



UNIVERSIDAD DE GUAYAQUIL

**FACULTAD DE CIENCIAS MATEMÁTICAS Y FÍSICAS
CARRERA DE INGENIERÍA EN SISTEMAS
COMPUTACIONALES**

**DESARROLLO Y EXPLOTACIÓN DE INFORMACIÓN
DE UNA BASE DE DATOS DE PATOLOGÍAS
PRESENTES EN LA CAVIDAD BUCAL CON
ÉNFASIS EN CARIES Y GINGIVITIS
A TRAVÉS DE HERRAMIENTAS
ORIENTADAS A MINERÍA DE
DATOS PARA LA FACULTAD
PILOTO DE ODONTOLOGÍA
DE LA UNIVERSIDAD
DE GUAYAQUIL**

PROYECTO DE TITULACIÓN

Previa a la obtención del Título de:

INGENIERO EN SISTEMAS COMPUTACIONALES

AUTOR:

ASCENCIO MORENO DANNY JONATHAN

TUTOR:

ING. LORENZO CEVALLOS TORRES, M. Sc

**GUAYAQUIL – ECUADOR
2020**



Presidencia
de la República
del Ecuador



Plan Nacional
de Ciencia, Tecnología,
Innovación y Saberes



SENESCYT
Secretaría Nacional de Educación Superior,
Ciencia, Tecnología e Innovación

REPOSITORIO NACIONAL EN CIENCIAS Y TECNOLOGÍA

FICHA DE REGISTRO DE TESIS

TÍTULO: DESARROLLO Y EXPLOTACIÓN DE INFORMACIÓN DE UNA BASE DE DATOS DE PATOLOGÍAS PRESENTES EN LA CAVIDAD BUCAL CON ÉNFASIS EN CARIES Y GINGIVITIS A TRAVÉS DE HERRAMIENTAS ORIENTADAS A MINERÍA DE DATOS PARA LA FACULTAD PILOTO DE ODONTOLOGÍA DE LA UNIVERSIDAD DE GUAYAQUIL

AUTOR: ASCENCIO MORENO
DANNY JONATHAN

REVISOR:
ING. JUAN AGUSTÍN SÁNCHEZ, MGS.

INSTITUCIÓN:
UNIVERSIDAD DE GUAYAQUIL

FACULTAD: FACULTAD DE CIENCIAS
MATEMÁTICAS Y FÍSICAS

CARRERA: INGENIERÍA EN SISTEMAS COMPUTACIONALES

FECHA DE PUBLICACIÓN:

Nº DE PÁGS:

ÁREA TEMÁTICA: INVESTIGATIVA, INFORMÁTICA

PALABRAS CLAVES: MACHINE LEARNING, WEKA, REDES BAYESIANAS, PREDICCIÓN, PATOLOGÍAS BUCALES

RESUMEN:

A partir de una base de datos provista por la FPO, adecuada mediante filtrados de datos, para poder aplicar técnicas de MACHINE LEARNING, precisamente redes bayesianas, se logró conocer el nivel de incidencia de algunas enfermedades que presentan los pacientes con respecto a dos de las patologías bucales de mayor relevancia: caries y gingivitis. La base obtenida fue procesada a través de la herramienta digital de minería de datos WEKA. La aplicación es de tipo software libre y dispone de una serie de herramientas que facilitan el manejo de los. El algoritmo empleado genera un gráfico que facilita la interpretación de la información. Los resultados observados por el modelo evidencian la situación actual de forma general de los pacientes permitiendo predecir con un grado moderado de certeza (68%) el comportamiento de las variables y generar un modelo predictivo, de acuerdo con los datos obtenidos y su correlación. Al estudiar el modelo generado se evidenció, la relación de las variables en la que hace énfasis el presente estudio, con respecto a las variables que el modelo señalo como principales implicadas en el desarrollo de estas patologías.

Nº DE REGISTRO:

Nº DE CLASIFICACIÓN:

ADJUNTO PDF

SI ☒

NO ☐

CONTACTO CON AUTOR:
ASCENCIO MORENO DANNY
JONATHAN

TELÉFONO: 0994798634
E-MAIL:
danny.ascenciom@ug.edu.ec

CONTACTO DE LA INSTITUCIÓN:
UNIVERSIDAD DE GUAYAQUIL,
CARRERA DE INGENIERÍA EN
SISTEMAS COMPUTACIONALES
DIRECCIÓN: VICTOR MANUEL
RENDÓN 429 Y BAQUERIZO
MORENO, GUAYAQUIL

NOMBRE:
AB. JUAN CHÁVEZ ATOCHA, ESP.

TELÉFONO:
2-307729

APROBACIÓN DEL TUTOR

En mi calidad de Tutor del trabajo de investigación, “Desarrollo y explotación de información de una base de datos de patologías presentes en la cavidad bucal con énfasis en caries y gingivitis a través de herramientas orientadas a minería de datos para la facultad piloto de odontología de la universidad de guayaquil elaborado por el Sr. Ascencio Moreno Danny Jonathan, de la Carrera de Ingeniería en Sistemas Computacionales, Facultad de Ciencias Matemáticas y Físicas de la Universidad de Guayaquil, previo a la obtención del Título de Ingeniero en Sistemas Computacionales, me permito declarar que luego de haber orientado, estudiado y revisado, la apruebo en todas sus partes.

Atentamente,

Ing. Lorenzo Cevallos Torres, M. Sc.
TUTOR

DEDICATORIA

A mis padres, ya que de no ser por ellos no hubiese podido concluir con esta última etapa. Por sus grandes enseñanzas, cariño y amor incondicional, por saber ser padres y no hacer nunca distinciones ni comparaciones y estar siempre ahí cuando más se los necesitó, los mejores padres del mundo.

A mi hermano José por la ayuda en primeras filas y que siempre mantiene los pies en la tierra a pesar de gozar de gran conocimiento, a mi hermano Javier y Johnny por el apoyo incondicional, por siempre mantener la unión familiar en todo momento.

AGRADECIMIENTO

A mi Dios por la gran familia
con la que me ha
bendecido.

A los maestros que gozan
de paciencia, aman enseñar
e impartir sus
conocimientos.

A los docentes de la carrera
que supieron guiarnos
hasta la meta.

TRIBUNAL DEL PROYECTO DE TITULACIÓN

Ing. Fausto Cabrera Montes, M.Sc.
DECANO DE LA FACULTAD
CIENCIAS MATEMÁTICAS Y FÍSICAS

Ing. Gary Reyes Zambrano, Mgs.
DIRECTOR DE LA CARRERA DE
INGENIERÍA EN SISTEMAS
COMPUTACIONALES

Ing. Lorenzo Cevallos Torres, M. Sc.
PROFESOR TUTOR DEL
PROYECTO
DE TITULACIÓN

Ing. Erick E. González Linch, M.Sc.
PROFESOR REVISOR DEL
PROYECTO
DE TITULACIÓN

Ab. Juan Chávez Atocha, Esp.
SECRETARIO

DECLARACIÓN EXPRESA

“La responsabilidad del contenido de este Proyecto de Titulación, me corresponden exclusivamente; y el patrimonio intelectual de la misma a la UNIVERSIDAD DE GUAYAQUIL”

DANNY JONATHAN ASCENCIO MORENO



**UNIVERSIDAD DE GUAYAQUIL
FACULTAD DE CIENCIAS MATEMÁTICAS Y FÍSICAS**

**CARRERA DE INGENIERÍA EN SISTEMAS
COMPUTACIONALES**

**DESARROLLO Y EXPLOTACIÓN DE INFORMACIÓN
DE UNA BASE DE DATOS DE PATOLOGÍAS
PRESENTES EN LA CAVIDAD BUCAL CON
ÉNFASIS EN CARIES Y GINGIVITIS
A TRAVÉS DE HERRAMIENTAS
ORIENTADAS A MINERÍA DE
DATOS PARA LA FACULTAD
PILOTO DE ODONTOLOGÍA
DE LA UNIVERSIDAD
DE GUAYAQUIL**

Proyecto de Titulación que se presenta como requisito para optar por el

Título de

INGENIERO EN SISTEMAS COMPUTACIONALES

Autor: ASCENCIO MORENO DANNY JONATHAN

C.I. 0927469379

TUTOR: Lorenzo Cevallos Torres, M. Sc

Guayaquil, marzo del 2020

CERTIFICADO DE ACEPTACIÓN DEL TUTOR

En mi calidad de Tutor del proyecto de titulación, nombrado por el Consejo Directivo de la Facultad de Ciencias Matemáticas y Físicas de la Universidad de Guayaquil.

CERTIFICO:

Que he analizado el Proyecto de Titulación presentado por el estudiante DANNY JONATHAN ASCENCIO MORENO, como requisito previo para optar por el título de Ingeniero en Sistemas Computacionales cuyo título es:

“Desarrollo y explotación de información de una base de datos de patologías presentes en la cavidad bucal con énfasis en caries y gingivitis a través de herramientas orientadas a la minería de datos para la facultad piloto de odontología de la universidad de guayaquil”

Considero aprobado el trabajo en su totalidad.

Presentado por:

DANNY JONATHAN ASCENCIO MORENO
C.I. 0927469379

Tutor: Lorenzo Cevallos Torres, M. Sc.

Guayaquil, marzo de 2020



**UNIVERSIDAD DE GUAYAQUIL
FACULTAD DE CIENCIAS MATEMÁTICAS Y FÍSICAS
CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES**

Autorización para Publicación de Proyecto de Titulación en Formato Digital

1. Identificación del Proyecto de Titulación

Nombre Alumno: DANNY JONATHAN ASCENCIO MORENO	
Dirección: URB. RENACER MZ 142 VII 24	
Teléfono: 0994798364	E-mail: danny.ascenciom@ug.edu.ec

Facultad: Ciencias Matemáticas y Físicas
Carrera: Ingeniería en Sistemas Computacionales
Proyecto de titulación al que opta: Investigación
Profesor guía:

Título del Proyecto de titulación: Desarrollo y explotación de información de una base de datos de patologías presentes en la cavidad bucal con énfasis en caries y gingivitis a través de herramientas orientadas a la minería de datos para la facultad piloto de odontología de la universidad de guayaquil
--

Tema del Proyecto de Titulación: (Palabras claves 5 a 8) Machine Learning, Weka, Redes Bayesianas, Predicción, Patologías Bucales

2. Autorización de Publicación de Versión Electrónica del Proyecto de Titulación

A través de este medio autorizo a la Biblioteca de la Universidad de Guayaquil y a la Facultad de Ciencias Matemáticas y Físicas a publicar la versión electrónica de este Proyecto de titulación.

Publicación electrónica:

Inmediata	<input checked="" type="checkbox"/>	Después de 1 año	<input type="checkbox"/>
-----------	-------------------------------------	------------------	--------------------------

Firma Alumno:

Danny Jonathan Ascencio Moreno

3. Forma de envío:

El texto del proyecto de titulación debe ser enviado en formato Word, como archivo .Doc. O .RTF y Puf para PC. Las imágenes que la acompañen pueden ser: .gif, .jpg o .TIFF.

DVDROM

☐

CDROM

☐

ÍNDICE GENERAL

APROBACIÓN DEL TUTOR.....	III
DEDICATORIA	IV
AGRADECIMIENTO.....	V
ÍNDICE GENERAL.....	XI
ABREVIATURAS	XIII
ÍNDICE DE CUADROS	XIV
ÍNDICE DE GRÁFICOS	XV
INTRODUCCIÓN	1
CAPÍTULO I.....	3
PLANTEAMIENTO DEL PROBLEMA	3
SITUACIÓN CONFLICTO NUDOS CRÍTICOS.....	4
CAUSAS Y CONSECUENCIAS DEL PROBLEMA	5
DELIMITACIÓN DEL PROBLEMA	6
FORMULACIÓN DEL PROBLEMA	6
Evaluación del Problema.....	6
OBJETIVOS	8
OBJETIVO GENERAL	8
OBJETIVOS ESPECÍFICOS.....	8
ALCANCES DEL PROBLEMA	8
JUSTIFICACIÓN E IMPORTANCIA	10
METODOLOGÍA DEL PROYECTO	11
CAPÍTULO II	13
MARCO TEÓRICO.....	13
ANTECEDENTES DEL ESTUDIO.....	13
FUNDAMENTACIÓN TEÓRICA.....	14
MODELO PREDICTIVO	14
HERRAMIENTAS PARA MINERÍA DE DATOS	15
CARACTERÍSTICAS DE LAS HERRAMIENTAS PARA MINERÍA DE DATOS	17
DEFINICIONES CONCEPTUALES	18
Machine Learning, aplicación.....	20
PROBABILIDADES CONDICIONALES Y TEOREMA DE BAYES	21
Redes Bayesianas.....	21
TÉRMINOS ODONTOLÓGICOS: CARACTERÍSTICAS.....	23
FUNDAMENTACIÓN LEGAL	25
COIP (Código Orgánico Integral Penal)	25
FOE (Federación Odontológica Ecuatoriana)	26
LEY ORGÁNICA DE EDUCACIÓN SUPERIOR.....	28
DECRETO 1014	28
SOBRE EL USO DEL SOFTWARE LIBRE	28
PREGUNTA CIENTÍFICA PARA RESPONDER	29
CAPÍTULO III	30

METODOLOGÍA DE LA INVESTIGACIÓN	30
DISEÑO DE LA INVESTIGACIÓN	30
Modalidad de la Investigación	30
Tipo de Investigación.....	31
Población y muestra	31
Descripción y codificación de las variables del estudio.....	31
Variables de la investigación	33
Instrumentos de Recolección de Datos	33
Entrevista a los administradores de la base.....	33
Lectura Científica.....	34
Filtrado de información de la Base de datos obtenida	34
Ingreso de datos a la aplicación Weka	38
Variables de entrada para la aplicación WEKA.....	40
Procesamiento de los datos ingresados, usando la técnica de redes bayesianas de minería de datos.....	50
CAPTULO IV	56
RESULTADOS CONCLUSIONES Y RECOMENDACIONES.....	56
Resultados	56
CONCLUSIONES	60
Anexos.....	66
ANEXO 1 – Cuadros de resumen con información de las variables para minería de datos en weka.	66
ANEXO 2 – Fichas médicas DE LA fpo PARA registrar LOS DATOS DE PACIENTES.	67
ANEXO 3 – conseNTIMIENTO PARA USO ACADÉMICO DE DATOS DE LA fpo	69

ABREVIATURAS

UG	Universidad de Guayaquil
FCMF	Facultad de Ciencias Físicas y Matemáticas
CISC	Carrera de Ingeniería en Sistemas Computacionales de la UG
FPO	Facultad Piloto de Odontalgia
ING.	Ingeniero/a
URL	Localizador de Fuente Uniforme
WWW	World Wide Web (red mundial)
XLSX	Extensión de archivo empleado en Microsoft Excel
CSV	Valores separados por coma (Comma separated values)
FOE	Federación Odontología Ecuatoriana
COIP	Código Orgánico Integral Penal

ÍNDICE DE CUADROS

Cuadro 1 - Descripción de causas y consecuencias	5
Cuadro 2 - Delimitación del problema.....	6
Cuadro 3 - Características de las Herramientas para minería de datos	18
Cuadro 4 - Variables de pacientes en las bases de datos.....	32
Cuadro 5 - Operacionalización de variables	33

ÍNDICE DE GRÁFICOS

Gráfico 1 – Metodología del Proyecto	11
Gráfico 2 – Interfaz de Orange.....	15
Gráfico 3 – Interfaz de Rapidminer	16
Gráfico 4 – Página principal de la interfaz de Weka	17
Gráfico 5 – Diente afectado por caries.....	24
Gráfico 6 – Vista de los campos hallados en la base de datos	35
Gráfico 7 - Espacios en blanco eliminados.	36
Gráfico 8 – Reconocimiento de diferentes ciudades en el campo de ubicación ...	36
Gráfico 9 - Espacios en blanco llenados	37
Gráfico 10 - Presentación de la información total lista para la aplicación de las técnicas de minería de datos.....	38
Gráfico 11 - Conversión de .xlsx a .csv	38
Gráfico 12 - Ajuste del carácter delimitador para la aplicación.....	39
Gráfico 13 - Documento listo para exportar a weka	39
Gráfico 14 – Variable 1.....	40
Gráfico 15 – Variable 2.....	40
Gráfico 16 – Variable 3.....	40
Gráfico 17 – Variable 4.....	41
Gráfico 18 – Variable 5.....	41
Gráfico 19 – Variable 6.....	41
Gráfico 20 – Variable 7.....	42
Gráfico 21 – Variable 8.....	42
Gráfico 22 – Variable 9.....	42
Gráfico 23 – Variable 10.....	43
Gráfico 24 – Variable 11.....	43
Gráfico 25 – Variable 12.....	43
Gráfico 26 – Variable 13.....	44
Gráfico 27 – Variable 14.....	44
Gráfico 28 – Variable 0.....	44
Gráfico 29 – Variable 16.....	45
Gráfico 30 – Variable 17.....	45
Gráfico 31 – Variable 18.....	45
Gráfico 32 – Variable 19.....	46
Gráfico 33 - Interfaz de weka.....	47
Gráfico 34 - Pasos para ejecutar el algoritmo de Bayes.....	48
Gráfico 35 – Ventana de Inicio	48
Gráfico 36 – Selección del archivo	49
Gráfico 37 – Ventana de explorador	50
Gráfico 38 – Red bayesiana	22
Gráfico 39 – Ejemplos de clasificadores bayesianos (a) simple, (b)TAN (c) BAN	23
Gráfico 40 – Pestaña “Select Attributes”	51
Gráfico 41 - Atributos rankeados con respecto a la variable ENF-PERIODON ..	51
Gráfico 42 - Atributos luego de la selección.....	52

Gráfico 43 - Selección del Clasificador	52
Gráfico 44 - Selección del algoritmo de búsqueda	53
Gráfico 45 - Resultados de la minería de datos.....	54
Gráfico 46 – Visualización de la red bayesiana	55
Gráfico 47 – Red bayesiana a partir de los datos	56
Gráfico 48 – Probabilidades de la variable CPO_T	57
Gráfico 49 - Probabilidades de la variable SEXO	57
Gráfico 50 - Probabilidades de la variable EDAD	58
Gráfico 51 - Probabilidades de la variable PROBLE-ACTUAL	58
Gráfico 52 - Probabilidades de la variable ANT-PERSONAL.....	59
Gráfico 53 - Probabilidades de la variable MALOC-GRADO	59
Gráfico 54 - DIABETES.....	60



**UNIVERSIDAD DE GUAYAQUIL
FACULTAD DE CIENCIAS MATEMÁTICAS Y FÍSICAS
CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES**

**DESARROLLO Y EXPLOTACIÓN DE INFORMACIÓN DE UNA BASE DE
DATOS DE PATOLOGÍAS PRESENTES EN LA CAVIDAD BUCAL CON
ÉNFASIS EN CARIES Y GINGIVITIS A TRAVÉS DE HERRAMIENTAS
ORIENTADAS A MINERÍA DE DATOS PARA LA FACULTAD PILOTO DE
ODONTOLOGÍA DE LA UNIVERSIDAD DE GUAYAQUIL**

Autor: Danny Jonathan Ascencio Moreno
Tutor: Ing. Lorenzo Cevallos Torres, M.Sc.

RESUMEN

A partir de una base de datos provista por la FPO, adecuada mediante filtrados de datos, para poder aplicar técnicas de MACHINE LEARNING, precisamente redes bayesianas, se logró conocer el nivel de incidencia de algunas enfermedades que presentan los pacientes con respecto a dos de las patologías bucales de mayor relevancia: caries y gingivitis. Estudios realizados previamente con redes bayesianas demostraron que los beneficios que ofrecen estos modelos son variados, desde una reducción del gasto médico, hasta la predicción de diferentes tipos de cáncer, dadas las características más notorias en cada caso. La base obtenida fue procesada a través de la herramienta digital de minería de datos WEKA. La aplicación es de tipo software libre y dispone de una serie de herramientas que facilitan el manejo de los datos presentes en la base de datos, como discretización de variables, conversión de tipos de datos, y un manejo de varios algoritmos encargados de ejecutar tareas de MACHINE LEARNING o aprendizaje automático. El algoritmo empleado genera un gráfico que facilita la interpretación de la información. Los resultados observados por el modelo evidencian la situación actual de forma general de los pacientes permitiendo predecir con un grado moderado de certeza (68%) el comportamiento de las variables y generar un modelo predictivo, de acuerdo con los datos obtenidos y su correlación. Al estudiar el modelo generado se evidenció, la relación de las variables en las que hace énfasis el presente estudio, con respecto a las variables que el modelo señaló como principales implicadas en el desarrollo de estas patologías.

