become constant realization and year after year have been increasing. According to the International Society of Plastic Surgery Aesthetics – ISAPS, Colombia has remained world-wide among the eight countries with the highest number of annual cosmetic surgeries. Cosmetic surgery requires, as an essential step, attendance at an appointment with the specialist, and only 40% of these, culminate in the completion of a surgical procedure. For this reason, specialists in plastic surgery need to identify the factors that influence the decision-making of patients regarding the practice of an aesthetic procedure. It is intended, through data mining, to discover the profile of the patients with greater probability of performing a plastic surgery and thus, to define the factors that influence the decision to undergo cosmetic surgeries as a strategy for the minimization of waiting time in the Allocation of valuation queries.

KEY WORDS: Data mining, profiling patients, prioritization appointment plastic surgery.

GLOSARIO

BRAQUITERIA: Es el tratamiento de tumores mediante el uso de material radioactivo.

CIRUGÍA ESTÉTICA: Es la rama de la cirugía plástica que se encarga de mejorar la apariencia estética de zonas del cuerpo-

CIRUGÍA MAXILOFACIAL: Es la rama de la cirugía que se encarga del tratamiento y corrección de las deformidades congénitas o adquiridas en los huesos del macizo facial.

MALFORMACIÓN CONGÉNITA: Es cualquier tipo de deformidad que se presenta al momento del nacimiento.

TABLA DE CONTENIDO

INTRODUCCIÓN	12
PARTE 1. FORMULACIÓN DEL PROYECTO Y REVISIÓN LITERARIA	14
1. DESCRIPCIÓN DEL PROBLEMA	15
2. JUSTIFICACIÓN	18
3. OBJETIVOS	20
4. METODOLOGÍA	21
5. MARCO TEÓRICO	24
5.1. LA CIRUGIA PLASTICA	24
5.2. MINERIA DE DATOS	26
5.2.1. ANÁLISIS Y TÉCNICAS	27
5.2.2. EVALUACIÓN DE LOS MODELOS DE MINERÍA DE DATOS	29
5.2.3. METODOLOGÍAS DE MINERÍA DE DATOS	31
5.2.4. PLATAFORMAS DE MINERÍA DE DATOS	31
6. MARCO LEGAL	33
7. ESTADO DEL ARTE	34
7.1. GENERALIDADES SOBRE ESTUDIOS DE CIRUGÍA PLÁSTICA.	34
7.2. MINERÍA DE DATOS APLICADA AL CAMPO DE LA SALUD.	35
7.3. MINERÍA DE DATOS APLICADA A OTROS CAMPOS.	37
7.4. DISCUSIÓN.	38
PARTE 2. DESARROLLO DE LA METODOLOGÍA CRISP-DM	40
8. DESCRIPCIÓN DEL NEGOCIO	41
9. COMPRENSIÓN DE LOS DATOS	43
9.1. DESCRIPCIÓN DE LOS DATOS	43
9.2. LIMPIEZA DE DATOS	46
9.3. ANÁLISIS ESTADÍSTICO DE LOS DATOS.	51
9.4. TRANSFORMACIONES	72
9.5. SELECCIÓN DE VARIABLES.	72

9.6. ANÁLISIS DE CORRELACIONES.	79
<u>10. MODELADO Y EVALUACIÓN</u>	<u>82</u>
10.1. EXPERIMENTO 1: SELECCIÓN DE FACTORES CON ANÁLISIS DE CORRELACIONES	83
10.2. EXPERIMENTO 2: SELECCIÓN DE FACTORES CON PCA	85
10.3. EXPERIMENTO 3: SELECCIÓN DE FACTORES CON ÁRBOLES DE DECISIÓN	86
10.4. EXPERIMENTO 4: SELECCIÓN DE FACTORES CON REGRESIÓN LOGÍSTICA	90
10.5. EXPERIMENTO 5: SELECCIÓN DE FACTORES CON REGLAS DE ASOCIACIÓN	94
10.6. EXPERIMENTO 6: PERFILAMIENTO DE LOS TIPOS DE PACIENTES CON CLUSTERING	98
<u>11. DESPLIEGUE</u>	<u>102</u>
11.1. FACTORES MÁS INFLUYENTES EN LA REALIZACIÓN DE UNA CIRUGÍA ESTÉTICA	102
11.2. SUGERENCIAS PARA EL CIRUJANO PLÁSTICO JOHN JAIRO VALENCIA QUINTERO	105
PARTE 3. CONCLUSIONES Y TRABAJO FUTURO	107
<u>12. CONCLUSIONES</u>	<u>108</u>
<u>13. TRABAJOS FUTUROS</u>	<u>110</u>
REFERENCIAS	112

LISTA DE ILUSTRACIONES

<i>Ilustración 1. Metodología CRISP_DM (IBM, 2012)</i>	21
<i>Ilustración 2. Áreas de aplicación de Minería de Datos (Piatetsky, 2014)</i>	27
<i>Ilustración 3. Herramientas de Minería de Datos (Piatetsky, 2015)</i>	32
<i>Ilustración 4. Selección - Campos Duplicidad</i>	47
<i>Ilustración 5. Resultados - Duplicidad</i>	47
<i>Ilustración 6. Preparación - Sexo</i>	48
<i>Ilustración 7. Datos Preparados - Sexo</i>	48
<i>Ilustración 8. Preparación - Estado Civil</i>	48
<i>Ilustración 9. Datos Preparados - Estado Civil</i>	48
<i>Ilustración 10. Preparación - Ocupación</i>	49
<i>Ilustración 11. Datos Preparados - Ocupación</i>	49
<i>Ilustración 12. Preparación - Cirugía Deseada</i>	49
<i>Ilustración 13. Datos Preparados - Cirugía Deseada</i>	49
<i>Ilustración 14. Preparación - Planificación</i>	50
<i>Ilustración 15. Datos Preparados - Planificación</i>	50
<i>Ilustración 16. Análisis Estadístico - Sexo</i>	51
<i>Ilustración 17. Análisis Estadístico - Edad</i>	52
<i>Ilustración 18. Análisis Estadístico - Estado Civil</i>	52
<i>Ilustración 19. Análisis Estadístico - Ocupación</i>	52
<i>Ilustración 20. Análisis Estadístico - Cirugía Deseada 1</i>	53
<i>Ilustración 21. Análisis Estadístico - Cirugía Deseada 2</i>	53
<i>Ilustración 22. Análisis Estadístico - AP_TIROIDES</i>	54
<i>Ilustración 23. Análisis Estadístico - AP_CORAZON</i>	54
<i>Ilustración 24. Análisis Estadístico - AP_SIDA</i>	55
<i>Ilustración 25. Análisis Estadístico - AP_RINONES</i>	55
<i>Ilustración 26. Análisis Estadístico - AP_VESICULA</i>	55
<i>Ilustración 27. Análisis Estadístico - AP_ARTRITIS</i>	56
<i>Ilustración 28. Análisis Estadístico - AP_PA</i>	56
<i>Ilustración 29. Análisis Estadístico - AP_SANGRADO</i>	56
<i>Ilustración 30. Análisis Estadístico - AP_PULMONES</i>	57
<i>Ilustración 31. Análisis Estadístico - AP_SISTNERVIOSO</i>	57
<i>Ilustración 32. Análisis Estadístico - AP_ESTOMAGO</i>	58
<i>Ilustración 33. Análisis Estadístico - AP_DIABETES</i>	58
<i>Ilustración 34. Análisis Estadístico - AP_ASMA</i>	58
<i>Ilustración 35. Análisis Estadístico - AP_LUPUS</i>	59
<i>Ilustración 36. Análisis Estadístico - AP_CANCER</i>	59
<i>Ilustración 37. Análisis Estadístico - AP_FIBROMIALGIA</i>	59
<i>Ilustración 38. Análisis Estadístico - AP_HIGADO</i>	60
<i>Ilustración 39. Análisis Estadístico - AP_ESCLERO</i>	60
<i>Ilustración 40. Análisis Estadístico - AP_OTRAS</i>	60

<i>Ilustración 41. Análisis Estadístico - MED_ASPIRINA.</i>	61
<i>Ilustración 42. Análisis Estadístico - MED_MOTRIN.</i>	61
<i>Ilustración 43. Análisis Estadístico - MED_IBUPROFENO.</i>	61
<i>Ilustración 44. Análisis Estadístico - MED_PARA_ARTRITIS.</i>	62
<i>Ilustración 45. Análisis Estadístico - MED_PARA_DORMIR.</i>	62
<i>Ilustración 46. Análisis Estadístico - MED_PARA_TOS.</i>	62
<i>Ilustración 47. Análisis Estadístico - MED_PARA_TIROIDES.</i>	63
<i>Ilustración 48. Análisis Estadístico - MED_HORMONAS.</i>	63
<i>Ilustración 49. Análisis Estadístico - MED_ANTICONCEPT.</i>	63
<i>Ilustración 50. Análisis Estadístico - MED_TRETINOINA.</i>	64
<i>Ilustración 51. Análisis Estadístico - MED_ANTIBIOTICOS.</i>	64
<i>Ilustración 52. Análisis Estadístico - MED_PARA_PRESION.</i>	64
<i>Ilustración 53. Análisis Estadístico - MED_PARA_DIABETES.</i>	65
<i>Ilustración 54. Análisis Estadístico - MED_DIURETICOS.</i>	65
<i>Ilustración 55. Análisis Estadístico - MED_FENOBARBITAL.</i>	65
<i>Ilustración 56. Análisis Estadístico - MED_DIGITALICOS.</i>	66
<i>Ilustración 57. Análisis Estadístico - MED_CORTISONA.</i>	66
<i>Ilustración 58. Análisis Estadístico - MED_PARA_TOS.</i>	66
<i>Ilustración 59. Análisis Estadístico - planificación.</i>	67
<i>Ilustración 60. Análisis Estadístico - AF_HA.</i>	67
<i>Ilustración 61. Análisis Estadístico - AF_EC.</i>	67
<i>Ilustración 62. Análisis Estadístico - AF_ART.</i>	68
<i>Ilustración 63. Análisis Estadístico - AF_COAG.</i>	68
<i>Ilustración 64. Análisis Estadístico - AF_ASM.</i>	68
<i>Ilustración 65. Análisis Estadístico - AF_DIAB.</i>	69
<i>Ilustración 66. Análisis Estadístico - AF_CAN.</i>	69
<i>Ilustración 67. Análisis Estadístico - AF_OTR.</i>	69
<i>Ilustración 68. Análisis Estadístico - HAB_CIGARRILLO.</i>	70
<i>Ilustración 69. Análisis Estadístico - HAB_LICOR.</i>	70
<i>Ilustración 70. Análisis Estadístico - HAB_CAFE.</i>	70
<i>Ilustración 71. Análisis Estadístico - HAB_PASTIL_DORMIR.</i>	71
<i>Ilustración 72. Análisis Estadístico - HAB_OTROS.</i>	71
<i>Ilustración 73. Análisis Estadístico - peso.</i>	71
<i>Ilustración 74. Análisis Estadístico - talla.</i>	72
<i>Ilustración 75. Análisis Estadístico - IMC.</i>	72
<i>Ilustración 77: Resultado del Experimento bajo la técnica Árbol de Decisiones</i>	88
<i>Ilustración 78. Cantidad de Datos por Clúster.</i>	99
<i>Ilustración 79. Porcentaje de Datos por Clúster.</i>	99

LISTA DE TABLAS

<i>Tabla 1. Descripción de Variables</i>	46
<i>Tabla 2. Limpieza de Datos.</i>	50
<i>Tabla 3. Descripción Estadística de los Datos.</i>	72
<i>Tabla 4. Análisis de Variables.</i>	78
<i>Tabla 5. Variables con correlaciones altas (mayor a 0.7).</i>	79
<i>Tabla 6. Variables con correlación baja (menor a 0.3).</i>	81
<i>Tabla 7. Variables para el modelamiento.</i>	83
<i>Tabla 8. Evaluación de Resultados - Experimento 1.</i>	84
<i>Tabla 9. Evaluación de Resultados - Experimento 2.</i>	85
<i>Tabla 10. Evaluación de Resultados - Experimento 3.</i>	87
<i>Tabla 11. Clasificación de las Ramas arrojadas - Experimento 3.</i>	89
<i>Tabla 12. Evaluación de Resultados - Experimento 4.</i>	93
<i>Tabla 13. Condiciones de las Variables - Experimento 4.</i>	94
<i>Tabla 14. Evaluación de Resultados - Experimento 5.</i>	96
<i>Tabla 15. Condiciones Claves - Experimento 5.</i>	97
<i>Tabla 16. Evaluación de Resultados - Experimento 6.</i>	101
<i>Tabla 17. Resultado Final.</i>	103

INTRODUCCIÓN

La extracción de información mediante minería de datos permite conocer detalles sobre el comportamiento de las variables que no se observan a simple vista y finalmente, lograr tomar decisiones en pro de los objetivos institucionales. En el área de la salud, la minería de datos ha contribuido a la obtención de resultados predictivos sobre el padecimiento de enfermedades como el cáncer, la leucemia e incluso, ha sido posible el reconocimiento de factores que influyen en el futuro diagnóstico de las mismas.

El problema que se presenta en éste caso, de manera específica, trata sobre la priorización en citas de valoración de los nuevos pacientes de cirugía plástica, es decir, existen épocas donde la agenda de los cirujanos es bastante ocupada, entre ellas, épocas vacacionales y festividades, las cuales son de preferencia por los pacientes nuevos ya que cuentan con la disponibilidad de tiempo requerida para la recuperación en caso de cirugía y para pacientes antiguos que asisten con la finalidad de revisar el avance de sus cirugías anteriores y manifestar inquietudes. existen épocas donde la agenda de los cirujanos es bastante ocupada, entre ellas, épocas vacacionales y festividades, las cuales son de preferencia por los pacientes nuevos ya que cuentan con la disponibilidad de tiempo requerida para la recuperación en caso de cirugía y para pacientes antiguos que asisten con la finalidad de revisar el avance de sus cirugías anteriores y manifestar inquietudes. Es entonces donde los pacientes nuevos requieren tiempos de espera demasiado extensos para obtener una valoración médica. Siendo la cita de valoración un requisito para la realización de una cirugía estética, se hace necesario identificar o establecer algunos criterios mínimos para la asignación prioritaria de las mismas.

La intencionalidad es aplicar minería de datos a historias clínicas de pacientes que han asistido a cita de valoración con un médico especialista en cirugía plástica y así, determinar las variables que influyen respecto a la decisión de realizarse un procedimiento quirúrgico. La solución permitirá entonces, al identificar las variables

que influyen en la realización de una cirugía estética, asignar citas, en los tiempos donde la agenda del cirujano se encuentra congestionada, de manera prioritaria; esto, según el resultado del modelado, con los pacientes que cumplen con las características identificadas.

El desarrollo del presente trabajo consta de tres partes: la formulación del proyecto y la revisión literaria donde se realizaron los estudios de entorno e identificación del problema y se localizaron estudios ya desarrollados mediante minería de datos en el área de la salud y otras áreas. En la segunda parte se desarrolla la metodología de minería de datos seleccionada para el estudio CRISP – DM. Por último, se presentan las conclusiones y los posibles trabajos futuros.

PARTE 1. FORMULACIÓN DEL PROYECTO Y REVISIÓN LITERARIA

1. DESCRIPCIÓN DEL PROBLEMA

En Colombia, las cirugías estéticas han venido en aumento en los últimos años (ISAPS International Survey on Aesthetic/Cosmetic, 2014), esto ha generado que los cirujanos plásticos aumenten el número de citas en sus consultorios con pacientes deseosos de mejorar algún aspecto “poco estético” de su cuerpo o con el cual no se sienten a gusto.

Es importante aclarar que en Colombia las cirugías estéticas no son cubiertas por las empresas prestadoras de salud ni por el sistema de salud público, a excepción de aquellas cirugías que hacen parte de un proceso médico donde se requieren procedimientos quirúrgicos encaminados a la recuperación de la función y/o la forma de secuelas provocadas por el sufrimiento de accidentes laborales, de tránsito, quemaduras o procesos pos-oncológicos y pos-bariátricos (Ministerio de La Salud, 1994). Ésta es entonces la razón por la cual la mayoría de especialistas en cirugía plástica, laboran de manera privada.

Para esto, muchos cirujanos plásticos, incluido el doctor John Jairo Valencia Quintero, son miembros de número de la Sociedad Colombiana de Cirugía Plástica, Estética y Reconstructiva; dicha sociedad es una comunidad médico científica, exclusivamente gremial y, aunque los especialistas no tienen la obligación de pertenecer a ella, es importante reconocer que ser socio genera un mayor respaldo para los pacientes ya que pueden corroborar el título académico de su cirujano, su historial académico, su hoja de vida laboral y los procedimientos para los cuales está capacitado.

Además, con el tiempo y la experiencia adquiridas por los especialistas, el número de pacientes se ha venido incrementando; adicionalmente, la cantidad de pacientes que se mantienen activamente asistiendo a consulta es amplio, debido al número de citas que implica todo el proceso de recuperación y al acompañamiento del especialista hacia el paciente (Arbelaéz Rodríguez, 2013).

La mayor captación de pacientes no se realiza por medios electrónicos (correo, redes sociales o página web), sino, se comunican de manera telefónica y como referidos por pacientes antiguos para solicitar ser atendidos. La cita de evaluación (primera cita) genera un consumo significativo de recursos por parte del especialista por dos motivos: 1) La cita de valoración resulta ser gratuita, 2) Ésta primera cita es la que mayor tiempo consume. Cuando se habla de consumo significativos de recursos se refiere al tiempo invertido del cirujano y éste representa dinero. La duración aproximada de una cita de valoración, según los tiempos registrados de forma manual en la agenda del cirujano corresponden en promedio a una hora y dicha duración no depende del tipo de cirugía que el paciente desea realizarse, sino, del objetivo expreso del médico, quien manifiesta que busca brindar a sus pacientes la información suficiente sobre los procedimientos además de resolver dudas e inquietudes, pues su lema es que: "Un paciente bien informado es un paciente seguro". Es importante también contar que en ésta cita además, se realiza la toma de fotografías pre operatorias y se exponen al paciente casos exitosos .

En el consultorio del cirujano estético John Jairo Valencia Quintero se presenta que, un paciente obtiene una cita de valoración en tiempos de espera inclusive de tres y cuatro semanas debido a la ocupación en la agenda del especialista y, luego de que dicha cita se lleva a cabo, el 60% de los pacientes toman la decisión de no realizarse el procedimiento estético; dicho porcentaje puede concluirse según las historias clínicas del cirujano plástico, además, cuando la secretaria contacta a los pacientes, éstos manifiestan que decidieron no realizarse la cirugía debido a algunos de éstos motivos: falta de recursos económicos, preferencia hacia otro cirujano plástico, falta de tiempo para esperar la disponibilidad del cirujano o falta de condiciones de salud óptimas para la cirugía.

Para superar los problemas mencionados, se requiere la definición de prioridades en la asignación de citas dependiendo de las características individuales de cada paciente. Y para aportar una solución a lo expuesto, se pretende realizar un estudio mediante minería de datos para conocer las variables que influyen en la

decisión del paciente para realizarse una cirugía estética y así, establecer parámetros de prioridades para la atención de los pacientes que desean asistir a citas de valoración.

2. JUSTIFICACIÓN

Actualmente, se está evidenciando la deserción de pacientes por la demora en asignación de citas de valoración, esto, muchas veces debido al aumento de las cirugías estéticas (Elliott, 2008) ; el tiempo que implica la espera ha generado en repetidas ocasiones que los pacientes pierdan el interés en ser evaluados, conllevando a la disminución en la calidad de la atención a los pacientes, quienes son los que, precisamente, reclaman atención, comunicación y disponibilidad (Pezoa, 2013).

La priorización de citas que se efectuará luego de conocer los factores que influyen en la realización de cirugías estéticas mediante minería de datos logrará:

- Aumentar los ingresos económicos proyectados por el especialista debido a la satisfacción de los pacientes al tener la posibilidad de conocer el criterio del cirujano.
- Brindar a los pacientes más propensos a realizarse la cirugía, la oportunidad de contar con la disposición por parte del especialista con la finalidad de establecer una relación de disponibilidad, confianza y seguridad, teniendo en cuenta siempre que éticamente lo más importante en la medicina es mantener una disposición al servicio y cuidado de los pacientes (Arriagadas & Ortiz, 2010).
- Conocer los patrones (Valenga, y otros, 2007) de los pacientes que están propensos, según la información contenida en su historia clínica, a operarse y así, establecer un orden de prioridad a la hora de asignar citas de valoración, procurando atenderles de manera prematura.
- Concentrar la agenda del especialista en las citas de seguimiento de sus pacientes y las citas de valoración en aquellos pacientes que tienen mayor probabilidad de continuar con el procedimiento estético y así lograr que el tiempo del especialista sea asignado lo más eficaz posible.

El objetivo del análisis mediante minería de datos está enfocado entonces en el desarrollo de un modelo que permita reconocer según el perfilamiento de los pacientes, la priorización necesaria para la correcta satisfacción de la demanda en citas de valoración y así disminuir los tiempos de espera para aquellos potenciales pacientes quirúrgicos.

3. OBJETIVOS

Objetivo General

Definir los factores que influyen en la realización de cirugías estéticas mediante técnicas de Minería de Datos, como estrategia para la priorización en la asignación de citas de valoración.

Objetivos Específicos

Los objetivos específicos nombrados a continuación hacen referencia a las etapas del proceso de minería de datos.

- Identificar las variables que hacen parte de la historia clínica en la realización de una cirugía estética.
- Preparar los datos aplicando limpieza y perfilamiento de datos.
- Diseñar un modelo de minería de datos para la gestión de priorización de citas.
- Definir una estrategia para la priorización de asignación de citas de valoración.

4. METODOLOGÍA

Para el proceso de Minería de Datos, se selecciona la metodología CRISP-DM, resaltando que se trata de una metodología flexible y se puede personalizar fácilmente según los objetivos del negocio (IBM, 2012).

Las etapas definidas por la metodología se presentan en la ilustración 1.

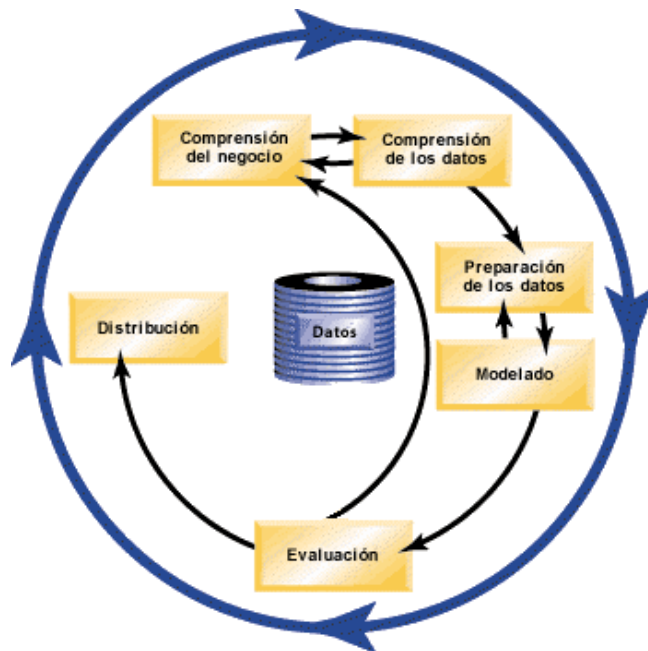


Ilustración 1. Metodología CRISP_DM (IBM, 2012)

✓ COMPRENSIÓN DEL NEGOCIO

En ésta fase se realizarán las entrevistas necesarias con el especialista correspondiente; en éste caso, el Cirujano Plástico, quien compartirá sus objetivos respecto a la minería de datos, los datos con los que cuenta y la explicación de los

mismos. Básicamente, la etapa está enfocada de manera cualitativa, donde sólo se tendrá la recolección de la información a través de entrevista.

✓ **COMPRESIÓN DE LOS DATOS**

En ésta etapa se revisarán entonces las bases de datos entregados por el especialista y se procederá al estudio de los mismos de manera detenida, la finalidad es conocer los valores posibles que pueden tomar las variables y los datos que se obtienen mediante la historia clínica. Es caracterizada de manera cuantitativa por el número de datos y variables.

✓ **PREPARACIÓN DE LOS DATOS**

Luego, se procede entonces a seleccionar los datos con los cuales se realizará el proceso de minería; es fundamental también realizar toda la tarea de limpieza, integración y formateo de datos; al igual que la fase anterior, es una etapa cuantitativa ya con un número de variables reducidas y datos, según la limpieza lo requiera.

✓ **MODELADO DE DATOS**

El modelamiento consiste en primera instancia en seleccionar la técnica de modelado, diseñar la evaluación, construir y evaluar el modelo; se considera una etapa cuantitativa por la cantidad de datos a balancear, la selección de las variables y los resultados de probabilidades arrojados por cada modelo.

✓ **EVALUACION**

Se decidirá entonces, según la evaluación de qué modelo es más útil para las necesidades del especialista y se hará una revisión del proceso para lograr esto. Ésta fase está caracterizada por ser cualitativa ya que se expondrá de manera asertiva la justificación sobre la técnica seleccionada.

