



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**

**FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN SISTEMAS
COMPUTACIONALES**

TEMA:

**“MINERIA DE DATOS APLICADA A LA DETECCIÓN DE
PATRONES DE RECLAMOS DEL SERVICIO DE AGUA
POTABLE EN LA CIUDAD DE GUAYAQUIL”.**

AUTORA:

VERA LETAMENDI MARIA MERCEDES

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL
TÍTULO DE INGENIERA EN SISTEMAS COMPUTACIONALES**

TUTOR:

ING. GALO ENRIQUE CORNEJO GOMEZ

GUAYAQUIL, ECUADOR

08 de marzo del 2018



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**

**FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN SISTEMAS
COMPUTACIONALES**

CERTIFICACIÓN

Certificamos que el presente trabajo de titulación, fue realizado en su totalidad por **Vera Letamendi María Mercedes**, como requerimiento para la obtención del título de **Ingeniera en Sistemas Computacionales**.

TUTOR



Docente Tutor

Ing. Galo Enrique Cornejo Gómez, Mgs.

DIRECTORA DE LA CARRERA



Ing. Beatriz del Pilar Guerrero Yépez, Mgs.

Guayaquil, a los 08 del mes de marzo del año 2018



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN SISTEMAS
COMPUTACIONALES**

DECLARACIÓN DE RESPONSABILIDAD

Yo, **Vera Letamendi María Mercedes**

DECLARO QUE:

El Trabajo de Titulación, **Minería de datos aplicada a la detección de patrones de reclamos del servicio de agua potable en la ciudad de Guayaquil**, previo a la obtención del título de **Ingeniero en Sistemas Computacionales**, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias bibliográficas. Consecuentemente este trabajo es de mi total autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

Guayaquil, a los 08 del mes de marzo del año 2018

LA AUTORA

M.A. Mercedes Vera Letamendi

Vera Letamendi María Mercedes



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**

**FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN SISTEMAS
COMPUTACIONALES**

AUTORIZACIÓN

Yo, Vera Letamendi María Mercedes

Autorizo a la Universidad Católica de Santiago de Guayaquil a la **publicación** en la biblioteca de la institución del Trabajo de Titulación, **Minería de datos aplicada a la detección de patrones de reclamos del servicio de agua potable en la ciudad de Guayaquil**, cuyo contenido, ideas y criterios son de mi exclusiva responsabilidad y total autoría.

Guayaquil, a los 08 del mes de marzo del año 2018

LA AUTORA:

M.A. Mercedes Vera Letamendi

Vera Letamendi María Mercedes



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**

**FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN SISTEMAS
COMPUTACIONALES**

TRIBUNAL DE SUSTENTACIÓN

Ing. Beatriz del Pilar Guerrero Yépez, Mgs.
DIRECTORA DE CARRERA

Ing. Ángela Olivia Yanza Montalván, Mgs.
COORDINADORA DEL ÁREA

Ing. Jorge Salvador Pesántes Méndez
OPONENTE



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**

**FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN SISTEMAS
COMPUTACIONALES**

TRABAJO DE TITULACIÓN

INFORME SOFTWARE ANTIPLAGIO

1.- DATOS SOBRE EL TRABAJO DE TITULACIÓN

PERÍODO: UTE B2017

OPCIÓN: PROPUESTA METODOLÓGICA

TÍTULO: MINERIA DE DATOS APLICADA A LA DETECCIÓN DE PATRONES DE RECLAMOS DEL SERVICIO DE AGUA POTABLE EN LA CIUDAD DE GUAYAQUIL

ESTUDIANTE: VERA LETAMENDI MARIA MERCEDES

2.- DESARROLLO

2.1 Observación sobre el porcentaje alcanzado

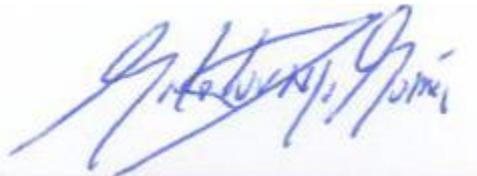
El documento se encuentra apropiadamente referenciado con citas bibliográficas.

2.2 Reporte de Software Antiplagio

| URKUND | |
|----------------|--|
| Documento | VERA LETAMENDI_FINAL.docx (D35881386) |
| Presentado | 2018-02-23 17:59 (-05:00) |
| Presentado por | galo.cornejo@cu.ucsg.edu.ec |
| Recibido | galo.cornejo.ucsg@analysis.orkund.com |
| Mensaje | RV: DOCUMENTO FINAL Mostrar el mensaje completo |
| | 0% de estas 36 páginas, se componen de texto presente en 0 fuentes. |

Fecha de elaboración: 23-02-2018

Firma:



AGRADECIMIENTO

Esta celebración es la oportunidad de demostrar mi agradecimiento a Dios, que ha hecho posible cumplir una experiencia más en mi vida, dándome la oportunidad de vivir, por fortalecer mi corazón, por darme sabiduría y por haber puesto en mi camino a aquellas personas que han sido mi apoyo y ayuda durante todo el periodo del estudio realizado.

Glenda Letamendi y Humberto Vera, mis padres, por ser mi motor en lo que llamo travesía (vida), sé que lucharon por mi bienestar, mi educación y mi salud. No conozco a nadie en este mundo a quienes les deba más amor y agradecimiento.

Ma. Paula, mi hermana, porque no importa que tan lejos estés, siempre me acuerdo de ti y jamás olvides que como las ramas de un árbol, crecemos en diferentes direcciones pero nuestra raíz es una sola, así la vida de cada una siempre será parte esencial de la vida de la otra.

Omayra, Erick y amigos, ¿Qué somos? espero haberles sacado una sonrisa en su rostro, pues somos un instante que nos alegra la vida, locos con camisas de fuerza en cada abrazo, somos la historia que hará eco en la eternidad. Gracias a aquellas personas que hacen mi carcajada más sonora, mi sonrisa más brillante y mi vida mejor.

Dama de hierro, gracias a mí, agradezco todo lo que he recibido y lo que está por venir. He puesto primero a Dios, y todo ha caído en su lugar, mantengo mi fe en él. Su tiempo y plan es perfecto; trayendo consigo buenas cosas y nuevas alegrías.

DEDICATORIA

Este trabajo está dedicado a mis padres, por sembrar en mí el valor de lucha día a día por conseguir mis sueños y por mostrarme el camino hacia la superación.