Palabras clave: Machine Learning, Weka, Redes bayesianas, Predicción, Patologías bucales.



**UNIVERSIDAD DE GUAYAQUIL
FACULTAD DE CIENCIAS MATEMÁTICAS Y FÍSICAS
CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES**

**DEVELOPMENT AND EXPLOITATION OF INFORMATION FROM A
DATABASE OF PATHOLOGIES PRESENT IN THE ORAL CAVITY WITH
EMPHASIS ON CARIES AND GINGIVITIS USING TOOLS ORIENTED TO
DATA MINING FOR THE PILOT FACULTY OF THE UNIVERSITY OF
GUAYAQUIL**

Author: Danny Jonathan Ascencio Moreno
Tutor: Ing. Lorenzo Cevallos Torres, M.Sc.

ABSTRACT

From a database provided by the FPO, fitted through data filters, to be able to apply MACHINE LEARNING techniques, specifically Bayesian networks, was possible to know the level of incidence of some diseases that patients present with respect to the two oral pathologies of greater relevance: cavities and gingivitis. Previous studies with Bayesian networks showed that the benefits offered by these models are wide, from a reduction in medical expenses, to the prediction of different types of cancer, given the most notable characteristics in each case. The base obtained was processed through the digital data mining tool WEKA. The application is an open source that has a series of tools which facilitate the handling of the data present in the database, such as discretization of variables, conversion of data types, and handling of several algorithms in charge of operations tasks of machine learning. The algorithm used generates a graph that makes easier the interpretation of the information. The results observed by the model show the current situation of the general form of the affected patients, predicted with a moderate degree of certainty (68%) the behavior of the variables and generate a predictive model, according to the data obtained and their correlation. When studying the generated model, was evident the relationship of the variables emphasized in the present study, with respect to the variables that the signal model as the main ones involved in the development of these pathologies.

Keywords: Machine Learning, Weka, Bayesian networks, Prediction, Oral pathologies.

INTRODUCCIÓN

En la facultad piloto de odontología (FPO), existe gran cantidad de información almacenada la cual se aprovechó para aplicar técnicas de minerías de datos siguiendo la rama de Machine Learning y aplicarla al campo de la Odontología, diversos estudios dan evidencia de buenos resultados aportados por estas herramientas en diferentes ámbitos, ya sea marketing, economía e incluso la misma salud humana. Se hace énfasis en las patologías conocidas como caries y gingivitis, que son las de mayor incidencia en la salud bucal de los pacientes tratados.

La caries es una patología bucal que se forma debido a varios factores, entre los más mencionados están las bacterias bucales que afectan la cavidad oral debido a la mala higiene de la cavidad bucal, o el abuso de azúcares, el efecto que causan las caries es un deterioro de los minerales que conforman el recubrimiento de los dientes y podría dar lugar a la pérdida de las piezas dentales afectadas.

La gingivitis, afecta a las encías del paciente esto debido a una mala alimentación, lo cual produce la aparición de placa y de no darle la atención del caso se desarrollará en un sarro, esto a su vez impide que la limpieza sea efectiva y queden bacterias que desgatarán las encías, y la gingivitis podría evolucionar a una periodontitis, dando como resultado la pérdida de dientes ya que la caries también se sumaría a los factores previos.

Para el análisis de esta problemática es menester mencionar algunas de sus causas. Entre las cuales se identificaron el no contar con información estandarizada para el ingreso de datos en las bases, puesto que hay elementos vacíos, algunos redundantes, datos que no tienen relación con sus respectivas filas. Esto dificulta el procesamiento de la información por las herramientas para minería de datos, y el resultado obtenido no será del todo acertado.

El modelo probabilístico conocido como Redes Bayesianas ha sido muy utilizado en el campo de la medicina, y fue el aplicado en la herramienta seleccionada para la minería de datos, dado que los diagnósticos médicos derivan de un extenso análisis de tratamientos posibles junto con la vasta experiencia acumulada y el

conocimiento reunido para tratar una determinada patología o enfermedad, el modelo probabilístico tomaría para su aprendizaje la gran cantidad de datos recopilados durante un año de pacientes tratados y determinaría su diagnóstico en base a las relaciones de las patologías y los factores que más tienden a elevar la probabilidad de presentar las patologías objetivo.

La distribución del presente trabajo se encuentra detallada de la siguiente manera: En el **Capítulo I**, se realiza el planteamiento del problema, la situación actual, causas y consecuencias, delimitación, formulación y evaluación del problema. Objetivo general, objetivos específicos, alcances del problema, justificación e importancia de la investigación.

En el **Capítulo II**, se desarrolla lo correspondiente al “Marco Teórico”. Conceptos básicos de las patologías objetivo, herramientas tecnológicas digitales, antecedentes del estudio, fundamentación teórica, fundamentación legal, pregunta científica a contestar.

En el **Capítulo III**, se especifica la propuesta investigativa, la metodología de la investigación su modalidad y tipo, la población que será considerada en el estudio. Descripción y codificación de las variables del estudio, variables de la investigación, instrumentos de recolección de datos, filtrado de la base de datos, ingreso de datos a la aplicación Weka.

En el **Capítulo IV**, se presentan los resultados obtenidos, las conclusiones, recomendaciones y bibliografía utilizadas durante el desarrollo del tema.

CAPÍTULO I

EL PROBLEMA

PLANTEAMIENTO DEL PROBLEMA

Ubicación del Problema en un contexto

En la FPO de la Universidad de Guayaquil, ubicada en la Ciudadela Universitaria de la misma ciudad, se brinda servicio a la comunidad en distintas especialidades como:

- Clínica del adulto mayor.
- Cirugía menor.
- Cirugía.
- Operatoria dental.
- Odontopediatría.
- Prostodoncia.

Las especialidades mencionadas generan una gran cantidad de datos de los pacientes, los cuales se encuentran almacenados en fichas médicas, recetas e imágenes. A pesar de la gran cantidad de información que se maneja no se cuenta con un método para procesar eficientemente los recursos recopilados, pues la información no está apta para su procesamiento directo mediante herramientas actuales, ya que primero hay que depurarla. La falta de conocimiento sobre herramientas digitales usadas en el área de la odontología también se suma al problema, independientemente del grado de preparación o experiencia que tenga el profesional odontólogo.

La caries dental y la gingivitis tienen un impacto impresionante en la salud y bienestar humano (Puttipan, Wanachantararak, & Khongkhunthian, 2017). Según (Gerson, Fernanda, & Claudio, 2016) la caries dental se ubica en el primer lugar de las principales patologías que aquejan a la población actual y ha sido conocida a lo largo de la historia de la humanidad, siendo en un principio no diferenciada

como un problema grave debido al desconocimiento de las causas, factores de riesgo y curso de esta patología. Por su lado, la gingivitis se considera la segunda alteración bucodental en cuanto a morbilidad y afecta a más de tres cuartas partes de la población (Muñoz Cruzatty, Moreira, & Oliveira del Rio, 2017)

Los modelos de predicción se están empleando en el sector médico para ayudar a prevenir riesgos, evitar posibles desgracias futuras y también en la toma de decisiones, simultáneamente muchos de esos modelos de predicciones evidencian una falta de calidad. (Wolfson, Vock, Johnson, & Connor, 2014)(Collins, Reitsma, Altman, & Moons, 2015)

Se requiere que los registros sean confiables para facilitar la labor de las herramientas de minería de datos y las áreas en las cuales atacar estas anomalías. Actualmente no se cuenta con un modelo de datos para la Facultad Piloto de Odontología (FPO) que se beneficie de la gran cantidad de datos generados y de esta manera ayude a diferenciar e identificar las patologías estudiadas, las de mayor incidencia y por ende cuál de ellas es la causante de una mayor afectación entre los pacientes tratados, así mismo no existe un modelo de validación que ayude a predecir el comportamiento de las variables, lo cual dificulta el desarrollo de tecnologías informáticas que trabajen con estos datos reales para la correcta identificación de las enfermedades en las cuales centrar la mayor atención.

SITUACIÓN CONFLICTO NUDOS CRÍTICOS

En la FPO se emiten datos de pacientes que, de ser tratados con las herramientas adecuadas, pueden ayudar a tomar decisiones en cuanto al curso de las enfermedades en las que hace énfasis la presente investigación (caries y gingivitis). La informática está presente en todos los ámbitos actualmente, sin embargo, la FPO no está haciendo uso de estos elementos; la lenta incorporación de estos recursos en el aspecto médico merma oportunidades de optimizar procesos.

Se puede apreciar, según las visitas realizadas en la FPO, que no cuentan con modelos de predicción y los datos no son homogéneos, esto complica la implementación de estas herramientas digitales (minería de datos). Los datos no estandarizados ni sistematizados con los que se cuenta en la actualidad no son idóneos para ser procesados con las herramientas tecnológicas adecuadas ya que al momento de generar información estadística e inferencias primero se los debe depurar y organizar en un modelo adecuado que garantice la confiabilidad de los datos provistos.

CAUSAS Y CONSECUENCIAS DEL PROBLEMA

En el cuadro 1, se encuentran las causas y consecuencias importantes de los inconvenientes para detectar factores de las patologías que causan afectaciones en la cavidad bucal.

Cuadro 1 - Descripción de causas y consecuencias

Causas	Consecuencias
Información no estandarizada	Datos inconclusos, elementos que no concuerdan con su campo en la base de datos
Poco uso de herramientas digitales	Posible pérdida de información, dificultad al importar datos en un formato adecuado que permita su tratamiento por técnicas de minería de datos
Desconocimiento de las herramientas para minería de datos	No se emplean herramientas especializadas para aprovechar la gran cantidad de datos que se generan
Ausencia de un modelo predictivo	Se desconoce la proyección a futuro de las patologías tratadas, mermando posible información relevante a tomar en cuenta en decisiones posteriores.

Elaborado por: Danny Ascencio Moreno.

Fuente: Datos propios de la investigación.

DELIMITACIÓN DEL PROBLEMA

Cuadro 2 - Delimitación del problema

Campo	Minería de datos
Área	Tecnología de Información
Aspecto	Investigativo
Tema	Desarrollo y explotación de información utilizando una base de datos sobre patologías presentes en la cavidad bucal con énfasis en caries y gingivitis a través de herramientas orientadas a minería de datos para la Facultad Piloto de Odontología de la Universidad de Guayaquil.

Elaborado por: Danny Ascencio Moreno

Fuente: Datos propios de la investigación.

FORMULACIÓN DEL PROBLEMA

¿De qué manera la creación de un modelo predictivo de minería de datos ayudará a detectar los factores más relevantes que afectan la cavidad bucal en niños y adultos en la Facultad Piloto de Odontología de Guayaquil?

EVALUACIÓN DEL PROBLEMA

En esta investigación detallamos algunos aspectos seleccionados para poder evaluar la problemática existente, a través de los cuales, se ha determinado el problema.

Delimitado:

Es delimitado porque el estudio está dirigido a obtener datos de las enfermedades de la cavidad bucal, la información podrá ser utilizada por estudiantes y profesionales odontológicos.

Claro:

La tecnología debe ser usada en distintos ámbitos de la sociedad, es importante que las personas que trabajan en el campo odontológico empiecen a tomar en cuenta las soluciones informáticas.

Al aplicar estas alternativas innovadoras se pretende tener información que identifique de una manera eficaz las enfermedades presentadas en la cavidad bucal.

Concreto:

Se trata de un tema en particular que involucra al área odontológica, la cual, aplicando recursos innovadores tecnológicos, tendrán como resultado información que ayudará a detectar las afectaciones.

Original:

Con la información obtenida en la base de datos y el uso correcto de la herramienta de minería de datos, se permitirá tener datos más precisos de las enfermedades en cuestión y a su vez determinar aspectos importantes para tomar acciones que provoquen, en un futuro, la reducción de afectados por esta problemática.

Factible:

La solución brindada es el desarrollo de un modelo de datos predictivo, que permita conocer la tendencia de las patologías según las variables que definen a cada paciente, empleando herramientas de minerías de datos, para obtener información veraz que pueda ser utilizada en futuras investigaciones.

Variables:

Para fines del presente trabajo de titulación se han identificado las siguientes variables

Variable Dependiente: Control de Historias Clínicas.

Variable Independiente: Modelo de Datos Predictivo

OBJETIVOS

OBJETIVO GENERAL

Desarrollar un modelo de datos que permita predecir el comportamiento de las variables mediante herramientas especializadas en minería de datos para identificar los factores más relevantes causantes de las patologías con énfasis en Caries y Gingivitis.

OBJETIVOS ESPECÍFICOS

- Recopilar información de diversas fuentes para el levantamiento de la base de datos.
- Seleccionar la base de datos acorde con la información obtenida de la investigación.
- Explotar la base de datos mediante herramientas de minerías de datos, para detectar los factores de riesgo donde se presenten patologías de la cavidad bucal.
- Presentar la información mediante un reporte (artículo) a la Facultad Piloto de Odontología, para dar a conocer los descubrimientos realizados.

ALCANCES DEL PROBLEMA

Para el desarrollo de esta propuesta se ha seleccionado la Facultad Piloto de Odontología de la Universidad de Guayaquil, la cual se encuentra ubicada en la Ciudadela Universitaria de la ciudad de Guayaquil, por la facilidad de acceso a los datos.

Mediante las revisiones de fichas médicas y entrevistas realizadas se procederá a la recolección de elementos que permita levantar la base de datos, conocer los datos que intervienen en la generación de dicha información.

Analizadas las herramientas empleadas para el levantamiento de información, se procede a la selección de los datos que serán procesados, de igual manera a su depuración, para que estén en condiciones de trabajar con la herramienta de minería de datos seleccionada.

Para la minería de datos se hace énfasis en un modelo predictivo, que refleje una proyección de la situación actual presentada por la base de datos seleccionada, de lo cual se busca obtener datos relevantes como evolución de las patologías estudiadas. Un ejemplo de esto lo representarían los lugares vulnerables donde hacen aparición en gran medida las afectaciones, también las edades frecuentes y si ese número incrementase con el pasar de los años, según la tendencia de la información obtenida.

Los resultados se darán a conocer mediante un artículo en una revista indexada, de tal manera que la Facultad Piloto de Odontología pueda tomar sus dediciones con respecto a la situación actual y a futuro.

JUSTIFICACIÓN E IMPORTANCIA

Una base de datos correctamente estructurada admite ordenar, clasificar y seleccionar información que requieran los sistemas o personas que estén trabajando con dicha base. Por el volumen de los datos es indispensable contar con un modelo encargado de administrar los recursos recopilados para trabajar de forma más eficiente con la información. Por ejemplo, se podría conocer mediante una correcta recopilación de datos, tendencias en los hábitos alimenticios y costumbres de la población afectada, posibles casos de enfermedades que estén relacionadas de manera indirecta con los síntomas de las patologías enfocadas y problemas no considerados que derivan de estas afectaciones. La presente investigación, enfocada en la minería de datos aplicada a la ciencia de la salud, servirá como aporte al momento de implementar nuevas tecnologías a futuro, facilitando el reconocimiento de los síntomas presentados, para hacer de la identificación y diferenciación, una tarea más eficiente.

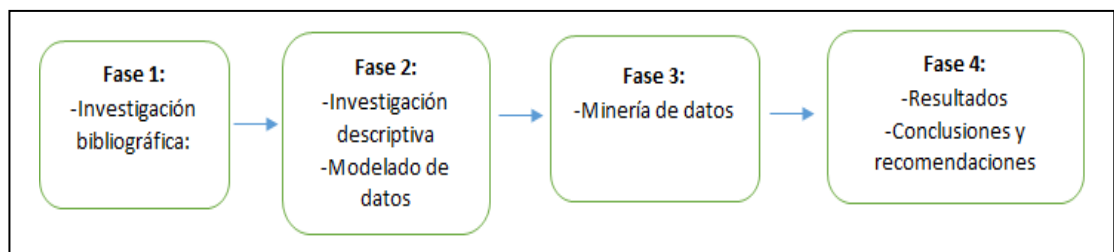
Actualmente las herramientas tradicionales no son las óptimas para el procesamiento de grandes volúmenes de datos, particularmente por la heterogeneidad que se presenta en algunas bases de datos (Anabella De Battista Patricia Cristaldo, Lautaro Ramos, Juan Pablo Nuñez, Soledad Retamar, Daniel Bouzenard, Norma Edith Herrera, 2016). La medicina está dependiendo cada vez mas de nuevas tecnologías, las cuales tienen gran impacto en sus prácticas. Un aspecto importante de estas tecnologías es que generan gran cantidad de datos digitales los cuales son almacenados, desafortunadamente esos datos no son aprovechados hasta que se analizan con el fin de descubrir conocimiento nuevo que podría beneficiar a los pacientes y comunidades (Cios, Krawczyk, Cios, & Staley, 2019).

De la información tratada se beneficiarán directamente los doctores de la FPO de la Universidad de Guayaquil mediante informes emitidos por las herramientas para el procesamiento de datos, permitiendo centrar su atención en las características poco evidentes de las enfermedades que afectan a la población tratada. Es crucial que iniciativas de este tipo sean replicadas en otros campos de la ciencia para que, aprovechando el gran impacto que esta teniendo la tecnología en diversos

ámbitos, faciliten tomar decisiones cruciales que incluso podrían dar paso a nuevos estudios que ayuden a mejorar la calidad de vida de las personas.

METODOLOGÍA DEL PROYECTO

Gráfico 1 – Metodología del Proyecto



Elaborado por: Danny Ascencio Moreno

Fuente: Elaboración Propia.

Fase 1:

La **investigación bibliográfica** es la primera etapa del proceso que proporciona el conocimiento de las investigaciones ya existentes, de un modo sistemático, a través de una amplia búsqueda de: información, conocimientos y técnicas sobre una cuestión determinada.

Fase 2:

La **investigación descriptiva, o método descriptivo de investigación**, es el procedimiento usado en ciencia para describir las características del fenómeno, sujeto o población a estudiar. Al contrario que el método analítico, no describe por qué ocurre un fenómeno, sino que se limita a observar lo que ocurre sin buscar una explicación.

Una vez adquirido los conocimientos mediante bibliografía y observación de los procesos llevados a cabo, se procede al modelado de la base de datos con los elementos e información necesaria para luego ser procesada por herramientas de minerías de datos.

Fase 3:

En la **fase de minería de datos** se procede a realizar la explotación de la base de datos con herramientas especializadas para estos procesos, realizando todo tipo de pruebas y adquiriendo la mayor cantidad de información en base al modelo de datos diseñado.

Fase 4:

En el apartado de **resultados**, obtenido los datos ya procesados por las herramientas seleccionadas se procede a interpretarlos para luego dar un reporte más exacto o tener sugerencias para realizar mejoras.

CAPÍTULO II

MARCO TEÓRICO

Los datos estadísticos generados a partir de bases de datos han sido desde siempre de gran ayuda a la hora de la toma de decisiones, ya sea al momento de implementar nuevas tecnologías, en el ámbito gerencial, al momento de decidir qué estrategias son más adecuadas para obtener mejores ganancias y como en el presente estudio, en el campo médico.

Existe una gran variedad técnicas aplicadas en minería de datos para extraer información de las bases disponibles, cada una tiene una función específica dependiendo del tipo de estudio que se quiera realizar; en el contexto de la odontología interesa mucho conocer causas o factores no considerados que la minería de datos podría exhibir y llegar a tener gran relevancia.

ANTECEDENTES DEL ESTUDIO

Los datos que resaltan en la investigación acerca de la salud provienen de fuentes tales como las historias clínicas, pruebas de laboratorio, censo de población, registros de enfermedades, etc. Se incorporan otras fuentes de información ya sean los obtenidos por medio de aplicaciones (dispositivos que se integran al diario vivir: celulares, relojes...), las redes sociales o las plataformas "nube", entre otras (Casallas & Jiménez, 2017).