✓ DESPLIEGUE

Se realiza la debida toma de decisiones según el resultado del proceso de minería y se establecen criterios que lleven a una toma de decisiones acertadas. Consiste entonces en la elaboración de un informe cualitativo donde se justifican los resultados encontrados y se realizan sugerencias al especialista para la toma de decisiones.

Cabe aclarar que el alcance está delimitado al proceso de minería de datos y la entrega de resultados de la misma, pero, es el especialista quien debe tomar las decisiones pertinentes, según los resultados del modelo, para la mejora de sus procesos.

5. MARCO TEÓRICO

Se describen a continuación cada una de las áreas que conforman el análisis del marco referencia: Cirugía Plástica y Minería de Datos.

5.1. LA CIRUGIA PLASTICA

La cirugía plástica es una de las especialidades quirúrgicas de la medicina y es la encargada de la reparación y reconstrucción de defectos y malformaciones congénitas además de intervenir algunos aspectos físicos causados por la conciencia de no belleza según el contexto sociocultural del individuo (Coll, 1989), añadido a eso, como consecuencia del desarrollo de la cirugía plástica, se da aparición a la cirugía estética que es la encargada de realizar intervenciones quirúrgicas para mejorar el aspecto físico que no corresponde a los conceptos de belleza dentro de un concepto sociocultural específico (Arias, González, García, & González, 2008).

Dentro de los procedimientos estéticos, las cirugías que más frecuentemente se realizan son:

- Mamoplastia: Cirugía de la mama, la cual puede ser de disminución o de aumento mediante implante de prótesis o autoinjerto de grasa.
- Mastopexia: Es la cirugía mediante la cual se realiza el levantamiento de la mama.
- Abdominoplastia: Cirugía de abdomen, frecuentemente realizada en mujeres luego del parto.
- Liposucción: Es el procedimiento en el cual se extrae grasa mediante cánulas y consiste en dar forma estética a varias zonas del cuerpo.
- Otoplastia: Es la reducción o cambio en la forma de las orejas.

- Rinoplastia: Cirugía estética y funcional de la nariz

Estos procedimientos estéticos, sólo deben ser realizados por un especialista en cirugía plástica, quien, luego de realizar sus estudios en medicina general, opta a su especialidad, la cual tiene una duración por lo menos de cuatro años.

Identificando el profesional que debe llevar a cabo entonces dichos procedimientos y antes de que éste sea llevado a cabo, se requiere la elaboración de la historia clínica; la cual consta de: La anamnesis, que es el proceso mediante el cual el profesional de la salud obtiene información del paciente respecto a su identificación, estado civil, ocupación, lugar de nacimiento, motivo de la consulta, los síntomas y en general toda aquella información que contribuya al esclarecimiento de un diagnóstico y, del examen físico que permite mediante la observación, palpación, percusión y auscultación recoger mayor información tendiente a reafirmar el diagnóstico del paciente. Éste proceso se complementa con las ayudas diagnósticas, como son los exámenes de sangre, las imágenes diagnósticas, entre otros (Ministerio de La Protección Social, 2007).

En el quehacer del especialista en cirugía plástica, se requieren además, una serie de pasos para llevar a cabo un procedimiento estético (Bustamante, Chui, Coca, & Flóres, 2011). En primera instancia y como en todo procedimiento médico, es necesaria la elaboración de la historia clínica, ésta contiene los datos personales (anamnesis) y el examen físico que realiza el especialista en cada consulta al paciente; la historia clínica es el documento al cual se recurre en el proceso médico para indicar o contraindicar alguna procedimiento, medicamento o intervención futura (Bustamante, Chui, Coca, & Flóres, 2011).

A pesar de que la elaboración de la historia clínica se construye desde la primera cita a la que el paciente asiste con el especialista, ésta se alimenta en cada una de las consultas posteriores. Para esto, están definidas tres tipos de citas:

- **Cita de valoración:** La cual es asignada de manera telefónica; éste es el primer momento en que el paciente es atendido por el cirujano, manifiesta

sus inconformidades (físicas) y el cirujano le ofrece desde su perspectiva profesional, posibles procedimientos quirúrgicos a realizar. Es importante aclarar que ésta cita es la que requiere mayor tiempo de consulta y en la cual es especialista define si puede suplir las necesidades o expectativas del paciente (Manrique, 2005).

- **Cita Pre Operatoria:** Es una cita previamente acordada, se lleva a cabo antes de la cirugía, con el fin de resolver dudas o inquietudes por parte del paciente y re afirmar las expectativas del mismo sobre el procedimiento a realizar.
- **Citas de Post Operatorias:** Son aquellas que se realizan después de la cirugía y durante los según el paciente lo requiera y el cirujano lo considere pertinente.

Las citas pre operatorias y post operatorias dependerán directamente de los pacientes que están a punto de realizarse un procedimiento estético o de aquellos que ya se lo realizaron y requieren revisiones por parte del especialista, ya que son los pacientes que, según él, tienen prioridad sobre la agenda. Sin embargo, las citas de valoración dependen de la disponibilidad de la agenda del médico una vez se han asignado las citas pre y post operatorias.

Es por ésta dependencia entonces que, el proceso de minería de datos, estará enfocado en la definición de una estrategia para la priorización en la asignación de citas de valorización mediante el hallazgo del comportamiento de los datos anamnéscos y de examen físico contenidos en la historia clínica y que influyen en la decisión de realizarse un procedimiento estético (Lizasoain & Joaristi, 2012).

5.2. MINERIA DE DATOS

Es el tratamiento de los datos con la finalidad de identificar patrones, tendencias o desviaciones que permiten analizar desde un concepto más amplio, el comportamiento del negocio y finalmente, tomar decisiones asertivas (Pérez &

Santin, 2007). Este proceso incluye la aplicación de algoritmos que permiten dicha identificación a grandes cantidades volúmenes de datos.

La minería de datos ha sido usada en muchas áreas; entre ellas, el análisis de relacionamiento con los clientes, la banca, la educación, la salud, la detección de fraudes y la ciencia.

A continuación, en la **ilustración 2**, el ranking entregado por el periódico KDnuggets, sobre la industria donde ha sido aplicada la minería de datos:

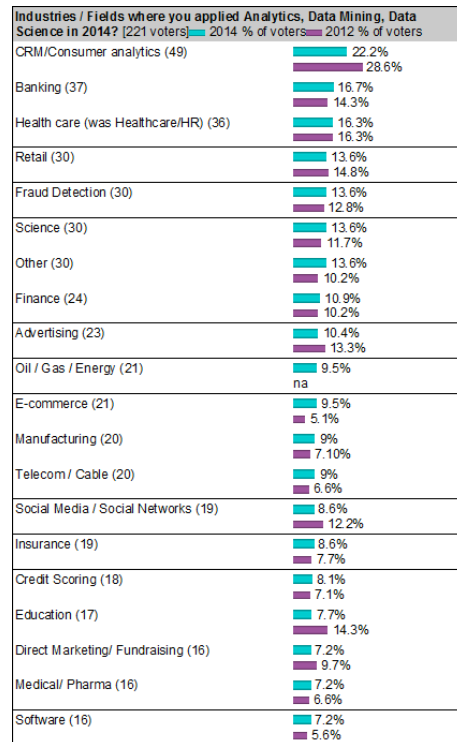


Ilustración 2. Áreas de aplicación de Minería de Datos (Piatetsky, 2014)

5.2.1. Análisis y Técnicas

Dentro de la minería se pueden realizar dos tipos de análisis: predictivos y descriptivos.

Análisis predictivo: Como su nombre lo indica, permite realizar la predicción sobre futuros comportamientos mediante técnicas supervisadas, los análisis predictivos son:

- **Clasificación:** Es el estudio categórico de una variable objetivo previamente definida, y luego permite predecir la categoría de nuevos elementos.
- **Predicción continua:** Es el estudio de los datos con el objetivo de predecir eventos numéricos futuros.

Algunos de los algoritmos más usados en las técnicas supervisadas son (Pérez & Santin, 2007):

- **Árboles de Decisión:** Éstos algoritmos son representados en forma de árbol y permiten mediante un nivel jerárquico evaluar las condiciones a tener en cuenta para la toma de una decisión.
- **Métodos de Regresión:** Identifica las relaciones que existen entre registros de datos y luego de dicha identificación crea patrones de predicción.
- **Redes Neuronales:** Son algoritmos de autoaprendizaje mediante entrenamiento con conjuntos de datos imitando el funcionamiento cerebral de los seres vivos.
- **Máquinas de Soporte Vectorial:** Realiza un aprendizaje creando un hiperplano de separación óptima de las categorías que se desean predecir.
- **Métodos basados en Ejemplos:** Entre ellos, está el algoritmo de K-VECINOS, que consiste en hallar el vecino más cercano y agrupa los datos según sus atributos y la cercanía con otros registros de datos.

Análisis descriptivo: Brinda la posibilidad de realizar segmentaciones, definir reglas de negocio o establecer perfiles específicos mediante técnicas no supervisadas, los análisis descriptivos son:

- **Agrupamiento/clustering:** Organiza grandes cantidades de información en grupos de datos con características comunes.

- **Asociación:** Genera reglas de asociación luego de estudiar los datos, es decir, la probabilidad de que un dato se comporte de manera específica dado el comportamiento de los otros datos.

Algunos de los algoritmos más usados en las técnicas no supervisadas son (Pérez & Santin, 2007):

- **K-means:** Busca similitudes entre los datos mediante medidas de distancia, de tal manera que datos muy cercanos pertenecen a un mismo clúster.
- **Apriori:** Crea reglas de asociación buscando las co-ocurrencias de registros en el conjunto de datos.

5.2.2. Evaluación de los Modelos de Minería de Datos

Aunque existen técnicas supervisadas y no supervisadas, la evaluación del modelo es indispensable, dicha evaluación depende de la precisión del mismo y de los resultados arrojados; el ideal entonces no será sólo el que brinde la mayor precisión, sino, el que esté en mayor medida, orientado a responder la pregunta planteada en el problema o en el objetivo principal (Pautsch, 2009).

- **Evaluación de análisis predictivos (técnicas supervisadas).**

Lo más usual en las técnicas supervisadas consiste en la división de los datos en dos conjuntos; uno de entrenamiento, requerido para construir y entrenar el modelo y un segundo conjunto, llamado el conjunto de prueba con el cual se evalúa la efectividad del modelo. Según la cantidad de datos con la que se cuenta, existen varias técnicas para efectuar dicha división (Formia, 2012).

La precisión, es una de las técnicas de evaluación de hipótesis más usual para evaluar los clasificadores, pero, algunas otras son:

- Porcentaje de Error de Muestra (o inversamente el porcentaje de Acierto (Accuracy), o el Alcance).
- La Precisión o la Especificidad (Recall, Precision, Specificity) o
- Técnicas más elaboradas como el Área Bajo la Curva (ROC).

Para la aplicación de éstas técnicas puede utilizarse el conjunto total de datos para entrenamiento y para validación o realizar la división antes mencionada. También, la validación cruzada puede aplicarse con la finalidad de evitar sesgo en las particiones de los datos e igualmente, a ésta validación se puede aplicar o no solapamiento en los subconjuntos. Otras técnicas de evaluación incluyen la combinación de hipótesis, tales el como el Bootsting, Bagging o Ramdomization (Dasso & Funes, 2012).

- **Evaluación de análisis descriptivos (técnicas no supervisadas).**

La agrupación o clustering puede ser validada de manera interna o externa. La validación externa de clustering requiere conocer previamente cuál es la partición adecuada de los datos donde los índices suelen tener medidas de similitud luego de aplicar el algoritmo. Para este tipo de validación existen diversos tipos de medidas de similitud, algunos son: recuento de pares, teoría de la información y número de ediciones. La validación interna consiste en estudiar los datos y cómo se agrupan, es decir, evalúa la partición en base a los datos y las distancias entre ellos basados en medidas de compacticidad y separabilidad, para este tipo de validación existe una gran cantidad de índices, la mayoría se basa en la cohesión de los clusters y la separación entre los mismos, algunos son: Calinski-Harabasz, Davies-Bouldin, Dunn, Silhouette, R-Squared, entre otros. Tanto en validación externa como en interna, dada una partición se calculan índices que miden la calidad o bondad de dicha partición. Si el valor devuelto por el índice es inusualmente alto (o bajo), se puede deducir que la partición representa la estructura interna de los datos. El objetivo principal es determinar cuándo un valor es inusualmente alto (o bajo) y cuando no. Lo que se utiliza es la comparación de

índices de distintas particiones. Se utiliza, por tanto, una aproximación comparativa donde no se analiza el valor del índice en términos absolutos sino en términos relativos.

5.2.3. Metodologías de Minería de Datos

Existen varias metodologías para la aplicación de minería de datos; entre ellas, CRISP-DM, SEMMA y KDD. De estas metodologías se resalta CRISP-DM con una alta aplicación en diferentes ámbitos como la banca, finanzas, salud, educación, etc. CRISP-DM (IBM, 2012) es una metodología que incluye los procesos normales de ejecución de proyectos, las actividades y las relaciones para que todas las fases sean llevadas a cabo de manera ordenada en pro del logro de objetivos de proyecto. Consiste entonces, en primera instancia, en entender las necesidades del negocio, posteriormente en el entendimiento de los datos para identificar las variables claves y las inconsistencias; luego, en la preparación de los datos y todo lo que esto conlleva (calidad, selección de variables, entre otras); después, en la aplicación y selección del algoritmo más adecuado evaluando resultados y finalmente, en la retroalimentación al negocio con los hallazgos identificados.

5.2.4. Plataformas de Minería de Datos

Existen varias plataformas para la minería de datos; su uso, dependerá exclusivamente de la tarea que se tenga y los recursos con los que cuente el experto; no todas las herramientas o plataformas de minería de datos contienen todos los algoritmos; además de estar limitadas por el pago de licencias de funcionamiento y por las características de diseño para su uso.

Entre algunas de las plataformas mencionadas en la lista de el periódico KDnuggets (**Ilustración 3**), se encuentra WEKA (*Waikato Environment for Knowledge Analysis*); weka es la herramienta de minería de datos desarrollada por la universidad de Waikato, contiene una gran variedad de algoritmos, su uso es

libre y presenta un ambiente de trabajo amigable, generando una gran usabilidad (Azoumana, 2013).

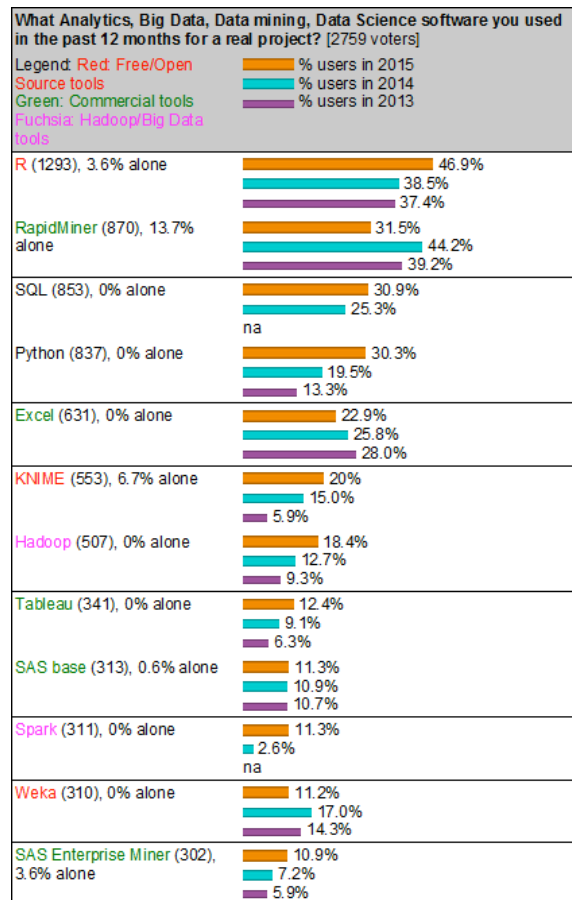


Ilustración 3. Herramientas de Minería de Datos (Piatetsky, 2015)

6. MARCO LEGAL

La protección de datos es un tema indispensable a tener en cuenta en trabajos de minería porque están involucradas grandes cantidades de información que requieren de un trato cuidadoso y respetuoso con los dueños de la misma.

Los modelos de negocio que involucran en su quehacer la relación directa con otras personas deben estar orientados al respeto de la privacidad de las personas (Aponte, 2007). En el contexto Colombiano se requiere tener en cuenta el respeto por la siguiente normatividad:

- Ley 23 de 1981 (Congreso De Colombia, 1981) donde se establecen de manera clara y específica los aspectos comportamentales y actitudinales del médico en la relación con el paciente y la conducta decorosa, respetuosa y digna, necesarias para hacer del proceso médico, un acto humanista.
- En la Resolución 1995 de 1999 (Ministerio de La Salud, 1999) se encuentra la normatividad necesaria para el manejo de las historias clínicas donde se especifican los parámetros necesarios a tener en cuenta desde la apertura de la misma, las agregaciones médicas, la custodia y los aspectos legales que conlleva la tenencia de los datos que allí se encuentran. Por otra parte, la Resolución 2346 de 2007 (Ministerio de La Protección Social, 2007) donde se establece la obligatoriedad de los campos que debe contener la historia clínica, los tipos de valoraciones médicas y el personal médico habilitado para acceder a ella.

Por último, es necesario siempre respetar la Ley sobre protección de datos (Congreso de Colombia, 2012) donde se definen los parámetros por los cuales una persona u organización puede acceder a la información, además, de los requerimientos a cumplir cuando se consulta, modifica o se extrae en un contexto específico.

7. ESTADO DEL ARTE

La revisión literaria se ha podido categorizar en tres grupos o secciones; entre ellos, algunas generalidades sobre la cirugía plástica, trabajos de minería de datos aplicada al campo de la salud y minería de datos aplicada en otros campos. A continuación, se hace referencia a ellos.

7.1. GENERALIDADES SOBRE ESTUDIOS DE CIRUGÍA PLÁSTICA.

A continuación, se mencionan algunos estudios que, aunque no están enmarcados directamente dentro de modelos de minería de datos, han logrado establecer patrones o reglas dentro de la realización de cirugías estéticas.

El estudio realizado por (Abreu de la Torre, Tumer, González, Macias Betancourt, & Leal, 2000), se identifica una regla de comportamiento de variables, concluyendo que las mujeres con edades entre los 56 y 65 años, casadas, que viven en zona urbana y requieren de su que hacer profesional una apariencia agradable, eran las personas que predominaban en su media poblacional estudiada; además de predominar un trastorno de personalidad histérico.

En la investigación realizada por (Olivares, 2007), se logra resaltar la importancia a nivel psicológico que tiene la reconstrucción mamaria en aquellas mujeres que, por causa de cáncer, han sufrido amputación mamaria, se realiza un estudio social y se sugieren la revisión no sólo de casos post operatorios sino, pre operatorios; cabe resaltar que éste artículo logra establecer una relación directa entre el comportamiento psicológico de las pacientes y la realización de una cirugía estética.

En la investigación realizada por (Palma & Solís, 2009) , se describen de manera específica, las repercusiones que a nivel psicológico genera la realización de una

cirugía estética, el nivel de bienestar o de conformidad consigo mismo luego del procedimiento y la sensación de insatisfacción con su aspecto anterior; dicha asociación es importante para aclarar que es posible identificar la relación que entre los aspectos psicológicos de un paciente y la decisión de realizarse un procedimiento estético.

En el estudio descriptivo realizado por (Bermúdez, 2010), se encontró que el exhibicionismo, los intentos de suicidio y los deseos inconscientes referentes a la feminidad son características propias de la mayoría de mujeres con múltiples cirugías estéticas y logra, luego de identificar éstas características, definir a éste tipo de pacientes dentro de un grupo de conductas específicas que deben ser tratadas.

En el trabajo de grado presentado por (Moreno & Zambrano, 2013), se encontró que no existe una relación directa entre autoestima y distorsión de la imagen corporal con la realización de cirugías estéticas, de igual forma, se sugiere a los cirujanos plásticos trabajar con equipos multidisciplinarios que puedan orientar a sus pacientes desde el punto de vista psicológico y no limitarse a mejorar exclusivamente, inconformidades físicas.

En el artículo como trabajo de grado producido por (Calderón, 2014), se identificó en un grupo de cinco mujeres, una relación directa entre la baja autoestima, el entorno social y la realización de cirugías estéticas y se sugiere, realizar la inclusión de aspectos psicológicos en la valoración para la realización de cirugías estéticas.

Finalmente, en el artículo escrito por (Rochefort, Berner, Castillo, & Vidal, 2014), se sugiere estudiar la posibilidad de excluir a un paciente de una cirugía estética según factores psiquiátricos; se puede entonces concluir la importancia que tiene el conocimiento de la relación que existe entre los factores psicológicos y psíquicos y la realización de cirugías estéticas.

7.2. MINERÍA DE DATOS APLICADA AL CAMPO DE LA SALUD.

La minería de datos ha abarcado diferentes tipos de negocios permitiendo la mejora de los procesos, esto, porque brinda el conocimiento sobre el valor real de los datos y revela conductas que no son visible a primera vista. En el campo de la medicina, sin embargo, por la carencia de sistemas de información centralizados, el tema de minería de datos aunque se ha venido trabajando de una manera más lenta, no se ha detenido debido a la necesidad imperante de prolongar y mejorar la calidad de vida. A continuación, se presentan algunos casos.

En el trabajo realizado por (Reparaz, y otros, 2008) se ejecutaron algoritmos de árboles de decisión como ID3 y J48 con la intencionalidad de caracterizar grupos o generar reglas de asociación y predecir la eficacia de la braquiterapia en el tratamiento de cáncer de próstata; se realizó la lectura de las reglas de asociación según las variables implicadas y se sugirió ejecutar más adelante algoritmos de red de bayes para contrastar o evaluar los resultados anteriormente obtenidos.

En el artículo desarrollado por (Timarán & Yépez, 2012), se concluyó que usando técnicas de clasificación basada en árboles de decisión; es posible definir patrones de supervivencia en mujeres con cáncer de cuello uterino. Se usaron algunos algoritmos de asociación para conocer la relación entre los factores socio económicos y clínicos asociados a éste grupo de población. Se hallaron reglas de clasificación obtenidas mediante WEKA con el Algoritmo J48.

En el estudio elaborado por (Rodríguez & Vallejo, 2012), se reconoce la importancia de la minería de datos y su aplicación para la detección y el diagnóstico de pacientes con cáncer, enfocándose específicamente en el cáncer de mama, el uso de la mamografía y su interpretación. Estadísticamente se recurrió al método PCA y PLS y en cuanto a minería de datos, se realizó la selección de atributos, se implementaron algoritmos como árboles de decisión y reglas de asociación para encontrar relaciones no identificadas a primera vista.

En el trabajo realizado por (Acosta, Rosete, & Rodríguez, 2012) , se utilizó metodología CRISP-DM 1.0 y se aplicó minería de datos a encuestas realizadas a

pacientes diabéticos o con riesgo de desarrollarla. Se logró identificar y describir las características de los datos a emplear para clasificar pacientes de una localidad específica en Cuba y se identificaron atributos relevantes a la investigación. Se identificaron problemas con la calidad de los datos y se tomaron medidas para mejorarlos. Las herramientas de software estadísticos fueron Microsoft Excel y Weka. Finalmente, se detalla cómo la medicina es uno de los campos donde más ha crecido el uso de la minería de datos.

En la revisión realizada por (Jiménez & Mateos, 2013) se realizó el modelado mediante red de enlaces de neuronas, específicamente modelo de redes de kohonen con datos particionados y un mapa de salida 3x2 neuronas que genera tres perfiles por grupo para encontrar las características comunes de pacientes diabéticos; finalmente el modelo demostró ser válido tanto en casos típicos como atípicos.

En la tesis escrita por (Franco, 2014) se realizó un estudio de minería de datos mediante algoritmos de agrupación; el algoritmo jerárquico (CHAMELEON) y el algoritmo particional k-Means para conocer la comorbilidad asociada a la diabetes mellitus tipo 2 (DM2) en Colombia; el modelo estaba enfocado en dos partes; primero identificar el diagnóstico de comorbilidad y luego el tipo de diagnóstico. Fueron usadas finalmente redes bayesianas y árboles de decisión con un desempeño bueno y facilitaron de manera gráfica la identificación de las características de los pacientes que presentaban éste diagnóstico.

7.3. MINERÍA DE DATOS APLICADA A OTROS CAMPOS.

Otros casos aplicados a la minería de datos y que pueden resultar útiles se presentan a continuación.

En el estudio realizado por (Valenga, y otros, 2007), se encontró que, por medio de una herramienta de distribución libre, los autores describen un proyecto de minería

de datos analizando en el mundo de la información criminal los homicidios dolosos cometidos en la República Argentina. Aplicada a la inteligencia criminal; la minería de datos es un área bastante nueva y ha tenido un enorme desarrollo; especialmente en los Estados Unidos. Algunas técnicas utilizadas fueron la agrupación de datos o clustering y la Clasificación. El software usado fue WEKA 3.5.3 y los algoritmos aplicados fueron K-means y posteriormente el algoritmo de inducción, J48.