ÍNDICE GENERAL

| | |
|---|----|
| CAPÍTULO I..... | 4 |
| 1.1. PLANTEAMIENTO DEL PROBLEMA..... | 4 |
| 1.2. HIPÓTESIS Y/O PREGUNTAS DE INVESTIGACIÓN..... | 5 |
| 1.3. OBJETIVOS..... | 5 |
| 1.4. JUSTIFICACIÓN Y ALCANCE..... | 6 |
| CAPÍTULO II..... | 8 |
| MARCO TEÓRICO, CONCEPTUAL Y LEGAL..... | 8 |
| 2.1 ELEMENTOS TEÓRICOS..... | 8 |
| 2.2 CONCEPTOS..... | 19 |
| 2.3 BASE LEGAL..... | 22 |
| MARCO CONTEXTUAL..... | 26 |
| 2.4 GLOBAL ¿DÓNDE SE VA A DESARROLLAR?..... | 26 |
| 2.5 INFRAESTRUCTURA TECNOLÓGICA..... | 27 |
| CAPÍTULO III..... | 28 |
| METODOLOGÍA Y RESULTADOS DE LA INVESTIGACIÓN..... | 28 |
| 3.1. METODOLOGÍA DE INVESTIGACIÓN..... | 28 |
| 3.2. TIPO Y MÉTODO DE INVESTIGACIÓN..... | 28 |
| 3.3. POBLACIÓN, MUESTRA Y MUESTREO..... | 29 |
| 3.4. TÉCNICAS E INSTRUMENTOS PARA RECOLECCIÓN DE INFORMACIÓN..... | 29 |
| 3.5. HERRAMIENTA PARA PROCESAMIENTO DE INFORMACIÓN | |

| | | |
|-------------------------------------|---|----|
| 3.6. | ANÁLISIS DE RESULTADOS | 33 |
| CAPÍTULO IV | | 38 |
| PROPUESTA | | 38 |
| 4.1. | INTRODUCCIÓN | 38 |
| 4.2. | METODOLOGÍA..... | 39 |
| 4.3. | SELECCIÓN DE LA HERRAMIENTA PARA IMPLEMENTAR LA METODOLOGIA | 40 |
| 4.4. | DESARROLLO DE LA METODOLOGIA | 41 |
| 4.5. | ELECCION DEL ALGORITMO, GENERACION DE REGLAS Y DESCRIPCION DEL MODELO PREDICTIVO | 47 |
| 4.6. | ALGORITMOS EN MINERIA DE DATOS | 50 |
| CONCLUSIONES Y RECOMENDACIONES..... | | 64 |
| REFERENCIAS BIBLIOGRAFICAS | | 66 |
| ANEXOS | | 69 |

RESUMEN

El presente trabajo denominado “**Minería de datos aplicada a la detección de patrones de reclamos del servicio de agua potable en la ciudad de Guayaquil**” realiza un diagnóstico a la Jefatura de Atención al Usuario de la EMAPAG EP, respecto al proceso de Recurso de Apelación – Tercera Instancia con la finalidad de la aplicación de la metodología KDD (“Descubrimiento de Conocimiento en Base de Datos”) para descubrir un modelo predictivo válido, útil y entendible que especifiquen patrones de acuerdo a la información; utilizando la herramienta KNIME como plataforma tecnológica dentro de la investigación y la ejecución de los algoritmos de clasificación fundamentados en inteligencia artificial como: árbol de decisión, naive bayes y weka j48 que satisfacen las necesidades dependiendo de los objetivos del caso de estudio en base a los datos proporcionados por la entidad.

El problema a resolver va de la mano con el cumplimiento del objetivo principal de la entidad el servicio a la comunidad de la Ciudad de Guayaquil; supervisando a la operadora Interagua Cía. Ltda., en el desempeño de sus funciones como ente privado a los sectores que utilizan el servicio de Agua Potable y Alcantarillado de la ciudad, la EMAPAG EP, busca la satisfacción de los usuarios y a través de este objetivo tener un crecimiento relevante en el sector público. Dentro del proceso de Atención al Usuario, una de las variables más relevantes es el tiempo que la analista de reclamos emplea para dar solución a los diferentes casos presentados por los usuarios. Cabe recalcar que estos datos son la materia prima que se utilizan en la metodología KDD.

El estudio de esta investigación es para determinar patrones respecto a los estados de las resoluciones que se maneja en la jefatura como: Revocar, Confirmar y Revocar Parcialmente, resultados que dependerán de las variables relevantes del proceso; realizándose un análisis de los datos.

Palabras Claves: minería de datos, metodología KDD, base de datos, patrones, resoluciones, proceso.

ABSTRACT

The present research work called "Data mining applied to the detection of complaint patterns of the drinking water service in the city of Guayaquil" makes a diagnosis in favor of the EMAPAG EP's Customer Service Office, regarding the Appeal - Third Instance with the purpose of applying the KDD methodology (Discovery of Knowledge in Database) to discover a valid, useful and understandable predictive model that specifies patterns according to the information; using the KNIME tool as a technological platform within the research and execution of classification algorithms based on artificial intelligence such as: decision tree, naive bayes and weka j48 that satisfy the needs of the case study objectives based on the data provided by the entity.

The problem to solve goes hand in hand with the fulfillment of the main objective of the entity the service to the community of the City of Guayaquil supervising the operator Interagua Cía. Ltda., In the performance of its functions as a private entity to the sectors that use the city's Drinking Water and Sewage service, EMAPAG EP, looks for the satisfaction of users through this objective to have a relevant growth in the public sector Within the process of Customer Service, one of the most relevant variables in the time that the claims analyst uses to solve the different cases presented by users. It should be noted that these data are the raw material used in the KDD methodology.

The study of this investigation is to determine patterns with respect to the states of the resolutions that are handled in the head office such as: Revoke, Confirm and Partially Revoke, results that depend on the relevant variables of the process; performing a data analysis.

Key terms: data mining, KDD methodology, database, patterns, resolutions, process.

INTRODUCCIÓN

La provisión del servicio de agua potable y alcantarillado para las comunidades es un imperativo social que debe ser contemplado no solamente en las normativas y regulaciones oficiales municipales y estatales, sino que también debe ser observado que efectivamente exista tal servicio, ya que es un recurso al que tienen derecho todas las personas, sin distinción de espacio físico o ubicación geográfica.

Además, en caso de que hubiera algún nivel de incumplimiento en la entrega del servicio o insatisfacción de los usuarios, la empresa responsable debe atender los reclamos que surjan, así como absolver cualquier inquietud y consulta que se presente en cuanto al servicio recibido.

En la ciudad de Guayaquil, y bajo el control de la Empresa Municipal de Agua Potable y Alcantarillado de Guayaquil (TI, 2017), el Concesionario International Water Services Guayaquil (Veolia, 2017) debe ofrecer los siguientes servicios:

- **Agua potable:** opera en todos los ciclos del proceso para hacerla apta para el consumo humano, desde el uso de químicos e hidráulicos para su potabilización, contando con instalaciones de bombeo y tratamiento, para distribuirla a través de redes y conexiones;
- **Alcantarillado sanitario:** es un sistema compuesto por colectores, conducción y tratamiento primario hasta su disposición final; y,
- **Alcantarillado de aguas lluvias:** que incluye un sistema de tuberías y sumideros para facilitar el desalojo de las aguas lluvias para transportarlas para su tratamiento final, evitando así inundaciones.