La integración de minería de datos y modelos predictivos ha dado grandes resultados en ámbitos de la salud como los presentados a continuación:

- El uso de la minería de datos de manera efectiva podría contribuir en la disminución del 8% del gasto médico en Estados Unidos gracias al uso de Big Data, como se menciona en (Mehta & Pandit, 2018).
- Una de las aplicaciones más comunes de la minería de datos en medicina y cuidado de la salud es predecir diferentes tipos de cáncer de seno el cual ha llamado la atención de muchos científicos. (Bahmani, Jamshidi, & Shaltook, 2019)

- La medicina genómica intenta construir estrategias individualizadas para el diagnóstico o la terapia, toma de decisiones, utilizando la información genómica de los pacientes. El análisis de big data descubre patrones ocultos, correlaciones desconocidas y otras ideas a través del análisis de varios conjuntos de datos a gran escala (He, Ge, & He, 2017).

FUNDAMENTACIÓN TEÓRICA

El avance de la tecnología trae consigo tendencias como, **machine learning**, **big data**, **data warehouse** entre otros, que están siendo incluidos en diversos campos y de manera específica en el presente estudio, relacionado con la medicina. Sin embargo, hay impedimentos a la hora de emplear abiertamente estos métodos, debido a la falta de conocimiento del gran beneficio que dichos métodos representan por parte de las otras áreas, ya que cada una se enfoca en su ámbito sin incluirse o relacionarse con las demás. A pesar de todo esto, estudios de esta índole abren una puerta al conocimiento de los beneficios que esta integración trae consigo (Escobar, Harold, Alcivar, & Puris, 2016; Martínez-García, Dalgo-Flores, Herrera-López, Analuisa-Jiménez, & Velasco-Acurio, 2019)

MODELO PREDICTIVO

Se lo conoce como un modelo de datos que se basa en estadísticas inferenciales, es un mecanismo que predice el comportamiento de las variables utilizando como entrada las características de ésta. (Timón, 2017)

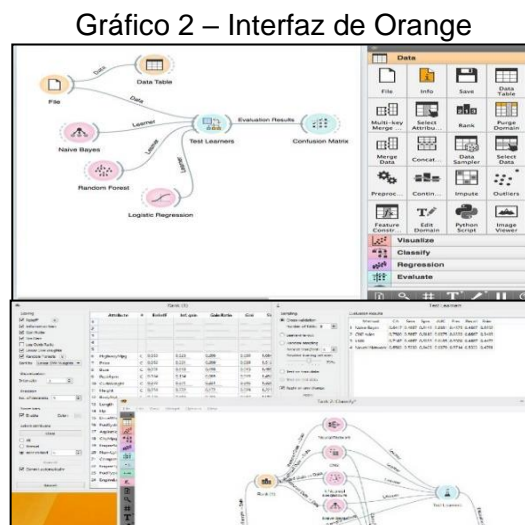
El modelo con el que se trabajará en el desarrollo de la tesis presente se lo conoce como **redes Bayesianas**, dicho modelo ofrece versatilidad, transparencia y la facilidad de analizar las probabilidades de cada una de las variables. Demuestran un rendimiento bastante elevado en lo referente al proceso de inferencia, algo que no se ve reflejado en otras técnicas (Paulino Flores & Huayna Dueñas, 2019)

HERRAMIENTAS PARA MINERÍA DE DATOS

ORANGE

Orange es una herramienta tipo software libre que se emplea en **data mining** y **machine learning** está desarrollado en Python, esta herramienta contiene técnicas de clasificación y evaluación automatizadas. Se pueden cargar logs pre procesados únicamente, pero solo soporta archivos de hasta 3 Mb. Funciona con varios tipos de archivos: *.tab, *.txt, *.data, *.dat, *.rda, *.rdo. Es una herramienta completa, aunque falla a la hora de cargar archivos de más de 3 Mb (Naik & Samant, 2016). En el

Gráfico 2 se puede apreciar la interfaz de Orange:

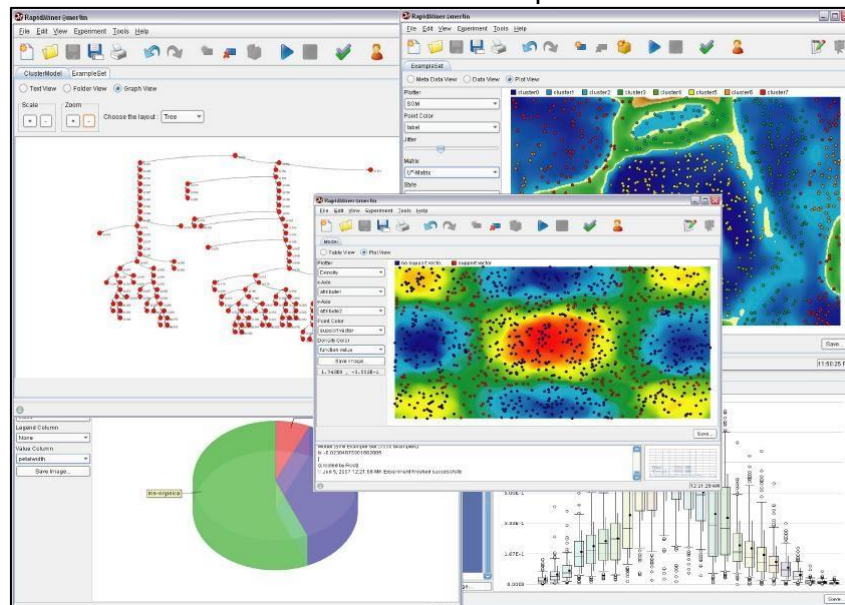


Elaborado por: Danny Ascencio Moreno
Fuente: Orange, 2016

RAPIDMINER

RapidMiner es un software especializado para ciencias de datos. Entre las aplicaciones de minería de datos llegó a situarse entre los primeros lugares de las más utilizadas para problemas reales. Su efectividad tanto en el ámbito comercial como en la educación, permite importar datos, prepararlos y visualizarlos, además dispone de extensiones que se le pueden instalar, entre las que se encuentra Weka que permite ejecutar algoritmos en RapidMiner (Lanzarini & Villa Monte, 2017). Su interfaz se puede apreciar en el Gráfico 3.

Gráfico 3 – Interfaz de Rapidminer

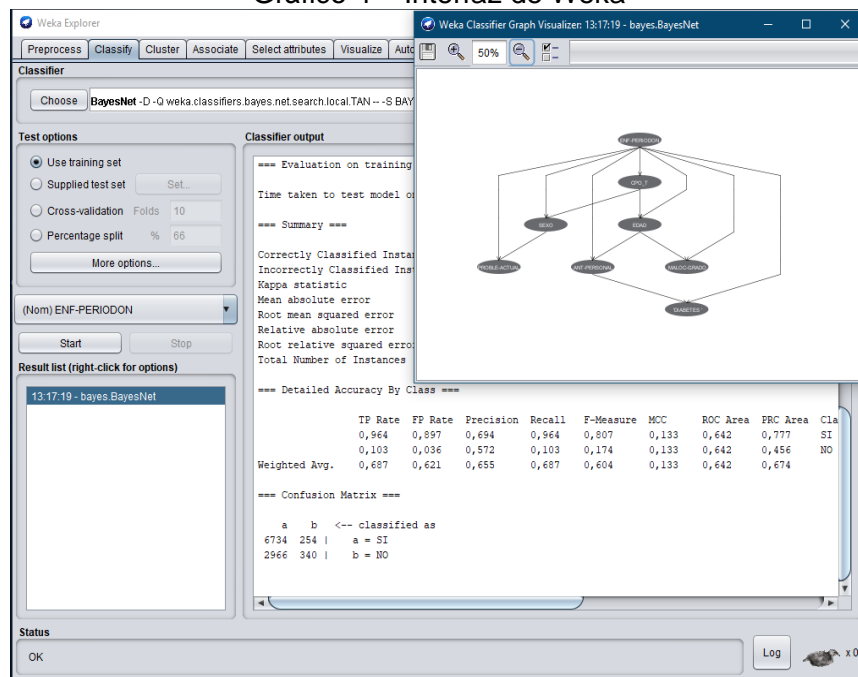


Elaborado por: Danny Ascencio Moreno
Fuente: Baoss, 2016

WEKA

La principal tarea de la minería de datos es explorar la gran cantidad de datos desde diferentes puntos de vista, clasificarlos y finalmente inventariarlos. Weka es quizá una de las mejores herramientas de código abierto para realizar minería de datos. Esta desarrollado en su totalidad en Java por la universidad de Waikato. Una de sus ventajas es que permite guardar los datos en diferentes formatos, como binary serialized instances, *.csv y *.arff (Attribute-Relation File Format) para después exportarlos a otra herramienta que disponga de alguna técnica de aprendizaje automático que Weka no posea. Adicionalmente cuenta con una gran librería de técnicas de aprendizaje automático (Kulkarni & Kulkarni, 2016; Naik & Samant, 2016). Su interfaz se puede apreciar en el Gráfico 4.

Gráfico 4 – Interfaz de Weka



Elaborado por: Danny Ascencio Moreno
Fuentes: Aplicación Weka

CARACTERÍSTICAS DE LAS HERRAMIENTAS PARA MINERÍA DE DATOS

Para la selección de la herramienta a utilizarse es de suma importancia, tener en cuenta varios factores entre los cuales destacan con carácter primordial: el tipo de archivo que aceptara como entrada el software y la interpretación de sus datos, priorizar uso de software libre, el aporte que desempeñara para acercar el estudio a los objetivos de la investigación, así como los resultados esperados y uno de los puntos más importantes es que la herramienta cuente con los algoritmos necesarios, en este caso se trabajan redes bayesianas, algunos de las herramientas seleccionadas solo constan con el algoritmo de naives bayes. En el Cuadro 3, se muestra de forma comparativa, algunas características de dichas herramientas.

Cuadro 3 - Características de las Herramientas para minería de datos

Herramienta	RapidMiner	Weka	Orange
Tipo de software	Comercial (30 días gratis)	Libre	Libre
Plataforma	Windows Linux	Todos	Windows MacOS Linux
Carga de archivos	xls, csv, access, arff, trf, database, spss, stata, dbase, bibtex, dasylab, url	csv, arff, db, url	tab, txt, data, dat, rda, rdo
Soporte	Para 200.000 usuarios	Grande	Moderado
Técnica Utilizada	Clustering, Árboles de decisión, Redes neuronales, Naive Bayes.	Clustering, Progresión, Predicción, Redes Bayesianas	Árboles de decisión, redes neuronales, Naive Bayes

Elaborado por: Danny Ascencio Moreno

Fuente: Sánchez, G. G., Ávila, S. D., & de la Rosa, J. L., 2005

Luego del análisis del cuadro comparativo, se eligió para este trabajo la herramienta weka, la cual se encargará del procesamiento de los datos por sus múltiples ventajas ya que cumple con lo requerido para la investigación.

DEFINICIONES CONCEPTUALES

MINERÍA DE DATOS

El proceso principal del cual se obtendrán los resultados que ayudaran a la toma de decisiones una vez procesada la información, siempre y cuando esos datos hayan sido correctos e inalterados, el resultado será interpretado según los criterios utilizados al momento de la explotación de la información.

La Minería de Datos es el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información no estructurada (interna y externa a la compañía) en información estructurada, para su explotación directa o para su análisis y conversión en conocimiento y así dar soporte a la toma de decisiones sobre el negocio o en el presente proyecto, hacia medidas para mejorar la salud bucodental de los pacientes tratados (Anabella De Battista Patricia Cristaldo, Lautaro Ramos, Juan Pablo Nuñez, Soledad Retamar, Daniel Bouzenard, Norma Edith Herrera, 2016)

AGRUPAMIENTO O CLUSTERING

El Clustering o análisis de clusters procedimiento de agrupación de una serie de vectores según criterios habitualmente de distancia; se tratará de disponer los vectores de entrada de forma que estén más cercanos aquellos que tengan características similares (Jiménez Márquez, 2019)

LICENCIA OPEN SOURCE

Es una licencia de software que permite acceder al código fuente y archivos binarios de un software para modificarlos y redistribuirlos sin ninguna restricción ni pagos al autor del software original (COCA DE PABLO, 2019)

BASES DE DATOS

Representan una colección de datos estructurados y organizados de tal manera que faciliten el acceso a ellos para los sistemas gestores de bases de datos, debido a normas establecidas para la creación de las bases se asegura que la información no esté duplicada, sean íntegra, y segura. En el presente trabajo la base de datos se encuentra en formatos de tablas en Excel las cuales se procesó para poderla incorporar al ambiente de trabajo según los requerimientos que la aplicación de minería de datos exige. (Esteban Schab, Ramiro Rivera, Luciano Bracco, Facundo Coto, Patricia Cristaldo, Lautaro Ramos, Natalia Rapesta, Juan Pablo Núñez, Soledad Retamar, Carlos Casanova, Anabella De Battista, , 2018)

MACHINE LEARNING

Es una extensión de la Inteligencia Artificial encargada de generar algoritmos con la capacidad de aprender sin la necesidad de programarlos de forma explícita. Para la persona encargada de desarrollar, no le tomará mucho tiempo el tomar en cuenta todos los posibles escenarios ni sus excepciones. Solo se debe cargar el algoritmo con una gran cantidad de datos para que aprenda y sepa que tarea realizar en cada uno de los casos que enfrente. (Judith Sandoval, 2018)

MACHINE LEARNING, APLICACIÓN

Acorde a la preparación de la base de datos con Weka, se realizó la minería de datos correspondiente, para poder clasificar se utilizan técnicas de machine learning. Machine learning significa aprendizaje de máquinas, que se traduce como aprendizaje automático. Son técnicas con las cuales los sistemas aprenden por sí mismos, de manera automática, identifican patrones complejos en grandes volúmenes de datos. Ejemplos de estas técnicas se observan en los automóviles de conducción automática de Google, recomendaciones que llegan a diario en nuestras redes o bandejas de entradas de correos guiándose por nuestras preferencias, detecciones de fraudes. (Especial, Spark, Gt, E, & Gls, 2018)

En este estudio, para poder identificar los factores que son de mayor relevancia que influyen en el desarrollo de gingivitis y caries, se aplicó un algoritmo que ejecuta técnicas de Machine Learning y es conocido como Redes Bayesianas.

PROBABILIDADES CONDICIONALES Y TEOREMA DE BAYES

En la teoría de probabilidades y debido a la demanda de diversos campos de la ciencia como la medicina o biología en general, es de sumo interés establecer relaciones de dependencia con el objetivo de diagnosticar o determinar causas de eventos.

Para esto, se suele recurrir a la llamada probabilidad condicional, definida en (Montgomery & Runger, 1994) como:

La probabilidad condicional de un evento B, dado un evento A tal que $P(A) > 0$, se denota con $P(B|A)$ y se calcula como:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Así, dos variables nominales o categóricas, pueden relacionarse mediante una tabla de contingencia y calcular las probabilidades condicionales en el margen, permitiendo generar conclusiones inmediatas.

Con la potencialidad de extender su alcance, se suele utilizar el llamado “Teorema de Bayes” (Montgomery & Runger, 1994), que especifica lo siguiente:

Sean E_1, E_2, \dots, E_k , k eventos mutuamente excluyentes y exhaustivos, y B , uno de dichos eventos tal que $P(B) > 0$, entonces:

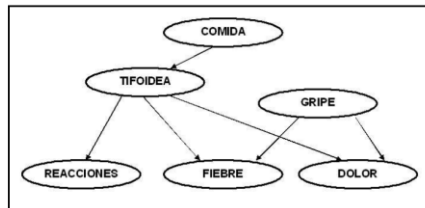
$$P(E_1|B) = \frac{P(B|E_1) P(E_1)}{P(B|E_1) P(E_1) + P(B|E_2) P(E_2) + \dots + P(B|E_k) P(E_k)}$$

Esta teoría, es la base de los algoritmos utilizados por las redes bayesianas.

REDES BAYESIANAS

Se constituyen en una representación gráfica que indica las dependencias entre diversas variables para generar un razonamiento probabilístico (Sucar, 2011). Un ejemplo de red bayesiana se presenta en el Gráfico 5.

Gráfico 5 – Red bayesiana



Fuente: (Sucar, 2011)

Los arcos con flechas indican las relaciones entre los nodos, de esta manera, dado que existe un arco que parte del nodo “COMIDA” hacia el nodo “TIFOIDEA”, se infiere una relación de dependencia entre el tipo de alimentación y la probabilidad de adquirir tifoidea.

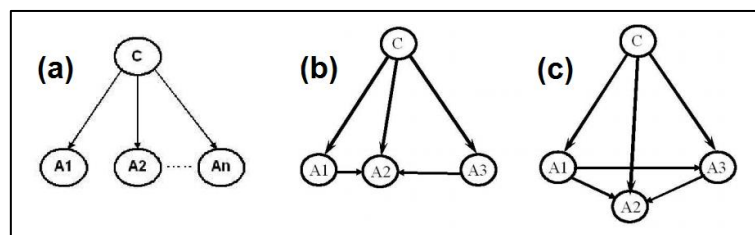
Este tipo de representación no necesariamente mostrará una dependencia absoluta, pero presenta probabilidades calculadas que por lo general son cercanas a la unidad y que permiten generar evidencia de relaciones que, hasta el momento previo de la aplicación, sólo resultaban del conocimiento empírico.

Para simplificar la construcción de una red, es posible utilizar clasificadores que producen resultados basados en la maximización de medidas evaluadoras como la precisión, especificidad, sensibilidad, entre otros, dependiendo del objetivo que se persigue con la creación del modelo en la investigación.

Entre los clasificadores más comunes, tenemos:

- Clasificador bayesiano simple: Conformados por una clase C y atributos $A_1, A_2, A_3, \dots A_k$ que son dependientes de dicha clase.
- Clasificador bayesiano simple aumentado en un árbol (TAN). Que mantiene a los atributos en el mismo nivel, pero genera relaciones condicionales probabilísticas entre ellos.
- Clasificador bayesiano simple aumentado en una red (BAN). Que genera un nuevo sub-nivel de la red, entendiéndose como nuevas dependencias convergentes entre los mismos atributos de los que se dispone.

Gráfico 6 – Ejemplos de clasificadores bayesianos (a) simple, (b) TAN (c) BAN



Fuente: (Sucar, 2011)

TÉRMINOS ODONTOLÓGICOS: CARACTERÍSTICAS

CAVIDAD BUCAL

Es la abertura corporal situada en región inferior de la cara por donde el ser humano ingiere los alimentos. Esta constituye el inicio del aparato digestivo y está constituido por algunos elementos como los dientes, la lengua, el paladar, entre otras (Marín, 2015).

SALIVA

Es un líquido alcalino, claro y viscoso. Sus funciones principales son: humedecer y lubricar los alimentos; arrastrar desechos alimenticios, lo que contribuye a una limpieza de los dientes (Marín, 2015).

PIEZAS DENTALES

Se las puede definir como un cuerpo duro que, engastado en las mandíbulas del hombre y muchos animales, queda descubierto, para servir como órgano de masticación o de defensa (Chávez-Reátegui & Manrique, 2016).

CARIES

La caries dental es la destrucción de los tejidos de los dientes causada por la presencia de ácidos producidos por las bacterias de la placa depositada en las superficies dentales, para la detención de esta patología existe un sistema que se lo conoce como **índice de caries dental**, la gravedad de la enfermedad se da a notar a través de indicadores de nivel poblacional, el índice empleado para cuantificar la gravedad de las caries es CPOD y CEOD en los niños. El símbolo **C**

significa el número de dientes con presencia de caries no restauradas, **P** se refiere a los dientes perdidos y **O** son los dientes restaurado (Flores Ortega, 2018)

En el Gráfico 7 se aprecia un diente afectado por esta enfermedad, la imágenes sobre Caries, tambien pueden ser tratadas por el proceso de mineria de datos para obtener informacion relevante.

Gráfico 7 – Diente afectado por caries



Elaborado por: Dozenist

Fuente: Copyright 2006 por Dozenist, Madrid

GINGIVITIS

La gingivitis, muy relacionada con la enfermedad periodontal, es otra de las patologías que tiene como característica principal inflamación en las encías, las cuales presentan un cambio en su coloración, pasando de un rosa pálido a un rojo brillante, otra de las características es el sangrado que presenta la zona afectada. Esto se ocasiona por la acumulación de placa dental alrededor de la parte que representa el margen gingival (Marín & Pelaez, 2015).