En el artículo escrito por (Sposito, Echeverry, Ryckeboer, & Bossero, 2010), se presenta el proceso de minería de datos mediante las herramientas de software MS SQL Server para el almacenamiento de los datos, SPSS para el procesamiento y Weka para encontrar el algoritmo más asertivo para la clasificación de las características que influyen en el rendimiento académico y en la deserción de los estudiantes de la Universidad Nacional de la Matanza; para esto, se ejecutaron los algoritmos FT y J48, siendo más efectivo el FT con un ROC de 77,86%; se concluye entonces que fue posible determinar dichas características y obtener resultados válidos.

Finalmente, y según el estudio realizado a los documentos anteriormente citados, se concluye que es posible identificar mediante técnicas de minería de datos, la caracterización conductual de variables y la elaboración de clustering y por ende, realizar segmentaciones de mercado identificando las variables que influyen en la realización de cirugías estéticas.

7.4. DISCUSIÓN.

En el campo general de la salud se encontró que, lo más usual, es el uso de algoritmos de árboles de decisión y las reglas de asociación para la identificación de patrones que influyen en el posible padecimiento de enfermedades (Timarán & Yépez, 2012).

Se evidenció también el uso de las metodologías KDD y CRISP-DM para llevar a cabo el proceso de minería de datos; especialmente en el área de la salud se resalta CRISP-DM. Además, Weka es una de las herramientas de mayor uso debido a que es una herramienta libre y cuenta con la mayoría de algoritmos usados en los procesos de minería de datos, también, se identificó que su uso fue efectivo para encontrar patrones de comportamiento; entre ellos, los patrones de supervivencia en mujeres con cáncer de cuello uterino.

Los mayores hallazgos de antecedentes en el contexto de cirugías estéticas están enfocados en el estudio de factores psicológicos; abarcando desde el perfil de los pacientes que se realizan cirugías estéticas hasta las implicaciones que conllevan las mismas; son investigaciones de tipo estadístico realizadas por psicólogos, lo que explica la carencia en la aplicación de algoritmos de minería de datos.

A pesar de los métodos estadísticos aplicados con éxito; se identificó en las referencias encontradas en el estado del arte sobre los estudios aplicados a la búsqueda de factores que influyen en una búsqueda o variable objetivo determinada, que la metodología CRISP – DM es la mejor opción debido a la claridad que existe en cada una de sus etapas, la forma cómo está estructurada y los resultados en el despliegue que involucran todo el proceso; además, algoritmos como árboles de decisión, regresión logística, PCA y reglas de asociación son los más óptimos y adecuados para la identificación de factores y de relevancias de variables en relación a una única variable objetivo (Rodríguez & Vallejo, 2012),

Ésta variable principal, que hace parte de los datos con los cuales se realizará el presente estudio ,se refiere a si un paciente se realizó o no un procedimiento estético y el objetivo es conocer los factores que influyen en dicha decisión; es entonces la metodología CRISP – DM y los algoritmos mencionados los que aplican para lograr lo deseado.

PARTE 2. DESARROLLO DE LA METODOLOGÍA CRISP-DM

8. DESCRIPCIÓN DEL NEGOCIO

En ésta fase se describe el contexto en el cual los datos fueron obtenidos.

✓ **La Clínica Las Américas**

Clínica Las Américas nació como un proyecto de un grupo de profesionales de la salud quienes se reunieron en 1989 para crear una Sociedad que gestara una clínica con una proyección nacional e internacional a la que denominaron Promotora Médica Las Américas y que hoy es conocida como Grupo Empresarial Las Américas (LAS AMÉRICAS).

Abrió sus puertas el 28 de octubre de 1993 con servicios ambulatorios que incluían quirófanos y algunas ayudas diagnósticas; con el transcurrir de los años siguió creciendo hasta convertirse en la institución de referencia que es hoy tanto para la ciudad y el país. Igualmente es considerada en el mercado internacional como una institución de alta complejidad que ofrece servicios especializados e integrales. Está ubicada en Medellín, Colombia y brinda servicios de salud de alta y mediana complejidad en 77 especialidades y subespecialidades. La misión es: *“Existimos para prestar servicios integrales de salud con calidad, contribuir a la formación del talento humano y fomentar la investigación”*. Y la visión es: *“Seremos reconocidos en esta década por los mejores resultados clínica, con servicio”*.

✓ **Torre Médica Las Américas**

La Torre Médica Las Américas está ubicada en la Clínica las Américas y es uno de los pocos espacios en la ciudad que cubre todas las áreas de la salud, en éste edificio se encuentran los consultorios médicos de profesionales y especialistas socios de la clínica, esto, permitiendo la generación de interconsulta entre los

profesionales y ofreciéndole al paciente una alta confiabilidad por tratarse de médicos.

✓ **Cirugía Plástica**

Ésta especialidad es una de las ofrecidas en la Torre Médica Las Américas y en la cual está formado el Cirujano Plástico John Jairo Valencia Quintero. Es importante resaltar que los datos a estudiar corresponden explícitamente a pacientes que desean o que se realizaron algún procedimiento estético.

✓ **Cirujano Plástico John Jairo Valencia Quintero**

Se graduó como Médico y Cirujano en Marzo de 1989 en la Universidad de Antioquia. Realizó su práctica rural en el Hospital Regional del municipio de Santa Fe de Antioquia. Entre los años de 1992 y 1996 hizo su especialización en Cirugía Plástica Reconstructiva y Estética en la Universidad de Antioquia.

Trabajó en el Instituto de los Seguros Sociales desde 1996 hasta 2002. En 1997 laboró en el Hospital General de Medellín y en el Hospital Pablo Tobón Uribe desde el año 2000 al 2003. Jornadas quirúrgicas múltiples en la Clínica Urabá en el municipio de Apartadó y en el Hospital la Merced del municipio de Ciudad Bolívar. En estas instituciones se dedicó principalmente a la cirugía reconstructiva, trauma facial y cirugía de la mano mientras alternaba con su práctica privada en la Clínica las Américas donde se vinculó como socio en 1996.

Desde el año 2004 se dedica exclusivamente a la práctica privada en la Clínica las Américas, orientado especialmente a la Cirugía Estética. Durante su ejercicio como especialista ha realizado más de 4500 cirugías en el campo de la cirugía reconstructiva del paciente quemado, reconstrucción post-oncológica, cirugía maxilofacial, reconstrucción post-traumática, cirugía de malformaciones congénitas, cirugía de la mano y cirugía estética.

9. COMPRENSIÓN DE LOS DATOS

9.1. DESCRIPCIÓN DE LOS DATOS

En la historia clínica está la recopilación de los datos de los pacientes que asisten a consultas de valoración con el especialista John Jairo Valencia Quintero. En la **Tabla 1** se describen los campos que ésta contiene.

VARIABLE	TIPO DE VARIABLE	DESCRIPCIÓN
PACIENTE		
N° Historia	Numérico	Corresponde a un número consecutivo asignado por la auxiliar de oficina.
Fecha	Fecha	Corresponde al día en que el paciente asiste por primera vez al médico.
Primer Nombre	Categórica	Datos personales del paciente.
Segundo Nombre	Categórica	
Apellidos	Categórica	
Sexo	Categórica	Género del paciente.
Fecha Nacimiento	Fecha	Fecha de nacido.
Edad Cirugía	Categórica	Corresponde a la edad que tenía el paciente al asistir a consulta de valoración.
Estado Civil	Categórica	Corresponde al estado civil que tenía el paciente al asistir a consulta de valoración.
Ocupación	Categórica	Actividad a la cual se dedica el paciente.
Entidad	Categórica	Institución por la cual es remitido el paciente; puede ser alguna eps.
Dirección Residencia	Categórica	Datos personales del paciente y de la persona que lo acompaña a la cita de
Teléfono	Categórica	

e-mail	Categórica	valoración.
Nombre acompañante	Categórica	
Teléfono Acompañante	Categórica	
Persona Responsable	Categórica	
Teléfono Responsable	Categórica	
Parentesco	Categórica	
Cirugía Deseada	Categórica	Hacen referencia a las cirugías que el paciente manifiesta, desea realizarse.
Otra Cirugía Deseada	Categórica	
ANTECEDENTES PATOLOGICOS		
Tiroides	Categórica	Éstas variables hacen referencia a los antecedentes patológicos presentados por el paciente al momento de asistir a la cita de valoración, son de tipo categórico, siendo las categorías (SI/NO).
Corazón	Categórica	
Sida HIV	Categórica	
Riñones	Categórica	
Vesícula	Categórica	
Artritis	Categórica	
Propensión al sangrado	Categórica	
Presión Arterial	Categórica	
Pulmones	Categórica	
Sistema Nervioso	Categórica	
Estomago	Categórica	
Diabetes	Categórica	
Asma	Categórica	
Lupus	Categórica	
Cáncer	Categórica	
Fibromialgia	Categórica	
Hígado	Categórica	
Escleroderma	Categórica	
Otras	Categórica	
MEDICAMENTOS		
Aspirina	Categórica	Éstas variables hacen referencia al consumo de éstos medicamentos al momento de asistir a la cita de valoración, son de tipo categórico, siendo las categorías (SI/NO).
Motrín	Categórica	
Ibuprofeno	Categórica	
Para Artritis	Categórica	
Para Dormir	Categórica	
Para la Tos	Categórica	

Para Tiroides	Categórica	
Hormonas	Categórica	
Anticonceptivos	Categórica	
Tretinoína	Categórica	
Antibióticos	Categórica	
Para la Presión	Categórica	
Para la Diabetes	Categórica	
Diuréticos	Categórica	
Fenobarbital	Categórica	
Digitálicos	Categórica	
Cortisona	Categórica	
Para Perder peso	Categórica	
ANTECEDENTES GINECO		
Edad Menstruación	Categórica	Corresponde a la edad en la cual se presentó la primera menstruación y al número de días del ciclo.
Ciclo	Texto	
Método de Planificación	Categórica	Hace referencia al método de planificación utilizado por el paciente.
ANTECEDENTES FAMILIARES		
Hipertensión Arterial	Categórica	Hacen referencia al padecimiento por parte de un familiar en primer o segundo grado de consanguinidad de las enfermedades en mención, son de tipo categórico, siendo las categorías (SI/NO).
Enfermedad Coronaria	Categórica	
Artritis	Categórica	
Transtornos de Coagulación	Categórica	
Asma	Categórica	
Diabetes	Categórica	
Cáncer de mama	Categórica	
Otro	Categórica	
HABITOS		
Cigarrillo	Categórica	Hace referencia a los hábitos que tiene el paciente al momento de asistir a la cita de valoración, son de tipo categórico, siendo las categorías (SI/NO).
Licor	Categórica	
Café	Categórica	
Píldoras para dormir	Categórica	
Otros	Categórica	

EVALUACIÓN FÍSICA		
Tensión Arterial	Texto	Corresponden a la cuantificación mediante instrumento de medición de la tensión arterial y los latidos cardiacos por minuto.
Frecuencia Cardiaca	Texto	
Peso	Numérico	Corresponden al peso y la talla tomadas por el cirujano plástico en la cita de valoración al paciente.
Talla	Numérico	
IMC	Numérico	Es el índice de masa corporal y representa la relación entre el peso y la talla.
Cara	Texto	Corresponden a la evaluación física realizada por el cirujano plástico al paciente.
Cuello	Texto	
Tórax	Texto	
Abdomen	Texto	
Extremidades	Texto	
Operado	Categorica	Indica que el paciente estuvo en consulta de valoración y se realizó o no una cirugía estética con el cirujano plástico John Jairo Valencia Quintero, es del tipo categorico, siendo las categorías (SI/NO).

Tabla 1. Descripción de Variables

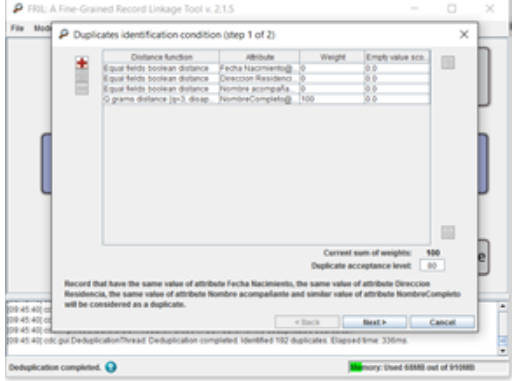
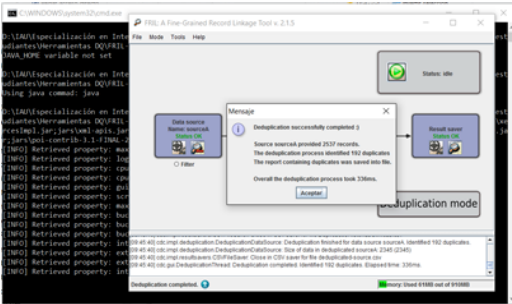
9.2. LIMPIEZA DE DATOS

Para la limpieza de Datos; se realizó el estudio mediante dos herramientas de software:

- ✓ **FRIL:** Es un software mediante el cual se logran detectar duplicados no idénticos mediante reglas de asociación de caracteres.

- ✓ **DQAnalyzer:** Es un software para realizar el perfilamiento de los datos; entrega, entre otras cosas, descripción estadística de los mismos, identificando las características individuales de los datos.

En la **tabla 2**, se pueden observar los hallazgos sobre duplicidad y perfilamiento de datos mediante los softwares mencionados y el resultado del tratamiento a dichos hallazgos.

SOFTWARE	HALLAZGO	TRATAMIENTO
FRILL	 <p style="text-align: center;">Ilustración 4. Selección - Campos Duplicidad.</p>  <p style="text-align: center;">Ilustración 5. Resultados - Duplicidad.</p>	<p>Mediante la revisión de los campos: primer nombre, segundo nombre, apellido y fecha de nacimiento, se encontró que en la base de datos habían 192 registros duplicados; dichos registros fueron revisados uno a uno con el cirujano; se descubrieron casos de error en digitación, duplicidad en elaboración de historia clínica por diferentes procedimientos en fechas diferentes; éstos registros fueron unificados.</p>

DQAnalyzer

PREPARACIÓN_DATOS_TÉISIS

Columns

Basic Frequency Domains Masks Quantiles Groups

Frequency Analysis

Value	Count	Percentage
F	1413	55,70%
f	974	38,39%
m	93	3,67%
M	56	2,21%
Martinez	1	0,04%

Ilustración 6. Preparación - Sexo.

DATOS_V8

Columns

Basic Frequency Domains Masks Quantiles Groups

Frequency Analysis

Value	Count	Percentage
F	2309	94,32%
M	139	5,68%

Ilustración 7. Datos Preparados - Sexo.

PREPARACIÓN_DATOS_TÉISIS

Columns

Frequency Analysis

Value	Count	Percentage
	53	2,09%
soltera	1067	42,06%
casada	760	29,96%
separada	116	4,57%
soltero	94	3,71%
Soltera	87	3,43%
u. libre	65	2,56%
Casada	53	2,09%
viuda	51	2,01%
casado	48	1,89%
divorciada	34	1,34%
Union Libre	14	0,55%
Separada	12	0,47%
casda	10	0,39%
Soltero	8	0,32%
u libre	8	0,32%
Casado	6	0,24%
unión libre	6	0,24%
casad	5	0,20%
separado	5	0,20%
Viuda	3	0,12%

Ilustración 8. Preparación - Estado Civil.

La tabla es mucho más extensa, sólo se ejemplifica la falta de homogenización.

DATOS_V8

Columns

Basic Frequency Domains Masks Quantiles Groups

Frequency Analysis

Value	Count	Percentage
soltero	1234	50,41%
casado	861	35,17%
separado	128	5,23%
union_libre	91	3,72%
viudo	56	2,29%
desconocido	44	1,80%
divorciado	34	1,39%

Ilustración 9. Datos Preparados - Estado Civil.

PREPARACIÓN_DATOS_TESIS

Columns

Basic Frequency Domains Masks Quantiles Groups

Frequency Analysis

Value	Count	Percentage
	284	11,19%
ama de casa	473	18,64%
estudiante	342	13,48%
comerciante	163	6,42%
independiente	143	5,64%
asesor	140	5,52%
empleado	128	5,05%
docente	79	3,11%
administrador	73	2,88%
abogado	62	2,44%
secretario	60	2,36%
estilista	43	1,69%
contador	38	1,50%
auxiliar administración	34	1,34%
enfermero	30	1,18%
auxiliar enfermería	24	0,95%
ingeniero	24	0,95%
diseñador	23	0,91%
auxiliar contable	20	0,78%

Ilustración 10. Preparación - Ocupación.

La tabla es mucho más extensa, sólo se ejemplifica la falta de homogenización.

DATOS_V8

Columns

Frequency Analysis

Value	Count	Percentage
ama_casa	454	18,55%
estudiante	334	13,64%
desconocida	258	10,54%
área_financiera	236	9,64%
vendedor	179	7,31%
área_administrativa	163	6,66%
independiente	139	5,68%
ciencias_salud	129	5,27%
empleado	124	5,07%
docente	78	3,19%
ciencias_sociales	77	3,15%
ingeniero	66	2,70%
artes	60	2,45%
área_belleza	54	2,21%
oficios_varios	40	1,63%
operario	20	0,82%
servidor_público	16	0,65%
pensionado	15	0,61%
ciencias_naturales	4	0,16%
servicio_doméstico	2	0,08%

Ilustración 11. Datos Preparados - Ocupación.

PREPARACIÓN_DATOS_TESIS

Columns

Basic Frequency Masks Quantiles Groups

Frequency Analysis

Value	Count	Percentage
	86	3,39%
aumento de senos	340	13,40%
liposucción	284	11,19%
rinoplastia	115	4,53%
abdominoplastia	90	3,55%
desea aumento mamario	59	2,33%
desea liposucción	55	2,17%
desea implantes mamarios	50	1,97%
blefaroplastia	46	1,81%
implantes mamarios	40	1,58%
desea abdominoplastia	35	1,38%
desea liposucción	26	1,02%
desea implantes	25	0,99%
liposcultura	24	0,95%
desea rinoplastia	20	0,79%
evaluación	19	0,75%
desea CSI de parpados	18	0,71%
desea CSI abdominal	17	0,67%
	17	0,67%

Ilustración 12. Preparación - Cirugía Deseada.

La tabla es mucho más extensa, sólo se ejemplifica la falta de homogenización.

DATOS_V8

Columns

Frequency Analysis

Value	Count	Percentage
aum_mamario	814	33,25%
liposucción	591	24,14%
abdominoplastia	264	10,78%
rinoplastia	233	9,52%
cirugia_facial	167	6,82%
mastopexia	94	3,84%
valoración_general	81	3,31%
cambio_implantes	58	2,37%
lesión	53	2,17%
ritidoplastia	32	1,31%
aum_glúteos	19	0,78%
otoplastia	18	0,74%
aum_implantes	9	0,37%
ginecomastia	5	0,20%
cirugia_mano	3	0,12%
ginecoplastia	2	0,08%
todoplastia	2	0,08%
bichectomía	1	0,04%
braquioplastia	1	0,04%
mastectomía	1	0,04%

Ilustración 13. Datos Preparados - Cirugía Deseada.

PREPARACIÓN_DATOS_TÉISIS

Columns

Basic Frequency Domains Masks Quantiles Groups

Frequency Analysis

Value	Count	Percentage
	1555	61,29%
no	241	9,50%
tubectomia	146	5,75%
no planifica	118	4,65%
AO	89	3,51%
DIU	53	2,09%
vasectomia	34	1,34%
ciclofen	29	1,14%
tubectomia	24	0,95%
Yasmin	21	0,83%
histerectomia	20	0,79%
Yadel	16	0,63%
dianf	12	0,47%
inyección	12	0,47%
pastillas	12	0,47%
preservativo	9	0,35%
SC	9	0,35%
vasectomia	9	0,35%
histerectomia	7	0,28%

Ilustración 14. Preparación - Planificación.

La tabla es mucho más extensa, sólo se ejemplifica la falta de homogenización.

DATOS_V8

Columns

Basic Frequency Domains Masks Quantiles Groups

Frequency Analysis

Value	Count	Percentage
desconocido	1191	48,65%
ninguno	523	21,36%
tubectomia	187	7,64%
met_oral	173	7,07%
pareja	137	5,60%
met_dispositivo	95	3,88%
met_inyectable	62	2,53%
vasectomia	42	1,72%
histerectomia	26	1,06%
preservativo	12	0,49%

Ilustración 15. Datos Preparados - Planificación.

Tabla 2. Limpieza de Datos.

Datos Nulos

Se encontraron 975 registros sin dato en la columna peso y 1265 registros sin dato en la columna talla; para tal caso, se halló la media de ambas columnas, siendo 63,4 para peso y 160 para talla; con ellos se sustituyeron los valores nulos,

Se encontraron 88 registros sin dato en la columna ocupación y se decidió crear una nueva categoría llamada “desconocida” para efectos de conservación del valor único de la verdad en los datos.

Datos Atípicos

En la columna planificación se encontraron 4 datos de tipo date (quizá por la última fecha en que se presentó la menstruación); dichas filas fueron eliminadas.

En la columna de peso se encontraron 3 valores muy bajos, no acordes según la talla del paciente; dichas filas fueron eliminadas.

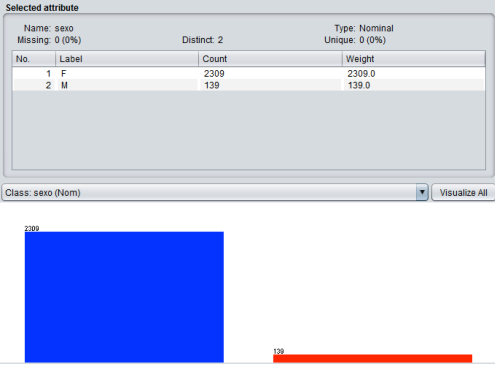
En la columna ocupación se encontraron 5 datos con nombres de empresas; dichas filas fueron eliminadas.

Registros Duplicados

Luego del análisis y corrección de los registros hallados como duplicados mediante FRIL; se considera la base de datos libre de duplicidad. Se realiza además, tal como se muestra en la **tabla 2**, la homogenización de varias columnas, por ejemplo, la columna ocupación, donde se encontraron datos no homogéneos tales como: maestra, docente, licenciada y profesora, entre otros.

9.3. ANÁLISIS ESTADÍSTICO DE LOS DATOS.

A continuación, en la **Tabla 3** se pueden observar las descripciones estadísticas de cada uno de los datos.

ATRIBUTO	DESCRIPCIÓN
 <p>Ilustración 16. Análisis Estadístico - Sexo.</p>	<p>La variable “sexo” contiene dos valores posibles, “F” para femenino con un total de registros de 2.309 y “M” para masculino con un total de registros de 139.</p>
	<p>La variable “edad” contiene una edad mínima de 1 año y edad máxima de 84 años.</p>

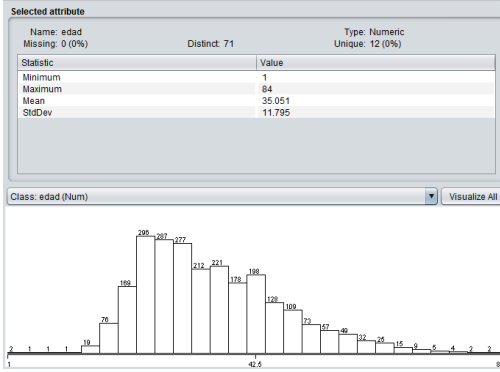


Ilustración 17. Análisis Estadístico - Edad.

En la curva se observa la media de 35,051 años.

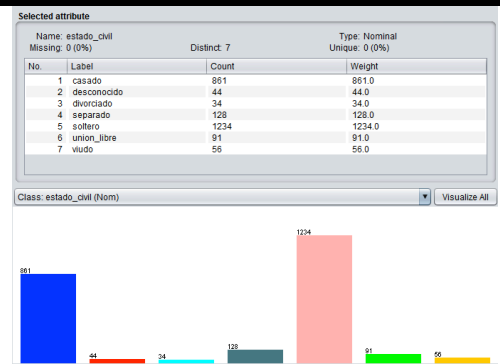


Ilustración 18. Análisis Estadístico - Estado Civil.

La variable “estado_civil” contiene un total de 7 categorías, donde la categoría “soltero” presenta el mayor número de registros, en total 1.234, seguido por la categoría “casado” con 861 registros y la categoría con menos registros es “divorciado”, con 34.

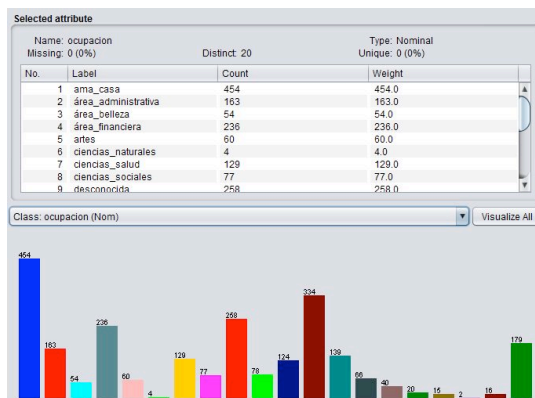


Ilustración 19. Análisis Estadístico - Ocupación.