Para asegurarse de que INTERAGUA CIA. LTDA., cumpla con las obligaciones contractuales, sobre todo en lo que respecta a la atención de consultas y reclamos, se ha dispuesto el proceso de Tercera instancia-Recurso de Apelación bajo la política de brindar con calidad la atención oportuna a los usuarios que la solicitan; todo ello bajo su compromiso con la mejora continua en sus procesos, con el cumplimiento de los requerimientos de los usuarios y requisitos legales aplicables y con el apoyo de

profesionales competentes y utilizando infraestructura adecuada para el cumplimiento de sus obligaciones y la satisfacción del usuario final.

El trabajo ha sido realizado incluyendo la estructura que se muestra a continuación:

En el capítulo I, se encuentra de manera detallada la problemática, la cual está relacionada con el objeto de estudio, su delimitación y su hipótesis a comprobar.

En el capítulo II, se mencionan los elementos teóricos, conceptos que intervienen en el proceso objeto de estudio, así mismo, se introduce el aspecto legal y el marco contextual de la investigación.

En el capítulo III, hace referencia a la metodología de la investigación donde se detalla explícitamente las características de la investigación, la técnica y herramienta que se empleó para el levantamiento de la información con el desarrollo del análisis de las respuestas obtenidas sobre el caso de estudio.

En el capítulo IV, hace referencia a la propuesta de la investigación, la aplicación de las etapas del desarrollo del proceso en la herramienta KNIME para la generación del modelo predictivo y el análisis de resultados.

CAPÍTULO I

EL PROBLEMA

En este capítulo se encuentra el detalle del problema observado en la entidad; teniendo relación directa con el planteamiento estructural del mismo, la definición de los objetivos a cumplir dentro de la investigación, la delimitación y la propuesta de solución al problema.

1.1. PLANTEAMIENTO DEL PROBLEMA

Si bien INTERAGUA Cía. Ltda., como parte de su funcionamiento, lleva un registro de reclamos de sus usuarios, atención ofrecida y estándares de soluciones presentadas, EMAPAG EP ha buscado mejorar los indicadores de atención al usuario, haciendo inclusiva su participación en la toma de decisiones del recurso de apelación del servicio de agua potable en la ciudad de Guayaquil.

Por ello, y con base a esos indicadores de gestión, la EMAPAG EP, como ente regulador de la concesionaria INTERAGUA, busca determinar similitudes en reclamos, en un período específico, para predecir la resolución óptima de los mismos a corto, mediano o largo plazo, con la debida justificación del tiempo utilizado para ofrecer soluciones, considerando que el tiempo de respuesta de dichos reclamos sea en el menor tiempo posible para el usuario y el uso de los diferentes recursos utilizados por parte de la empresa no se exceda de lo necesario.

Por tanto, desde la mirada de EMAPAG EP, se deberá determinar la efectividad en la atención a los usuarios, en cuanto a operatividad, frecuencia y similitud de reclamos, tiempos de solución, razones de los reclamos, y finalmente el nivel de satisfacción de la comunidad beneficiaria del servicio de la concesionaria.

1.2. HIPÓTESIS Y/O PREGUNTAS DE INVESTIGACIÓN

A través de esta investigación se buscará confirmar o negar la hipótesis y a responder la pregunta de investigación que se plantea:

Hipótesis:

Un modelo predictivo para mejorar los tiempos de respuestas a reclamos de Tercera Instancia-Recurso de Apelación presentados por los usuarios del servicio del servicio de agua potable en la ciudad de Guayaquil, que se generan por alta facturación en las planillas en periodos mensuales.

- **Variable independiente:** modelo predictivo
- **Variable dependiente:** identificación de patrones de reclamos de usuarios por la alta facturación de consumo de agua.

Pregunta de investigación:

¿Las técnicas de minería de datos permiten generar un modelo predictivo de solución para reclamos presentados por los usuarios del servicio de agua potable de la ciudad de Guayaquil?

1.3. OBJETIVOS

1.3.1. Objetivo General

Analizar los patrones de reclamos del servicio de agua potable por medio del uso de técnicas de minería de datos, que contribuya a facilitar la generación de informes respecto al monitoreo de la productividad de los recursos que intervienen en el proceso de Tercera Instancia-Recurso de Apelación.

1.3.2. Objetivos Específicos

- Realizar un diagnóstico del proceso de Tercera Instancia-Recurso de Apelación, con el fin de establecer los puntos críticos a ser solucionados;
- Diseñar la vista minable para la revisión de posibles escenarios respecto al caso de estudio, que permita analizar de manera predictiva el comportamiento de resultados;

- Evaluar los resultados obtenidos a partir de la aplicación de las técnicas de minería de datos, que facilite el diseño de un modelo para generar soluciones efectivas; y,
- Generar el modelo predictivo con empleo de los algoritmos necesarios respecto a la solución, para detectar patrones de reclamos y optimizar recursos.

1.4. JUSTIFICACIÓN Y ALCANCE

Esta investigación tiene el propósito de ofrecer, a través de la minería de datos, una herramienta válida para el análisis de procesos relativos a atención de reclamos y satisfacción de los usuarios, en el campo en que se desarrolla este estudio.

Los resultados podrán ser utilizados para analizar el nivel de eficiencia con que está desarrollando INTERAGUA, sus procesos y la importancia que da a reclamos de sus usuarios, así como los tiempos de atención; los beneficiarios, por un lado, son los integrantes de la comunidad, y, por otro lado, la misma empresa INTERAGUA, y la EMAPAG EP, podrán identificar cómo se optimiza el uso de recursos.

El presente estudio analizará el proceso utilizado para *“Tercera instancia-Recurso de Apelación”*; para la determinación del resultado más probable por cada tipo de reclamo establecido mediante patrones encontrados en la investigación. La investigación comprenderá únicamente a la EMAPAG EP, específicamente al área correspondiente la Jefatura de Atención al Usuario.

Cabe recalcar que se fundamentará en el uso de técnicas de minería de datos, con el objetivo de aplicarlas a una vista minable creada a partir de las bases de datos de reclamos comerciales correspondiente a los años 2015, 2016 y 2017 para obtener resultados, es decir, extraer patrones, tendencias y regularidades para describir y comprender mejor los datos y para predecir comportamientos futuros; utilizándolas respecto a su relación con el caso de estudio indicado.

Este estudio se enmarca dentro de la línea *investigación y desarrollo de nuevos servicios o productos* de la carrera Ingeniería en Sistemas Computacionales de la Facultad de Ingeniería de la Universidad Católica de Santiago de Guayaquil.