Elaborado por: Kirsten J Wade, Alison M Meldrum

Fuente: (J & M, 2011)

FUNDAMENTACIÓN LEGAL

COIP (CÓDIGO ORGÁNICO INTEGRAL PENAL)

Artículos del Código Penal vigente que abordan la mala práctica médica.

Art. 424 de la Constitución de la República establece que “La constitución es la norma suprema y prevalece sobre cualquier otra del ordenamiento jurídico. Las normas y los actos del poder público deberán mantener conformidad con las disposiciones constitucionales; en caso contrario carecerán de eficacia jurídica”.

Art. 156 indica que los médicos, enfermeras, farmacéuticas, practicantes, empleados de casas de salud o propietarios de farmacias o droguerías que, desobedeciendo órdenes de autoridad competente, paralizaren los servicios o se abstuvieren de prestar su colaboración a los que necesitaren de ellos, serán reprimidos con prisión de uno a cinco años y multa de treinta y cinco a setenta dólares de los Estados Unidos de Norteamérica. Se aplicará el máximo de las penas previstas en este artículo a los miembros de las organizaciones profesionales que hubieren incitado a la comisión de tales hechos, si éstos se hubieren consumado.

Art. 346 indica que el médico, cirujano o practicante que, por favorecer a alguno hubiere certificado falsamente enfermedades o imposibilidades propias para dispensar de un servicio debido legalmente, o de cualquiera otra obligación impuesta por la Ley, o para exigir o reclamar un derecho, será reprimido con prisión de seis meses a dos años y multa de seis a sesenta y dos dólares de los Estados Unidos de Norte América si ha sido movido por dones o promesas, será reprimido con un prisión de uno a cinco años, a más de la multa antes indicada.

Art. 434 indica que, cuando los actos previstos en los artículos anteriores fueron cometidos por imprudencia, o por negligencia, o por impericia en el propio arte o profesión, o por inobservancia de los reglamentos u ordenanzas, se impondrá multa de ocho a setenta dólares de los Estados Unidos de Norte América, si no

resultare enfermedad o muerte de alguna persona; y prisión de seis meses a cinco años, si resultare enfermedad o muerte.

Art. 435 indica que, será reprimido con prisión de seis meses a dos años y multa de ocho a setenta y siete dólares de los Estados Unidos de Norte América, el que violare las medidas adoptadas por las autoridades competentes para impedir la introducción o propagación de una epidemia.

Art. 436 indica que, los médicos, botánicos, o cualquier persona que, por falta de precaución o de cuidado, recetaren, despacharen o suministraren medicamentos que comprometan gravemente la salud, serán reprimidos con prisión de seis meses a un año; si hubieren causado enfermedad que parezca o fuere incurable, la prisión será de unos a tres años; y en caso de haber producido la muerte, la prisión será de tres a cinco años.

Art. 456 prescribe que, si las sustancias administradas voluntariamente, que pueden alterar gravemente la salud, han sido dadas sin intención de causar la muerte, pero la han producido, se reprimirá al culpado con reclusión menor de tres a seis años.

Art. 457 prescribe que, en la infracción mencionada en el artículo anterior, se presumirá la intención de dar la muerte si el que administró las sustancias nocivas es médico, farmacéutico o químico; o si posee conocimientos en dichas profesiones, aunque no tenga los títulos o diplomas para ejercerlas.

FOE (FEDERACIÓN ODONTOLÓGICA ECUATORIANA)

Según el código de ética de la FOE (Federación Odontológica Ecuatoriana) nos indica en sus artículos lo siguiente:

Art. 1 El presente Código, de acuerdo con la disposición Transitoria Tercera de la Ley de Federación Odontológica Ecuatoriana (FOE), para el Ejercicio, Defensa y

Perfeccionamiento Profesional, establece las normas de conducta para el Odontólogo en sus relaciones con:

1. El Estado;
2. La Sociedad;
3. La Federación Odontológica Ecuatoriana;
4. Entre Odontólogos y Profesionales afines;
5. Pacientes; y,
6. Personal Para odontológico.

Art. 2 El presente artículo nos indica las normas de acción que debe tener un profesional en la Odontología según el código de ética profesional para odontólogos de la FOE.

1. El profesional Odontólogo es una persona que brinda servicios a la Sociedad y debe someterse a las exigencias que se presenten de la naturaleza y dignidad humana.

2. El Odontólogo debe propender a la incrementación de sus conocimientos y de su cultura general, desempeñando una intachable honestidad en el desempeño de su actividad, para así, poder mantener y aumentar el prestigio de su profesión.

3. El odontólogo en las actividades que realiza deberá emplear sus conocimientos de manera que brinde seguridad al paciente y no de manera ilegal o inmoralmemente. Deberá hacer respetar su profesión y procederá en todo momento con la debida prudencia y honorabilidad que la sociedad exige.

4. En la realización del ejercicio de su profesión no tendrá preferencias de raza, nacionalidad, religión, nivel socioeconómico, ni de convicciones políticas.

5. La conducta que muestre el odontólogo tendrá una gran influencia y repercutirán en el prestigio de la profesión y deberá ser defendida de todo comentario desfavorable.

En el capítulo III de la FOE se prescriben dos artículos que se regulan las relaciones de los odontólogos con el estado, entidades públicas y privadas.

Art. 3 Ningún Odontólogo permitirá que sus servicios profesionales, su nombre o su silencio faciliten o hagan posible la práctica ilegal de la odontología, sea con carácter particular o en entidades públicas, semipúblicas o privadas; el incumplimiento de esta disposición será considerada como falta grave. Se sancionará con lo dispuesto en el literal d. del Art. 25 de la Ley de Federación.

En caso de reincidencia se aplicará lo indicado en el literal e. Del mismo artículo de la Ley.

Art. 4 El Odontólogo que se halle en desempeño de sus funciones profesionales en instituciones de derecho público semi-público o privado y se le exigieren actividades que no correspondan a las obligaciones inherentes al cargo, podrá negarse a cumplirlas solicitando la intervención del respectivo Colegio Odontológico Provincial.

LEY ORGÁNICA DE EDUCACIÓN SUPERIOR

Art. 32.- Programas informáticos. - Las empresas que distribuyan programas informáticos tienen la obligación de conceder tarifas preferenciales para el uso de las licencias obligatorias de los respectivos programas, a favor de las instituciones de educación superior, para fines académicos.

Las instituciones de educación superior obligatoriamente incorporarán el uso de programas informáticos con software libre.

DECRETO 1014

SOBRE EL USO DEL SOFTWARE LIBRE

Art. 1: Establecer como política pública para las entidades de administración Pública central la utilización del Software Libre en sus sistemas y equipamientos informáticos.

Art. 2: Se entiende por software libre, a los programas de computación que se pueden utilizar y distribuir sin restricción alguna, que permitan el acceso a los códigos fuentes y que sus aplicaciones puedan ser mejoradas.

Estos programas de computación tienen las siguientes libertades:

- Utilización de programa con cualquier propósito de uso común.
- Distribución de copias sin restricción alguna.
- Estudio y modificación de programa (Requisito: código fuente disponible)
- Publicación del programa mejorado (Requisito: código fuente disponible)

Art. 3: Las entidades de la administración pública central previa a la instalación del software libre en sus equipos, deberán verificar la existencia de capacidad técnica que brinde el soporte necesario para este tipo de software.

Art. 4: Se faculta la utilización de software propietario (no libre) únicamente cuando no exista una solución de software libre que supla las necesidades requeridas, o cuando esté en riesgo de seguridad nacional, o cuando el proyecto informático se encuentre en un punto de no retorno.

Art. 5: Tanto para software libre como software propietario, siempre y cuando se satisfagan los requerimientos.

Art. 6: La subsecretaría de Informática como órgano regulador y ejecutor de las políticas y proyectos informáticos en las entidades de Gobierno Central deberá realizar el control y seguimiento de este Decreto.

Art. 7: Encargue de la ejecución de este decreto los señores Ministros Coordinadores y el señor Secretario General de la Administración Pública y Comunicación.

PREGUNTA CIENTÍFICA PARA RESPONDER

¿El uso de un modelo predictivo para la gestión de la información recopilada por los profesionales odontólogos, ayudaría a la toma de decisiones, clasificación y manejo de datos en lo referente a las patologías estudiadas tales como caries y gingivitis?

CAPÍTULO III

METODOLOGÍA DE LA INVESTIGACIÓN

Se procuró que la información sea lo más reciente posible, para tener una correcta visión de la situación real del tema propuesto, en el caso de la investigación presente, mediante un modelo de datos y haciendo uso de la técnica conocida como minería de datos, se obtuvo un conjunto de resultados, los cuales pueden influir positivamente en la toma de decisiones.

DISEÑO DE LA INVESTIGACIÓN

MODALIDAD DE LA INVESTIGACIÓN

El diseño del presente trabajo es no experimental de tipo transeccional ya que en el estudio no se manipulan de manera intencional las variables y se realiza en un periodo determinado (2018 - 2019).

La investigación bibliográfica del trabajo realizado está constituida por artículos con un enfoque similar al propuesto. Así, la disminución del gasto médico gracias a los modelos predictivos y la revisión de trabajos relacionados con la predicción de posibles tipos de cáncer de senos, son estudios que constituyen una base valedera para el desarrollo del trabajo presente.

Por lo antes expuesto, este tipo de investigación puede definirse como sugiere Matos Ayala (2018):

“La **investigación bibliográfica** o documental consiste en la revisión de material bibliográfico existente con respecto al tema a estudiar. Se trata de uno de los principales pasos para cualquier investigación e incluye la selección de fuentes de información. Se le considera un paso esencial porque incluye un conjunto de fases que abarcan la observación, la indagación, la interpretación, la reflexión y el análisis para obtener bases necesarias para el desarrollo de cualquier estudio.”

TIPO DE INVESTIGACIÓN

Exploratorio

La investigación se basó en información proporcionada por la FPO tales como, datos en los que se aplicaron técnicas de minería de datos, para evidenciar su comportamiento.

El tema propuesto en el presente trabajo no ha sido muy explotado en los diversos campos de estudios, teniendo gran potencial por la facilidad de generar puntos de vista claros con información local, cabe recalcar que toda la información se debió pasar por un proceso y que hubo elementos que no fueron de gran ayuda a la hora de ejecutar el algoritmo. Con la investigación se pretende darle otro enfoque a los datos que genera la Facultad y utilizarlos para mejorar el servicio que prestan.

POBLACIÓN Y MUESTRA

Los datos corresponden a pacientes de la Facultad Piloto de Odontología. Se trata de pacientes (varones y mujeres) con edades comprendidas desde los 4 a 97 años que se realizan chequeos, revisiones, operaciones, en la ciudadela Universitaria (Guayaquil), se lograron identificar reglas que determinaron sus patrones con respecto a patologías de gran incidencia en dicha población las cuales son Caries y Gingivitis.

La muestra consta de 10298 pacientes que acudieron a la FPO en un periodo comprendido entre mayo 2018 a marzo 2019, por conveniencia para el estudio se trabajó con la totalidad de los datos.

DESCRIPCIÓN Y CODIFICACIÓN DE LAS VARIABLES DEL ESTUDIO

A continuación, se muestra una breve descripción de cada variable de las cuales se nutre la base de datos.

Cuadro 4 - Variables de pacientes en las bases de datos

No	Variable	Descripción
1	Número de historia clínica	Número único ligado a cada paciente. No existen valores repetidos a lo largo de toda la base
2	Cédula	Número conformado por 10 dígitos para identificación de cada persona. No existen valores repetidos a lo largo de toda la base
3	Sexo	Género del paciente atendido. Masculino (M), Femenino (F)
4	Edad	Número que representa la edad del paciente
5	Ubicación sectorial	Sector de la ciudad de Guayaquil en la cual habita el paciente, o lugar del cual proviene. Se registra una mezcla de sectores de Guayaquil, otras ciudades y provincias del Ecuador además de otros países.
6	Problema actual	Condición de sintomático (SI) o asintomático (NO) del paciente
7	Antecedentes Personales	Descripción de posibles afecciones previas. Codificado como SI/NO
8	Alergia a antibiótico	Condiciones verificadas con una "X", o un espacio en blanco si no se presenta. Codificada como SI/NO
9	Alergia a anestesia	
10	Hemorragias	
11	VIH/SIDA	
12	Tuberculosis	
13	Asma	
14	Diabetes	
15	Hipertensión	
16	Enfermedad Cardíaca	
17	Examen estomatognático	Línea de texto con la descripción del resultado
18	Enfermedad periodontal	Campo que admite cuatro niveles: no (no presenta), leve, moderada y severa.
19	Grado de Maloclusión	Grado de maloclusión de dientes: 0 (no presenta), 1, 2, 3
20	C, P, O / Total	Números de Caries (C), dientes perdidos (P) y obturados (O) en un paciente, además de la suma por paciente (Total)
21	Plan de tratamiento	Marcas indicadas con "X" de la prescripción médica sugerida. Comprende profilaxis, periodoncia, endodoncia, etc.
22	Observación	Línea de texto que indica observaciones adicionales
23	Fecha	Fecha de atención en el formato Día/Mes/Año
24	Diagnóstico revisado por / Alumnos del diagnóstico	Líneas de texto indicando el nombre del doctor encargado de revisar los resultados, así como el nombre de los alumnos practicantes

Elaborado por: Danny Ascencio Moreno

Fuente: Datos de la investigación

VARIABLES DE LA INVESTIGACIÓN

En el Cuadro 5 se definen las variables que conforman la investigación realizada, con el fin de establecer las relaciones entre ellas adecuadamente.

Cuadro 5 - Operacionalización de variables

Tipo de variable	Independiente	Dependiente
Variable	Modelo Datos	Presencia de Caries, Gingivitis
Indicadores	Clasificación y manejo de la información recopilada por odontólogos.	Determinación de diagnósticos y nivel de afectación de las patologías. Modelo predictivo en base a los datos históricos recopilados por la FPO Detección de antecedentes que influyan en la aparición de las patologías estudiadas
Técnicas y/o Instrumentos	Fichas médicas, Odontogramas	Revisión en la base de datos adquirida, depuración de la base de datos, uso de WEKA para establecer el modelo predictivo

Elaborado por: Danny Ascencio Moreno

Fuente: Información de la Investigación del Proyecto

INSTRUMENTOS DE RECOLECCIÓN DE DATOS

Entrevista a los administradores de la base

Se realizó varias visitas a la FPO, y se dialogó con el Dr. Patricio Ventimilla, para conocer acerca de los procedimientos que realizan los médicos para la recolección de datos y también las patologías que estudian o registran. El Dr. Manuel Gonzales, aclaró otras dudas adicionales del formato con el que guardan la base de datos en los registros, también la cantidad de extensiones que dispone la FPO, indicó las distintas salas de atención con que disponen. Con esa información recopilada, se procedió a realizar una carta de aprobación para acceder a la base de datos al Decano el Dr. José Fernando Franco Valdiviezo y posterior hablar con el Dr. Fausto Pilco para que permita el acceso al archivo.

Se utilizó este método para obtener una mejor comprensión de los temas tratados, despejar dudas y generar conocimientos acerca del tema y sus procedimientos o métodos a seguir por los médicos para la creación de documentos que almacenen los datos de los pacientes y diagnóstico con respecto a las patologías tratadas. Las entrevistas realizadas con los médicos ofrecen una idea clara de los procesos que emplean en estas áreas, esclarecen ideas tomadas de documentos científicos para poder modelarlo en tablas las cuales, guardará los datos más relevantes para su posterior explotación.

Las preguntas ayudan a que el contexto del tema no se pierda de vista, es esto lo que diferencia claramente, en dicha conversación dos roles claros, los cuales son el entrevistador y el medico a cargo del equipo de trabajo con el cual se esclarecen los elementos que interactuaran con el modelo de datos.

Lectura Científica

Esta se define como un estado de lectura avanzado, su objetivo principal se denota en el conocimiento de la trascendencia y veracidad de un enunciado o texto, en otras palabras, se refiere a la interpretación y crítica de fuentes científicas, la cual permite una correcta interpretación de lo leído bajo tres condiciones las cuales son espacio, tiempo y persona.

La lectura científica aportará elementos adicionales e inclusive sirve de ayuda para generar preguntas que realizar para una próxima entrevista, con los elementos que interactúan directamente con los datos que se están recopilando.

FILTRADO DE INFORMACIÓN DE LA BASE DE DATOS OBTENIDA

Una vez obtenida la base de datos (en formato Excel) se procedió a su respectiva depuración.

Grafico N. 8 Base de datos en Excel, datos en bruto

Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

La base de datos obtenida se presenta con una gran cantidad de información faltante y fuera de lugar, con elementos que no corresponden a su respectiva columna, por lo cual se procedió a descartar filas (pacientes) que no tienen información consistente a la investigación.

Gráfico 8 – Vista de los campos hallados en la base de datos

1	UNIVERSIDAD DE GUAYAQUIL									
2	FACULTAD PILOTO DE ODONTOLOGIA									
3	DEPARTAMENTO DE DIAGNÓSTICO									
4										
5	DIAGNÓSTICOS, INDICES DE C.P.O. y PLAN DE TRATAMIENTO									
6	REPORTE DIARIO									
7										
8	SEMESTRE : 1ero (MAYO 14 a SEPTIEMBRE 11 DEL 2018)									
9										
10	N° H.C.	CEDULA	SEXO	EDAD	UBICACIÓN SECTORIAL	ASINTOMATICO	PROBLEMA ACT	SINTOMATICO (Dolor)	ANTECEDENTES PERSONALES / (Referencia personal y familiar)	NO REFIERE
31	85808	850037391	F	22	CENTRO	X				X
32	86467	922829239	F	35	SUR	X				X
33	86836	908319304	F	36	SUR	X				X
34	89764	130325843	F	72	NORTE			X		X

Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

Se eliminaron filas en blanco y se unieron unas con otras para mantener continuidad en los datos.

Gráfico 9 - Espacios en blanco eliminados.

Nº	IDMOS	LOGOTIPO	T	E	SIN FONIA		X		I		X	ETNOLOGIO ATMI	MODERADO	
95	104699	602029662	F	51	NORTE	X						X	SIN PATOLOGIA APARENTE	LEVE
96	11	F:9 *M:2												
97														
98														
99	85242	923302194	M	38	NOROESTE	X						X	SIN PATOLOGIA APARENTE	LEVE
100	85678	1703970655	M	63	NOROESTE	X						X	SIN PATOLOGIA APARENTE	MODERADO
101	89419	930966346	M	26	NORTE	X						X	11 ATM CHASQUIDO	MODERADO

Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

Algunos datos encontrados no concuerdan con lo descrito por su columna, tal es el caso de ubicación sectorial, en el que constan nombres de países (Venezuela) y ciudades (New York), si bien es cierto que estos elementos forman parte de la población tratada por la FPO, no se deberían establecer en esa columna. Se procedió a fusionar todos los sectores de Guayaquil y desechar las filas con localidades con menos de 50 personas.

Gráfico 10 – Reconocimiento de diferentes ciudades en el campo de ubicación

UBICACIÓN SECTORIAL
NORTE
SUROESTE
NORTE
SUR OESTE
MACHALA
VENEZUELA+
CENTRO
SUR
NORTE
NORTE
NORTE

Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

Los elementos deben ser homogéneos, ya que la aplicación considera un “SUR” diferente a “sur”, o “Sur”. Se realizó la estandarización de las categorías presentes de forma manual, utilizando la base de datos. Además, como se muestra en el Gráfico 11, los espacios en blanco se eliminan o se reemplazan con un dato que se pueda procesar por la herramienta seleccionada para realizar la Minería de Datos, la herramienta seleccionada (Weka) no admite celdas vacías.

Gráfico 11 - Espacios en blanco llenados

[illegible]

Elaborado por: Danny Ascencio Moreno
Fuente: Información de la base de datos

Posterior a esto, se pudo notar un grupo de variables que no aportan información importante al estudio que se plantea, tal es el caso de las variables:

- Historia-clínica y número de cédula por no representar secuencias ni repetirse.
- Planes de tratamiento, que no se considera al no conocer si el paciente siguió o no la prescripción indicada, por lo cual, al no haber seguimiento, la variable se vuelve técnicamente irrelevante.
- Alumnos o doctores involucrados, dado que el objetivo es evaluar los datos y las riquezas que puedan ofrecer, no conocer el desempeño de los doctores o los practicantes.
- C, P, O, dado que están consideradas en la variable Total (CPO), la cual fue reorganizada en cuatro rangos (R1 menor a 8, R2 entre 8 y 16, R3 entre 16 y 24, R4 mayor a 24).
- Observación y examen estomatognático, porque no se presentan con una codificación uniforme. Su recodificación, puede ser sujeto de investigaciones posteriores.