La variable “ocupación” contiene un total de 20 categorías, donde la categoría “ama_de_casa” presenta el mayor número de registros, en total 454, seguido por la categoría “estudiante” con 334 registros y la categoría con menos registros es “servicio_doméstico”, con 2.

La variable “cirugía_deseada1” contiene un total de 20

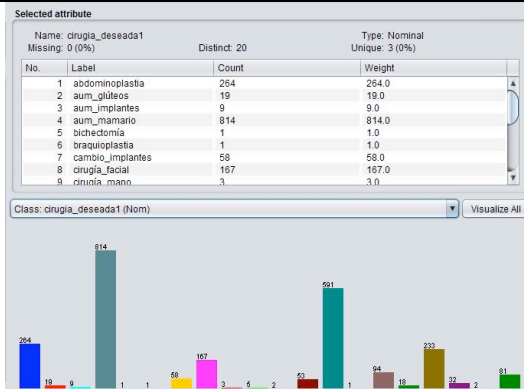


Ilustración 20. Análisis Estadístico - Cirugía Deseada 1.

categorías, donde la categoría “aum_mamario” presenta el mayor número de registros, en total 814, seguido por la categoría “liposucción” con 591 registros y las categorías con menos registros son “bichectomía”, “braquioplastia” y “mastectomía”, con sólo 1 registro.

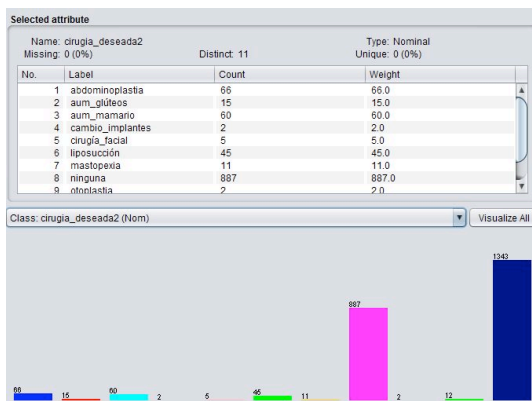


Ilustración 21. Análisis Estadístico - Cirugía Deseada 2.

La variable “cirugía_deseada2” contiene un total de 11 categorías, donde la categoría “única” presenta el mayor número de registros, en total 1.343, queriendo decir esto que la paciente sólo estuvo interesada en la “cirugía_deseada1” y pudo habérsela o no realizado; seguido por la categoría “ninguna” con 887 registros, que significa que a pesar de que la paciente estuvo interesada en la “cirugía_deseada1”, no se realizó ninguna cirugía y las categorías con menos registros son “cambio_implantes”, y “otoplastia”, con sólo 2 registros.

La variable “AP_TIROIDES” contiene dos valores posibles, “SI” que indica que el paciente ha presentado antecedentes patológicos de tiroides con un total de 117 registros y “NO” que indica que el paciente no ha

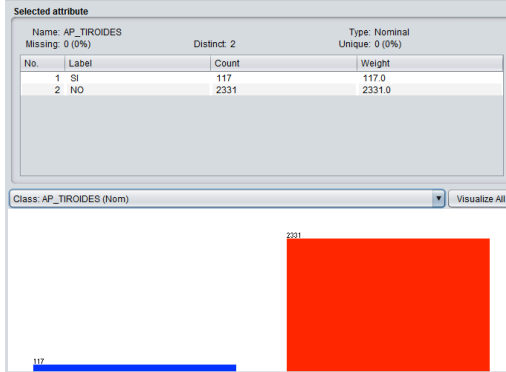


Ilustración 22. Análisis Estadístico - AP_TIROIDES.

presentado antecedentes patológicos de tiroides con un total de 2.331 registros.

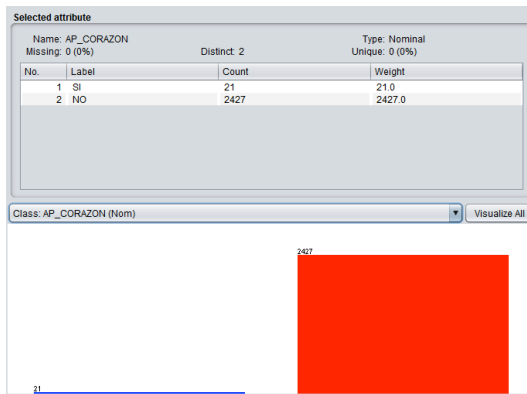


Ilustración 23. Análisis Estadístico - AP_CORAZON.

La variable “AP_CORAZON” contiene dos valores posibles, “SI” que indica que el paciente ha presentado antecedentes patológicos de enfermedades coronarias con un total de 21 registros y “NO” que indica que el paciente no ha presentado antecedentes patológicos de enfermedades coronarias con un total de 2.427 registros.

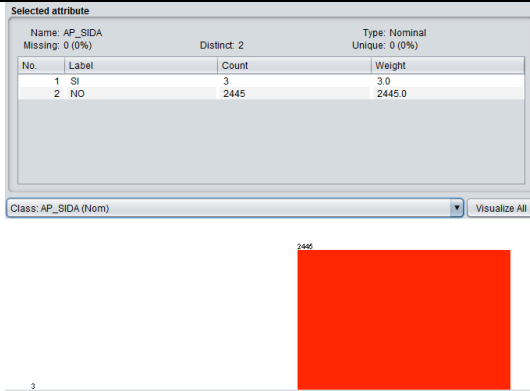


Ilustración 24. Análisis Estadístico - AP_SIDA.

La variable “AP_SIDA” contiene dos valores posibles, “SI” que indica que el paciente ha presentado antecedentes patológicos de SIDA con un total de 3 registros y “NO” que indica que el paciente no ha presentado antecedentes patológicos de SIDA con un total de 2.445 registros.

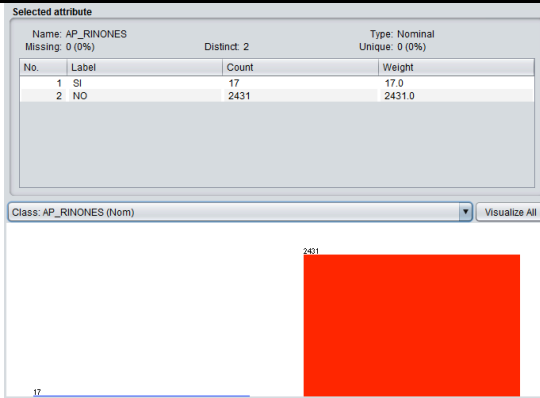


Ilustración 25. Análisis Estadístico - AP_RINONES.

La variable “AP_RINONES” contiene dos valores posibles, “SI” que indica que el paciente ha presentado antecedentes patológicos de enfermedades relacionadas al funcionamiento de los riñones con un total de 17 registros y “NO” que indica que el paciente no ha presentado antecedentes patológicos de enfermedades relacionadas al funcionamiento de los riñones con un total de 2.431 registros.

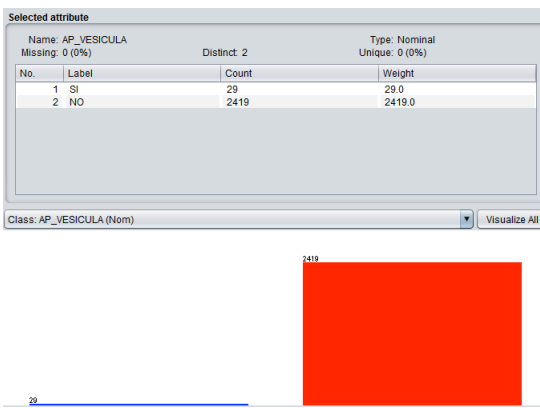


Ilustración 26. Análisis Estadístico - AP_VESICULA.

La variable “AP_VESICULA” contiene dos valores posibles, “SI” que indica que el paciente ha presentado antecedentes patológicos de enfermedades relacionadas al funcionamiento de la vesícula con un total de 29 registros y “NO” que indica que el paciente no ha presentado antecedentes patológicos de enfermedades relacionadas al funcionamiento de la vesícula con un total de 2.419 registros.

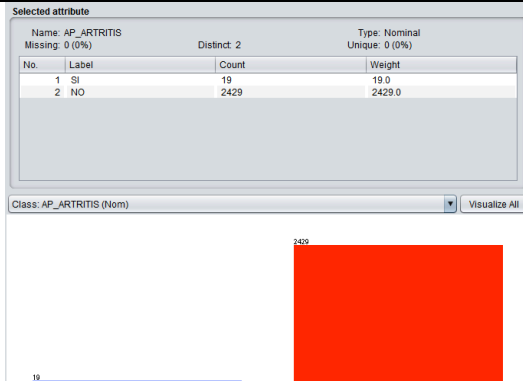


Ilustración 27. Análisis Estadístico - AP_ARTRITIS.

La variable “AP_ARTRITIS” contiene dos valores posibles, “SI” que indica que el paciente ha presentado antecedentes patológicos de artritis con un total de 19 registros y “NO” que indica que el paciente no ha presentado antecedentes patológicos de artritis con un total de 2.429 registros.

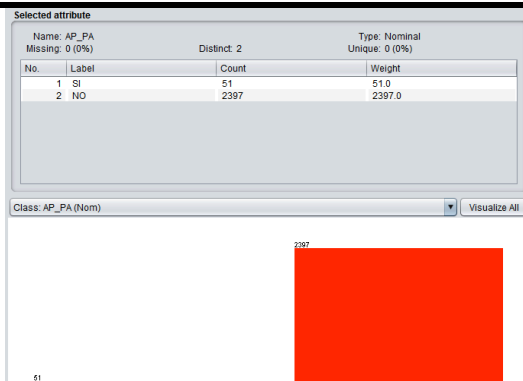


Ilustración 28. Análisis Estadístico - AP_PA.

La variable “AP_PA” contiene dos valores posibles, “SI” que indica que el paciente ha presentado antecedentes patológicos de presión arterial con un total de 51 registros y “NO” que indica que el paciente no ha presentado antecedentes patológicos de presión arterial con un total de 2.397 registros.

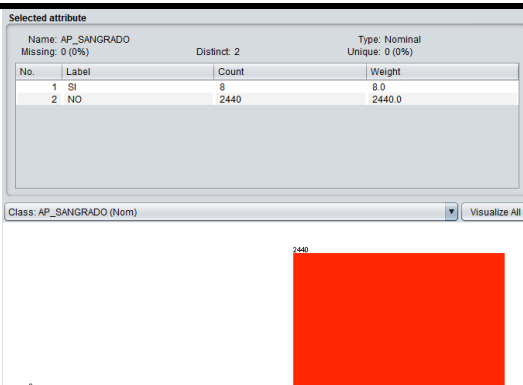


Ilustración 29. Análisis Estadístico - AP_SANGRADO.

La variable “AP_SANGRADO” contiene dos valores posibles, “SI” que indica que el paciente ha presentado antecedentes patológicos de sangrado con un total de 8 registros y “NO” que indica que el paciente no ha presentado antecedentes patológicos de sangrado con un total de 2.440 registros.

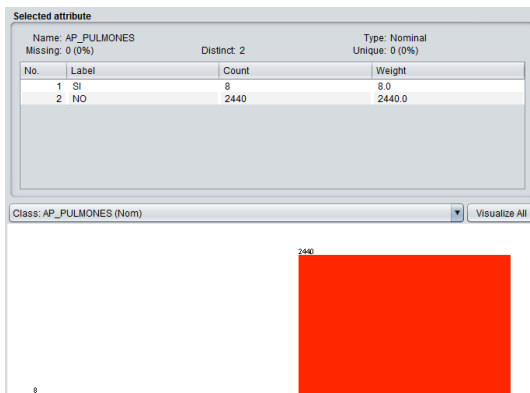


Ilustración 30. Análisis Estadístico - AP_PULMONES.

La variable “AP_PULMONES” contiene dos valores posibles, “SI” que indica que el paciente ha presentado antecedentes patológicos de enfermedades relacionadas al funcionamiento de los pulmones con un total de 8 registros y “NO” que indica que el paciente no ha presentado antecedentes patológicos de enfermedades relacionadas al funcionamiento de los pulmones con un total de 2.440 registros.

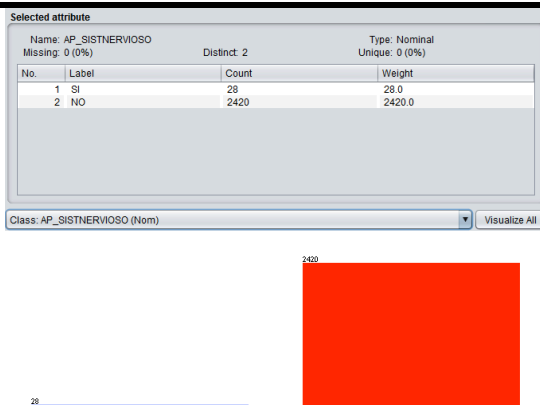


Ilustración 31. Análisis Estadístico - AP_SISTNERVIOSO.

La variable “AP_SISTNERVIOSO” contiene dos valores posibles, “SI” que indica que el paciente ha presentado antecedentes patológicos de enfermedades relacionadas al funcionamiento del sistema nervioso con un total de 28 registros y “NO” que indica que el paciente no ha presentado antecedentes patológicos de enfermedades relacionadas al funcionamiento del sistema nervioso con un total de 2.420 registros.

La variable “AP_ESTOMAGO” contiene dos valores posibles, “SI” que indica que el paciente ha presentado antecedentes patológicos de enfermedades relacionadas al funcionamiento del sistema digestivo con un total de 53 registros y “NO” que indica

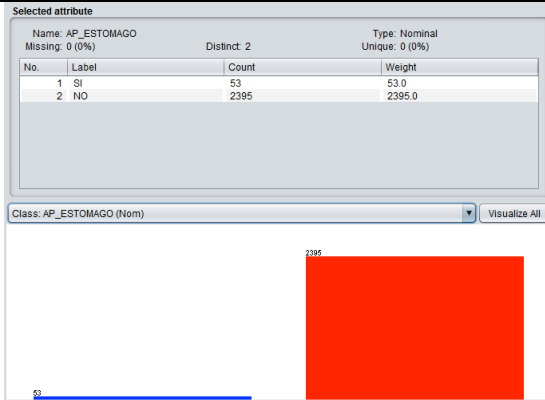


Ilustración 32. Análisis Estadístico - AP_ESTOMAGO.

que el paciente no ha presentado antecedentes patológicos de enfermedades relacionadas al funcionamiento del sistema digestivo con un total de 2.395 registros.

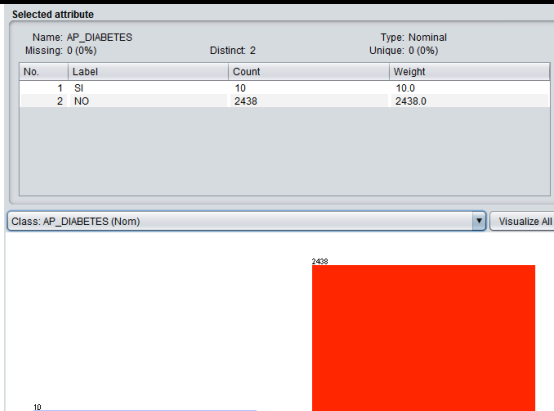


Ilustración 33. Análisis Estadístico - AP_DIABETES.

La variable “AP_DIABETES” contiene dos valores posibles, “SI” que indica que el paciente ha presentado antecedentes patológicos de diabetes con un total de 10 registros y “NO” que indica que el paciente no ha presentado antecedentes patológicos de diabetes con un total de 2.438 registros.

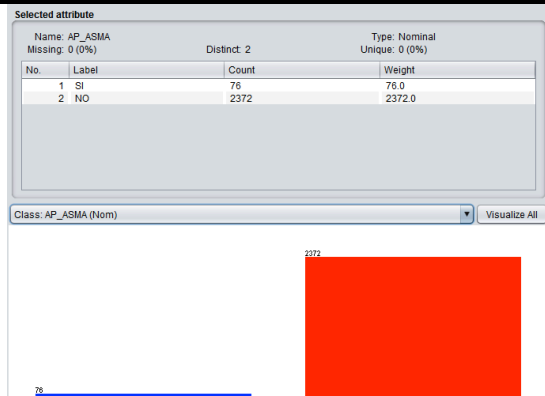


Ilustración 34. Análisis Estadístico - AP_ASMA.

La variable “AP_ASMA” contiene dos valores posibles, “SI” que indica que el paciente ha presentado antecedentes patológicos de asma con un total de 76 registros y “NO” que indica que el paciente no ha presentado antecedentes patológicos de asma con un total de 2.372 registros.

Selected attribute

Name: AP_LUPUS
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 1 (0%)

No.	Label	Count	Weight
1	SI	1	1.0
2	NO	2447	2447.0

Class: AP_LUPUS (Nom) Visualize All

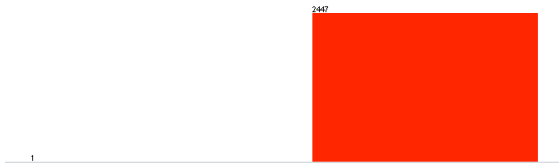


Ilustración 35. Análisis Estadístico - AP_LUPUS.

La variable “AP_LUPUS” contiene dos valores posibles, “SI” que indica que el paciente ha presentado antecedentes patológicos de lupus con un total de 1 registro y “NO” que indica que el paciente no ha presentado antecedentes patológicos de lupus con un total de 2.447 registros.

Selected attribute

Name: AP_CANCER
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	SI	14	14.0
2	NO	2434	2434.0

Class: AP_CANCER (Nom) Visualize All

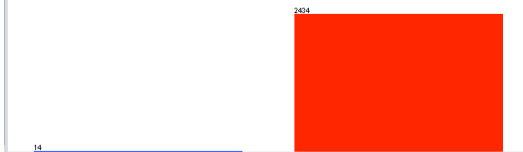


Ilustración 36. Análisis Estadístico - AP_CANCER.

La variable “AP_CANCER” contiene dos valores posibles, “SI” que indica que el paciente ha presentado antecedentes patológicos de cáncer con un total de 14 registros y “NO” que indica que el paciente no ha presentado antecedentes patológicos de cáncer con un total de 2.434 registros.

Selected attribute

Name: AP_FIBROMIALGIA
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	SI	10	10.0
2	NO	2438	2438.0

Class: AP_FIBROMIALGIA (Nom) Visualize All

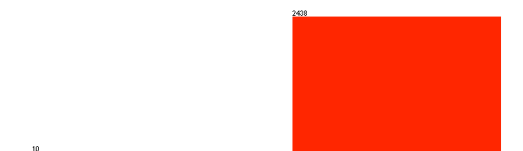


Ilustración 37. Análisis Estadístico - AP_FIBROMIALGIA.

La variable “AP_FIBROMIALGIA” contiene dos valores posibles, “SI” que indica que el paciente ha presentado antecedentes patológicos de fibromialgia con un total de 10 registros y “NO” que indica que el paciente no ha presentado antecedentes patológicos de fibromialgia con un total de 2.438 registros.

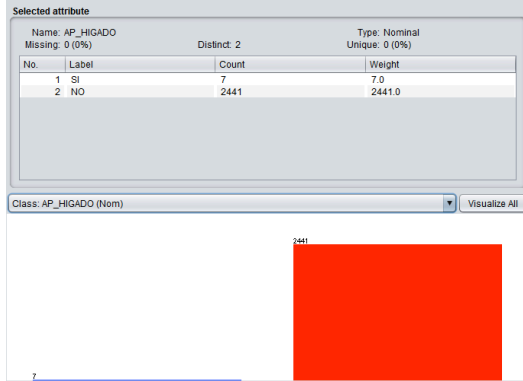


Ilustración 38. Análisis Estadístico - AP_HIGADO.

La variable “AP_HIGADO” contiene dos valores posibles, “SI” que indica que el paciente ha presentado antecedentes patológicos de enfermedades relacionadas al funcionamiento del hígado con un total de 7 registros y “NO” que indica que el paciente no ha presentado antecedentes patológicos de enfermedades relacionadas al funcionamiento del hígado con un total de 2.441 registros.

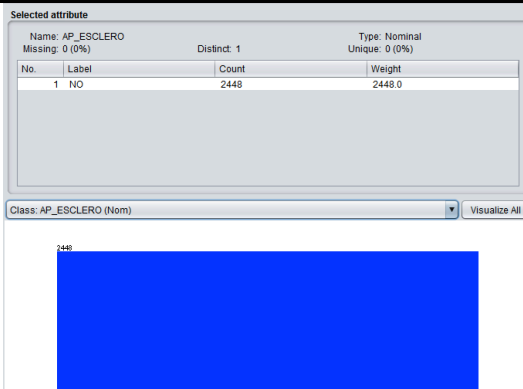


Ilustración 39. Análisis Estadístico - AP_ESCLERO.

La variable “AP_ESCLERO” contiene sólo un valor posible, “NO” que indica que los pacientes no han presentado antecedentes patológicos de esclerosis con un total de 2448 registros.

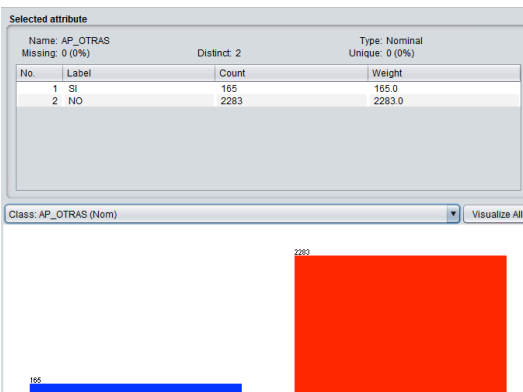


Ilustración 40. Análisis Estadístico - AP_OTRAS.

La variable “AP_OTRAS” contiene dos valores posibles, “SI” que indica que el paciente ha presentado antecedentes patológicos de otras enfermedades con un total de 165 registros y “NO” que indica que el paciente no ha presentado antecedentes patológicos de otras enfermedades con un total de 2.283 registros.

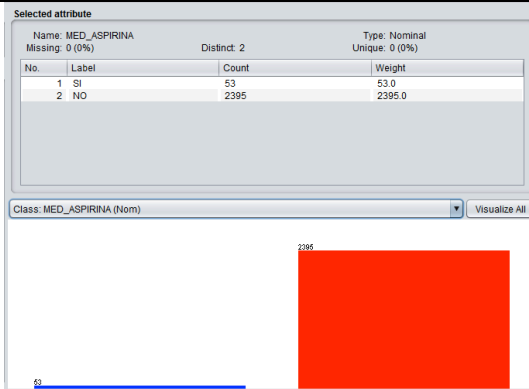


Ilustración 41. Análisis Estadístico - MED_ASPIRINA.

La variable “MED_ASPIRINA” contiene dos valores posibles, “SI” que indica que el paciente actualmente consume aspirina con un total de 53 registros y “NO” que indica que el paciente actualmente no consume aspirina con un total de 2.395 registros.

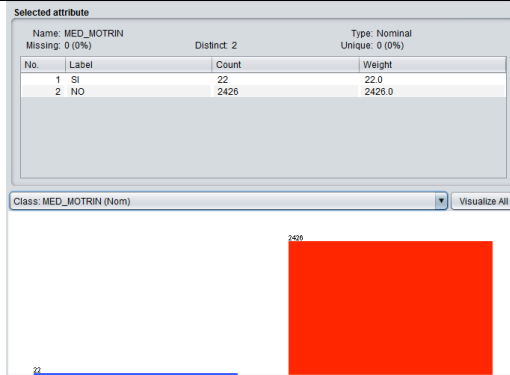


Ilustración 42. Análisis Estadístico - MED_MOTRIN.

La variable “MED_MOTRIN” contiene dos valores posibles, “SI” que indica que el paciente actualmente consume motrín con un total de 22 registros y “NO” que indica que el paciente actualmente no consume motrín con un total de 2.426 registros.

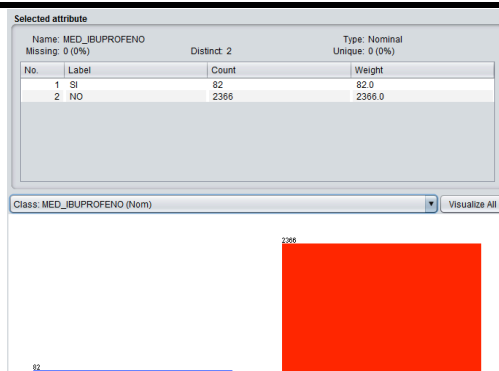


Ilustración 43. Análisis Estadístico - MED_IBUPROFENO.

La variable “MED_IBUPROFENO” contiene dos valores posibles, “SI” que indica que el paciente actualmente consume ibuprofeno con un total de 82 registros y “NO” que indica que el paciente actualmente no consume ibuprofeno con un total de 2.366 registros.