CAPÍTULO II

MARCO TEÓRICO, CONCEPTUAL Y LEGAL

Para abordar el aspecto de manejo de datos que intervienen en los procesos administrativos de las empresas, y diseñar proyecciones de cumplimiento y de satisfacción de los usuarios, es importante presentar, en este capítulo, un acercamiento a las diferentes teorías, principios y opiniones de expertos sobre este tema. Así mismo, en un segundo apartado se ofrece algunas conceptualizaciones de términos que están relacionados con el ámbito de estudio y, finalmente, se hace un resumen de algunas normativas que sustentan el funcionamiento de las empresas que ofrecen servicio a la comunidad, como es el caso de la EMAPAG EP e INTERAGUA CIA. LTDA.

2.1 ELEMENTOS TEÓRICOS

El avance de la tecnología hace que todos los procesos deban ser revisados y actualizados, así mismo, independientemente de la naturaleza del negocio o empresa, se hace necesario mantener almacenada la información resultante de su funcionamiento.

Con el tiempo, la información generada se convierte en un elemento importante que sienta las bases para una mejora permanente y que da pautas, además, para corregir errores pasados y optimizar los recursos. El análisis de los procesos y la información contenida en una base de datos debe ser realizado de forma ordenada y sistemática, clasificada por sectores de información o áreas de negocios. (Troche Clavijo, 2014)

La información crece continuamente y es difícil procesarla manualmente, además, “la información en sí misma tiene pocas ventajas, su sistematización, incorporación y utilización son los elementos que aportan su valor añadido: el conocimiento” (Rodríguez León, Lorenzo, & Matilde, 2016, p. 44), por ello y como consecuencia de un manejo adecuado de información, se facilita una toma de decisiones óptima.

La toma de decisiones dentro de las organizaciones trasciende por el empleo de los datos en proyecciones administrativas. La era digital maneja tres enfoques: estratégico, táctico y tecnológico; aplicando e implementando la analítica de datos, por lo antes mencionado y por un entorno económico, tecnológico y empresarial en el que

actualmente vivimos es considerado una variable relevante al momento de decidir cambiar o seguir con las mismas estrategias de negocio como consecuencia otorga responsabilidades a las entidades como la de ejecutar investigaciones de comportamientos en diferentes situaciones que adopta el mercado respecto a su rutina diaria. Sin embargo, la EMAPAP EP, específicamente la Jefatura de Atención al Usuario, carece del uso de un modelo predictivo que resuelva los reclamos en tercera instancia para minimizar los tiempos de respuesta de los mismo y aprovechar el uso de los recursos en otros procedimientos dentro de la jefatura, esta investigación tiene como objetivo el descubrimiento de nuevos conocimientos a partir de los datos manejados en la institución.

Las empresas que llevan un esquema bajo patrones aplicado a la gestión de clientes en el futuro usaran el conocimiento para hacer crecer el negocio; utilizando y tomando como referencia la ecuación de la figura 1.

Por eso se plantea la siguiente ecuación («El descubrimiento de Patrones», 2016):



Figura 1. Esquema de patrones. Autoria propia

Los patrones de datos para la gestión de clientes:

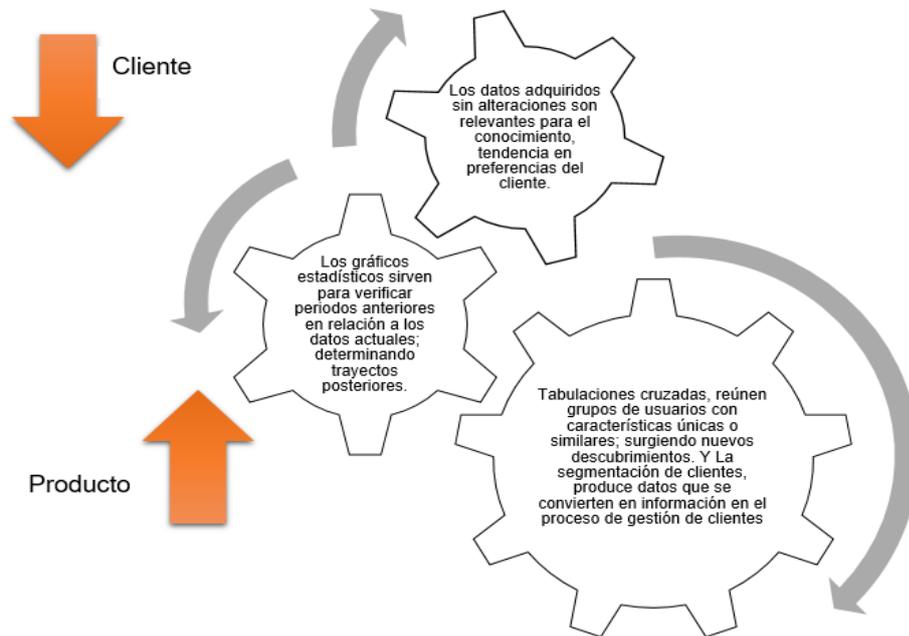


Figura 2. Gestion de Clientes, autoria propia.

Los patrones de personas en la gestión de clientes, incluye a: personas que trabajan para la empresa y los que no lo hacen para la misma por medio de la evaluación de comportamientos para la verificación del programa organizacional de trabajo a largo plazo.

Los patrones de procesos en la gestión de clientes, interviene la madurez de los procesos impulsa, la transformación de la gestión de clientes y las expectativas de los clientes, generan la evolución de la entidad por la satisfacción a ellos.

Para lograr entender la administración empresarial es necesario el descubrimiento de nuevos patrones de datos, personas y procesos para la toma de decisión gerencial.

Dentro de la creación de un proyecto de minería de datos se puede destacar sus fases de manera general (Roman Carrillo & Virseda Benito, s. f., p. 3) :

1. Filtrado de datos. _ eliminación de valores no válidos.
2. Selección de variables. _ variables predominantes en el problema.
3. Extracción de conocimiento. _ modelo de conocimiento.
4. Interpretación y evaluación. _ validación de resultados.

En el marco teórico de la investigación se hará referencia a los elementos teóricos más relevantes que formaron parte del producto final. Teniendo en cuenta el aporte de artículos científicos, revistas tecnológicas, trabajos de titulación, libros, etc., mencionare lo siguiente: *¿Cuál es la finalidad de las entidades para generar información?*, este suceso ha generado cuatro factores relevantes:

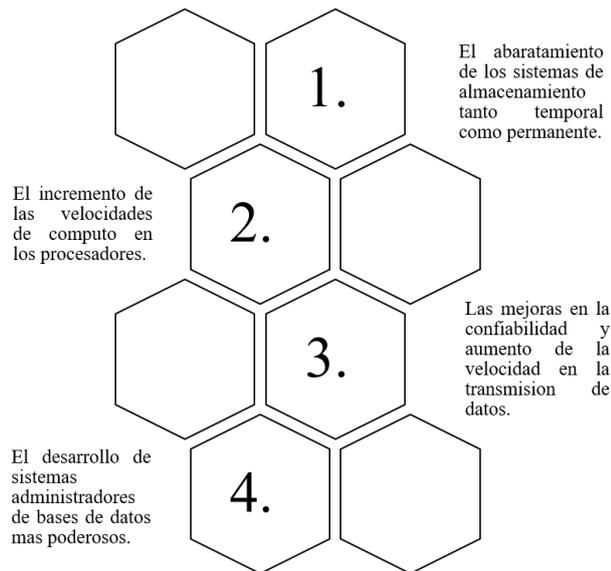


Figura 3. Factores relevantes en la generación de información. Información tomada de (Roman Carrillo & Virseda Benito, s. f., p. 1)