Una vez depurada, la base de datos tiene el siguiente aspecto:

Gráfico 12 - Presentación de la información total lista para la aplicación de las técnicas de minería de datos

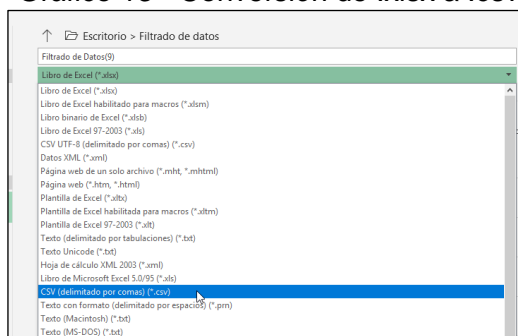
FECHA	SEXO	EDAD	CIUDAD	PROBLE-AC	ACTI-ANT	PERSON-ALER	ANTIBI	ALER-ANESTI	HEMORRAGI	VHI-SIDA	TUBERCULO	ASMA	DIABETES	HIPERTENSIC	ENF-CARDIA	ANT-FAMILI	MALOC-GRAT	CPO_T	ENF-PERIODON
9/4/2018	F	48	DAULE	ASINTOMATI	SI	NO	NO	NO	NO	NO	NO	NO	SI	NO	NO	NO	2 RANGOS	SI	
9/4/2018	F	20	GUAYAQUIL	ASINTOMATI	SI	NO	NO	NO	NO	NO	NO	NO	SI	NO	NO	NO	1 RANGOS	NO	
9/4/2018	F	15	GUAYAQUIL	SINTOMATI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	SI	1 RANGOS	NO	
9/4/2018	F	16	GUAYAQUIL	ASINTOMATI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	1 RANGOS	NO	
9/4/2018	M	73	GUAYAQUIL	ASINTOMATI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	3 RANGOS	SI	
9/4/2018	F	13	GUAYAQUIL	ASINTOMATI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	2 RANGOS	NO	
9/4/2018	M	66	GUAYAQUIL	ASINTOMATI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	1 RANGOS	SI	
9/4/2018	M	21	DURAN	ASINTOMATI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	2 RANGOS	SI	
9/4/2018	F	50	GUAYAQUIL	ASINTOMATI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	1 RANGOS	SI	
9/4/2018	M	28	GUAYAQUIL	ASINTOMATI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	2 RANGOS	NO	
9/4/2018	F	67	DURAN	ASINTOMATI	SI	SI	NO	NO	NO	NO	NO	NO	SI	SI	NO	NO	1 RANGOS	SI	
9/4/2018	M	25	GUAYAQUIL	ASINTOMATI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	1 RANGOS	SI	
9/4/2018	F	47	GUAYAQUIL	ASINTOMATI	SI	SI	NO	NO	NO	NO	NO	NO	NO	NO	NO	SI	2 RANGOS	NO	
9/4/2018	F	24	GUAYAQUIL	ASINTOMATI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	2 RANGOS	SI	
9/4/2018	F	32	GUAYAQUIL	ASINTOMATI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	2 RANGOS	SI	
9/4/2018	F	10	GUAYAQUIL	SINTOMATI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	2 RANGOS	NO	
9/4/2018	F	11	GUAYAQUIL	ASINTOMATI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	1 RANGOS	NO	
9/4/2018	M	12	DAULE	ASINTOMATI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	1 RANGOS	NO	
10/4/2018	F	9	GUAYAQUIL	ASINTOMATI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	2 RANGOS	NO	
10/4/2018	F	4	GUAYAQUIL	ASINTOMATI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	2 RANGOS	NO	
10/4/2018	M	23	GUAYAQUIL	ASINTOMATI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	2 RANGOS	NO	
10/4/2018	F	21	GUAYAQUIL	ASINTOMATI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	2 RANGOS	NO	
10/4/2018	M	15	DURAN	ASINTOMATI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	2 RANGOS	NO	
10/4/2018	F	22	GUAYAQUIL	ASINTOMATI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	1 RANGOS	NO	
10/4/2018	F	20	GUAYAQUIL	ASINTOMATI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	1 RANGOS	NO	

Elaborado por: Danny Ascencio Moreno
Fuente: Información de la base de datos

INGRESO DE DATOS A LA APLICACIÓN WEKA

Luego de preparar los datos, el archivo está listo para trabajar con la herramienta para la minería de datos, ahora es necesario convertir el archivo .xlsx a .csv que corresponde al formato de lectura de información externa en Weka.

Gráfico 13 - Conversión de .xlsx a .csv



Elaborado por: Danny Ascencio Moreno
Fuente: Información de la base de datos

El documento se almacenará con un formato que está delimitado por “;” se debe reemplazar ese carácter por la “,” que es el delimitador con el que trabaja Weka

Gráfico 14 - Ajuste del carácter delimitador para la aplicación

SEXO,EDAD,UBICACION-SEC,Problema-Actual,ANT-PERSONAL,ALER. ANTIBIO,ALER. ANESTESIA ;HEMORRAGIAS ;VIH-SIDA;TUBERCULOSIS ;ASMA;t

F,48,Daule,ASINTOMATICO,SI,NO,NO,NO,NO,NO,NO,SI,NO,NO,MODERADA,2,5,13,1,19,9-abr-18

F,20,SUR,ASINTOMATICO,SI,NO,NO,NO,NO,NO,NO,SI,NO,NO,NO,NO,1,9,0,0,9,9-abr-18

F,15,norte,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,SI,NO,1,12,1,1,14,9-abr-18

F,16,SUR,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,1,13,0,0,13,9-abr-18

M,73,sur,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,MODERADA,3,1,23,1,25,9-abr-18

F,13,Norte,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,2,3,0,0,3,9-abr-18

M,66,Suroeste,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,MODERADA,1,16,6,6,28,9-abr-18

M,21,Duran,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,LEVE,2,12,1,1,14,9-abr-18

F,50,norte,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,MODERADA,1,11,5,6,22,9-abr-18

M,28,Norte,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,2,1,0,16,17,9-abr-18

F,67,Duran,ASINTOMATICO,SI,SI,NO,NO,NO,NO,SI,SI,NO,NO,LEVE,1,13,5,1,19,9-abr-18

M,25,noroeste,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,MODERADA,1,10,1,1,12,9-abr-18

F,47,SUR,ASINTOMATICO,SI,SI,NO,NO,NO,NO,NO,NO,NO,SI,NO,2,11,10,3,24,9-abr-18

F,24,SUR,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,LEVE,2,5,1,9,15,9-abr-18

F,32,NOR-OESTE,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,LEVE,2,8,0,2,10,9-abr-18

F,10,Norte,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,2,4,0,0,4,9-abr-18

F,11,NOR-ESTE ;ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,1,3,0,0,3,9-abr-18

M,12,Daule,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,1,8,0,0,9-abr-18

F,9,sur,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,2,1,0,0,1,10-abr-18

F,4,SUR,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,2,0,0,0,0,10-abr-18

M,23,SUR,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,2,3,3,2,8,10-abr-18

F,21,SUR,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,1,7,4,1,12,10-abr-18

M,15,Duran,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,2,1,0,2,3,10-abr-18

Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

El reemplazar los caracteres “;” por “,” permite que la aplicación diferencie cada atributo que se encuentra delimitado por el carácter especificado. Dejando listo el archivo para el ingreso de datos a la aplicación que se encargará de la minería de datos.

Gráfico 15 - Documento listo para exportar a weka

SEXO,EDAD,UBICACION-SEC,Problema-Actual,ANT-PERSONAL,ALER. ANTIBIO,ALER. A

F,48,Daule,ASINTOMATICO,SI,NO,NO,NO,NO,NO,NO,SI,NO,NO,MODERADA,2,5,13,1,19,9-abr-18

F,20,SUR,ASINTOMATICO,SI,NO,NO,NO,NO,NO,NO,SI,NO,NO,NO,NO,1,9,0,0,9,9-abr-18

F,15,norte,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,SI,NO,1,12,1,1,14,9-abr-18

F,16,SUR,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,1,13,0,0,13,9-abr-18

M,73,sur,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,MODERADA,3,1,23,1,25,9-abr-18

F,13,Norte,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,2,3,0,0,3,9-abr-18

M,66,Suroeste,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,MODERADA,1,16,6,6,28,9-abr-18

M,21,Duran,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,LEVE,2,12,1,1,14,9-abr-18

F,50,norte,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,MODERADA,1,11,5,6,22,9-abr-18

M,28,Norte,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,2,1,0,16,17,9-abr-18

F,67,Duran,ASINTOMATICO,SI,SI,NO,NO,NO,NO,SI,SI,NO,NO,LEVE,1,13,5,1,19,9-abr-18

M,25,noroeste,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,MODERADA,1,10,1,1,12,9-abr-18

F,47,SUR,ASINTOMATICO,SI,SI,NO,NO,NO,NO,NO,NO,NO,SI,NO,2,11,10,3,24,9-abr-18

F,24,SUR,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,LEVE,2,5,1,9,15,9-abr-18

F,32,NOR-OESTE,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,LEVE,2,8,0,2,10,9-abr-18

F,10,Norte,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,2,4,0,0,4,9-abr-18

F,11,NOR-ESTE ;ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,1,3,0,0,3,9-abr-18

M,12,Daule,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,1,8,0,0,9-abr-18

F,9,sur,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,2,1,0,0,1,10-abr-18

F,4,SUR,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,2,0,0,0,0,10-abr-18

M,23,SUR,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,2,3,3,2,8,10-abr-18

F,21,SUR,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,1,7,4,1,12,10-abr-18

M,15,Duran,ASINTOMATICO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,NO,2,1,0,2,3,10-abr-18

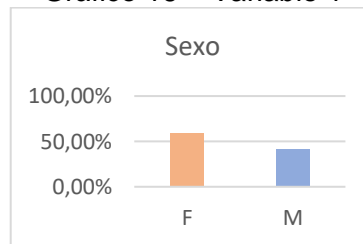
Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

Así queda listo el archivo para trabajar con Weka. Todos los gráficos presentados a continuación son de elaboración propia con base en la información disponible, además, las tablas correspondientes se encuentran en el Anexo 1. De aquí, las variables resultantes analizadas de forma independiente tienen las siguientes características.

VARIABLES DE ENTRADA PARA LA APLICACIÓN WEKA

Gráfico 16 – Variable 1

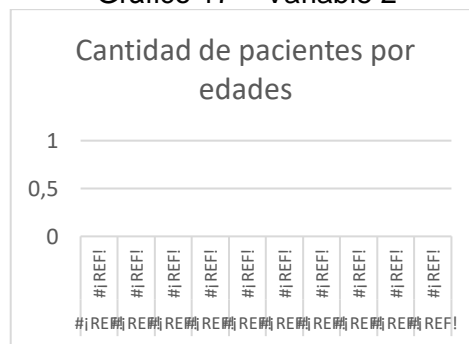


Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

Se aprecia que el 58% de la población tratada corresponde al género femenino, mientras que el 42% restante, al género masculino. Prácticamente 6 de cada 10 personas son de sexo femenino.

Gráfico 17 – Variable 2

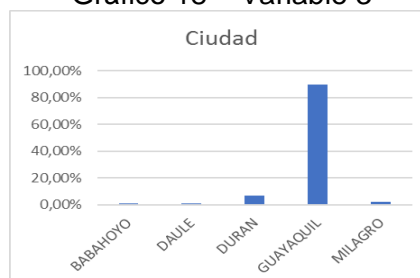


Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

Se evidencia una mayor afluencia de pacientes que comprenden edades entre 13 a 60 años, principalmente de personas con edades comprendidas entre 22 y 31 años.

Gráfico 18 – Variable 3

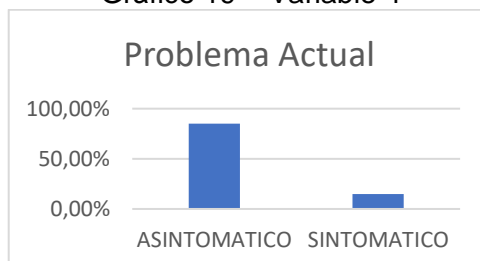


Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

El 89% de los pacientes tratados corresponden a la ciudad de Guayaquil, aunque la FPO ha tenido visita de diversas ciudades, las más significativas (≥ 50 pacientes) son Durán, Milagro, Daule y Babahoyo.

Gráfico 19 – Variable 4

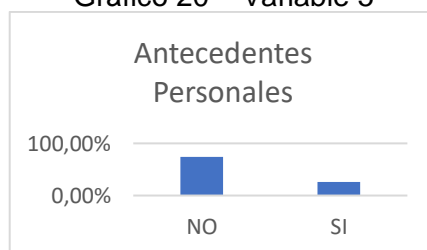


Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

En su mayoría (85%), los pacientes no presentan dolor al momento de su revisión médica.

Gráfico 20 – Variable 5

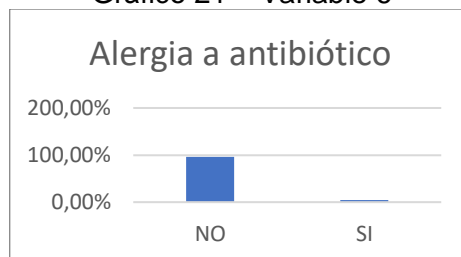


Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

Las personas examinadas en su gran mayoría, un 73% para ser mas precisos, no presentan antecedentes personales.

Gráfico 21 – Variable 6

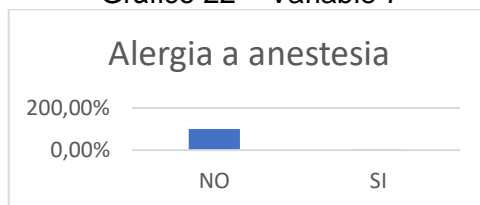


Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

El 95% de los pacientes examinados no tienen una reacción alérgica a algún tipo de antibiótico.

Gráfico 22 – Variable 7

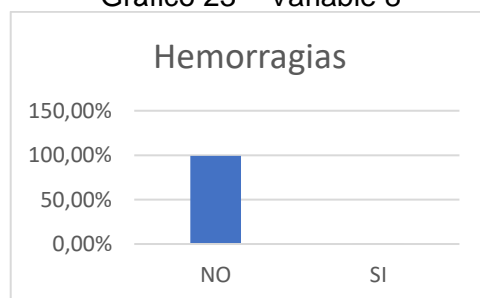


Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

De los examinados el 99% no presentan alergia a la anestesia.

Gráfico 23 – Variable 8

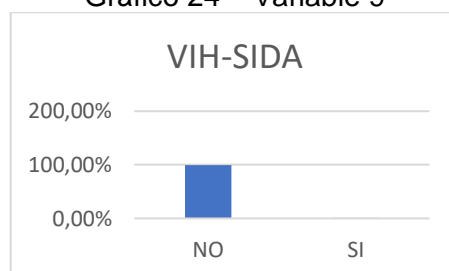


Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

Sólo el 0.63% de los examinados presento algún tipo de hemorragia.

Gráfico 24 – Variable 9

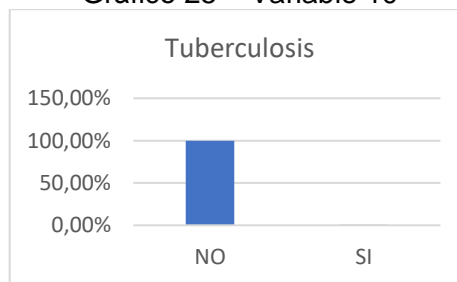


Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

Un 0.12% de la población tratada está infectada por VIH.

Gráfico 25 – Variable 10

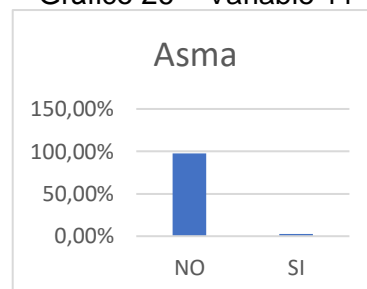


Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

El porcentaje de tuberculosis en los pacientes tratados es relativamente bajo, solo un 0.29%

Gráfico 26 – Variable 11

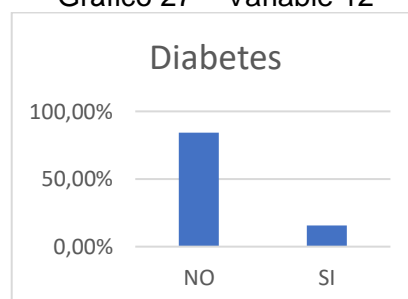


Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

Solo un 2% de las personas tratadas tiene asma

Gráfico 27 – Variable 12

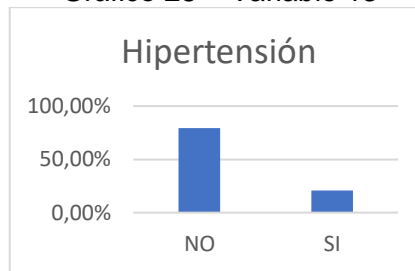


Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

EL 15% de los pacientes tienen diabetes,

Gráfico 28 – Variable 13

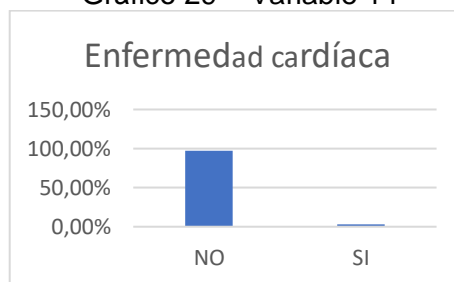


Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

Apenas un 20% de los pacientes tratados tienen hipertensión

Gráfico 29 – Variable 14

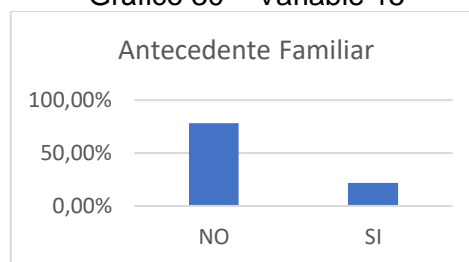


Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

El 2% posee algún tipo de enfermedad cardíaca

Gráfico 30 – Variable 15

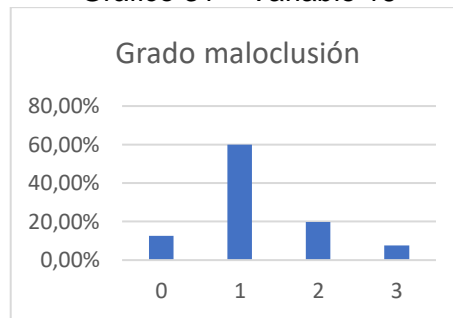


Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

El 21% de la población analiza tiene algún tipo de antecedente familiar

Gráfico 31 – Variable 16

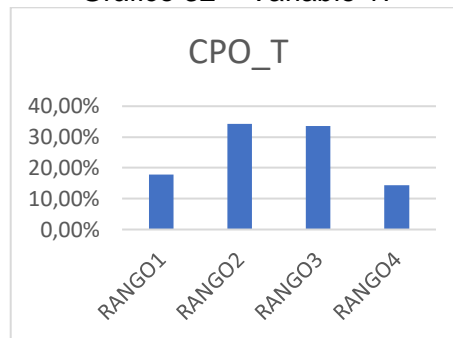


Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

Maloclusión en grado 1 es lo más común hallado en los pacientes tratados con un 59% de la población.