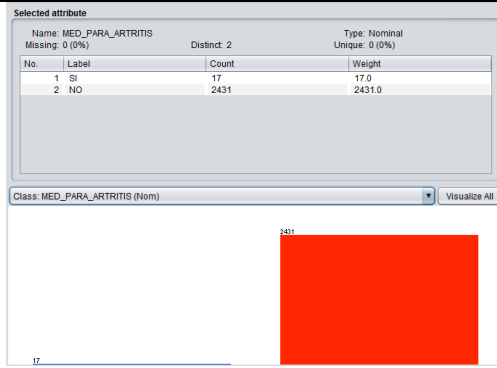


Ilustración 44. Análisis Estadístico - MED_PARA_ARTRITIS.

La variable “MED_PARA_ARTRITIS” contiene dos valores posibles, “SI” que indica que el paciente actualmente consume medicamentos para el tratamiento de artritis con un total de 17 registros y “NO” que indica que el paciente actualmente no consume medicamentos para el tratamiento de artritis con un total de 2.431 registros.

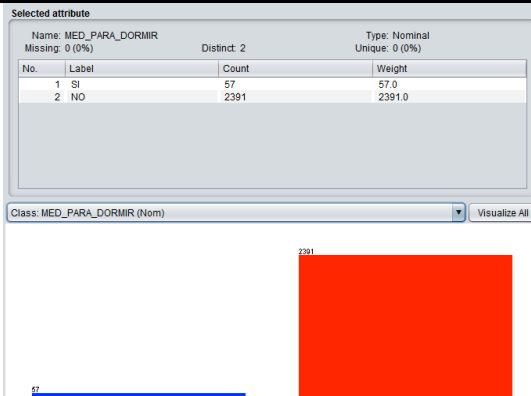


Ilustración 45. Análisis Estadístico - MED_PARA_DORMIR.

La variable “MED_PARA_DORMIR” contiene dos valores posibles, “SI” que indica que el paciente actualmente consume medicamentos para inducir al sueño con un total de 57 registros y “NO” que indica que el paciente actualmente no consume medicamentos para inducir al sueño con un total de 2.391 registros.

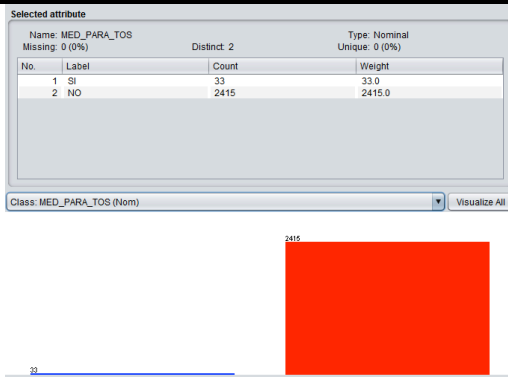


Ilustración 46. Análisis Estadístico - MED_PARA_TOS.

La variable “MED_PARA_TOS” contiene dos valores posibles, “SI” que indica que el paciente actualmente consume medicamentos para la tos con un total de 33 registros y “NO” que indica que el paciente actualmente no consume medicamentos para la tos con un total de 2.415 registros.

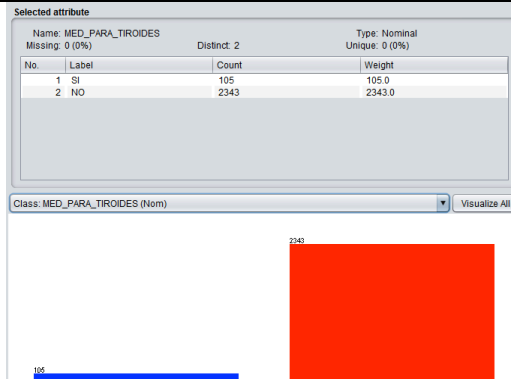


Ilustración 47. Análisis Estadístico - MED_PARA_TIROIDES.

La variable “MED_PARA_TIROIDES” contiene dos valores posibles, “SI” que indica que el paciente actualmente consume medicamentos para el tratamiento de la tiroides con un total de 105 registros y “NO” que indica que el paciente actualmente no consume medicamentos para el tratamiento de la tiroides con un total de 2.343 registros.

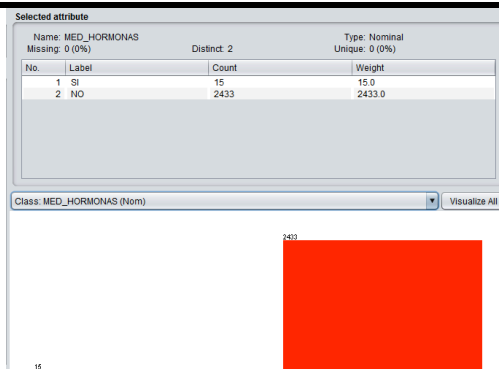


Ilustración 48. Análisis Estadístico - MED_HORMONAS.

La variable “MED_HORMONAS” contiene dos valores posibles, “SI” que indica que el paciente actualmente consume medicamentos hormonales con un total de 15 registros y “NO” que indica que el paciente actualmente no consume medicamentos hormonales con un total de 2.433 registros.

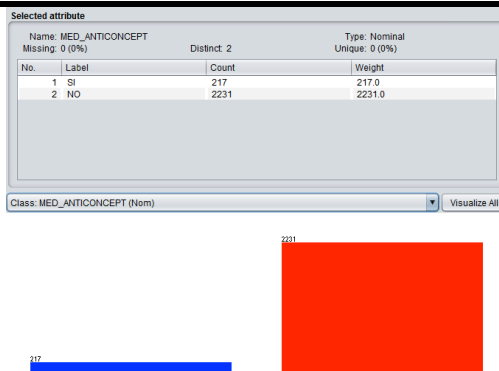


Ilustración 49. Análisis Estadístico - MED_ANTICONCEPT.

La variable “MED_ANTICONCEPT” contiene dos valores posibles, “SI” que indica que el paciente actualmente consume anticonceptivos con un total de 217 registros y “NO” que indica que el paciente actualmente no consume anticonceptivos con un total de 2.231 registros.

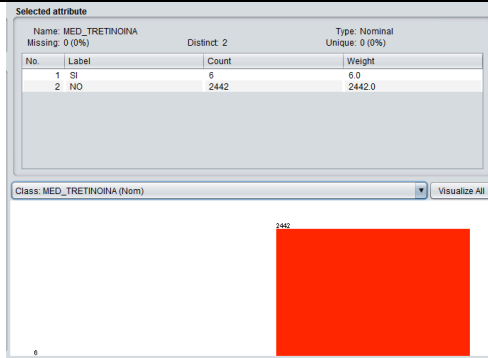


Ilustración 50. Análisis Estadístico - MED_TRETINOINA.

La variable “MED_TRETINOINA” contiene dos valores posibles, “SI” que indica que el paciente actualmente consume tretinoína con un total de 6 registros y “NO” que indica que el paciente actualmente no consume tretinoína con un total de 2.442 registros.

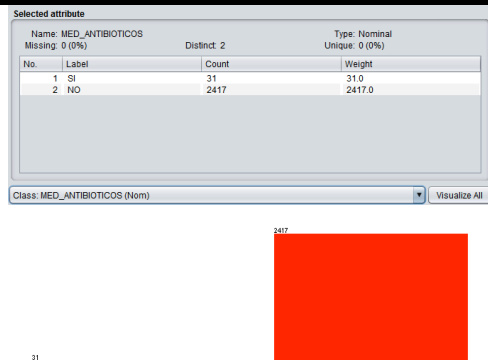


Ilustración 51. Análisis Estadístico - MED_ANTIBIOTICOS.

La variable “MED_ANTIBIOTICOS” contiene dos valores posibles, “SI” que indica que el paciente actualmente consume antibióticos con un total de 31 registros y “NO” que indica que el paciente actualmente no consume antibióticos con un total de 2.417 registros.

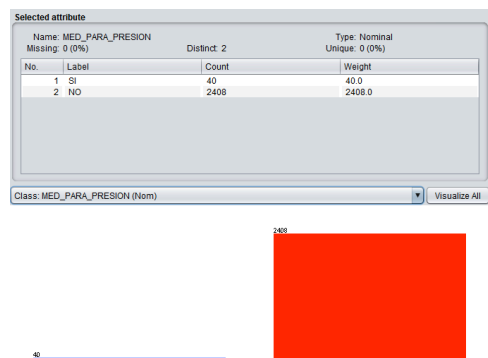


Ilustración 52. Análisis Estadístico - MED_PARA_PRESION.

La variable “MED_PARA_PRESION” contiene dos valores posibles, “SI” que indica que el paciente actualmente consume medicamentos para el tratamiento de la presión arterial con un total de 40 registros y “NO” que indica que el paciente actualmente no consume medicamentos para el tratamiento de la presión arterial con un total de 2.408 registros.

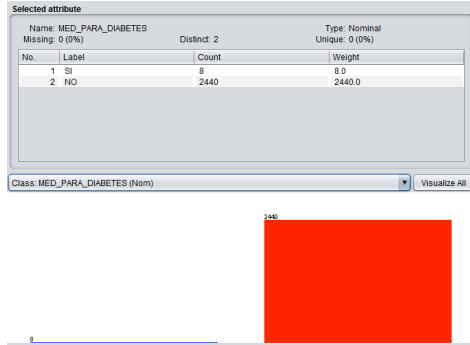


Ilustración 53. Análisis Estadístico - MED_PARA_DIABETES.

La variable “MED_PARA_DIABETES” contiene dos valores posibles, “SI” que indica que el paciente actualmente consume medicamentos para el tratamiento de la diabetes con un total de 8 registros y “NO” que indica que el paciente actualmente no consume medicamentos para el tratamiento de la diabetes con un total de 2.440 registros.

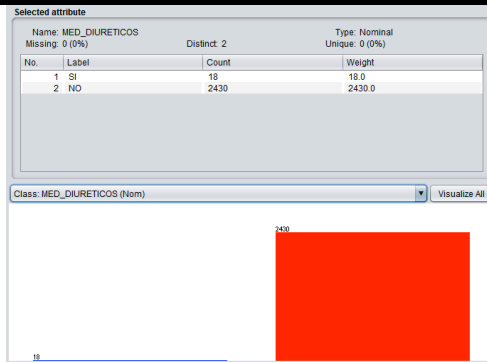


Ilustración 54. Análisis Estadístico - MED_DIURETICOS.

La variable “MED_DIURETICOS” contiene dos valores posibles, “SI” que indica que el paciente actualmente consume medicamentos de éste tipo con un total de 18 registros y “NO” que indica que el paciente actualmente no consume medicamentos de éste tipo con un total de 2.430 registros.

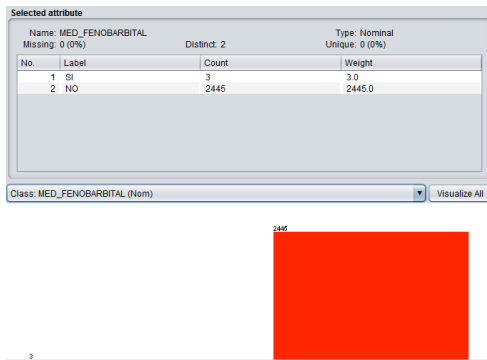


Ilustración 55. Análisis Estadístico - MED_FENOBARBITAL.

La variable “MED_FENOBARBITAL” contiene dos valores posibles, “SI” que indica que el paciente actualmente consume fenobarbital con un total de 3 registros y “NO” que indica que el paciente actualmente no consume fenobarbital con un total de 2.445 registros.

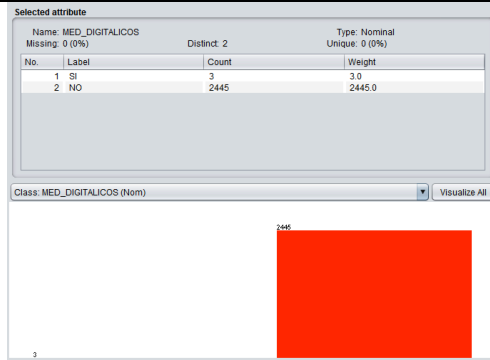


Ilustración 56. Análisis Estadístico - MED_DIGITALICOS.

La variable “MED_DIGITALICOS” contiene dos valores posibles, “SI” que indica que el paciente actualmente consume medicamentos de éste tipo con un total de 3 registros y “NO” que indica que el paciente actualmente no consume medicamentos de éste tipo con un total de 2.445 registros.

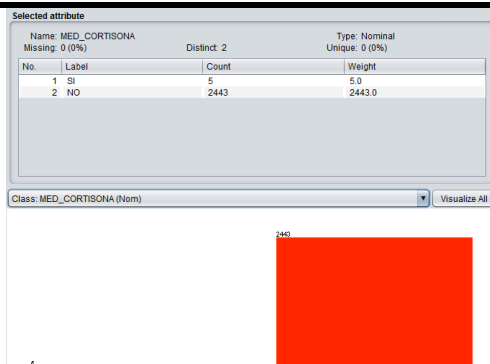


Ilustración 57. Análisis Estadístico - MED_CORTISONA.

La variable “MED_CORTISONA” contiene dos valores posibles, “SI” que indica que el paciente actualmente consume cortisona con un total de 5 registros y “NO” que indica que el paciente actualmente no consume cortisona con un total de 2.443 registros.

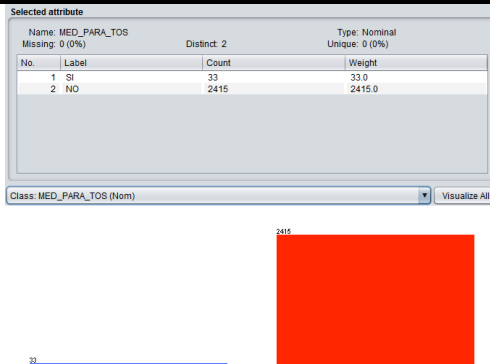


Ilustración 58. Análisis Estadístico - MED_PARA_TOS.

La variable “MED_PARA_TOS” contiene dos valores posibles, “SI” que indica que el paciente actualmente consume medicamentos para el control de la tos con un total de 33 registros y “NO” que indica que el paciente actualmente no consume medicamentos para el control de la tos con un total de 2.415 registros.

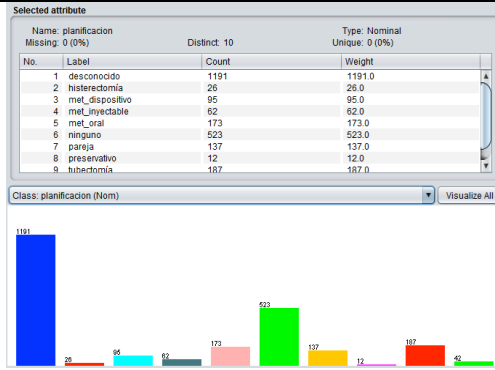


Ilustración 59. Análisis Estadístico - planificación.

La variable “ocupación” contiene un total de 10 categorías, donde la categoría “desconocido” presenta el mayor número de registros, en total 1.191, seguido por la categoría “ninguno” con 523 registros y la categoría con menos registros es “preservativo”, con 12.

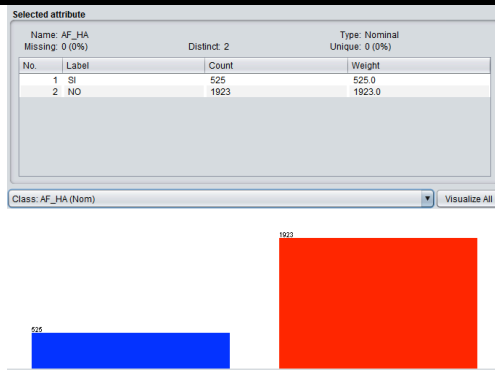


Ilustración 60. Análisis Estadístico - AF_HA.

La variable “AF_HA” contiene dos valores posibles, “SI” que indican que el paciente tiene antecedentes familiares de hipertensión arterial con un total de 525 registros y “NO” que indica que el paciente actualmente no tiene antecedentes familiares de hipertensión arterial con un total de 1.923 registros.

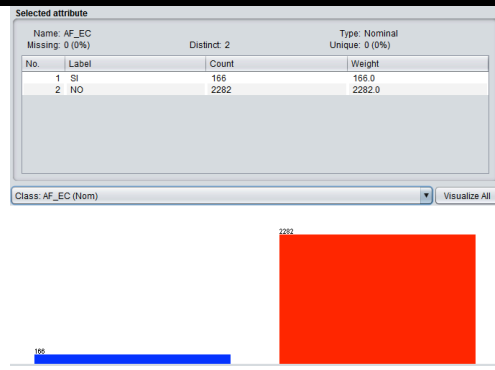


Ilustración 61. Análisis Estadístico - AF_EC.

La variable “AF_EC” contiene dos valores posibles, “SI” que indican que el paciente tiene antecedentes familiares sobre enfermedades coronarias con un total de 166 registros y “NO” que indica que el paciente actualmente no tiene antecedentes sobre enfermedades coronarias con un total de 2.282 registros.

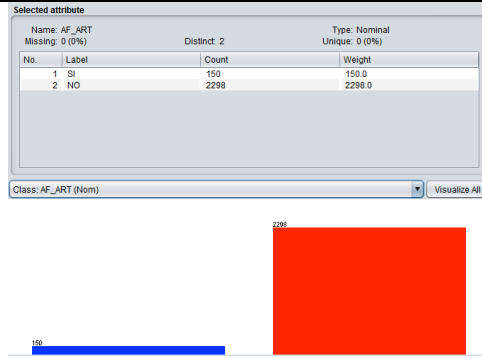


Ilustración 62. Análisis Estadístico - AF_ART.

La variable “AF_ART” contiene dos valores posibles, “SI” que indican que el paciente tiene antecedentes familiares de artritis con un total de 150 registros y “NO” que indica que el paciente actualmente no tiene antecedentes familiares de artritis con un total de 2.298 registros.

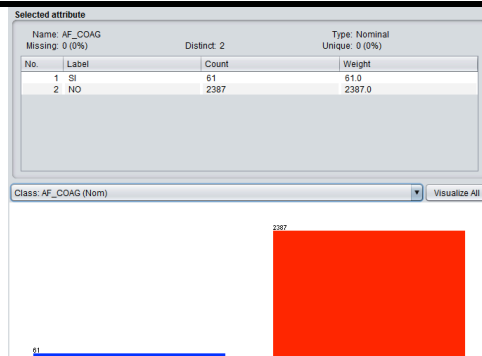


Ilustración 63. Análisis Estadístico - AF_COAG.

La variable “AF_COAG” contiene dos valores posibles, “SI” que indican que el paciente tiene antecedentes familiares sobre problemas de coagulación de sangre con un total de 61 registros y “NO” que indica que el paciente actualmente no tiene antecedentes familiares sobre problemas de coagulación de sangre con un total de 2.387 registros.

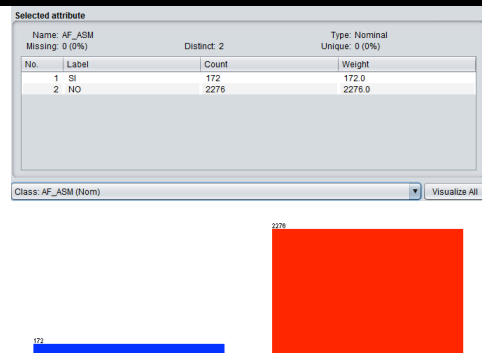


Ilustración 64. Análisis Estadístico - AF_ASM.

La variable “AF_ASM” contiene dos valores posibles, “SI” que indican que el paciente tiene antecedentes familiares de asma con un total de 172 registros y “NO” que indica que el paciente actualmente no tiene antecedentes familiares de asma con un total de 2.275 registros.

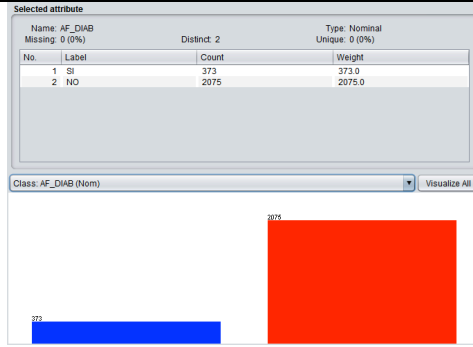


Ilustración 65. Análisis Estadístico - AF_DIAB.

La variable “AF_DIAB” contiene dos valores posibles, “SI” que indican que el paciente tiene antecedentes familiares de diabetes con un total de 373 registros y “NO” que indica que el paciente actualmente no tiene antecedentes familiares de diabetes con un total de 2.075 registros.

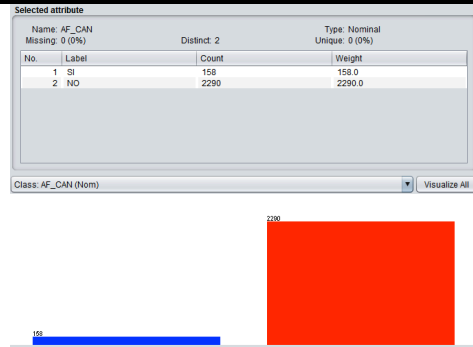


Ilustración 66. Análisis Estadístico - AF_CAN.

La variable “AF_CAN” contiene dos valores posibles, “SI” que indican que el paciente tiene antecedentes familiares de cáncer con un total de 158 registros y “NO” que indica que el paciente actualmente no tiene antecedentes familiares de cáncer con un total de 2.290 registros.

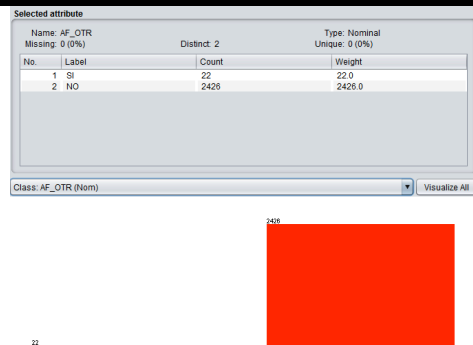


Ilustración 67. Análisis Estadístico - AF_OTR.

La variable “AF_OTR” contiene dos valores posibles, “SI” que indican que el paciente tiene antecedentes familiares de otras enfermedades con un total de 22 registros y “NO” que indica que el paciente actualmente no tiene antecedentes familiares de otras enfermedades con un total de 2.426 registros.

No.	Label	Count	Weight
1	SI	115	115.0
2	NO	2333	2333.0

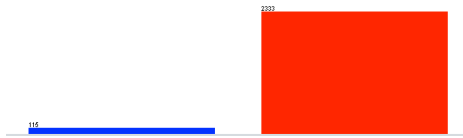


Ilustración 68. Análisis Estadístico - HAB_CIGARRILLO.

La variable “HAB_CIGARRILLO” contiene dos valores posibles, “SI” que indican que el paciente tiene el hábito de fumar cigarrillo con un total de 115 registros y “NO” que indica que el paciente no tiene el hábito de fumar cigarrillo con un total de 2.333 registros.

No.	Label	Count	Weight
1	SI	101	101.0
2	NO	2347	2347.0

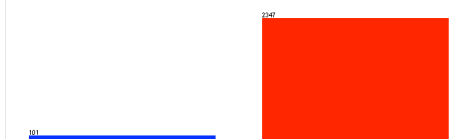


Ilustración 69. Análisis Estadístico - HAB_LICOR.

La variable “HAB_LICOR” contiene dos valores posibles, “SI” que indican que el paciente tiene el hábito de consumir licor con un total de 101 registros y “NO” que indica que el paciente no tiene el hábito de consumir licor con un total de 2.347 registros.

No.	Label	Count	Weight
1	SI	159	159.0
2	NO	2289	2289.0

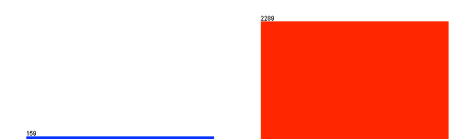


Ilustración 70. Análisis Estadístico - HAB_CAFE.

La variable “HAB_CAFE” contiene dos valores posibles, “SI” que indican que el paciente tiene el hábito de consumir café con un total de 159 registros y “NO” que indica que el paciente no tiene el hábito de consumir café con un total de 2.289 registros.

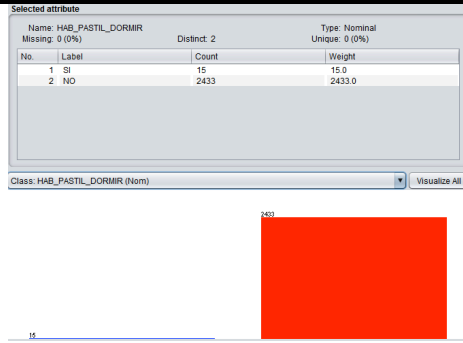


Ilustración 71. Análisis Estadístico - HAB_PASTIL_DORMIR.

La variable “HAB_PASTIL_DORMIR” contiene dos valores posibles, “SI” que indican que el paciente tiene el hábito de consumir pastillas para inducir al sueño con un total de 15 registros y “NO” que indica que el paciente no tiene el hábito de consumir pastillas para inducir al sueño con un total de 2.433 registros.