Para una mejor comprensión, se puede conceptualizar a la **minería de datos** como un proceso de la gestión de grandes cantidades de datos que busca descubrir nuevos conocimientos a partir de la detección de patrones, incluye mecanismos tecnológicos como: inteligencia artificial, estadística y sistema de base de datos («e-Ciencias de la Información», 2016). Sin embargo, la extracción de la información es parte de la etapa de análisis en bruto donde su ejecución traerá un beneficio único para la organización, la conversión de los datos en conocimiento futuro. Los modelos predictivos hallados deben demostrar lo significativo que son mediante el patrón de comportamiento resultante.

El objetivo del proceso analítico en los datos es el descubrimiento de patrones, esto hará crecer el análisis y se podrá visualizar como trabaja una entidad a través del tiempo y de los cambios evolutivos que ha tenido para una mejora continua en el proceso de planificación estratégica de negocio y el diseño de futuros cambios.

La minería de datos es utilizada en el campo de la inteligencia artificial, enfocada al descubrimiento de nuevos conocimientos, donde se pretende visualizar información a partir de datos específicos, donde la extracción y transformación de su organización surge a partir de la extracción de información; aplicando la minería de datos, como un proceso dentro de la metodología KDD (Knowledge Discovery from Databases) (Roman Carrillo & Virseda Benito, s. f., p. 2), es decir, descubrimiento de conocimientos en base de datos, donde se maneja una gran cantidad de datos que al ser procesados y convertidos en información pueden ser utilizados para posibles proyección futuras e incentivar la toma de decisiones gerenciales a través de los resultados obtenidos; partiendo de la detección de patrones respecto a variables que intervienen en la investigación de los proyectos.

La información utilizada en el procedimiento antes mencionado debe tener como característica principal la integridad de los datos, para que su uso en la implementación de algún modelo predictivo resulte de un alto nivel de confianza en su rendimiento final. Cabe recalcar que la información obtenida en los grandes almacenes de datos denominados por los peritos en materia tecnológica “base de datos”, a menudo puede producir vistas a partir de selecciones específicas en una búsqueda, es decir, el uso de la clasificación por sector de la información. La minería de datos tiene subprocesos como: extracción de datos, limpieza de datos, selección de características, algoritmos y análisis de resultados.

Las técnicas de minería de datos, se utilizan para descubrir patrones de comportamientos de ciertos grupos, los mismos que comparten características similares; hablando de los datos obtenidos del almacén de datos para lograr resultados que puedan justificar dicho comportamiento de la información.

En la figura 4 se ilustra la relación que existe entre dato, información y conocimiento. Se observa igualmente el volumen que presenta en cada nivel y el valor que los responsables de las decisiones le dan en esa jerarquía. El área interna dentro del triángulo representa los objetivos que se han propuesto. La separación del triángulo representa la estrecha unión entre dato e información, no así entre la información y el conocimiento. La minería de datos, trabaja extrayendo, analizando y decidir respecto

a la guía de, conductas, asociaciones, ordenamiento, que puedan generar algún modelo que nos permita comprender mejor el dominio para ayudar en una posible toma de decisión. (Roman Carrillo & Virseda Benito, s. f., p. 2)

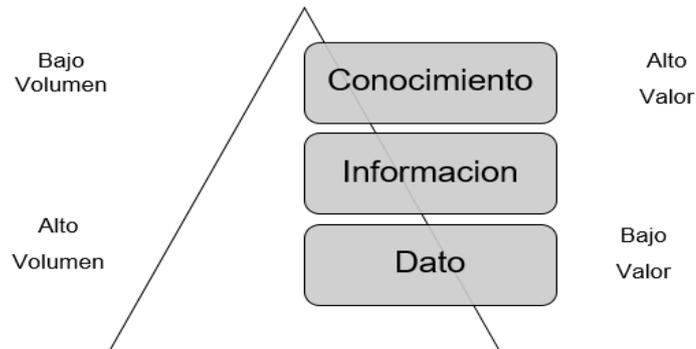


Figura 4. *Relación entre dato, información y conocimiento.*

La finalidad de utilizar minería de datos es obtener un modelo por medio de la creación de un modelo a partir de los datos, que permita finalmente la predicción con la aplicación de las técnicas de minería de datos, herramientas que están basadas en algoritmos de inteligencia artificial para el fin antes mencionado. En esta investigación se utilizarán métodos de descubrimiento, donde se hallarán patrones; siendo de carácter descriptivo o predictivo.

El producto obtenido de la aplicación de algoritmos de minería de datos, preverá comportamientos futuros de algún tipo de situación que maneje gran cantidad de datos, estos se clasifican en (García, 2017):

- De regresión: Para la identificación de tendencias. La regresión lineal es la más empleada, aunque también son frecuentes las parabólicas, polinómicas y logarítmicas. Sin embargo, estos algoritmos son bastante ineficientes en espacios multidimensionales en los que intervienen más de dos variables. En estos casos suele ser mejor utilizar algoritmos basados en modelos probabilísticos o estadísticos.
- De clasificación: Para predecir variables discretas o para determinar atributos comunes. Algunos algoritmos de agrupamiento o clustering como

los K-means y los K-medoids suelen dar buenos resultados cuando se buscan formas de agrupación basadas en la distancia entre puntos que representan la distancia media entre elementos de un grupo. Mediante técnicas iterativas se pueden buscar soluciones que converjan rápidamente hacia un óptimo local, incluso en grandes conjuntos de datos.

- De segmentación: Para organizar los datos en categorías y subcategorías. Los árboles de decisión son una técnica muy empleada tanto en clasificación como en segmentación. En determinados problemas de este tipo también serán de utilidad algunos tipos de redes neuronales e incluso los algoritmos de soporte vectorial (SVM).
- De series temporales: Para analizar una secuencia de eventos e identificar patrones con una finalidad predictiva. En series sencillas algunas técnicas de regresión pueden ser de utilidad, aunque tanto las máquinas SVM como el deep learning funcionan muy bien en este tipo de tareas.
- De asociación: Para buscar correlaciones y vínculos no lineales en conjuntos de datos. Esto permitirá descubrir reglas de asociación que podrán emplearse con carácter explicativo o predictivo. Los modelos de asociación basados en casos permiten agrupar los elementos que comparten una etiqueta o propiedad en un espacio de búsqueda determinado. Una vez creadas las reglas de asociación y establecida la probabilidad de ocurrencia entre elementos relacionados, el modelo permite elaborar predicciones basadas en la concurrencia de los elementos asociados. Un ejemplo típico de este tipo de predicciones sería sugerir a un cliente la compra de un nuevo producto en base a las últimas compras que ha realizado.
- De interpretación: para extraer significados en el ámbito del text mining. La búsqueda de información significativa en textos es una de las tareas más complejas del data mining. Este tipo de algoritmos combinan técnicas de deep learning con lingüística computacional. Un ejemplo de este tipo sería el análisis dinámico del sentimiento del mercado partiendo de una base de noticias financieras en tiempo real.