Gráfico 32 – Variable 17

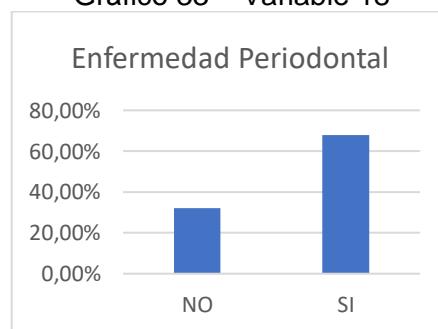


Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

Con respecto al CPO_T, se divide en 4 rangos los cuales se distribuyen de la siguiente manera: RANGO1 representa menos de 8 CPO y comprende el 17.79%, el RANGO2 representa de 9 a 16 CPO siendo un 34.22%, RANGO3 de 17 a 23 CPO, identificando a un 33.58% y el RANGO4 los valores restantes con un 14.41%

Gráfico 33 – Variable 18

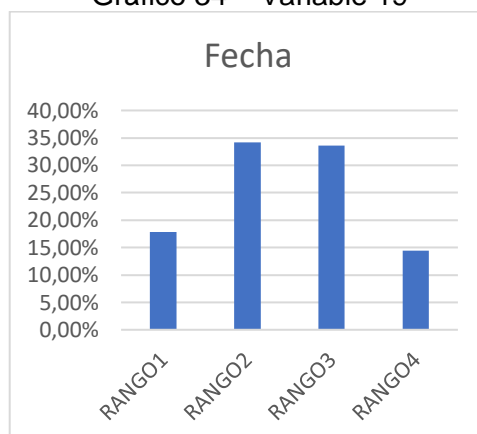


Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

La enfermedad periodontal se refleja en un 67.88% de los pacientes tratados, permitiendo suponer que una gran cantidad de la población se despreocupa de su higiene bucal o no presta atención a los síntomas mínimos.

Gráfico 34 – Variable 19



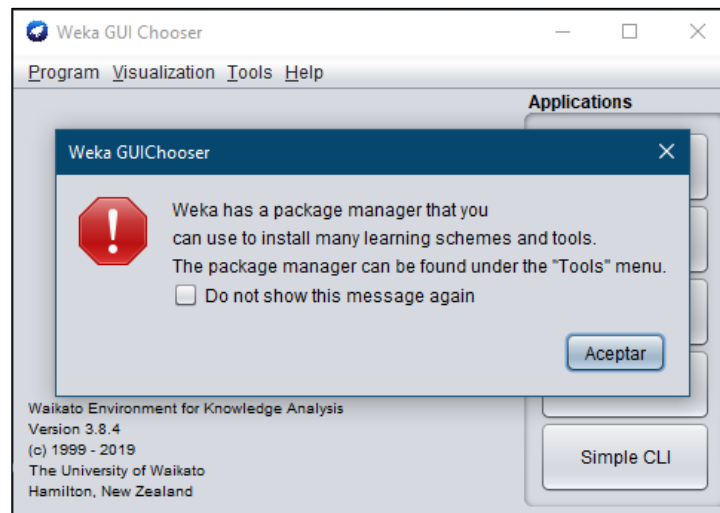
Elaborado por: Danny Ascencio Moreno

Fuente: Información de la base de datos

Se puede evidenciar que el mes de mayor afluencia fue el de noviembre, poca recepción en los meses de abril, incrementando en el mes de mayo, y llegando a un pico en el mes de junio, para decaer gradualmente hasta agosto, bajando de manera drástica en los meses de septiembre y octubre.

Al abrir el programa Weka aparecerá una ventana como se muestra en el **Gráfico N.19**, Se procede a aceptar la ventana que aparece al iniciar el software, la cual indica que se pueden adicionar paquetes con esquemas para aprendizaje que no estén instalados por defecto, en caso de querer adicionar un paquete a la aplicación se realiza desde la pestaña “Tools”

Gráfico 35 - Interfaz de weka



Elaborado por: Danny Ascencio Moreno

Fuente: Datos de la aplicación weka

Una vez iniciado el programa, se presenta la interfaz de Weka con todos sus beneficios, de los cuales es importante para la investigación el primer botón que se muestra en el Gráfico 36, el cual desplegará la ventana **Explorer** un entorno visual, para el uso de los paquetes.

Experimenter Se centra en la automatización de tareas para realizar experimentos a gran escala.

Simple CLI Es una consola de comandos para invocar directamente con java los paquetes de weka.

KnowledgeFlow Con esta opción se puede generar proyectos de minería de datos por medio de flujos de información.

Workbench Es una colección de algoritmos de aprendizaje automático y herramientas de preprocesamiento de datos que incluye prácticamente todos los algoritmos.

Gráfico 36 - Pasos para ejecutar el algoritmo de Bayes



Elaborado por: Danny Ascencio Moreno
Fuente: Datos de la aplicación weka

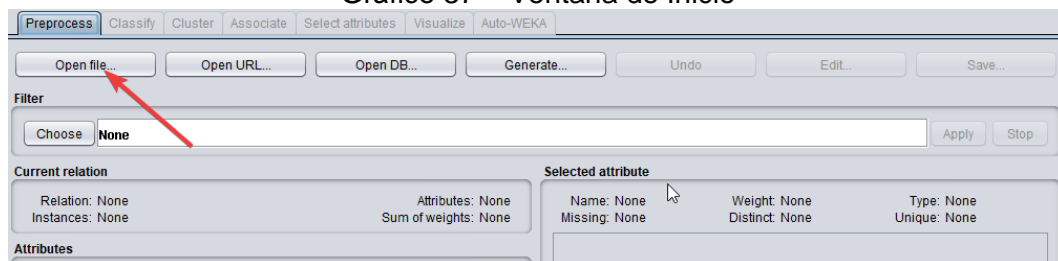
Se presenta una ventana con una serie de pestañas que no están activas por el momento, a continuación, se muestran cuatro botones activos, los cuales permitirán el ingreso de la información a procesarse.

El botón “**Open URL...**” permitirá el ingreso de información a través de una dirección web

El botón “**Open DB...**” admitirá el ingreso a través de una base de datos

El botón “**Open File...**” recibirá un archivo como fuente de datos y es la opción que se escogió para ingresar la base de datos previamente depurada en formato .CSV y delimitado por comas.

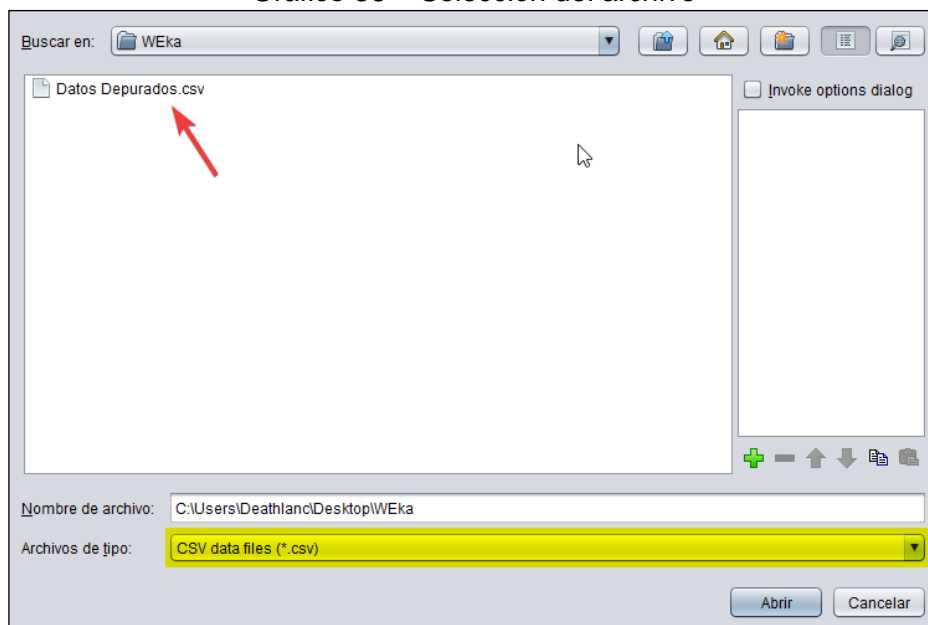
Gráfico 37 – Ventana de Inicio



Elaborado por: Danny Ascencio Moreno
Fuente: Datos de la aplicación weka

En esta ventana la herramienta permite navegar por los archivos del Pc para acceder a el documento deseado. En la parte inferior escoger el tipo de archivo que se va a ingresar para que la aplicación la muestre en pantalla, siempre y cuando el formato coincida con el seleccionado.

Gráfico 38 – Selección del archivo



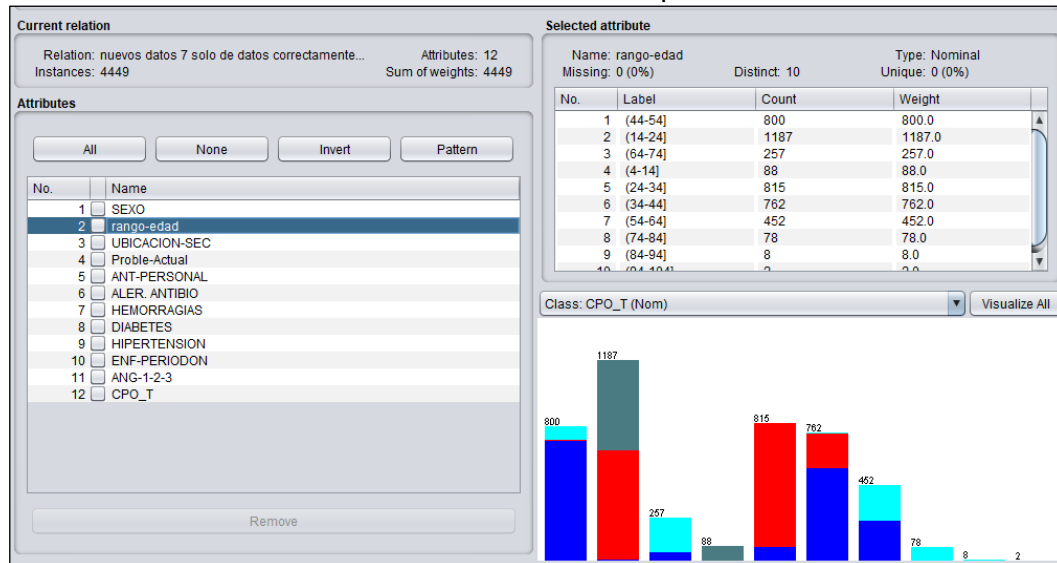
Elaborado por: Danny Ascencio Moreno
Fuente: Datos de la aplicación weka

Finalmente, la presentación de la información se da como se observa en el Gráfico 39, si el archivo tiene el formato adecuado internamente, weka reconocerá cada uno de los atributos y los presentará con su debido tipo de dato, la herramienta divide la ventana en 3 secciones, de las cuales en la primer ventana se pueden hacer ediciones en la información, como seleccionar un elemento, seleccionar todos los elementos, desmarcar todos los seleccionados, y por ultimo ingresar un patrón de datos, en la parte inferior se brinda la opción de eliminar una variables en caso de que sea necesario.

En la segunda sección se muestran elementos correspondientes al tipo de dato tales como; Nombre, Porcentaje de perdida en de datos, distinción de clases, tipo de variable.

En la tercera pestaña se presenta la información de forma gráfica junto con un botón para visualizar todos los elementos de esta manera.

Gráfico 39 – Ventana de explorador

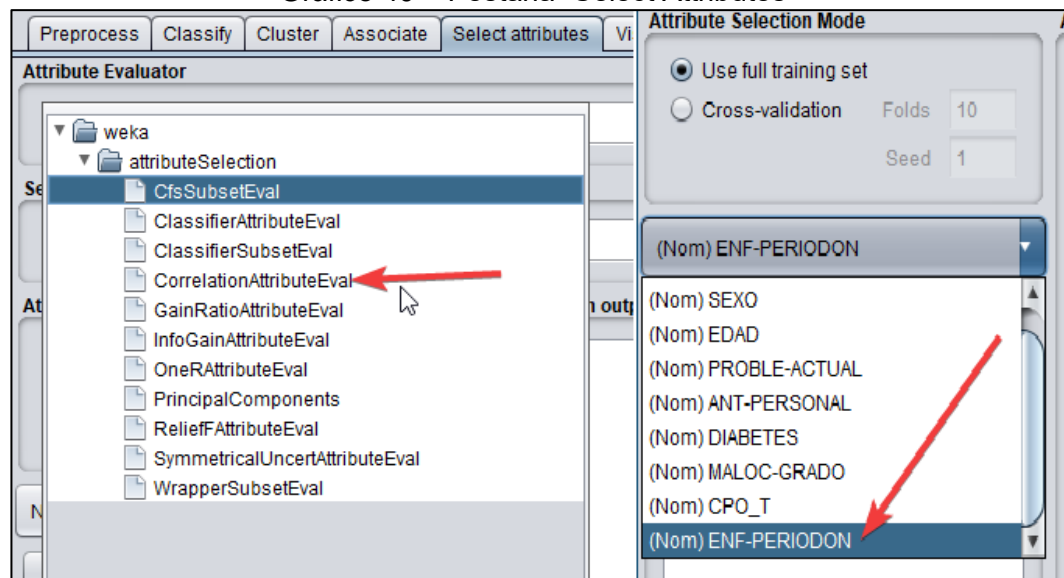


Elaborado por: Danny Ascencio Moreno
 Fuente: Datos de la aplicación weka

PROCESAMIENTO DE LOS DATOS INGRESADOS, USANDO LA TÉCNICA DE REDES BAYESIANAS DE MINERÍA DE DATOS

En la pestaña “Select Attributes” se encuentran herramientas para seleccionar las variables que mejor convengan para dar los mejores resultados a la hora de aplicar el clasificador NetBayes. En este caso se seleccionó La opción de “CorrelationAttributeEval” el cual arrojó una lista que se puede apreciar en el Gráfico 41 con los elementos tipo ranking de los cuales se escogieron los 8 primeros, por la variedad de datos que mostraron poseer.

Gráfico 40 – Pestaña “Select Attributes”



Elaborado por: Danny Ascencio Moreno
Fuente: Datos de la aplicación weka

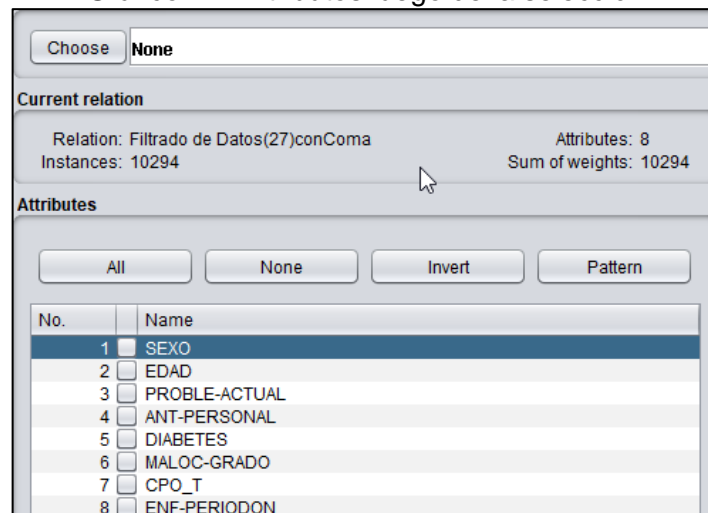
Gráfico 41 - Atributos rankeados con respecto a la variable ENF-PERIODON

Correlation Ranking Filter		
Ranked attributes:		
0.06533	2	EDAD
0.06522	1	SEXO
0.0569	17	CPO_T
0.02526	12	DIABETES
0.02412	4	PROBLE-ACTUAL
0.02384	5	ANT-PERSONAL
0.02369	16	MALOC-GRADO
0.01904	13	HIPERTENSION
0.0174	9	VHI-SIDA
0.01725	14	ENF-CARDIACA
0.01636	11	ASMA
0.01017	10	TUBERCULOSIS
0.00961	3	CIUDAD
0.00757	6	ALER-ANTIBIO
0.00729	15	ANT-FAMILIAR
0.0023	8	HEMORRAGIAS
0.00118	7	ALER-ANESTESIA
Selected attributes: 2,1,17,12,4,5,16,13,9,14,11,10,3,6,15,8,7 : 17		

Elaborado por: Danny Ascencio Moreno
Fuente: Datos de la aplicación weka

Presentación de los atributos con los que se trabajará el proceso de minería de datos, a continuación, se seleccionará el clasificador y se ejecutará el programa que mostrará el grafo generado y las instancias correctamente clasificadas, así como su contraparte las instancias incorrectamente clasificadas.

Gráfico 42 - Atributos luego de la selección



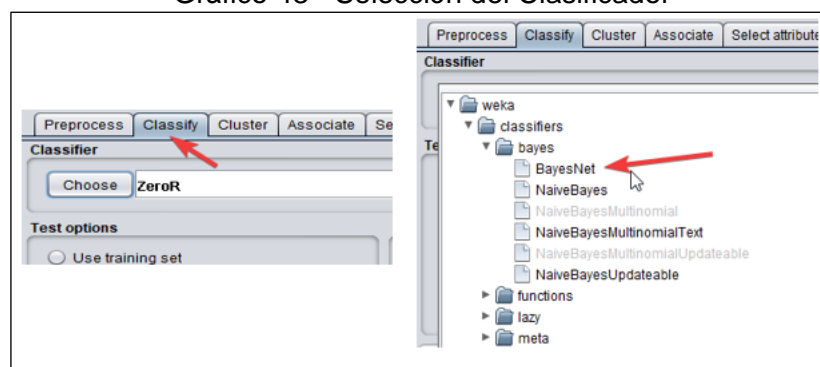
Elaborado por: Danny Ascencio Moreno

Fuente: Datos de la aplicación weka

En la pestaña **classify** se seleccionará el algoritmo con el que se trabajará los datos. Al inicializar dicha ventana se evidencia la presencia de otros métodos estadístico los cuales a breves rasgos son:

Bayes: Método basado en el modelo establecido por Bayes
Funciones, Métodos “matemáticos”, redes neuronales, regresiones.