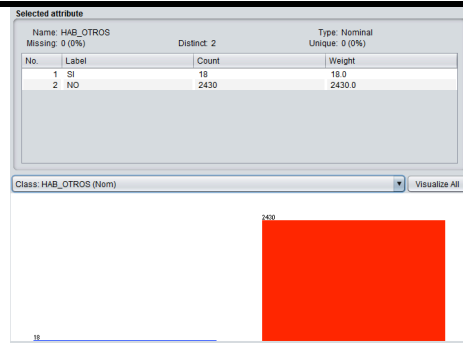


Ilustración 72. Análisis Estadístico - HAB_OTROS.

La variable “HAB_OTROS” contiene dos valores posibles, “SI” que indican que el paciente tiene otros hábitos no saludables con un total de 18 registros y “NO” que indica que el paciente no tiene hábitos no saludables con un total de 2.430 registros.



Ilustración 73. Análisis Estadístico - peso.

La variable “peso” contiene un peso mínimo de 11 kilogramos y peso máximo de 117 kilogramos. En la curva se observa que la media se encuentra en los 63,421 kilogramos.

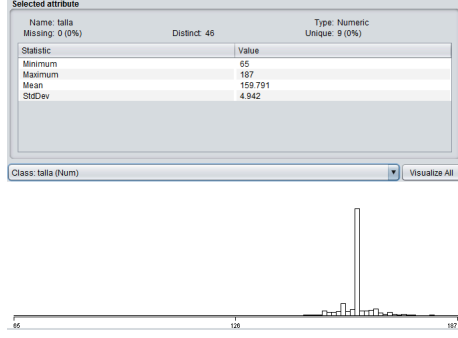
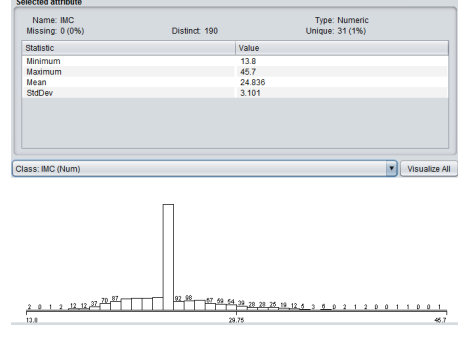
 <p>Ilustración 74. Análisis Estadístico - talla.</p>	<p>La variable “talla” contiene una talla mínima de 65 centímetros y una talla máxima de 187 centímetros.</p> <p>En la curva se observa la media de 159,791.</p>
 <p>Ilustración 75. Análisis Estadístico - IMC.</p>	<p>La variable “IMC” contiene un índice de masa corporal mínimo de 13,8 y un índice de masa corporal máximo de 45,7.</p> <p>En la curva se observa una media de 24.</p>

Tabla 3. Descripción Estadística de los Datos.

9.4. TRANSFORMACIONES

Las variables no requirieron ninguna transformación; a excepción de la necesidad que se presentó para ejecutar el algoritmo “A priori”, el cual no permite variables numéricas; para esto entonces, desde el negocio, se categorizaron las variables: edad e IMC.

9.5. SELECCIÓN DE VARIABLES.

Para efectos de minería, como se ilustra en la **Tabla 4**, se eliminaron columnas que, según el criterio del cirujano plástico y el análisis de datos, se identificaron como no aptas, redundantes o irrelevantes porque no inciden en la toma de decisión frente a la realización de cirugía estética:

VARIABLE	TIPO DE VARIABLE	ANÁLISIS	JUSTIFICACIÓN
PACIENTE			
N° Historia	Numérico	Irrelevante	Ya que corresponde a un número consecutivo asignado por la auxiliar de oficina.
Fecha	Fecha	Irrelevante	Ya que corresponde al día en que el paciente asiste por primera vez al médico.
Primer Nombre	Categórica	Irrelevante	Ya que éste tipo de datos no generan resultados de minería relevantes para el negocio.
Segundo Nombre	Categórica	Irrelevante	
Apellidos	Categórica	Irrelevante	
Sexo weka: sexo	Categórica	Relevante	Corresponde al género del paciente.
Fecha Nacimiento	Fecha	Irrelevante	Presenta redundancia con la variable Edad.
Edad Cirugía weka: edad	Categórica	Relevante	Corresponde a la edad que tenía el paciente al asistir a consulta de valoración.
Estado Civil weka: estado_civil	Categórica	Relevante	Corresponde al estado civil que tenía el paciente al asistir a consulta de valoración.
Ocupación weka: ocupación	Categórica	Relevante	Actividad a la cual se dedica el paciente.
Entidad	Categórica	Irrelevante	Institución por la cual es remitido el paciente; resulta irrelevante ya que el 90% de los pacientes asisten por consulta privada.
Dirección Residencia	Categórica	Irrelevante	Ya que éste tipo de datos no generan resultados de minería relevantes para el negocio.
Teléfono	Categórica	Irrelevante	
e-mail	Categórica	Irrelevante	
Nombre acompañante	Categórica	Irrelevante	
Teléfono Acompañante	Categórica	Irrelevante	
Persona Responsable	Categórica	Irrelevante	

Teléfono Responsable	Catagórica	Irrelevante	
Parentesco	Catagórica	Irrelevante	
Cirugía Deseada weka: cirugía_deseada1	Catagórica	Relevante	Hacen referencia a las cirugías que el paciente manifiesta, desea realizarse.
Otra Cirugía Deseada weka: cirugía_deseada2	Catagórica	Relevante	
ANTECEDENTES PATOLOGICOS			
Tiroides	Catagórica	Irrelevante	<p>Por regla de negocio se realizó el respectivo estudio de las variables junto al cirujano plástico y se llegó a la conclusión que por definición, se presenta redundancia entre los antecedentes patológicos del paciente y los medicamentos que éste consume.</p> <p>Un ejemplo claro, es la relación que existe entre un paciente que presenta antecedentes patológicos de tiroides y consume medicamentos para el control de la misma.</p> <p>También es importante aclarar, que la presencia de alguno de éstos antecedentes patológicos, no es impedimento para la realización de una cirugía estética.</p>
Corazón	Catagórica	Irrelevante	
Sida HIV	Catagórica	Irrelevante	
Riñones	Catagórica	Irrelevante	
Vesícula	Catagórica	Irrelevante	
Artritis	Catagórica	Irrelevante	
Propensión al sangrado	Catagórica	Irrelevante	
Presión Arterial	Catagórica	Irrelevante	
Pulmones	Catagórica	Irrelevante	
Sistema Nervioso	Catagórica	Irrelevante	
Estomago	Catagórica	Irrelevante	
Diabetes	Catagórica	Irrelevante	
Asma	Catagórica	Irrelevante	
Lupus	Catagórica	Irrelevante	
Cáncer	Catagórica	Irrelevante	
Fibromialgia	Catagórica	Irrelevante	
Hígado	Catagórica	Irrelevante	
Escleroderma	Catagórica	Irrelevante	
Otras	Catagórica	Irrelevante	
MEDICAMENTOS			
Aspirina weka: MED_ASPIRINA	Catagórica	Relevante	Hace referencia al consumo de éste medicamento actualmente.

Motrin	Categórica	Redundante	Por definición, el ibuprofeno y el motrín contienen los mismos componentes, por tal razón, se presenta redundancia.
Ibuprofeno weka: MED_IBUPROFENO	Categórica	Relevante	Hace referencia al consumo de éstos medicamento actualmente.
Para Artritis weka: MED_PARA_ARTRITIS	Categórica	Relevante	
Para Dormir weka: MED_PARA_DORMIR	Categórica	Relevante	
Para la Tos weka: MED_PARA_TOS	Categórica	Relevante	
Para Tiroides weka: MED_PARA_TIROIDES	Categórica	Relevante	
Hormonas weka: MED_HORMONAS	Categórica	Relevante	
Anticonceptivos	Categórica	Redundante	Por definición, la variable presenta redundancia con el método de planificar, se asume entonces que en el método, la paciente indica si consume anticonceptivos orales.
Tretinoína weka: MED_TRETINOINA	Categórica	Relevante	Hace referencia al consumo de éste medicamento actualmente.

Antibióticos weka: MED_ANTIBIOTICOS	Categórica	Relevante	
Para la Presión weka: MED_PARA_PRESION	Categórica	Relevante	
Para la Diabetes weka: MED_PARA_DIABETES	Categórica	Relevante	
Diuréticos weka: MED_DIURETICOS	Categórica	Relevante	
Fenobarbital weka: MED_FENOBARBITAL	Categórica	Relevante	
Digitálicos weka: MED_DIGITALICOS	Categórica	Relevante	
Cortisona weka: MED_CORTISONA	Categórica	Relevante	
Para Perder peso weka: MED_PARA_PESO	Categórica	Relevante	
ANTECEDENTES GINECO			
Edad Menstruacion	Categórica	Irrelevante	Corresponde a la edad en la cual se presentó la primera menstruación y al número de días del ciclo. El cirujano plástico decidió que éstas variables no fuesen incluidas en el proceso de minería porque las considera irrelevantes frente a la realización de una cirugía estética.
Ciclo	Texto	Irrelevante	

Método de Planificación weka: PLANIFICACION	Categórica	Relevante	Hace referencia al método de planificación utilizado por el paciente e influye en los posibles componentes hormonales del método.
ANTECEDENTES FAMILIARES			
Hipertensión Arterial weka: AF_HA	Categórica	Relevante	Hacen referencia al padecimiento por parte de un familiar en primer o segundo grado de consanguineidad de las enfermedades en mención.
Enfermedad Coronaria weka: AF_EC	Categórica	Relevante	
Artritis weka: AF_ART	Categórica	Relevante	
Transtornos de Coagulación weka: AF_COAG	Categórica	Relevante	
Asma weka: AF_ASM	Categórica	Relevante	
Diabetes weka: AF_DIAB	Categórica	Relevante	
Cáncer de mama weka: AF_CAN	Categórica	Relevante	
Otro weka: AF_OTR	Categórica	Relevante	
HABITOS			
Cigarrillo weka: HAB_CIGARRILLO	Categórica	Relevante	Hace referencia al hábito de fumar cigarrillo en el paciente.
Licor weka: HAB_LICOR	Categórica	Relevante	Hace referencia al hábito de consumir licor por el paciente.
Café weka: HAB_CAFE	Categórica	Relevante	Hace referencia al hábito de consumir cafeína.
Pildoras para dormir weka: HAB_PASTIL_DORMIR	Categórica	Relevante	Hace referencia al hábito de consumir medicamentos para inducir al sueño.

Otros weka: HAB_OTROS	Categórica	Relevante	Hace referencia a la existencia de otros hábitos no saludables en el paciente.
EVALUACIÓN FÍSICA			
Tension Arterial	Texto	No apto	Corresponden a la cuantificación mediante instrumento de medición de la tensión arterial y los latidos cardiacos por minuto; por el tipo de dato; no son aptos para el proceso de minería.
Frecuencia Cardiaca	Texto	No apto	
Peso	Numérico	Redundante	Corresponden al peso y la talla tomadas por el cirujano plástico en la cita de valoración al paciente; presentan redundancia con el índice de masa corporal ya que éste es la relación directa entre peso y talla.
Talla	Numérico	Redundante	
IMC weka: IMC	Numérico	Relevante	Es el índice de masa corporal y representa la relación entre el peso y la talla.
Cara	Texto	No apto	Corresponden a la evaluación física realizada por el cirujano plástico al paciente; por el tipo de dato; no son aptos para el proceso de minería.
Cuello	Texto	No apto	
Tórax	Texto	No apto	
Abdomen	Texto	No apto	
Extremidades	Texto	No apto	
operado weka: operado	Categórica	Relevante	Indica que el paciente estuvo en consulta de valoración y se realizó o no una cirugía estética con el cirujano plástico John Jairo Valencia Quintero.

Tabla 4. Análisis de Variables.

9.6. ANÁLISIS DE CORRELACIONES.

El análisis de correlaciones se utiliza para determinar variables con contenidos redundantes o correlacionados. En la **Tabla 5** se presenta la matriz de correlaciones de variables altamente correlacionadas; es importante aclarar que la matriz original fue 101 filas por 101 columnas.

VARIABLE 1	VARIABLE 2	CORRELACIÓN
planificacion=pareja	sexo=M	0.99
cirugia_deseada2=ninguna	operado=NO	0.97
MED_CORTISONA=NO	MED_FENOBARBITAL=NO	0.77
MED_CORTISONA=NO	MED_DIGITÁLICOS=NO	0.77
MED_TRETINOINA=NO	MED_FENOBARBITAL=NO	0.71
MED_TRETINOINA=NO	MED_DIGITÁLICOS=NO	0.71

Tabla 5 . Variables con correlaciones altas (mayor a 0.7).

Por regla de negocio se define que, usualmente por como está estructurada la historia clínica, son las mujeres quienes responden a la pregunta sobre métodos de planificación ya que éstos inciden directamente sobre el sistema endocrino; siendo así, la primera y más alta correlación será ignorada para efectos de reducción de variables

Por regla de negocio además, se identificó que cuando no se deseó una segunda cirugía “cirugia_deseada2=ninguna” sí se construyó un correlación con los pacientes que no se realizaron una cirugía estética “operado=NO” ; pero ambas variables se mantienen ya que éstas son las únicas categorías correlacionadas.

Las correlaciones encontradas entre medicamentos (MED) no se tendrán en cuenta ya que la prescripción médica de cada uno de ellos es independiente del otro, es decir, el fenobarbital es usado como anticonvulsionante, la tretinoína es prescrito para tratamiento de acné, los medicamentos digitálicos se usan en el tratamiento de insuficiencias cardíacas congestivas y la cortisona es un

medicamento utilizado como antiinflamatorio para el tratamiento de la artritis; motivo entonces por el cual el uso de uno no se relaciona con el otro.

Por otra parte, respecto a las correlaciones, en la **Tabla 6**, se presentan las variables con una correlación baja con relación a la variable objetivo.

VARIABLE 1	VARIABLE 2	CORRELACIÓN
estado_civil=desconocido	operado=NO	0
ocupacion=empleado	operado=NO	0
ocupacion=oficios_varios	operado=NO	0
ocupacion=operario	operado=NO	0
ocupacion=vendedor	operado=NO	0
cirugia_deseada1=cirugía_mano	operado=NO	0
MED_PARA_ARTRITIS=NO	operado=NO	0
estado_civil=soltero	operado=NO	0.01
estado_civil=viudo	operado=NO	0.01
ocupacion=ciencias_sociales	operado=NO	0.01
ocupacion=ingeniero	operado=NO	0.01
cirugia_deseada1=ginecoplastia	operado=NO	0.01
cirugia_deseada2=cambio_implantes	operado=NO	0.01
MED_ANTIBIOTICOS=NO	operado=NO	0.01
MED_PARA_DIABETES=NO	operado=NO	0.01
AF_ART=NO	operado=NO	0.01
estado_civil=casado	operado=NO	0.02
ocupacion=área_belleza	operado=NO	0.02
ocupacion=área_financiera	operado=NO	0.02
ocupacion=docente	operado=NO	0.02
MED_ASPIRINA=NO	operado=NO	0.02
MED_HORMONAS=NO	operado=NO	0.02
MED_TRETINOINA=NO	operado=NO	0.02
MED_PARA_PRESION=NO	operado=NO	0.02
MED_CORTISONA=NO	operado=NO	0.02
planificacion=preservativo	operado=NO	0.02
AF_COAG=NO	operado=NO	0.02
AF_CAN=NO	operado=NO	0.02
ocupacion=área_administrativa	operado=NO	0.03
ocupacion=desconocida	operado=NO	0.03
cirugia_deseada1=lesión	operado=NO	0.03
cirugia_deseada1=ritidoplastia	operado=NO	0.03

MED_FENOBARBITAL=NO	operado=NO	0.03
MED_DIGITALICOS=NO	operado=NO	0.03
MED_PARA_PESO=NO	operado=NO	0.03
AF_ASM=NO	operado=NO	0.03
Edad	operado=NO	0.04
ocupacion=servicio_doméstico	operado=NO	0.04
cirugia_deseada1=aum_implantes	operado=NO	0.04
MED_PARA_DORMIR=NO	operado=NO	0.04
MED_PARA_TOS=NO	operado=NO	0.04
MED_DIURETICOS=NO	operado=NO	0.04
ocupacion=artes	operado=NO	0.05
HAB_PASTIL_DORMIR=NO	operado=NO	0.05
HAB_OTROS=NO	operado=NO	0.05
MED_IBUPROFENO=NO	operado=NO	0.06
AF_EC=NO	operado=NO	0.06
estado_civil=divorciado	operado=NO	0.07
IMC	operado=NO	0.08
HAB_CIGARRILLO=NO	operado=NO	0.08
cirugia_deseada1=liposucción	operado=NO	0.08
AF_DIAB=NO	operado=NO	0.08
HAB_CIGARRILLO=NO	operado=NO	0.08
IMC	operado=NO	0.08
HAB_LICOR=NO	operado=NO	0.11
HAB_CAFE=NO	operado=NO	0.11
HAB_LICOR=NO	operado=NO	0.11
HAB_CAFE=NO	operado=NO	0.11
AF_HA=NO	operado=NO	0.15
planificacion=ninguno	operado=NO	0.2
cirugia_deseada1=valoración_general	operado=NO	0.24

Tabla 6. Variables con correlación baja (menor a 0.3).

Se encuentran bajas correlaciones con la variable objetivo en la categoría “NO”, es decir “operado=NO” y varias categorías de las variables restantes, dichas categorías no se hacen relevantes ya que el objetivo principal del proceso de minería es conocer las variables que influyen en la categoría “SI” de la variable “operado”; sumando a ello, las correlaciones se presentan entre algunas categorías de cada variable y no en la totalidad de las mismas.

10.MODELADO Y EVALUACIÓN

El objetivo de esta fase es aplicar las diversas técnicas y algoritmos sobre el conjunto de datos para obtener la información oculta y los patrones implícitos en ellos, de esta manera, al aplicar las diferentes técnicas de minería de datos, se intentará encontrar las respuestas buscadas en la presente investigación.

Los algoritmos a aplicar para determinar los factores que más influyen en la realización de una cirugía estética son :

- Selección de factores con análisis de correlaciones
- Selección de factores con PCA
- Selección de factores con Árboles de Decisión
- Selección de factores con regresión logística
- Reglas de asociación

Se pretende, con cada uno de los algoritmos, hallar los factores más relevantes de cada uno y finalmente, realizar una votación entre todas las técnicas, revisando el número de veces que el factor fue hallado como el más relevante; así entonces, se definirán los factores más influyentes como los que aparecieron el mayor número de veces.

Adicionalmente, se realiza una agrupación de los pacientes para identificar perfiles de usuarios del cirujano John Jairo Valencia Quintero.

Premisas:

- La herramienta seleccionada para el modelado es Weka 3.8.0
- El éxito de las historias clínicas analizadas como objeto de estudio en la muestra de datos, se medirá por la realización de una cirugía estética.
- Los datos analizados se componen de 38 variables seleccionadas, numeradas en la **Tabla 7**.
- La cantidad de registros es de 2448.

VARIABLES			
1	Sexo	20	MED_DIGITALICOS
2	Edad	21	MED_CORTISONA
3	estado_civil	22	MED_PARA_PESO
4	Ocupación	23	Planificación
5	cirugia_deseada1	24	AF_HA
6	cirugia_deseada2	25	AF_EC
7	MED_ASPIRINA	26	AF_ART
8	MED_IBUPROFENO	27	AF_COAG
9	MED_PARA_ARTRITIS	28	AF_ASM
10	MED_PARA_DORMIR	29	AF_DIAB
11	MED_PARA_TOS	30	AF_CAN
12	MED_PARA_TIROIDES	31	AF_OTR
13	MED_HORMONAS	32	HAB_CIGARRILLO
14	MED_TRETINOINA	33	HAB_LICOR
15	MED_ANTIBIOTICOS	34	HAB_CAFE
16	MED_PARA_PRESION	35	HAB_PASTIL_DORMIR
17	MED_PARA_DIABETES	36	HAB_OTROS
18	MED_DIURETICOS	37	IMC
19	MED_FENOBARBITAL	38	Operado

Tabla 7. Variables para el modelamiento.

10.1. EXPERIMENTO 1: SELECCIÓN DE FACTORES CON ANÁLISIS DE CORRELACIONES

Objetivo del experimento: Analizar, mediante una correlación, las variables de historia clínica que más influyen en la realización de una cirugía estética.

Evaluación de resultados:

RANKED	VARIABLE	VALOR
1	cirugia_deseada2	0.82628
2	AF_HA	0.14981
3	HAB_LICOR	0.10978
4	HAB_CAFE	0.10828
5	IMC	0.08274
6	AF_DIAB	0.07954

7	AF_OTR	0.07823
8	HAB_CIGARRILLO	0.07627
9	Planificación	0.07005
10	cirugia_deseada1	0.06545

Tabla 8. Evaluación de Resultados - Experimento 1.

Interpretación de los resultados

De acuerdo a los resultados arrojados por el análisis de correlaciones, las 10 variables valoradas en la **Tabla 8**, juegan un papel importante dentro del modelo, ya que sus pesos son los más altos.

Se puede concluir entonces que la variable “cirugia_deseada2” tiene la mayor incidencia, un 82,628% sobre la variable estudiada, queriendo decir esto que, la segunda cirugía que el paciente manifiesta que desea realizarse en la consulta de valoración, es determinante para saber si realmente se realizará o no una cirugía. Es importante resaltar que también la variable “cirugia_deseada1” hace parte de los pesos mayores y, aunque su peso no es tan alto como el de la anterior, tiene una influencia del 6,545%.

Dentro del ranking también se encontró que variables referentes a los antecedentes familiares de los pacientes influyen sobre la variable estudiada; antecedentes de Hipertensión Arterial en un 14,981%, de Diabetes en un 7,954% y otros antecedentes con un 7,823%.

Hábitos como el de consumir licor con una influencia de 10,978%, consumir café con 10,828% y fumar con 7,627% fueron también relevantes y son evidentes desde el punto de vista médico ya que, quienes tengan dichos hábitos, deben pasar por períodos de abstinencia específicos para ser aptos para un procedimiento quirúrgico.

La influencia de la variable “IMC” (Índice de Masa Corporal) con un 8,274% es también indiscutible ya que, por lo general, muchos pacientes que desean o se

realizan liposucción y/o abdominoplastia presentan un índice de masa corporal alto.

Por último, la variable “Planificación” presentó una influencia del 7% sobre la variable estudiada.

10.2. EXPERIMENTO 2: SELECCIÓN DE FACTORES CON PCA

Objetivo del experimento: Analizar las variables de las historias clínicas que más influyen en la realización de una cirugía estética mediante PCA (*Principal Component Analysis* o Análisis de Componentes principales)

Modelado: Principal Component (PCA)

Evaluación de resultados:

RANKED	VARIABLE
1	IMC
2	cirugia_deseada1=cirugía_facial
3	cirugia_deseada1=aum_mamario
4	cirugia_deseada1=bichectomía
5	cirugia_deseada1=braquioplastia
6	cirugia_deseada1=aum_implantes
7	cirugia_deseada1=aum_glúteos
8	cirugia_deseada1=abdominoplastia
9	ocupacion=servicio_doméstico
10	ocupacion=servidor_público

Tabla 9. Evaluación de Resultados - Experimento 2.

Interpretación de los resultados:

Haciendo un top 10 sobre las variables que arrojó el método PCA en la **Tabla 9**, se encuentra que la variable más importante en este método de análisis, es el “IMC” (Índice de Masa Corporal); tiene coherencia en cuanto que, muchos de los

pacientes que se realizan cirugías estéticas corporales tienen un índice de masa corporal alto.

Las siguientes 7 variables de mayor relevancia pertenecen a la primera cirugía que el paciente manifiesta que desea realizarse; algunas categorías son: cirugía facial, aumento mamario, bichectomía, braquioplastia, aumento de implantes, aumento de glúteos y abdominoplastia; hasta el momento sobre éstas no se puede concluir nada en cuanto los siguientes algoritmos nos arrojen más datos.

Las últimas dos variables del top 10 son ocupaciones de los pacientes; servicio doméstico y servidor público; al igual que sobre la categoría anterior; no se puede concluir hasta no tener un avance; llama la atención que por estrato socio-económico, dichas ocupaciones son opuestas.

10.3. EXPERIMENTO 3: SELECCIÓN DE FACTORES CON ÁRBOLES DE DECISIÓN

Objetivo del experimento: Analizar las variables de las historias clínicas que más influyen en la realización de una cirugía estética mediante árboles de decisión.

Modelado: Árboles de decisión (J48), validación cruzada con 10 grupos y balanceo de datos para mejorar la calidad del aprendizaje.

Evaluación de resultados:

DETAILED ACCURACY BY CLASS				
TP Rate	FP Rate	Precision	Recall	Class
0,966	0,022	0,979	0,996	SI
0,978	0,004	0,996	0,978	NO
0,987	0,013	0,978	0,987	Weighted Avg.
F-Measure	MCC	ROC Area	PRC Area	Class
0,978	0,975	0,995	0,99	SI
0,978	0,975	0,995	0,995	NO

0,987	0,975	0,995	0,992	Weighted Avg.
CONFUSIÓN MATRIX				
A	b	Classified as		
1524	6	a=SI		
33	1496	b=NO		
Correctly Classified Instances		3020	98.7251 %	
Incorrectly Classified Instances		39	1.2749 %	

Tabla 10. Evaluación de Resultados - Experimento 3.