En su estudio científico (Gil, Ledis, & Valencia, 2012, p. 2) definen que:

Con el término “descubrimiento de conocimiento en bases de datos”, Knowledge Discovery Databases o KKD, fue designado el proceso que pretende obtener conocimiento a partir de datos almacenados en bodegas (data warehouse). En dicho proceso se incluyen la preparación de los datos, el análisis estadístico, el algoritmo usado para la minería de datos y la evaluación e interpretación de los mismos, obteniendo como resultado el descubrimiento de conocimiento.

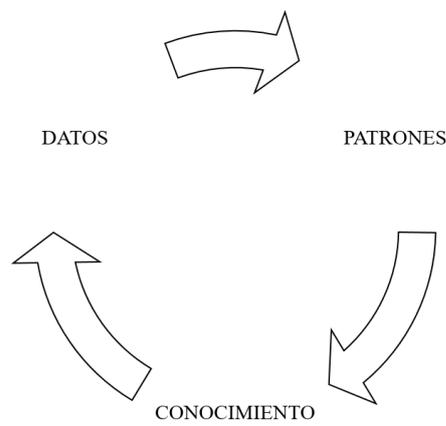


Figura 5. *Proceso de descubrimiento del conocimiento (KDD)*

El proceso KDD (Knowledge Discovery in Databases, Descubrimiento de conocimientos en bases de datos), que se muestra en la Figura 6, según (Pacheco Aguila & Fernández Segredo, 2015) permite interacciones entre la herramienta informática y el usuario final, repetitivo; comprometiendo pasos con la participación del usuario en la toma de decisiones a partir de la extracción de información relevante o patrones.

Se resume en las siguientes nueve etapas, según Pacheco Águila & Fernández Segredo (2015) : Recolección de datos y dominio del problema, Selección y limpieza de datos, Transformación de los datos, Reducción de datos y proyección, Elección de la técnica de minería de datos, Elección del algoritmo de minería de datos, Minería de datos, Interpretación y la utilización del conocimiento obtenido.

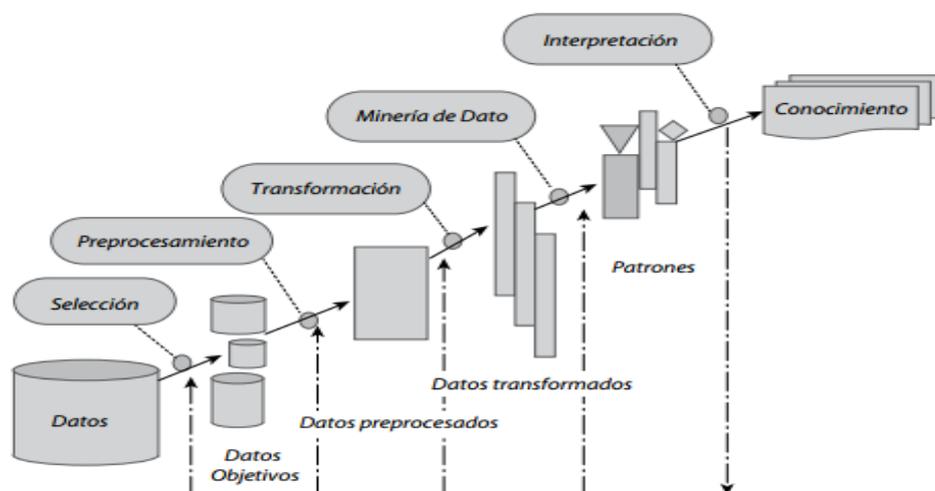


Figura 6. Etapas del proceso de la metodología KDD. Tomado de (Alvarado Pérez, Caicedo Zambrano, Hernández Arteaga, Hidalgo Troya, & Timarán Pereira, 2016)

Las etapas de la metodología KDD (Alvarado Pérez et al., 2016, pp. 63-86):

En la *etapa de selección*, una vez identificado el conocimiento relevante y prioritario y definidas las metas del proceso KDD, desde el punto de vista del usuario final, se crea un conjunto de datos objetivo, seleccionando todo el conjunto de datos o una muestra representativa de este, sobre el cual se realiza el proceso de descubrimiento.

La selección de los datos varía de acuerdo con los objetivos del negocio.

En la *etapa de pre-procesamiento/limpieza (data cleaning)* se analiza la calidad de los datos, se aplican operaciones básicas como la remoción de datos ruidosos, se seleccionan estrategias para el manejo de datos desconocidos (missing y empty), datos nulos, datos duplicados y técnicas estadísticas para su reemplazo. En esta etapa, es de suma importancia la interacción con el usuario o analista.

En la *etapa de transformación/reducción de datos*, se buscan características útiles para representar los datos dependiendo de la meta del proceso. Se utilizan métodos de reducción de dimensiones o de transformación para disminuir el número efectivo de variables bajo consideración o para encontrar representaciones invariantes de los datos. Los métodos de reducción de dimensiones pueden simplificar una tabla de una base de datos horizontal o verticalmente. La reducción horizontal implica la eliminación de tuplas idénticas como producto de la sustitución del valor de un atributo

por otro de alto nivel, en una jerarquía definida de valores categóricos o por la discretización de valores continuos (por ejemplo, edad por un rango de edades). La reducción vertical implica la eliminación de atributos que son insignificantes o redundantes con respecto al problema, como la eliminación de llaves, la eliminación de columnas que dependen funcionalmente (por ejemplo, edad y fecha de nacimiento).

En la etapa de *minería de datos*, El objetivo de la etapa minería de datos es la búsqueda y descubrimiento de patrones insospechados y de interés, aplicando tareas de descubrimiento como clasificación. Las técnicas de minería de datos crean modelos que son predictivos o descriptivos. Los modelos predictivos pretenden estimar valores futuros o desconocidos. El proceso de descubrimiento de conocimiento en bases de datos de variables de interés, que se denominan variables objetivo, dependientes o clases, usando otras variables denominadas independientes o predictivas, como por ejemplo predecir para nuevos clientes si son buenos o malos basados en su estado civil, edad, género y profesión, o determinar para nuevos estudiantes si desertan o no en función de su zona de procedencia, facultad, estrato, género, edad y promedio de notas. Entre las tareas predictivas están la clasificación y la regresión. Los modelos descriptivos identifican patrones que explican o resumen los datos; sirven para explorar las propiedades de los datos examinados, no para predecir nuevos datos, como identificar grupos de personas con gustos similares o identificar patrones de compra de clientes en una determinada zona de la ciudad.