Gráfico 43 - Selección del Clasificador

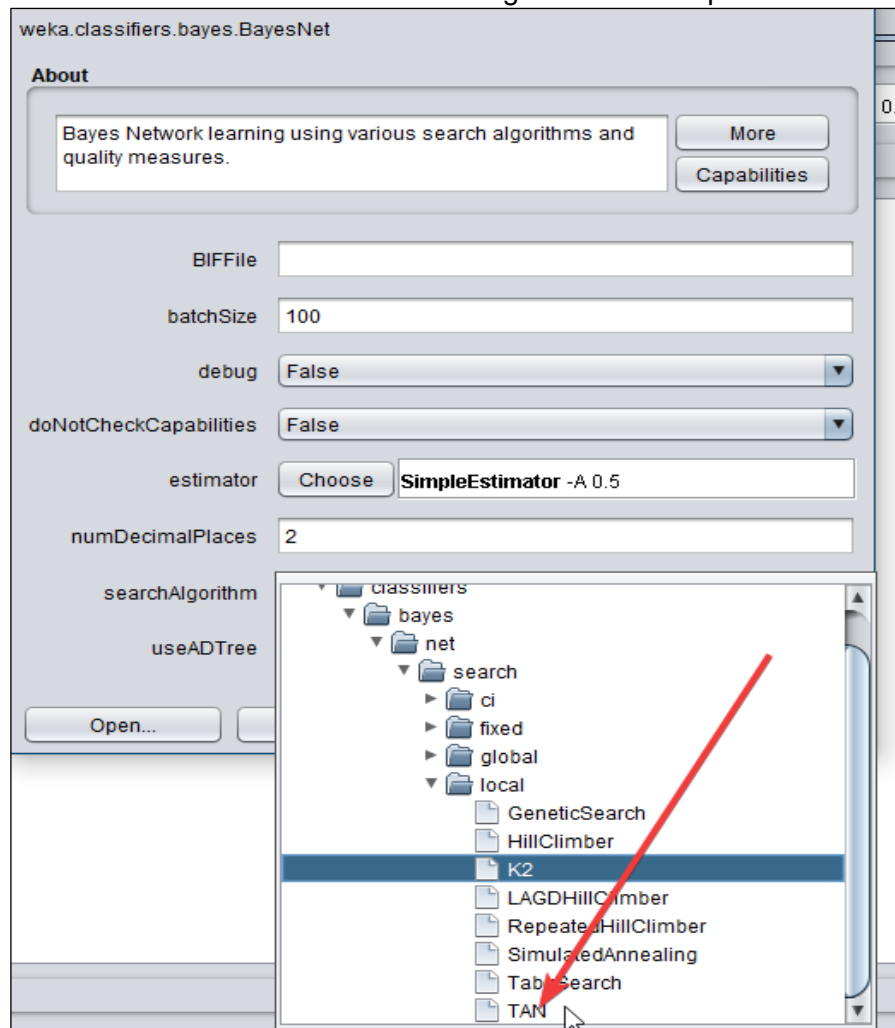


Elaborado por: Danny Ascencio Moreno

Fuente: Datos de la aplicación weka

Bayes dispone de algunos algoritmos previamente instalados de los cuales se escogerá "TAN", de manera predeterminada se encuentra seleccionado el algoritmo "K2"

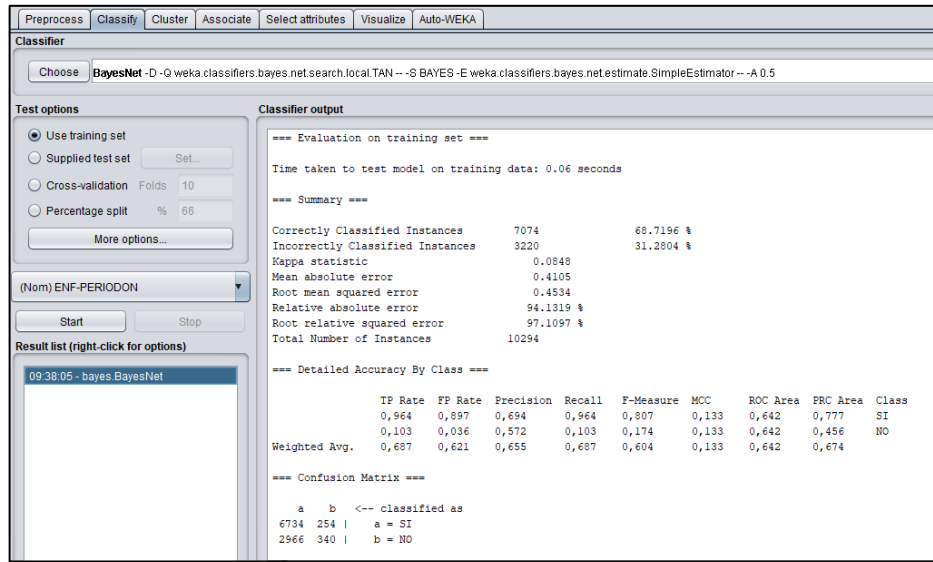
Gráfico 44 - Selección del algoritmo de búsqueda



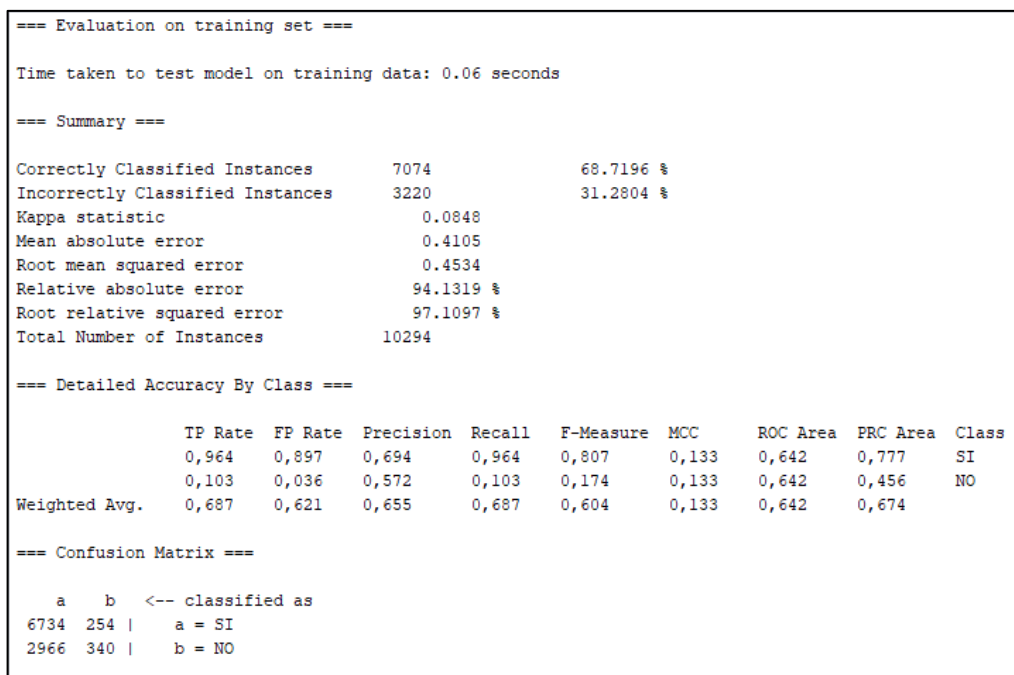
Elaborado por: Danny Ascencio Moreno
Fuente: Datos de la aplicación weka

En el Gráfico 45 (a) y (b) se aprecian los resultados de la minería de datos con sus valores y matriz de conjunción, instancias correctamente clasificadas y las incorrectamente clasificadas.

Gráfico 45 - Resultados de la minería de datos



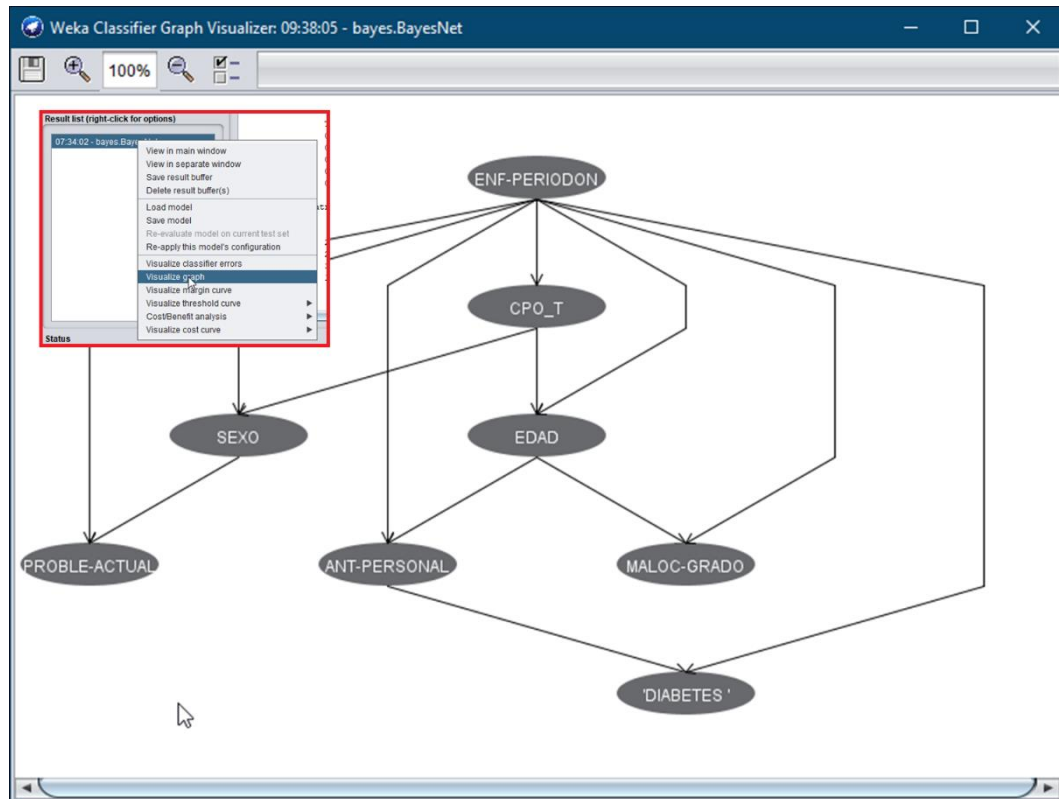
(a)



(b)

Se puede visualizar el gráfico de la red generada mediante las opciones que dispone el proceso realizado en el Gráfico 46.

Gráfico 46 – Visualización de la red bayesiana



Elaborado por: Danny Ascencio Moreno
Fuente: Datos de la aplicación weka

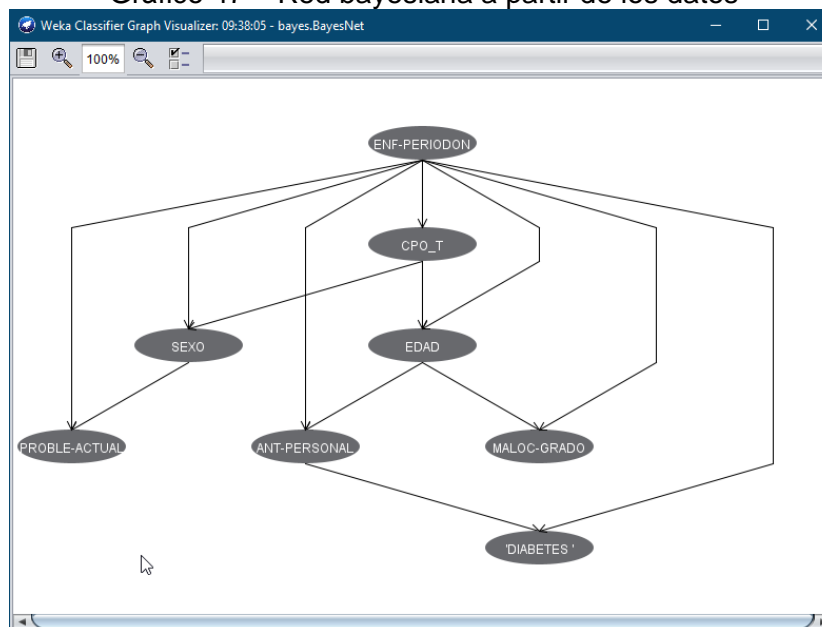
Los procedimientos utilizados en la presente investigación han sido avalados por el juicio de un experto en el área como se muestra en el Anexo 4.

CAPTULO IV

RESULTADOS CONCLUSIONES Y RECOMENDACIONES

RESULTADOS

Gráfico 47 – Red bayesiana a partir de los datos



Elaborado por: Danny Ascencio Moreno

Fuente: Datos de la aplicación weka

La red muestra las varias relaciones condicionales entre los nodos escogidos. Entre las relaciones directas más destacadas se tienen:

Influencia de la enfermedad periodontal (Relacionada con Gingivitis) sobre CPO_T (Indicador de Caries), de la misma manera, la edad sobre el grado de maloclusión, o los antecedentes personales sobre la diabetes.

Estas relaciones son consistentes con el sentido común, ya que las caries y gingivitis suelen estar relacionadas, la edad juega un papel importante en el grado de maloclusión de algún diente, las deformaciones se intensifican con la edad si no se corrigen, y es conocido además, que en el padecimiento de diabetes, juega un papel importante la genética del paciente, relacionada directamente con los antecedentes.

Gráfico 48 – Probabilidades de la variable CPO_T

ENF-PERIODON	RANGO3	RANGO2	RANGO4	RANGO1
SI	0,368	0,337	0,143	0,153
NO	0,268	0,354	0,146	0,231

Elaborado por: Danny Ascencio Moreno

Fuente: Datos de la aplicación weka

Considerando que se presente enfermedad periodontal, existe una probabilidad de 0.7 de tener una CPO_T entre 8 a 24 (Rango 2 o Rango 3). Dejando una probabilidad restante de 0.3 de estar en grado severo, o en el grado mínimo. Por otro lado, es relevante que en caso de no presentar enfermedad periodontal, la posibilidad de ser afectado por un grado severo de caries (Rango 4) es de apenas 14.6%, el menor posible entre los rangos. Esto ratifica la importancia de actuar cuando las afecciones sólo corresponden a una y no a varias, porque el problema se puede intensificar cuando se debe lidiar con más de un problema de salud.

Gráfico 49 - Probabilidades de la variable SEXO

ENF-PERIODON	CPO_T	F	M
SI	RANGO3	0,627	0,373
SI	RANGO2	0,54	0,46
SI	RANGO4	0,564	0,436
SI	RANGO1	0,487	0,513
NO	RANGO3	0,721	0,279
NO	RANGO2	0,611	0,389
NO	RANGO4	0,655	0,345
NO	RANGO1	0,563	0,437

Elaborado por: Danny Ascencio Moreno

Fuente: Datos de la aplicación weka

En general, las proporciones en el Gráfico 50 son similares a los porcentajes originales de Hombres (41%) y Mujeres (59%). Uno de los valores que resalta en la tabla, es $p = 0.721$, que corresponde a las mujeres sin enfermedad periodontal pero con un alto CPO (mayor a 24).

Gráfico 50 - Probabilidades de la variable EDAD

Probability Distribution Table For EDAD											
ENF-PERIODON	CPO_T	γ(41.2-50.5)°	γ(13.3-22.8)°	γ(69.1-78.4)°	γ(-inf-13.3)°	γ(59.8-69.1)°	γ(22.6-31.9)°	γ(31.9-41.2)°	γ(50.5-59.8)°	γ(78.4-87.7)°	γ(87.7-inf)°
SI	RANGO3	0.254	0.059	0.018	0	0.082	0.159	0.218	0.208	0.003	
SI	RANGO2	0.154	0.234	0.007	0.004	0.028	0.29	0.203	0.077	0.003	
SI	RANGO4	0.22	0.013	0.082	0.001	0.239	0.04	0.108	0.279	0.014	0.003
SI	RANGO1	0.064	0.464	0.003	0.026	0.002	0.324	0.098	0.018	0.001	
NO	RANGO3	0.228	0.104	0.019	0.001	0.061	0.192	0.221	0.171	0.002	0.003
NO	RANGO2	0.101	0.334	0.006	0.008	0.017	0.337	0.154	0.04	0.001	
NO	RANGO4	0.154	0.011	0.158	0.001	0.261	0.019	0.075	0.255	0.058	0.003
NO	RANGO1	0.029	0.533	0.002	0.051	0.005	0.319	0.05	0.007	0.003	0.003

Elaborado por: Danny Ascencio Moreno

Fuente: Datos de la aplicación weka

Aproximadamente el 80% de las personas con bajo CPO_T (Rango 1), tienen entre 13 y 32 años, independientemente de tener o no enfermedad periodontal. Sin embargo, el 74% de las personas con enfermedad periodontal y CPO_T alto (Rango 4) tienen 41 años o más, lo cual ligeramente contrasta con los pacientes mayores a 41 años que poseen alto CPO_T pero no tienen enfermedad periodontal (67%).

Es decir, la edad avanzada está relacionada con la presencia de caries en alta medida, y la enfermedad periodontal, aumenta la probabilidad de padecer de caries.

Gráfico 51 - Probabilidades de la variable PROBLE-ACTUAL

ENF-PERIODON	SEXO	ASINTOMATICO	SINTOMATICO
SI	F	0,823	0,177
SI	M	0,874	0,126
NO	F	0,844	0,156
NO	M	0,897	0,103

Elaborado por: Danny Ascencio Moreno

Fuente: Datos de la aplicación weka

Las personas tratadas en la FPO que presentan no enfermedad periodontal, y son del género masculino, son ligeramente más propensas (89.97%) a acudir sin síntomas a las citas médicas que las del género femenino (84.74%).

Gráfico 52 - Probabilidades de la variable ANT-PERSONAL

ENF-PERIODON	EDAD	SI	NO
SI	Y(41.2-50.5]Y	0,298	0,702
SI	Y(13.3-22.6]Y	0,177	0,823
SI	Y(69.1-78.4]Y	0,598	0,402
SI	Y(-inf-13.3]Y	0,23	0,77
SI	Y(59.8-69.1]Y	0,484	0,516
SI	Y(22.6-31.9]Y	0,178	0,822
SI	Y(31.9-41.2]Y	0,211	0,789
SI	Y(50.5-59.8]Y	0,377	0,623
SI	Y(78.4-87.7]Y	0,554	0,446
SI	Y(87.7-inf]Y	0,5	0,5
NO	Y(41.2-50.5]Y	0,295	0,705
NO	Y(13.3-22.6]Y	0,137	0,863
NO	Y(69.1-78.4]Y	0,564	0,436
NO	Y(-inf-13.3]Y	0,092	0,908
NO	Y(59.8-69.1]Y	0,539	0,461
NO	Y(22.6-31.9]Y	0,172	0,828
NO	Y(31.9-41.2]Y	0,231	0,769
NO	Y(50.5-59.8]Y	0,403	0,597
NO	Y(78.4-87.7]Y	0,5	0,5
NO	Y(87.7-inf]Y	0,5	0,5

Elaborado por: Danny Ascencio Moreno

Fuente: Datos de la aplicación weka

Las proporciones de pacientes con antecedentes personales varían significativamente con respecto a la edad, pero de forma leve con respecto a la presencia o ausencia de enfermedad periodontal. Los resultados se muestran en la tabla y permiten llegar a conclusiones similares a aquellas presentadas con el CPO_T, así, aproximadamente el 80% de las personas entre 13 y 32 años (inclusive hasta 41 años) no presentan antecedentes personales.

Gráfico 53 - Probabilidades de la variable MALOC-GRADO

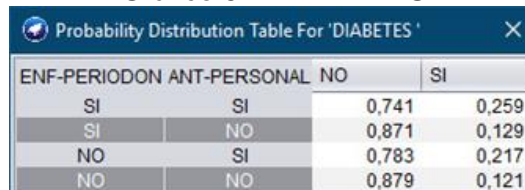
ENF-PERIODON	EDAD	G2	G1	G3	G0
SI	Y(41.2-50.5]Y	0,239	0,562	0,078	0,121
SI	Y(13.3-22.6]Y	0,187	0,704	0,092	0,018
SI	Y(69.1-78.4]Y	0,131	0,339	0,124	0,406
SI	Y(-inf-13.3]Y	0,224	0,618	0,118	0,039
SI	Y(59.8-69.1]Y	0,199	0,35	0,097	0,354
SI	Y(22.6-31.9]Y	0,204	0,674	0,09	0,031
SI	Y(31.9-41.2]Y	0,204	0,639	0,078	0,079
SI	Y(50.5-59.8]Y	0,21	0,496	0,074	0,221
SI	Y(78.4-87.7]Y	0,121	0,397	0,052	0,431
SI	Y(87.7-inf]Y	0,375	0,375	0,125	0,125
NO	Y(41.2-50.5]Y	0,239	0,539	0,065	0,156
NO	Y(13.3-22.6]Y	0,163	0,745	0,069	0,023
NO	Y(69.1-78.4]Y	0,092	0,141	0,015	0,752
NO	Y(-inf-13.3]Y	0,19	0,71	0,05	0,05
NO	Y(59.8-69.1]Y	0,075	0,187	0,08	0,658
NO	Y(22.6-31.9]Y	0,185	0,727	0,053	0,035
NO	Y(31.9-41.2]Y	0,228	0,657	0,054	0,061
NO	Y(50.5-59.8]Y	0,189	0,395	0,068	0,347
NO	Y(78.4-87.7]Y	0,044	0,074	0,015	0,868
NO	Y(87.7-inf]Y	0,083	0,083	0,083	0,75

Elaborado por: Danny Ascencio Moreno

Fuente: Datos de la aplicación weka

La probabilidad de una persona mayor a 60 años que no tenga enfermedad periodontal, presente maloclusión es baja (probabilidad de G0 es de 0.76), en personas más jóvenes tener o no enfermedad periodontal no es muy relevante a la hora de presentar maloclusión

Gráfico 54 - DIABETES



		ANT-PERSONAL	
		NO	SI
ENF-PERIODON	SI	0,741	0,259
	NO	0,871	0,129
		NO	SI
	SI	0,783	0,217
	NO	0,879	0,121

Elaborado por: Danny Ascencio Moreno

Fuente: Datos de la aplicación weka

En cuanto a la presencia de diabetes en función de los antecedentes personales y enfermedad periodontal no existen diferencias significativas.

CONCLUSIONES

- Se realizó con éxito la búsqueda y recopilación de información relevante de una base de datos actualizada, cuya utilización involucra a la FPO de la Universidad de Guayaquil.
- Se seleccionaron las bases de datos con las variables idóneas para el ejercicio, tomando en cuenta criterios médicos y métricas matemáticas de correlación, además de las consideraciones estadísticas correspondientes.
- Se llevó a cabo la explotación de la base de datos empleando la herramienta “redes bayesianas” de minería de datos para detectar factores importantes relacionados con caries y gingivitis.
- Se determinó que factores como la diabetes son ligeramente influyentes mientras que otros como CPO_T y EDAD lo son en gran medida.
- El modelo de redes bayesianas permite realizar predicciones para futuras observaciones, con un moderado poder predictivo (alrededor del 68.7% de acierto). Las inferencias y clasificaciones son veraces y presentan nuevas luces e información sobre la situación actual de caries y gingivitis en la población que rodea la FPO de la Universidad de Guayaquil.