Interpretación de resultados:

Como se observa en la **Tabla 10**, en este experimento, se tiene el 99.5% de datos clasificados correctamente. El área ROC para cada clúster es superior a 0.7, lo cual indica que el resultado tiene un alto porcentaje de confiabilidad. Sobre la diagonal principal de la matriz de confusión se tiene la mayor parte de los datos. Lo cual genera un resultado muy cercano a la realidad.

De acuerdo con los resultados de este experimento de selección de factores, por medio de un árbol de decisión, se puede concluir que la realización de una cirugía estética, es influenciada en gran medida por las variables que se presentan a continuación en la **Ilustración 4**.

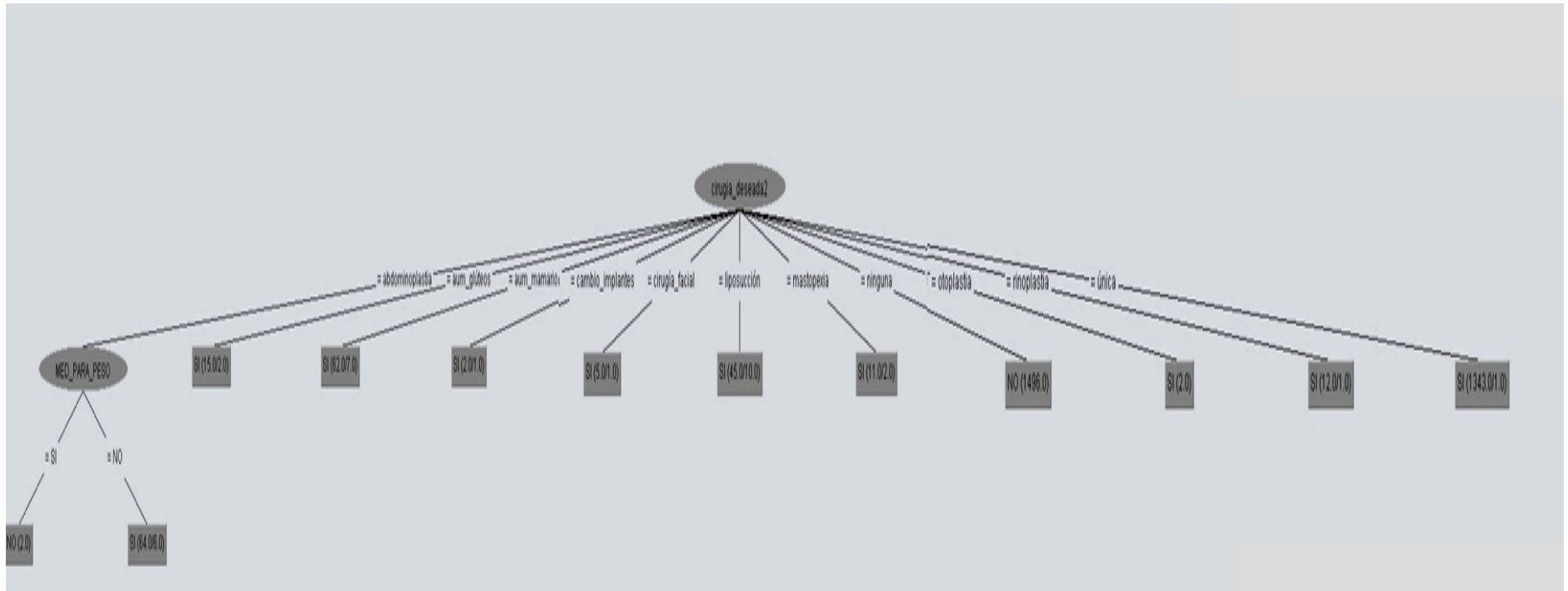


Ilustración 76: Resultado del Experimento bajo la técnica Árbol de Decisiones

RAMA	VARIABLE
1	cirugia_deseada2= ninguna (NO;1496)
2	cirugia_deseada2= única (SI;1343)
3	cirugia_deseada2=abdominoplastia; MED_PARA_PESO=SI (NO;2) MED_PARA_PESO=NO (SI;64)
4	cirugia_deseada2=aum_mamario (SI;62)
5	cirugia_deseada2=liposucción (SI;45)
6	cirugia_deseada2=aum_glúteos (SI;15)
7	cirugia_deseada2=rinoplastia (SI;12)
8	cirugia_deseada2=mastopexia (SI;11)
9	cirugia_deseada2=cirugía_facial (SI;5)
10	cirugia_deseada2=cambio_implantes (SI;2)
11	cirugia_deseada2=otoplastia (SI;2)

Tabla 11. Clasificación de las Ramas arrojadas - Experimento 3.

Rama Uno: Segunda cirugía deseada = ninguna: La primera rama, que mayor predominancia tiene en éste experimento, hace referencia a la no realización de una cirugía estética para los pacientes que en el campo “cirugía_deseada2” tienen como valor “ninguna”; esto es una obviedad según el negocio ya que, cuando un paciente no se realizaba una cirugía estética con el cirujano plástico, el campo cirugía_deseada2 era marcado como ninguna para no dejarlo nulo o vacío.

Rama Dos: Segunda cirugía deseada = única: La segunda rama, que mayor predominancia tiene también en éste experimento, hace referencia a la realización de una cirugía estética para los pacientes que en el campo “cirugía_deseada2” tienen como valor “única”; esto es un descubrimiento interesante porque quiere decir que aquellos pacientes que sólo manifestaron el deseo de realizarse una cirugía estética, finalmente se operaron.

Rama Tres: Segunda cirugía deseada = abdominoplastia y consumo de medicamentos para perder peso: La tercera rama es específica con el tipo de cirugía abdominoplastia y hace una relación directa con el consumo de

medicamentos para perder peso; indicando que los pacientes que manifiestan realizarse una abdominoplastia como segunda cirugía y no consumen medicamentos para perder peso, son propensos a realizarse alguna cirugía estética.

Rama cuatro, cinco, seis, siete, ocho, nueve, diez y once: Segunda cirugía deseada: Cada una de éstas ramas muestra la relevancia que tiene el tipo de cirugía que el paciente manifiesta como segunda opción en su deseo de realizarse alguna cirugía estética; llegando a la realización de la misma, entre ellas, de manera más común el aumento mamario y menos común, la otoplastia.

10.4. EXPERIMENTO 4: SELECCIÓN DE FACTORES CON REGRESIÓN LOGÍSTICA

Objetivo del experimento: Analizar las variables de las historias clínicas que más influyen en la realización de una cirugía estética mediante regresión logística.

Modelado: Simple logistic, con una validación cruzada de 10 grupos y balanceo de los datos.

Evaluación de resultados:

SimpleLogistic:

Class 0 :

0.19+
[edad]*0.02+
[estado_civil=desconocido]*-0.58+
[estado_civil=divorciado]*1.16+
[estado_civil=soltero]*0.47+
[estado_civil=union_libre]*-0.28+
[estado_civil=viudo]*-0.95+
[ocupacion=ama_casa]*-0.17+
[ocupacion=área_financiera]*-0.35+
[ocupacion=docente]*0.58+
[ocupacion=empleado] * 1.14 +
[ocupacion=estudiante] * -0.42 +

[ocupacion=independiente] * -0.52 +
 [ocupacion=ingeniero] * 0.55 +
 [ocupacion=oficios_varios] * -0.79 +
 [ocupacion=operario] * -0.95 +
 [cirugia_deseada1=aum_implantes] * 0.66+
 [cirugia_deseada1=cambio_implantes]*0.69+
 [cirugia_deseada1=cirugía_facial] * 0.5 +
 [cirugia_deseada1=liposucción] * -0.26 +
 [cirugia_deseada1=mastopexia]*0.81+
 [cirugia_deseada2=cambio_implantes] * -1.82 +
 [cirugia_deseada2=liposucción] * -0.58 +
 [cirugia_deseada2=mastopexia] * -0.42 +
 [cirugia_deseada2=ninguna] * -6.32+
 [cirugia_deseada2=otoplastia]*1.41+
 [cirugia_deseada2=rinoplastia] * 0.38 +
 [cirugia_deseada2=única] * 2.83 +
 [MED_PARA_ARTRITIS=NO] * 0.7 +
 [MED_PARA_TOS=NO] * 0.58 +
 [MED_PARA_TIROIDES=NO] * 1.44 +
 [MED_PARA_PRESION=NO] * -0.62 +
 [MED_PARA_PESO=NO] * 0.96 +
 [planificacion=histerectomía] *1.22+
 [planificacion=met_dispositivo] * 1.11 +
 [planificacion=met_inyectable] * 1.11 +
 [planificacion=met_oral] * -0.59 +
 [planificacion=ninguno] * -0.55+
 [planificacion=pareja] * -1.32 +
 [planificacion=preservativo] * 1.32 +
 [planificacion=vasectomía] * 1.4 +
 [AF_EC=NO] * 0.68 + [AF_ART=NO] * -0.67 +
 [AF_ASM=NO] * -0.3 +
 [AF_DIAB=NO] * 0.2 +
 [HAB_CIGARRILLO=NO] * -0.47 +
 [HAB_PASTIL_DORMIR=NO] * -0.64+
 [HAB_OTROS=NO] * -1.35

Class 1 :

-0.19 +
 [edad]*-0.02+
 [estado_civil=desconocido]*0.58+
 [estado_civil=divorciado]*1.16+
 [estado_civil=soltero]*0.47+
 [estado_civil=union_libre]*0.28+
 [estado_civil=viudo]*0.95+

[ocupacion=ama_casa]*0.17+
 [ocupacion=área_financiera]*0.35+
 [ocupacion=docente]*-0.58+
 [ocupacion=empleado]*1.14+
 [ocupacion=estudiante]*0.42+
 [ocupacion=independiente]*0.52+
 [ocupacion=ingeniero]*-0.55+
 [ocupacion=oficios_varios]*0.79+
 [ocupacion=operario]*0.95+
 [cirugia_deseada1=aum_implantes]*-0.66 +
 [cirugia_deseada1=cambio_implantes]*-0.69 +
 [cirugia_deseada1=cirugía_facial]*-0.5 +
 [cirugia_deseada1=liposucción]*0.26 +
 [cirugia_deseada1=mastopexia]*-0.81 +
 [cirugia_deseada2=cambio_implantes]*1.82 +
 [cirugia_deseada2=liposucción] *0.58 +
 [cirugia_deseada2=mastopexia]*0.42+
 [cirugia_deseada2=ninguna] *6.32+
 [cirugia_deseada2=otoplastia]*-1.41+
 [cirugia_deseada2=rinoplastia] * -0.38+
 [cirugia_deseada2=única]*-2.83+
 [MED_PARA_ARTRITIS=NO]*-0.7+
 [MED_PARA_TOS=NO]*-0.58+
 [MED_PARA_TIROIDES=NO]*-1.44+
 [MED_PARA_PRESION=NO]*0.62+
 [MED_PARA_PESO=NO]*-0.96+[
 planificacion=histerectomía]*-1.22+
 [planificacion=met_dispositivo]* -1.11+
 [planificacion=met_inyectable]*-1.11+
 [planificacion=met_oral]*0.59+
 [planificacion=ninguno]*0.55+
 [planificacion=pareja]*1.32+
 [planificacion=preservativo]*-1.32+
 [planificacion=vasectomía]*-1.4+
 [AF_EC=NO]*-0.68+
 [AF_ART=NO]*0.67+
 [AF_ASM=NO]*0.3+
 [AF_DIAB=NO]*-0.2+
 [HAB_CIGARRILLO=NO]*0.47+
 [HAB_PASTIL_DORMIR=NO]*0.64+
 [HAB_OTROS=NO]*1.35

DETAILED ACCURACY BY CLASS				
TP Rate	FP Rate	Precision	Recall	Class
0,99	0,019	0,981	0,99	SI
0,981	0,01	0,99	0,981	NO
0,986	0,014	0,986	0,986	Weighted Avg.
F-Measure	MCC	ROC Area	PRC Area	Class
0,986	0,971	0,998	0,998	SI
0,986	0,971	0,998	0,999	NO
0,986	0,971	0,998	0,998	Weighted Avg.
CONFUSIÓN MATRIX				
A	b	Classified as		
1515	15	a=SI		
29	1500	b=NO		
Correctly Classified Instances		3015	98.5616 %	
Incorrectly Classified Instances		44	1.4384 %	

Tabla 12. Evaluación de Resultados - Experimento 4.

Interpretación de resultados:

De acuerdo al resultado arrojado por el método, ver **Tabla 12**, se determina que la clase 0 tiene todas las condiciones para ser considerado el conjunto de variables con mayor posibilidad de ser exitoso. En la **Tabla 13**, se presentan las condiciones de las variables en el experimento.

CONDICIONES
Edad
estado civil= desconocido, divorciado, soltero, unión libre, viudo
ocupación= ama de casa, área financiera, docente, empleado, estudiante, independiente,ingeniero, oficios varios, operario
cirugia_deseada1=aum_implantes, cambio_implantes, cirugia facial, liposucción, mastopexia
cirugia_deseada2=cambio implantes, liposucción, mastopexia, ninguna, otoplastia, rinoplastia, única
MED_PARA_ARTRITIS=NO
MED_PARA_TOS=NO
MED_PARA_TIROIDES=NO
MED_PARA_PRESION=NO

MED_PARA_PESO=NO
planificación= histerectomía,met_dispositivo,met_inyectable,met_oral,ninguno,pareja,preservativo,vasectomía
AF_EC=NO
AF_ART=NO
AF_ASM=NO
AF_DIAB=NO
HAB_CIGARRILLO=NO
HAB_PASTIL_DORMIR=NO
HAB_OTROS=NO

Tabla 13. Condiciones de las Variables - Experimento 4.

En este experimento, se encuentra la principal influencia de algunas variables no identificadas en los algoritmos anteriores; la edad, que por descripción de los datos, logramos identificar que la mayoría de los pacientes del cirujano se encuentra entre los 20 y 35 años, así que se podría encontrar una relación directa entre ésta variable y la variable objetivo; la variable estado civil con excepción de las categorías casado y separado y la variable planificación con excepción de las categorías desconocido y tubectomía.

Aparece nuevamente la primera y segunda cirugía deseada como destacadas en éste algoritmo; además, los pacientes que no consumen medicamentos para la artritis, la tos, la tiroides, la presión y el peso; que no tienen antecedentes familiares de enfermedades coronarias, artritis, asma y diabetes y que no tienen hábitos de cigarrillo, pastillas para dormir ni ningún otro.

Se continúa a la espera de la aparición de éstas nuevas variables en los siguientes algoritmos.

10.5. EXPERIMENTO 5: SELECCIÓN DE FACTORES CON REGLAS DE ASOCIACIÓN

Objetivo del experimento: Descubrir las variables que tiene un comportamiento en común dentro del conjunto de datos de las historias clínicas de los pacientes.

Modelado: A priori

Para realizar la ejecución del algoritmo; la base de datos fue dividida en dos secciones, esto, debido a la cantidad de columnas y de registros y la no entrega de resultados por parte de Weka.

Dicha división fue realizada de la siguiente forma:

- **Parte 1 con las variables:** sexo, edad, estado_civil, ocupacion, cirugía_deseada1, cirugía_deseada2, planificación, AF_HA, AF_EC, AF_ART, AF_COAG, AF_ASM, AF_DIAB, AF_CAN, AF_OTR, HAB_CIGARRILLO, HAB_LICOR, HAB_CAFE, HAB_PASTIL_DORMIR, HAB_OTROS, IMC y operado.
- **Parte 2 con las variables:** MED_PARA_DIABETES, MED_PARA_PRESION, MED_PARA_DORMIR, MED_PARA_ARTRITIS, MED_PARA_PESO, MED_PARA_TIROIDES, MED_PARA_TOS, MED_FENOBARBITAL, MED_DIGITALICOS, MED_TRETINOINA, MED_ASPIRINA, MED_DIURETICOS, MED_ANTIIBIOTICOS, MED_HORMONAS, MED_CORTISONA, MED_IBUPROFENO y operado.

Evaluación de resultados:

A continuación, en la **Tabla 14**, se evidencian algunas reglas del algoritmo a priori:

BASE DE DATOS	REGLA
Parte 1	938. sexo=F cirugía_deseada2=ninguna AF_EC=NO AF_COAG=NO AF_OTR=NO 778 ==> operado=NO 778 conf:(1)
	942. cirugía_deseada2=ninguna AF_EC=NO AF_ART=NO HAB_CAFE=NO HAB_PASTIL_DORMIR=NO HAB_OTROS=NO 778 ==> operado=NO 778 conf:(1)

	19800. edad=juventud cirugía_deseada2=única AF_EC=NO AF_ASM=NO AF_CAN=NO HAB_PASTIL_DORMIR=NO HAB_OTROS=NO 776 ==> operado=SI 775 conf:(1)
	19994. cirugía_deseada2=única AF_HA=NO AF_EC=NO AF_ART=NO AF_COAG=NO AF_DIAB=NO HAB_CIGARRILLO=NO HAB_LICOR=NO HAB_PASTIL_DORMIR=NO HAB_OTROS=NO 775 ==> operado=SI 774 conf:(1)
Parte 2	6. MED_PARA_ARTRITIS=NO MED_PARA_TIROIDES=NO MED_TRETINOINA=NO MED_ANTIOTICOS=NO MED_PARA_DIABETES=NO MED_FENOBARBITAL=NO MED_CORTISONA=NO 2292 ==> operado=SI 1439 conf:(0.63)
	89. MED_PARA_ARTRITIS=NO MED_PARA_TIROIDES=NO MED_HORMONAS=NO MED_TRETINOINA=NO MED_ANTIOTICOS=NO MED_PARA_DIABETES=NO MED_FENOBARBITAL=NO MED_DIGITALICOS=NO 2284 ==> operado=SI 1433 conf:(0.63)
	120. MED_PARA_ARTRITIS=NO MED_PARA_TIROIDES=NO MED_HORMONAS=NO MED_ANTIOTICOS=NO MED_PARA_DIABETES=NO MED_FENOBARBITAL=NO MED_CORTISONA=NO 2286 ==> operado=SI 1434 conf:(0.63)
	19811. MED_ASPIRINA=NO MED_PARA_TIROIDES=NO MED_HORMONAS=NO MED_ANTIOTICOS=NO MED_PARA_PRESION=NO MED_PARA_DIABETES=NO MED_DIURETICOS=NO MED_FENOBARBITAL=NO MED_DIGITALICOS=NO MED_PARA_PESO=NO 2185 ==> operado=SI 1357 conf:(0.62)

Tabla 14. Evaluación de Resultados - Experimento 5.

Interpretación de resultados:

En la **primera parte**, el algoritmo entregó 20.000 reglas de asociación, se observaron algunas características específicas:

- No entregó reglas donde se evidenciara la aparición de las variables: “ocupación”, “cirugía_deseada1” y “planificación”.

- Todas las categorías relacionadas con antecedentes familiares son iguales a “NO”.
- Todas las categorías relacionadas con hábitos son iguales a “NO”.
- Todas las categorías relacionadas con sexo son iguales a “F”.
- La única categoría relacionada con IMC es “saludable”.
- La variable IMC, siempre aparece cuando el consecuente es “operado=SI”.
- Todas las reglas donde el consecuente es “operado=SI”, indican que “cirugia_deseada2=única”.
- Todas las reglas donde el consecuente es “operado=NO”, indican que “cirugia_deseada2=ninguna”.

En la **parte dos**, el algoritmo también entregó 20.000 reglas de asociación; con las características específicas:

- Todas las categorías de los medicamentos sin iguales a “NO”.
- El del consecuente siempre es “operado=SI”.

De acuerdo a los resultados arrojados por este método, las variables claves son las que se presentan a continuación en la **Tabla 15**.

CONDICIONES
Cirugia_deseada2="única"
Sexo femenino
Rango de edad = juventud
IMC saludable
No consumen medicamentos
No tienen antecedentes familiares
No tienen malos hábitos

Tabla 15. Condiciones Claves - Experimento 5.

Interpretación de los Datos:

Los resultados obtenidos en este experimento, son relativamente similares a los obtenidos mediante otros algoritmos puesto que continúa relacionando

la importancia de un paciente sano, que no consume medicamentos, no tienen antecedentes familiares ni tiene malos hábitos y considerando la segunda cirugía deseada.

10.6. EXPERIMENTO 6: PERFILAMIENTO DE LOS TIPOS DE PACIENTES CON CLUSTERING

Objetivo del experimento: Establecer los tipos de pacientes del cirujano John Jairo Valencia Quintero mediante un análisis de clustering.

Modelado: SimpleKMeans

Evaluación de resultados: Para determinar la cantidad de grupos se utilizan dos estrategias. La primera es utilizar el algoritmo EM para identificar la cantidad de clusters, lo cual arroja una cantidad de 6 grupos. Para verificar esta cantidad, se realizan varios experimentos con diferente cantidad de grupos para identificar la agrupación con menor cohesión, es decir con grupos más compactos. Se obtuvo lo siguiente:

Cantidad de clusters: 4
Cohesión: 8.405,8

Cantidad de clusters: 5
Cohesión: 7.749,7

Cantidad de clusters: 6
Cohesión: 7.665,3

Cantidad de clusters: 7
Cohesión: 7.402,1

Cantidad de clusters: 8
Cohesión: 7.508

De acuerdo a lo anterior se concluye que la mejor agrupación se obtiene con 7 clústers ya que crean grupos más compactos. A continuación, se presenta la cantidad de datos, porcentajes y el perfilamiento de los clústers:

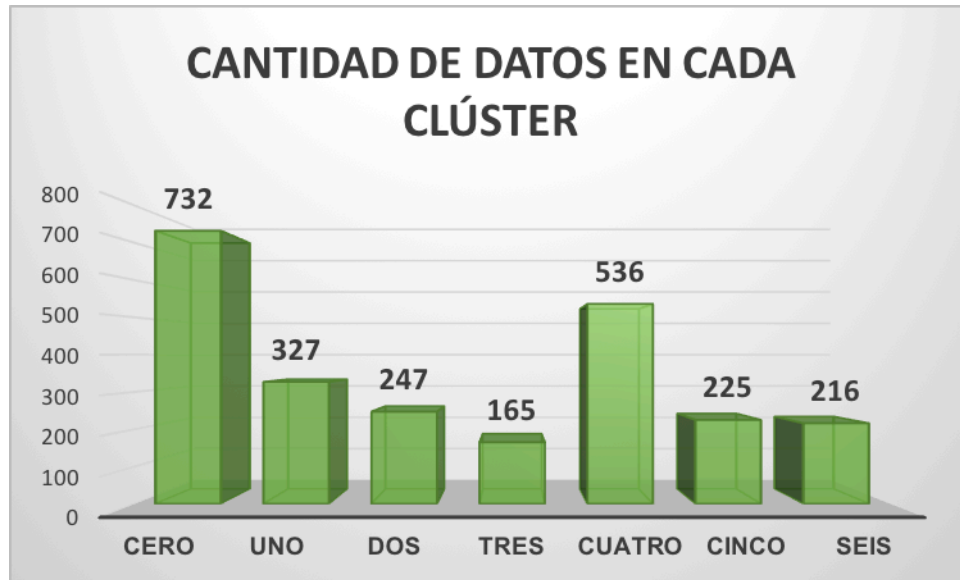


Ilustración 77. Cantidad de Datos por Clúster.

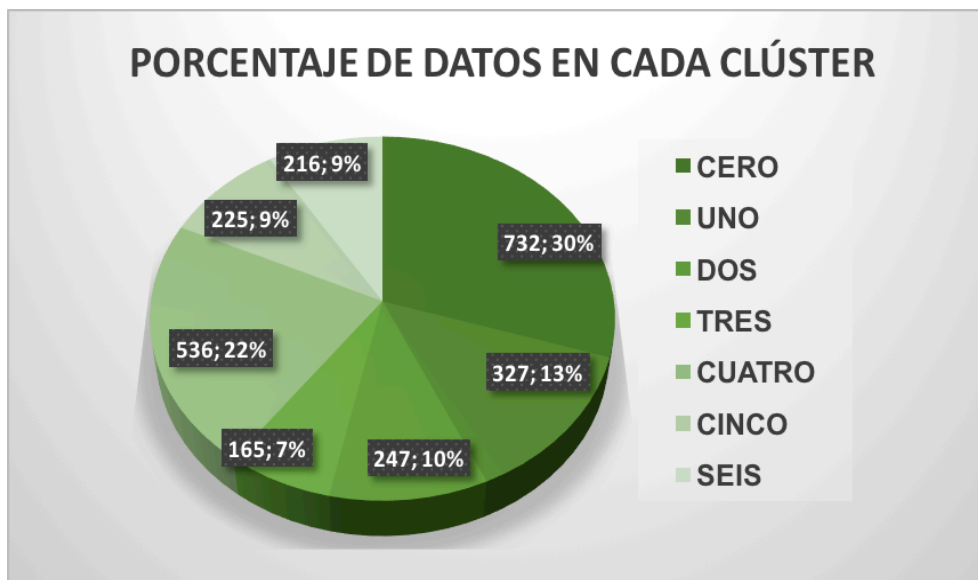


Ilustración 78. Porcentaje de Datos por Clúster.