Entre las tareas descriptivas se cuentan las reglas de asociación, los patrones secuenciales, los clustering y las correlaciones. Por lo tanto, la escogencia de un algoritmo de minería de datos incluye la selección de los métodos por aplicar en la búsqueda de patrones en los datos, así como la decisión sobre los modelos y los parámetros más apropiados, dependiendo del tipo de datos (categóricos, numéricos) por utilizar.

En la etapa de *interpretación/evaluación*, se interpretan los patrones descubiertos y posiblemente se retorna a las anteriores etapas para posteriores iteraciones. Esta etapa puede incluir la visualización de los patrones extraídos, la remoción de los patrones redundantes o irrelevantes y la traducción de los patrones útiles en términos que sean entendibles para el usuario. Por otra parte, se consolida el conocimiento descubierto para incorporarlo en otro sistema para posteriores acciones o, simplemente, para

documentarlo y reportarlo a las partes interesadas; también para verificar y resolver conflictos potenciales con el conocimiento previamente descubierto.

Las principales metodologías para el proceso de minería de datos son: KDD, CRISP-DM (proceso estándar, jerárquico (niveles de abstracción), de preferencia por la comprensión del negocio) y SEMMA (proceso de selección, exploración y modelamiento de grandes cantidades de datos para descubrir patrones de negocios desconocidos). («Metodología CRISP-DM - Parte 1», 2016).

Tabla 1

Comparación entre las metodologías aplicadas en minería de datos

| KDD | CRISP-DM | SEMMA |
|--------------------------------------|------------------------------|-----------------------------|
| Selección | Conocimiento el negocio | Selección |
| Pre- Procesamiento/Limpieza | | Exploración |
| Transformación/Reducción de datos | Conocimiento de los datos | Transformación/Variables |
| Minería de datos | | Modelamiento de datos |
| Interpretación/Evaluación | | Evaluación/Medidas/Reportes |

Nota: Comparación de las fases de las metodologías aplicadas en minería de datos. En la investigación se aplicará la metodología KDD, proceso que consta con etapas consecutivas; automatizando el procesamiento de los datos como consecuencia más periodo de tiempo en análisis y al descubrimiento de la correlación de los datos.

Se puede decir que según la Universidad Nacional Autónoma de México (UNAM) (s. f.), el reconocimiento de patrones es la ciencia que se encarga de la descripción y clasificación (reconocimiento) de objetos, personas, señales, representaciones, etc. Esta ciencia trabaja con base en un conjunto previamente establecido de todos los posibles objetos (patrones) individuales a reconocer.

La UNAM indica que:

“El reconocimiento automático, descripción, clasificación y agrupamiento de patrones son actividades importantes en una gran variedad de disciplinas científicas, como biología, sicología, medicina, visión por computador, inteligencia artificial, teledetección, etc. Un sistema de reconocimiento de patrones tiene uno de los siguientes

objetivos: Identificar el patrón como miembro de una clase ya definida (clasificación supervisada) y Asignar el patrón a una clase todavía no definida (clasificación no supervisada, agrupamiento o clustering).

El diseño de un sistema de reconocimiento de patrones se lleva a cabo normalmente en tres fases: la primera, adquisición y pre-proceso de datos, la segunda, extracción de características y la tercera, toma de decisiones o agrupamiento.” (Universidad Nacional Autónoma de México, s. f.)

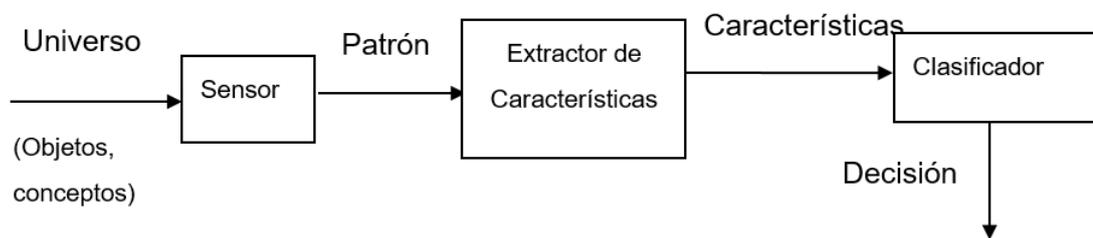


Figura 7. Esquema General de un Sistema de Detección de Patrones. Tomado de (Universidad Nacional Autónoma de México, s. f.)

2.2 CONCEPTOS

Para la mejor comprensión del trabajo de titulación, se establece algunos términos poco usuales para los lectores interesados en la investigación; dando a conocer las definiciones y su relevancia respecto a la minería de datos como herramienta estratégica y de apoyo para la toma de decisiones gerenciales.

El termino *Análisis Predictivo*, se incluye dentro del proceso de minería de datos relacionado con el futuro de las cosas (Aisemberg, 2013) explica que el mismo permite extraer conclusiones confiables sobre eventos futuros, a través de la aplicación de métodos estadísticos, matemáticos y de reconocimiento de patrones. La variable más relevante dentro del proceso es el predictor, con cierto nivel de fiabilidad.

Tabla 2

Diferencias entre BI y Análisis Predictivo

| Business Intelligence | Análisis Predictivo |
|--|---|
| Captura y analiza Descriptivo | Predice (Comportamientos y preferencias) Predictivo |
| Respecto al pasado | Respecto al futuro |
| Preguntas BI: <ul style="list-style-type: none"> • ¿Cuántos productos vendimos? • ¿Cuál es la tasa de morosidad? | Preguntas Análisis Predictivo: <ul style="list-style-type: none"> • ¿Qué clientes están por comprar? • ¿Qué clientes pagaran? |

Nota: Diferencias entre BI y Análisis Predictivo. La comparación establece un resultado favorable para la investigación en la cual se aplicará el análisis predictivo para la toma de decisiones gerenciales de la entidad.