RECOMENDACIONES

- Se aconseja realizar el estudio con variables acorde a la investigación las cuales desafortunadamente no existieron.
- Es imperante que se revisen los datos al momento del ingreso ya que un error en este proceso distorsiona los resultados de posibles análisis como el del presente estudio, generando la necesidad de una depuración exhaustiva de la base de datos.
- Es importante la recolección de datos como hábitos de consumo, estimación de la ingesta de azúcar en la dieta o frecuencia de los cuidados bucales, ya que pueden ser de mucha utilidad para alimentar el modelo en futuras ocasiones y sería de gran ayuda incluir en la tabulación, variables que se encuentran en las fichas, pero no están en la base, como los signos vitales.
- Se sugiere la implementación de una sectorización estandarizada para cada paciente en la base, la cual permita ligarlos a una ubicación geográfica de manera eficiente y acertada.

Bibliografía:

- Bahmani, E., Jamshidi, M., & Shaltooli, A. A. (2019). *Breast Cancer Prediction Using a Hybrid Data Mining Model*. 3, 327–331.
<https://doi.org/http://dx.doi.org/10.30630/joiv.3.4.240>
- Cios, K. J., Krawczyk, B., Cios, J., & Staley, K. J. (2019). *How the new technologies and data they generate are transforming medicine*.
- COCA DE PABLO, D. (2019). *GRADO EN ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS TRABAJO FIN DE GRADO LA CULTURA OPEN SOURCE EN LAS EMPRESAS OPEN SOURCE CULTURE IN COMPANIES AUTOR / A : DANIEL COCA DE PABLO DIRECTOR / A : MARÍA ELENA GARCÍA RUIZ*.
- Jiménez Márquez, I. (2019). *Diseño y evaluación de algoritmos automáticos de clustering para modelos de canal radio en interiores*. 112.
- Martínez-García, D. N., Dalgo-Flores, V. M., Herrera-López, J. L., Analuisa-Jiménez, E. I., & Velasco-Acurio, E. F. (2019). Avances de la inteligencia artificial en salud. *Dominio de Las Ciencias*, 5(3), 603.
<https://doi.org/10.23857/dc.v5i3.955>
- Paulino Flores, L. A., & Huayna Dueñas, A. M. (2019). Sistema Experto Probabilístico basado en Redes Bayesianas para la predicción del cáncer de cuello uterino. *Revista Peruana de Computación y Sistemas*, 2(1), 15–26. <https://doi.org/10.15381/rpcs.v2i1.16360>
- Matos Ayala, A. (2018). *Investigación Bibliográfica: Definición, Tipos, Técnicas*. Obtenido de <https://www.lifeder.com/investigacion-bibliografica>.
- Mehta, N., & Pandit, A. (2018). International Journal of Medical Informatics Concurrence of big data analytics and healthcare : A systematic review. *International Journal of Medical Informatics*, 114(March), 57–65.
<https://doi.org/10.1016/j.ijmedinf.2018.03.013>
- Judith Sandoval, L. (2018). *ENERO-DICIEMBRE 2018 Derechos Reservados • Escuela Especializada en Ingeniería ITCA-FEPADE*. 11, 36–40. Retrieved from http://redicces.org.sv/jspui/bitstream/10972/3626/1/Art6_RT2018.pdf
- Especial, A. E., Spark, A. A., Gt, S., E, S. G. T. F., & Gl, C. (2018). *Análisis de*

métodos para reconocimiento de estado de la luz de los semáforos aplicada para carros de conducción autónoma. 20–22.

Flores Ortega, P. W. (2018). Prevalencia de caries dental en estudiantes de nivel primario de la institución educativa particular Pestalozzi, distrito Nuevo Chimbote, provincia del Santa, región Áncash, año 2018. *Universidad Católica Los Ángeles de Chimbote*. Retrieved from <http://repositorio.uladech.edu.pe/handle/123456789/6349>

Schab, E., Rivera, R., Bracco, L., Coto, F., Cristaldo, P., Ramos, L., ... De Battista, A. (2018). Minería de datos y visualización de información. *XX Workshop de Investigadores En Ciencias de La Computación*, 325–329. Retrieved from http://sedici.unlp.edu.ar/bitstream/handle/10915/67382/Documento_completo.pdf-PDFA.pdf?sequence=1&isAllowed=y

Timón, C. E. (2017). “Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo - herramientas Open Source que permiten su uso. 26/27, (Principio activo y prestación ortoprotésica), 67. Retrieved from <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/59565/6/caresptimTFG0117memòria.pdf>

Casallas, A. F., & Jiménez, D. (2017). Big Data aplicado a la medicina. *Vol. 1 Núm. 2 (2017): Revista Avenir 2017,1,2.*

He, K. Y., Ge, D., & He, M. M. (2017). *Big Data Analytics for Genomic Medicine*. (Table 1), 1–18. <https://doi.org/10.3390/ijms18020412>

Lanzarini, L. C., & Villa Monte, A. (2017). *Enseñanza De La Minería De Datos Y El Impacto De Las Herramientas De Software Utilizadas*. (1), 109–121.

Puttipan, R., Wanachantararak, P., & Khongkhunthian, S. (2017). *Effects of Caesalpinia sappan on pathogenic bacteria causing dental caries and gingivitis*. 11(6), 316–322. <https://doi.org/10.5582/ddt.2017.01055>

Quintero Quintero, M. M., Hernández, F. A., Maldonado Reyes, D. A., García Loera, I. V., Funatsu Díaz, I. J. L., Barrientos-Lozano, L., ... Flores-Gracia, J. (2017). *Estudio De Herramientas De Minería De Datos Para La Tarea De Clasificación*. 66. Retrieved from www.latindex.org

- Muñoz Cruzatty, J. P., Moreira, E. M. T., & Oliveira del Rio, J. A. (2017). *Salud bucal de los adolescentes, en el centro de Salud Portoviejo, Ecuador*. 2(3), 81–90. <https://doi.org/10.23857/pc.v2i3.49>
- Naik, A., & Samant, L. (2016). Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime. *Procedia Computer Science*, 85(Cms), 662–668. <https://doi.org/10.1016/j.procs.2016.05.251>
- De Battista, A., Cristaldo, P., Ramos, L., Nuñez, J. P., Retamar, S., Bouzenard, D., & Herrera, N. E. (2016). Minería de Datos Aplicada a Datos Masivos. *XVIII Workshop de Investigadores En Ciencias de La Computación (WICC 2016, Entre Ríos, Argentina)*, 288–292. Retrieved from <http://sedici.unlp.edu.ar/handle/10915/52901>
- Escobar, T., Harold, E., Alcivar, M., & Puris, A. (2016). *Aplicaciones de Minería de Datos en Marketing*. 3(8), 503–512.
- Gerson, C., Fernanda, G., & Claudio, P. (2016). *ESTADO DE SALUD ORAL EN EL ECUADOR ORAL HEALTH STATUS IN ECUADOR*. 1(3), 65–70.
- Kulkarni, E. G., & Kulkarni, R. B. (2016). WEKA Powerful Tool in Data Mining General Terms. *International Journal of Computer Applications*, 5(Rtdm), 975–8887.
- Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). *Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement*. 1–10. <https://doi.org/10.1186/s12916-014-0241-z>
- Wolfson, J., Vock, D. M., Johnson, P. E., & Connor, P. J. O. (2014). *Data mining for censored time-to-event data : A Bayesian network model for predicting cardiovascular risk from electronic health record data*. 1–31
- J, K., & M, A. (2011). Gingivitis Control. *Gingival Diseases - Their Aetiology, Prevention and Treatment*. <https://doi.org/10.5772/24813>
- Sucar, L. E. (2011). Introduction to bayesian networks and influence diagrams. *Decision Theory Models for Applications in Artificial Intelligence: Concepts and Solutions*, 9–32. <https://doi.org/10.4018/978-1-60960-165-2.ch002>

Montgomery, D. C., & Runger, G. C. (1994). Applied Statistics and Probability for Engineers. In *European Journal of Engineering Education* (Vol. 19).
<https://doi.org/10.1080/03043799408928333>.

ANEXOS

ANEXO 1 – CUADROS DE RESUMEN CON INFORMACIÓN DE LAS VARIABLES PARA MINERÍA DE DATOS EN WEKA.

Sexo	Porcentaje
F	58.95%
M	41.05%

Enfermedad Periodontal	Porcentaje
NO	32.12%
SI	67.88%

Problema actual	Porcentaje
ASINTOMATICO	85.08%
SINTOMATICO	14.92%

CPO_T	Valores	Porcentaje
RANGO1	≤ 8	17.79%
RANGO2	(8, 16]	34.22%
RANGO3	(16, 24]	33.58%
RANGO4	> 24	14.41%

Fecha	Conteo
2018	8407
abr	88
may	789
jun	1460
jul	1247
ago	1190
sep	147
oct	213
nov	2409
dic	864
2019	1887
ene	1211
feb	670
mar	6

Nº	Rango de Edad	Conteo	Porcentaje
1	(0-13.3]	84	0.82%
2	(13.3-22.6]	2111	20.51%
3	(22.6-31.9]	2297	22.31%
4	(31.9-41.2]	1701	16.52%
5	(41.2-50.5]	1723	16.74%
6	(50.5-59.8]	1344	13.06%
7	(59.8-69.1]	721	7.00%
8	(69.1-78.4]	248	2.41%
9	(78.4-87.7]	59	0.57%
10	(87.7-inf]	6	0.06%

Ciudad	Porcentaje
BABAHOYO	0.83%
DAULE	0.74%
DURAN	6.91%
GUAYAQUIL	89.63%
MILAGRO	1.90%

Maloclusión (grado)	Porcentaje
0	12.67%
1	59.89%
2	19.80%
3	7.65%

Afección	Porcentaje	
	SI	NO
Antecedentes Personales	26.08%	73.92%
Alergia a antibiótico	4.28%	95.72%
Alergia a anestesia	0.20%	99.80%
Hemorragias	0.63%	99.37%
VIH-SIDA	0.12%	99.88%
Tuberculosis	0.29%	99.71%
Asma	2.67%	97.33%
Diabetes	15.77%	84.23%
Hipertensión	20.69%	79.31%
Enfermedad cardiaca	2.93%	97.07%
Antecedentes familiares	21.94%	78.06%

ANEXO 2 – FICHAS MÉDICAS DE LA FPO PARA REGISTRAR LOS DATOS DE PACIENTES.

ESTABLECIMIENTO		NOMBRE		APELLIDO		SEXO (M-F)		EDAD		N° HISTORIA CLÍNICA	

MENOR DE 1 AÑO	1 - 4 AÑOS	5 - 9 AÑOS PROGRAMADO	5 - 14 AÑOS NO PROGRAMADO	15 - 14 AÑOS PROGRAMADO	15 - 19 AÑOS	MAYOR DE 20 AÑOS	EMBARAZADA
----------------	------------	--------------------------	------------------------------	----------------------------	--------------	------------------	------------

1 MOTIVO DE CONSULTA ANOTAR LA CAUSA DEL PROBLEMA EN LA VERSIÓN DEL INFORMANTE

2 ENFERMEDAD O PROBLEMA ACTUAL REGISTRAR SÍNTOMAS, CRONOLOGÍA, LOCALIZACIÓN, CARACTERÍSTICAS, INTENSIDAD, CAUSA AFINENTE, SÍNTOMAS ASOCIADOS, EVOLUCIÓN, ESTADO ACTUAL

3 ANTECEDENTES PERSONALES Y FAMILIARES

1. ALERGIA ANTIBIÓTICO	2. ALERGIA ANESTESIA	3. HEMORRAGIAS	4. VIHUSIDA	5. TUBERCULOSIS	6. ASMA	7. DIABETES	8. HIPERTENSIÓN	9. ENF. CARDIACA	10. OTRO
------------------------	----------------------	----------------	-------------	-----------------	---------	-------------	-----------------	------------------	----------

4 SIGNOS VITALES

PRESIÓN ARTERIAL	FRECUENCIA CARDÍACA <small>min</small>	TEMPERATURA <small>°C</small>	F. RESPIRAT. <small>min</small>
------------------	--	-------------------------------	---------------------------------

5 EXAMEN DEL SISTEMA ESTOMATOGNÁTICO DESCRIBIR ABAJO LA PATOLOGÍA DE LA REGIÓN AFECTADA ANOTANDO EL NÚMERO

1. LABIOS	2. MEJILLAS	3. MAXILAR SUPERIOR	4. MAXILAR INFERIOR	5. LENGUA	6. PALADAR	7. PISO	8. CARRILLOS
9. GLÁNDULAS SALIVALES	10. ORO FARINGE	11. A. T. M.	12. GANGLIOS				

6 ODONTOGRAMA PINTAR CON AZUL PARA TRATAMIENTO REALIZADO - ROJO PARA PATOLOGÍA ACTUAL
MOVILIDAD Y RECESIÓN MARCAR "X" (1, 2 ó 3). SI APLICA

RECESIÓN

MOVILIDAD

VESTIBULAR

LINGUAL

VESTIBULAR

MOVILIDAD

RECESIÓN

MOVILIDAD

RECESIÓN

MOVILIDAD

RECESIÓN

MOVILIDAD

RECESIÓN

7 INDICADORES DE SALUD BUCAL

HIGIENE ORAL SIMPLIFICADA				ENFERMEDAD PERIODONTAL	MAL OCCLUSIÓN	FLUOROSIS
				LEVE	ÁNGULO I	LEVE
				MODERADA	ÁNGULO II	MODERADA
				SEVERA	ÁNGULO III	SEVERA

8 INDICES CPO-ceo

	C	P	O	TOTAL
D				
d	c	e	o	TOTAL

9 SIMBOLOGÍA DEL ODONTOGRAMA

* _{rojo} SELLANTE NECESARIO	⊗ PERDIDA (OTRA CAUSA)	≡ PRÓTESIS TOTAL
* _{azul} SELLANTE REALIZADO	△ ENDODONCIA	⊞ CORONA
X _{rojo} EXTRACCIÓN INDICADA	□ PRÓTESIS FIJA	○ azul OBTURADO
X _{azul} PERDIDA POR CARIES	(---) PRÓTESIS REMOVIBLE	○ rojo CARIES

ODONTOLOGÍA (1)

10 PLANES DE DIAGNÓSTICO, TERAPÉUTICO Y EDUCACIONAL

BIOMETRIA		QUIMICA SANGUINEA		RAYOS - X		OTROS	

11 DIAGNÓSTICO

	PRE= PRESUNTIVO DEF= DEFINITIVO	CIE	PRE	DEF		CIE	PRE	DEF
1					3			
2					4			

FECHA DE APERTURA	FECHA DE CONTROL	PROFESIONAL	FIRMA	NÚMERO DE HOJA
-------------------	------------------	-------------	-------	----------------

12 TRATAMIENTO

SESIÓN Y FECHA	DIAGNOSTICOS Y COMPLICACIONES	PROCEDIMIENTOS	PRESCRIPCIONES	CÓDIGO Y FIRMA
SESIÓN 1				CÓDIGO
FECHA				FIRMA
SESIÓN 2				CÓDIGO
FECHA				FIRMA
SESIÓN 3				CÓDIGO
FECHA				FIRMA
SESIÓN 4				CÓDIGO
FECHA				FIRMA
SESIÓN 5				CÓDIGO
FECHA				FIRMA
SESIÓN 6				CÓDIGO
FECHA				FIRMA
SESIÓN 7				CÓDIGO
FECHA				FIRMA
SESIÓN 8				CÓDIGO
FECHA				FIRMA
SESIÓN 9				CÓDIGO
FECHA				FIRMA

SNS-MSP / HCU-form.033 / 2008

ODONTOLÓGIA (2)

**ANEXO 3 – CONSENTIMIENTO PARA USO ACADÉMICO DE DATOS
DE LA FPO**

*Autocredencial
Dr. Pilco
Chavarrea*



UNIVERSIDAD DE GUAYAQUIL

ESPECIE VALORADA – NIVEL PREGRADO

Guayaquil, 03 de Julio del 2019

Doctor

JOSE FRANCO VALDIVIEZO

DECANO DE LA FACULTAD PILOTO DE ODONTOLOGIA

Ciudad.-

De mis consideraciones:

Yo, **ASCENCIO MORENO DANNY JONATHAN**, con CI. **0927469379**, **estudiante de la carrera de Ingeniería en SISTEMAS COMPUTACIONALES** de la **FACULTAD DE CIENCIAS FISICAS Y MATEMATICAS**, solicito a usted que se me conceda realizar una recopilación de información para formular un modelo de datos. Requiero de los datos clínicos con respecto a gingivitis, caries y problemas periodontales, los mismos que se reflejan en cada una de las historias clínicas, para desarrollar mi tesis **“DESARROLLO Y EXPLOTACION DE INFORMACION DE UNA BASE DE DATOS DE PATOLOGIAS PRESENTES EN LA CAVIDAD BUCAL CON ENFASIS EN CARIES, GINGIVITIS Y PROBLEMAS PERIODONTALES A TRAVEZ DE HERRAMIENTAS ORIENTADAS A MINERIA DE DATOS”**, con información obtenida en la **FACULTAD PILOTO DE ODONTOLOGIA** referente a administración donde se hallan todos los datos ingresados.

La presente solicitud será guiada por el Ing. Segundo Delgado M. Phd tutor de tesis.
De antemano le agradezco, por la atención brindada a la presente.

UNIVERSIDAD DE GUAYAQUIL
FACULTAD PILOTO DE ODONTOLOGIA
DR. FAUSTO PILCO CHAVARREA MSc.
COORDINADOR DEL OPTO. DE DIAGNOSTICO
REVISADO
FECHA: 03/07/2019
FIRMA: *[Firma]*
COORDINADOR DE FORMACION ACADEMICA DE LA FACULTAD PILOTO DE ODONTOLOGIA/D

ASCENCIO MORENO DANNY JONATHAN

CI. 0927469379

[Firma]
4 JUL 2019
146

ANEXO 4 – JUICIO DE EXPERTO

Validación de “Desarrollo y explotación de información de una base de datos de patologías presentes en la cavidad bucal con énfasis en caries y gingivitis a través de herramientas orientadas a minería de datos para la Facultad Piloto de Odontología de la universidad de Guayaquil”

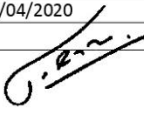
Juicio de Experto

Instrucciones para validar el desarrollo planteado

1. Lea para validar las alternativas.
2. Califique de acuerdo con su criterio tomando en consideración lo siguiente:
Funcionalidad y objetivos del proyecto.
3. Utilice la escala del 1 al 5, considerando que 1 es el valor menos significativo y 5 es el valor más significativo.
4. En caso de tener alguna sugerencia, ingresar el comentario en la columna “observación”.

Aspectos	Escala de valores					Observaciones
	En total acuerdo	De acuerdo	Parcialmente de acuerdo	En desacuerdo	En total desacuerdo	
	5	4	3	2	1	
¿Considera si es factible adoptar esta técnica en la FPO como ayuda en la toma de decisión, asumiendo que se dispone de la colaboración de la Carrera de Ingeniería en Sistemas?		X				
¿Considera Ud. que el filtrado y depuración aplicado a los datos se llevó a cabo de una manera correcta?	X					
¿Según su criterio, con base en los resultados de la investigación, considera adecuado la implementación de técnicas de minería de datos en el campo de la Odontología?	X					
¿En el marco de favorecer el uso de software libre y asegurar la replicabilidad del modelo planteado, considera adecuado la utilización de la herramienta Weka para el procesamiento de los datos?	X					
¿Considera Ud. que la investigación cumple con los objetivos propuestos?		X				

Validado Por:

Apellidos y Nombre:	Reyes Zambrano Gary Xavier
Cargo/Profesión:	Ingeniero en Sistemas
Fecha:	22/04/2020
Firma:	

Gracias por su colaboración