Nro.CLÚSTER	OPERADO	DESCRIPCIÓN DEL CLÚSTER
0	SI	Pacientes de sexo femenino, de estado civil soltero, de ocupación estudiante, con deseo de realizarse sólo aumento mamario, método de planificación desconocido, que no consumen ningún medicamento, no tienen antecedentes familiares y no tienen ningún hábito.
1	NO	Pacientes de sexo femenino, de estado civil soltero, de ocupación desconocida, con deseo de realizarse sólo aumento mamario, ningún método de planificación, que no consumen ningún medicamento, no tienen antecedentes familiares y no tienen ningún hábito.
2	NO	Pacientes de sexo femenino, de estado civil soltero, de ocupación ama de casa, con deseo de realizarse sólo liposucción, método de planificación desconocido, que no consumen ningún medicamento, no tienen antecedentes familiares y no tienen ningún hábito.
3	NO	Pacientes de sexo femenino, de estado civil casado, de ocupación en área administrativa, con deseo de realizarse sólo liposucción, método de planificación desconocido, que no consumen ningún medicamento y no tienen ningún hábito.
4	SI	Pacientes de sexo femenino, de estado civil casado, de ocupación ama de casa, con deseo de realizarse sólo abdominoplastia, método de planificación desconocido, que no consumen ningún medicamento, no tienen antecedentes familiares y no tienen ningún hábito.

5	NO	Pacientes de sexo femenino, de estado civil casado, de ocupación ama de casa, con deseo de realizarse sólo aumento mamario, método de planificación desconocido, que no consumen ningún medicamento, no tienen antecedentes familiares y no tienen ningún hábito.
6	SI	Pacientes de sexo femenino, de estado civil soltero, de ocupación vendedor, con deseo de realizarse sólo liposucción, método de planificación desconocido, que no consumen ningún medicamento, con antecedentes familiares de hipertensión arterial y diabetes y no tienen ningún hábito.

Tabla 16. Evaluación de Resultados - Experimento 6.

Interpretación de los resultados

Según los resultados de la evaluación presentados en la **Tabla 16**, se puede interpretar que los pacientes con más factibilidad de realizarse una cirugía estética con el cirujano John Jairo Valencia Quintero son:

- Pacientes de sexo femenino.
- Pacientes de estado civil soltero y casado
- Pacientes interesados en aumento mamario, liposucción o abdominoplastia.

11.DESPLIEGUE

Se revisan entonces en conjunto, los resultados entregados por cada algoritmo y se evalúa si con dichos resultados es posible llegar a una conclusión para definir la priorización de pacientes para la asignación de citas de valoración del cirujano plástico John Jairo Valencia Quintero.

11.1. FACTORES MÁS INFLUYENTES EN LA REALIZACIÓN DE UNA CIRUGÍA ESTÉTICA

A continuación, en la **Tabla 17**, se presentarán los resultados obtenidos con cada uno de los experimentos, esto con la finalidad de encontrar algunos patrones comunes que permiten a la vez concluir cuáles son los factores que más influencia tienen en la realización de una cirugía estética en pacientes que asisten a cita de valoración con el cirujano plástico John Jairo Valencia Quintero.

VARIABLE	EXPERIMENTOS DE MINERÍA DE DATOS					
	1. Corre/	2. PCA	3. Árb de Dec	4. Regr Logís	5. Asoc	Nro. VECES
Sexo					X	1
Edad				X	X	2
estado_civil				X		1
ocupacion		X		X		2
cirugia_deseada1	X	X		X		3
cirugia_deseada2	X		X	X	X	4
MED_ASPIRINA					X	1
MED_IBUPROFENO					X	1
MED_PARA_ARTRITIS				X	X	2
MED_PARA_DORMIR					X	1
MED_PARA_TOS				X	X	2
MED_PARA_TIROIDES				X	X	2
MED_HORMONAS					X	1

MED_TRETINOINA					X	1
MED_ANTIOTICOS					X	1
MED_PARA_PRESION				X	X	2
MED_PARA_DIABETES					X	1
MED_DIURETICOS					X	1
MED_FENOBARBITAL					X	1
MED_DIGITALICOS					X	1
MED_CORTISONA					X	1
MED_PARA_PESO				X	X	2
planificacion	X			X		2
AF_HA	X				X	2
AF_EC				X	X	2
AF_ART				X	X	2
AF_COAG					X	1
AF_ASM				X	X	2
AF_DIAB	X			X	X	3
AF_CAN					X	1
AF_OTR	X				X	2
HAB_CIGARRILLO	X			X	X	3
HAB_LICOR	X				X	2
HAB_CAFE	X				X	2
HAB_PASTIL_DORMIR				X	X	2
HAB_OTROS				X	X	2
IMC	X	X			X	3

Tabla 17. Resultado Final.

A continuación se presentan en su orden de importancia, los cinco factores que, según los resultados arrojados por los experimentos, son los que más influyen en la realización de cirugías estéticas:

- **Factor Uno: Segunda cirugía deseada:** Se concluye que éste factor es el más relevante en cuanto a la decisión de realizarse una cirugía estética. Son varias reglas de negocio que se podrían tener en cuenta y que se deben evaluar; por ejemplo:
 - Muchos pacientes manifiestan su deseo de realizarse dos procedimientos en una sólo cirugía, por disminuir costos;

presenta menos costos una sola que cirugía que dos (gastos de quirófano, anestesia, instrumentación y honorarios médicos).

- En la base de datos, cuando la segunda cirugía es igual a “única”, quiere decir que el paciente se realizó sólo la primera cirugía y cuando es igual a “ninguna”, quiere decir que el paciente no se operó.

- **Factor Dos: Primera cirugía deseada:** Se determina importante entonces conocer qué tipo de cirugía motiva al paciente a asistir a una cita de valoración; pues, según el análisis realizado mediante minería de datos, las citas de valoración que finalmente terminan en una cirugía son: las cirugías faciales, el aumento mamario, la bichectomía, el aumento de implantes, el aumento de glúteos y la abdominoplastia.
- **Factor Tres: Índice de Masa Corporal:** Éste factor es relevante y obvio según la regla de negocio ya que, para que un paciente médicamente pueda realizarse una cirugía estética debe tener un índice de masa corporal adecuado; es decir, no deberá presentar obesidad ni deberá estar por debajo del peso correspondiente a su talla; esto, por los riesgos que generan ambas condiciones para un procedimiento estético.
- **Factor Cuatro: Hábitos de Cigarrillo:** Es un factor importante pero también obvio ya que el cirujano plástico John Jairo Valencia Quintero, exige a los pacientes que tienen el hábito de fumar, abandonar dicho hábito por lo menos dos meses antes del procedimiento estético ya que fumar cigarrillo afecta el proceso de oxigenación de los tejidos y por ende, el proceso de cicatrización
- **Factor cinco: Antecedentes Familiares de Diabetes:** Los antecedentes familiares pueden ser importantes ya que aumentan el riesgo de que el paciente sufra una enfermedad específica; llama la

atención precisamente éste antecedente ya que con el paciente diabético se aumenta el riesgo de complicaciones en cirugía porque, en éste caso, el paciente presenta aporte de nutrientes y de oxígeno menor a los tejidos en reparación , además, éste tipo de paciente puede tener daños en otros órganos internos asociados a su enfermedad.

11.2. SUGERENCIAS PARA EL CIRUJANO PLÁSTICO JOHN JAIRO VALENCIA QUINTERO

Teniendo en cuenta los resultados obtenidos en este estudio y en cumplimiento del último objetivo propuesto en el anteproyecto, sobre la definición de una estrategia para la priorización de asignación de citas de valoración, se realizan las siguientes recomendaciones y sugerencias para el cirujano plástico John Jairo Valencia Quintero:

- Para el proceso de minería de datos se requiere una base de datos limpia, estructurada y normalizada de modo que, errores humanos no puedan comprometer la veracidad de los modelos, además, para la obtención de resultados coherentes es necesario el estricto conocimiento de la regla de negocio que permita la inicial identificación de correlaciones entre variables y redundancias; es por esto que se hace necesario definir desde el inicio nombres homogéneos para campos como ocupación, cirugías, planificación, entre otros.
- Ya que las citas de valoración son asignadas de manera telefónica a través de la secretaria del cirujano plástico, se sugiere compartir los resultados del presente análisis.

- Crear un modelo de priorización de citas, basado en las variables obtenidas como relevantes en el presente estudio que tenga en cuenta los cinco factores para asignar las citas de valoración; dicho modelo puede implementarse haciendo preguntas básicas mediante vía telefónica al paciente, tales como:
 - ¿Qué cirugías desea realizarse?
 - ¿Cuál es su peso y su talla? (Para hallar el IMC)
 - ¿Fuma usted cigarrillo?
 - ¿Hay en su familia antecedentes de diabetes?

Una vez obtenidas dichas respuestas, identificar, según el modelo, la priorización que debe darse al paciente para asignar la consulta de valoración.

- Implementar en su historia clínica campos relevantes como: nivel de escolaridad, nacionalidad, estrato socio-económico y número de hijos; esto, con la intención de contar con más factores que en un futuro puedan ser estudiados y que entreguen resultados más amplios.
- Para identificar con mayor precisión la relevancia que tienen los antecedentes familiares de diabetes, sería importante añadir en la historia clínica, la persona que remite o que referencia al paciente donde el cirujano plástico y el nivel de consanguinidad que se presenta entre ellos.
- Revisar los 7 clúster entregados sobre el modelo de perfilamiento de sus pacientes y logre así mayor conocimiento sobre los mismos.

PARTE 3. CONCLUSIONES Y TRABAJO FUTURO

12. CONCLUSIONES

Al analizar los resultados de los distintos experimentos en su conjunto, se destacan algunos factores que contribuyen a la realización de una cirugía estética con el médico John Jairo Valencia Quintero. Conocer éstos factores, son un punto de inicio para que, la rama de la cirugía plástica, identifique a los pacientes más propensos a someterse a un procedimiento estético una vez han asistido a una cita de valoración y lograr así, priorizar en su agenda a dichos pacientes. En éste sentido, los principales hallazgos del estudio son:

- Las cirugías que el paciente manifiesta que desea realizarse, son en primera opción, los factores más relevantes; esto tiene sentido en cuanto el paciente que asiste a una valoración tiene claro su deseo e intención de realizarse un procedimiento específico y el papel del cirujano plástico es reforzar la confianza del paciente brindando la información adecuada; además, es importante, tener en cuenta el factor social y cultural donde, muchas veces, las cirugías plásticas son motivadas por las necesidades del medio.
- El factor índice de masa corporal, que es la relación entre el peso y la estatura, es el segundo más relevante a la hora de un paciente decidir realizarse un procedimiento estético, esto es sobresaliente en la medida que un índice de masa corporal por fuera del rango de normalidad está asociada a mayor morbilidad en pacientes quirúrgicos, es decir, con índices de masa corporal por debajo del adecuado o con sobrepesos altos, se aumentan los riesgo de complicaciones operatorias y post operatorias.
- Otro factor importante, asociado al hábito del paciente que consume cigarrillo, es relevante en el estudio dadas las implicaciones que tiene

éste hábito en el proceso quirúrgico y post operatorio; el cigarrillo produce poca oxigenación en la sangre e influye negativamente en los procesos de cicatrización, es importante recalcar además, que éste hábito genera problemas a nivel sistémico, es decir, a nivel pulmonar, cardiaco y cerebral. La falta de cicatrización en un procedimiento estético puede generar deformidades a nivel cutáneo y también aumenta el riesgo de contraer infecciones .

- Se halló también la correlación entre la realización de un procedimiento estético y los antecedentes familiares de diabetes; dicha correlación no es clara dado que un paciente diabético bien controlado médicamente puede someterse a un procedimiento estético; dicha correlación merece estudiarse desde el área de la salud para determinar qué factores pueden afectar quirúrgicamente a un paciente con antecedentes diabéticos y así, quizá, definir medidas de prevención.
- Factores asociados a otro tipo de hábitos como el café, el licor o las pastillas para dormir también hicieron parte de las variables que influyen frente a la realización de un procedimiento estético; médicamente resultan ser importantes desde la evaluación comportamental que el cirujano hace al paciente; es decir, éste tipo de hábitos pueden estar asociados a conductas de dependencia no propias para un paciente quirúrgicos.

13. TRABAJOS FUTUROS

A continuación, se presentan algunas sugerencias de trabajo futuro para el cirujano plástico John Jairo Valencia Quintero, a partir de los resultados de la minería:

- Al conocer que los factores más influyentes en la realización de una cirugía estética, los cuales son: las cirugías que el paciente manifiesta que desea realizarse, el índice de masa corporal y hábitos como el cigarrillo, sumando a los antecedentes familiares de diabetes; se recomienda comenzar a realizar campañas informativas, a través de plegables en el consultorio, página web y páginas de redes sociales sobre la importancia de mantener un peso saludable antes de un procedimiento estético, lo relevante sobre no tener el hábito de fumar y los riesgos que éste implica sobre el proceso de cicatrización y, en caso de tener antecedentes familiares de diabetes; lo significativo de realizarse exámenes de diagnóstico de forma periódica.
- Dado el hallazgo anterior, y que médicamente no se encuentra una razón lógica para que los antecedentes familiares de diabetes influyan en la realización de una cirugía estética, se recomienda a la rama de cirugía plástica, hacer análisis y estudios más profundos sobre la real importancia de éste diagnóstico familiar en un paciente que desea someterse a una cirugía estética.
- Sobre la premisa anterior, es importante que el cirujano plástico John Jairo Valencia Quintero, tenga en cuenta que, ya que muchos de sus pacientes son referidos y suelen, por lo general, ser familiares, evalúe, en dichas historias clínicas específicas, los antecedentes familiares sobre diabetes y logre hallar una relación entre éstos.

- Se presenta además, dado que los datos ya están homogéneos y tiene calidad de datos, la posibilidad de realizar minería de datos, mediante análisis predictivo, a las historias clínicas de los pacientes que se encuentran en proceso de valoración con el fin de predecir si se realizarán una cirugía estética.

A continuación, se presentan trabajos futuros relacionados con los datos:

- Dado que los datos con los cuales se realizó el proceso de minería fueron digitados de manera exclusiva para el estudio, se sugiere realizar un control de calidad a dicha base de datos para revisión de historias clínicas duplicadas, datos digitados en columnas no correspondientes y datos atípicos.
- Se sugiere al cirujano plástico John Jairo Valencia Quintero, emplear términos homogéneos para las columnas, esto, con la finalidad de no generar duplicidad; por ejemplo, la columna ocupación contenía registros como: maestra, docente, licenciada, profesora; generando así, redundancia en las categorías.

REFERENCIAS

Arriagadas, J., & Ortiz, A. (2010). *ALGUNAS REFLEXIONES ETICAS SOBRE LA CIRUGÍA PLÁSTICA*. COMITÉ DE ÉTICA, CLÍNICA LAS CONDES, DEPARTAMENTO DE CIRUGÍA PLÁSTICA, CLÍNICA LAS CONDES, Las Condes, Chile.

Congreso De Colombia. (18 de FEBRERO de 1981). LEY 23 DE 1981. *Por la cual se dictan normas en materia de ética médica* . COLOMBIA.

Ministerio de La Salud. (08 de JULIO de 1999). RESOLUCIÓN NÚMERO 1995 DE 1999. *por la cual se establecen normas para el manejo de la Historia Clínica* . COLOMBIA.

Ministerio de La Protección Social. (11 de JULIO de 2007). RESOLUCIÓN NÚMERO 2346 DE 2007. *Por la cual se regula la práctica de evaluaciones médicas ocupacionales y el manejo y contenido de las historias clínicas ocupacionales* . COLOMBIA.

Congreso de Colombia. (17 de OCTUBRE de 2012). LEY ESTATUTARIA No.1581. *POR LA CUAL SE DICTAN DISPOSICIONES GENERALES PARA LA PROTECCIÓN DE DATOS PERSONALES* . COLOMBIA.

Arias, J., González, J., García, Á., & González, R. (2008). *RINOPLASTIA*. Sociedad Española de Cirugía Oral y Maxilofacial, Protocolos Clínicos.

Reparaz, D., Merlino, H., Rancan, C., Rodríguez, D., Britos, P., & García Martínez, R. (2008). *DETERMINACION DE LA EFICACIA DE LA BRAQUITERAPIA EN TRATAMIENTO DE CÁNCER BASADA EN MINERÍA DE DATOS*. Universidad de Buenos Aires., Facultad de Ingeniería, Buenos Aires.

Arbelaéz Rodríguez, D. (2013). *LA CIRUGÍA PLÁSTICA EN COLOMBIA: ¿UNA OBLIGACIÓN DE MEDIOS O DE RESULTADO?*

Abreu de la Torre, C., Tumer, O. F., González, R. A., Macias Betancourt, R., & Leal, E. (2000). *PERFIL DE PERSONALIDAD EN PACIENTES QUE SOLICITAN CIRUGÍA ESTÉTICA*. *CIRUGÍA PLÁSTICA* , 10 (3), 97-101.

Acosta, R., Rosete, A., & Rodríguez, A. (2012). *PREDICCIÓN DE PACIENTES DIABÉTICOS. PREPROCESADO PARA MINERÍA DE DATOS*. Ciudad Universitaria José Antonio Echeverría (CUJAE), Facultad de Ingeniería Informática, La Habana.

Aponte, E. J. (2007). *LA IMPORTANCIA DE LA PROTECCIÓN DE DATOS DE CARÁCTER PERSONAL EN LAS RELACIONES COMERCIALES*. Maracaibo, Venezuela.

Bermúdez, E. A. (2010). *ESTRUCTURACIÓN PSÍQUICA EN MUJERES CON MÚLTIPLES CIRUGÍAS ESTÉTICAS*. *Revista de Psicoanálisis, Psicoterapia y Salud Mental* , 3 (7).

Bustamante, G., Chui, M. R., Coca, I., & Flóres, B. (2011). PRE Y POSTOPERATORIO. *Revista de Actualización Clínica*, 15, 856-860.

Calderón, L. F. (2014). *FUNCIÓN PSICOSOCIAL DE LA CIRUGÍA ESTÉTICA: EL CASO DE CINCO MUJERES EN LA CIUDAD DE CALI*. UNIVERSIDAD DE SAN BUENAVENTURA-CALI, FACULTAD DE PSICOLOGÍA, Cali.

Coll, A. R. (1989). COSMETIC SURGERY. 195-199.

Dasso, A., & Funes, A. (2012). *AGREGACIÓN DE MÉTRICAS DE MINERÍA DE DATOS USANDO FUNCIONES DE LÓGICA CONTÍNUA*. Recuperado el 12 de Marzo de 2017, de http://sedici.unlp.edu.ar/bitstream/handle/10915/18896/Documento_completo.pdf?sequence=1

Formia, S. A. (2012). *EVALUACIÓN DE TÉCNICAS DE EXTRACCIÓN DE CONOCIMIENTO EN BASES DE DATOS Y SU APLICACIÓN EN LA DESERCIÓN DE ALUMNOS UNIVERSITARIOS*. Recuperado el 12 de Marzo de 2017, de http://sedici.unlp.edu.ar/bitstream/handle/10915/26772/Documento_completo.pdf?sequence=1

Franco, Á. M. (2014). *MODELO PARA ANÁLISIS DE RIESGO DE LA DIABETES MELLITUS 2 USANDO INTELIGENCIA DE NEGOCIOS Y MINERÍA DE DATOS*. Tesis, Universidad Nacional de Colombia, Cundinamarca, Bogotá.

IBM. (2012). *MANUAL CRISP_DM DE IBM SPSS MODELER*. EEUU.

ISAPS International Survey on Aesthetic/Cosmetic. (2014). PROCEDURES PERFORMED IN 2014.

Jiménez, M. I., & Mateos, R. (2013). PERFILES DE PACIENTES DIABÉTICOS QUE SUFREN REACCIONES ADVERSAS A MEDICAMENTOS. IDENTIFICACIÓN Y CARACTERIZACIÓN A TRAVÉS DE MINERÍA DE DATOS. *An. Real Acad. Farm.*, 80 (2), 274-321.

Lizasoain, L., & Joaristi, L. (2012). LAS NUEVAS TECNOLOGÍAS Y LA INVESTIGACIÓN EDUCATIVA. EL ANÁLISIS DE VARIABLES CATEGORIALES. *revista española de pedagogía* (251), 111-130.

Manrique, D. (2005). *SELECCIÓN DE PACIENTES Y VALORACIÓN PREOPERATORIA-PAPEL DEL CIRUJANO*. Distribuna, Bogotá.

Olivares, M. E. (2007). CIRUGÍA MAMARIA: ASPECTOS PSICOLÓGICOS. *PSICOONCOLOGÍA*, 4 (2-3), 447-464.

Moreno, H., & Zambrano, Y. (2013). *AUTOESTIMA Y DISTORSIÓN DE LA IMÁGEN CORPORAL EN MUJERES CON MAMOPLASTIA Y/O GLUTEOPLASTIA DE AUMENTO*. UNIVERSIDAD RAFAEL

URDANETA, FACULTAD DE CIENCIAS POLÍTICAS, ADMINISTRATIVAS Y SOCIALES. Maracaibo: Irma Peña.

Palma, S., & Solís, L. (2009). *ESTUDIO DE LAS REPRESENTACIONES SOCIALES DEL CUERPO DE MUJERES QUE SE HAN REALIZADO ALGÚN PROCEDIMIENTO DE CIRUGÍA ESTÉTICA*.

Universidad Academia de Humanismo Cristiano, Escuela de Psicología, Santiago.

Pautsch, J. G. (2009). *MINERÍA DE DATOS APLICADA AL ANÁLISIS DE DESERCIÓN EN LA CARRERA DE ANALISTA EN SISTEMAS DE COMPUTACIÓN*. Recuperado el 12 de Marzo de 2017, de

<http://exa.exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/TFAGermanPAUTSCHFinal.pdf>

Pezoa, M. (2013). *SATISFACCIÓN Y CALIDAD PERCIBIDA EN LA ATENCIÓN DE SALUD HOSPITALARIA: RANKING DE PRESTADORES INFORME GLOBAL DEPARTAMENTO*.

Superintendencia de Salud, Departamento de Estudios y Desarrollo, Chile.

Piatetsky, G. (Mayo de 2015). WHAT ANALYTICS, BIG DATA, DATA MINING, DATA SCIENCE SOFTWARE YOU USED IN THE PAST MONTHS FOR A REAL PROJECT? *R leads RapidMiner, Python catches up, Big Data tools grow, Spark ignites*, págs.

<http://www.kdnuggets.com/2015/05/poll-r-rapidminer-python-big-data-spark.html>.

Pérez, C., & Santin, D. (2007). *MINERÍA DE DATOS, TÉCNICAS Y HERRAMIENTAS*. Paraninfo.

Rocheffort, G., Berner, J., Castillo, P., & Vidal, H. P. (2014). IMPORTANCIA DE LA DETECCIÓN DEL TRASTORNO DISMÓRFICO CORPORAL EN LA CONSULTA DE CIRUGÍA ESTÉTICA. *Cirugía Plástica Ibero-latinoamericana*, 40 (3), 253-259.

Rodríguez, F., & Vallejo, N. (2012). *APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS PARA EL DIAGNÓSTICO PREMATURO DE CÁNCER*. Universidad Carlos III, Madrid, Madrid.

Sposito, O. M., Echeverry, M. E., Ryckeboer, H. L., & Bossero, J. (2010). *APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS PARA LA EVALUACIÓN DEL RENDIMIENTO ACADÉMICO Y LA DESERCIÓN ESTUDIANTIL*. Universidad Nacional de La Matanza, Departamento de Ingeniería e Investigaciones Tecnológicas, Buenos Aires.

Timarán, R., & Yépez, M. C. (2012). LA MINERÍA DE DATOS APLICADA AL DESCUBRIMIENTO DE PATRONES DE SUPERVIVENCIA EN MUJERES CON CÁNCER INVASIVO DE CUELLO UTERINO. *Universidad y Salud*, 14 (2), 117-129.

Valenga, F., Fernández, E., Merlino, H., Rodríguez, D., Procopio, C., Britos, P., y otros. (2007). *MINERÍA DE DATOS APLICADA A LA DETECCIÓN DE PATRONES DELICTIVOS EN ARGENTINA*. Universidad de Buenos Aires, Facultad de Ingenierías., Buenos Aires.

LAS AMÉRICAS, C. (s.f.). *CLÍNICA LAS AMÉRICAS*. Obtenido de <http://clinicalasamericas.lasamericas.com.co/>

Ministerio de La Salud. (05 de 08 de 1994). RESOLUCION NUMERO 5261 DE 1994.

Azoumana, K. (2013). ANÁLISIS DE LA DESERCIÓN ESTUDIANTIL EN LA UNIVERSIDAD SIMÓN BOLIVAR, FACULTAD INGENIERÍA DE SISTEMAS, CON TÉCNICAS DE MINERÍA DE DATOS. *Pensamiento Americano* , 41-51.

Elliott, A. (2008). *MAKING THE CUT: HOW COSMETIC SURGERY IS TRANSFORMING OUR LIVES*. Londres.