En su estudio (Aisemberg, 2013) indica que:

Modelo Predictivo, se utiliza para proyectar comportamientos futuros. Conceptualmente se encarga de la recopilación de datos para ser procesados mediante indicadores con un nivel de relevancia alto, se realiza la formulación del modelo estadístico, predicciones y la validación del modelo con los datos disponibles. Todo modelo predictivo es montado sobre una herramienta sofisticada o software a medida para las pruebas con los datos obtenidos. (párr. 3)

En su estudio de “Los modelos predictivos en el nuevo marketing” (Merodio, 2014) indica que:

Los *modelos predictivos* podemos definirlos como una serie de técnicas matemáticas que combinadas adecuadamente nos ayudan a obtener posibles comportamientos futuros, en base al análisis de datos históricos y actuales que crean modelos estadísticos cada vez más fiables. Esta herramienta lógicamente tecnológica es una oportunidad para el análisis de datos en el tiempo para la

mejora continua del presente y futuro, y se lo debemos a la unión de Big Data y modelos predictivos. (párr. 2)

El termino *Vista minable*, fue mencionado en el *Análisis de información y minería de datos para la toma de decisiones* según (diegocond, 2017) se centra en que: “Una vista minable se refiere a un conjunto de datos que según su relevancia son los más importantes dentro del contenedor de datos y asimismo, agrupar estos datos de una forma que se puedan procesar con más facilidad”.

El análisis de datos no es algo nuevo, como lo indica Osores Melisa, Editora asociada para América Latina de TechTarget S.A., en su blog: Desde hace años, las empresas tratan de aprovechar la información que reciben de sus usuarios, canales, proveedores, etc. con el fin de tomar mejores decisiones de negocios. Sin embargo, la cantidad, calidad, formato y fuentes de la información se han ido transformando en los últimos años, y las herramientas para analizarla se han ido adaptando a estos cambios.

Adicionalmente, sobre la minería de datos se puede ejecutar un conjunto de técnicas para realizar análisis predictivos y de tendencias; esto se conoce como analítica de datos (data analytics). («Big data, el siguiente paso en la analítica de datos», s. f.)

El modelo de madurez en la administración de la información nos da una mejor idea de cómo se organizan estos procesos en una organización.



Figura 8. Modelo de madurez en administración de la información. Tomado de («Modelo de madurez en administración de la información», s. f.)

En la ciudad de Guayaquil, la Empresa Municipal de Agua Potable y Alcantarillado de Guayaquil, EP («Tercera Instancia – EMAPAG | Sitio Web de EMAPAG-EP Empresa Municipal de Agua Potable y Alcantarillado de Guayaquil», 2017) tiene el compromiso de servicio a la comunidad en todas sus dependencias, como es de conocimiento público la entidad supervisa el trabajo que diariamente realiza la concesionaria, encargada de los servicios básicos tales como: agua potable, alcantarillado sanitario y alcantarillado pluvial (aguas lluvias).

El recurso de *Tercera Instancia*, la atención del reclamo es la responsabilidad que la entidad realiza a diario; dando cumplimiento a sus objetivos de la calidad del servicio.

La EMAPAG EP, analiza la respuesta de INTERAGUA, al reclamo del usuario, para garantizar sus derechos. Cuando un reclamo ingresado en dos instancias previas ante en INTERAGUA, no satisface al usuario, EMAPAG EP atenderá el requerimiento de recurso de apelación presentado por escrito, adjuntando los requisitos correspondientes. EMAPAG-EP analiza, investiga y revisa la respuesta que se emitió en las instancias, a fin de garantizar el fiel cumplimiento del Reglamento Interno de Manejo de los Servicios.

Cuando el reclamo del usuario ya ha sido atendido dos veces por INTERAGUA, y el usuario no está conforme se recomienda a los mismos utilizar el ***Recurso de Tercera Instancia***.

2.3 BASE LEGAL

Normativas, leyes y reglamentos que sustentan el área de mi investigación.

La EMAPAG, EP, dentro del marco legal contribuye con su servicio a la comunidad bajo los siguientes reglamentos relevantes que interviene en el proceso de ***Recurso de Apelación-Tercera Instancia***:

Reglamento Interno de Manejo de los Servicios de Agua Potable, Alcantarillado Sanitario y Drenaje Pluvial en el Catón Guayaquil, versión 09 de abril de 2010, respalda a los Art. 1 al 20 comprendidos en el Reglamento de Servicios (*Resolución No. 1.12.04.2020*):

- Anexo I._ Instructivo para el proceso de crítica interna a seguir por parte de la empresa, previo a la emisión de facturas.

- Anexo II. _ Infracciones y sanciones a los usuarios.
- Anexo III. _ Valor de los servicios administrativos, técnicos y especiales.
- Anexo IV. _ Instructivo que regula la conexión, medición y facturación de los servicios de agua potable y alcantarillado sanitario para inmuebles multifuncionales.
- Anexo V._ Factibilidad de servicios y nuevos desarrollos:
 - A.- Instructivo para la expedición de informes de factibilidad de los servicios de agua potable, alcantarillado sanitario y drenaje pluvial en proyectos de desarrollo inmobiliario, industrial, comercial o urbanístico.
 - B.- Instructivo para la revisión de estudios y/o diseños sobre infraestructura de agua potable, alcantarillado sanitario y drenaje pluvial de proyectos de desarrollo inmobiliario, industrial, comercial o urbanístico.
- Anexo VI. _ Garantía bancaria para recepción del Sistema de Depuración de Aguas Residuales Domésticas (SDARD).
- Reglamento de la Estructura Tarifaria para la Prestación de los Servicios de Agua Potable y Alcantarillado para el Cantón Guayaquil para el Segundo Quinquenio de la Concesión, dando cumplimiento a los siguientes objetivos:
 - Cubrir los costos totales anuales que corresponden a: los gastos de operación y mantenimiento; Los impuestos y gravámenes; los cargos de depreciación; el costo de capital aplicado al valor de las inversiones netas del mismo año; y los costos de regulación y control.
 - Asegurar la sostenibilidad financiera de la prestación de los servicios, con un adecuado plan de expansión, rehabilitación, metas de calidad y rentabilidad aceptable' c, La facturación por consumos registrados para permitir el control y la aplicación de políticas de gestión técnica y comercial- en beneficio directo al usuario.

En el portal web, el usuario cuenta con información correspondiente al procedimiento llevado en la EMAPAG EP, adicionalmente es de conocimiento público el Pliego Tarifario y Estructura, la tabla con los valores respecto al rango de consumo mensual:

Tabla 3

Pliego Tarifario y Estructura

| Tarifa Referencial | Cargo Variable de Agua Potable | | | Cargo Fijo | |
|---|---------------------------------|----------|---------------------|-------------------------|---------------------|
| | Rango de Consumo m ³ | Usuarios | US\$/M ³ | Diámetro de la conexión | US\$/M ³ |
| Octubre 2017 US\$0,590 m ³ | 0-15 | 287.537 | 0,3068 | 1/2" | 1,19 |
| | 16-30 | 158.808 | 0,4543 | 3/4" | 7,97 |
| | 31-60 | 62.006 | 0,643 | 1" | 20,49 |
| | 61-100 | 9.541 | 0,836 | 1 1/2" | 34,15 |
| | 101-300 | 3.657 | 0,939 | 2" | 34,15 |
| | 301-2,500 | 855 | 1,492 | 3" | 56,93 |
| | 2,501-5,000 | 58 | 1,931 | 4" | 170,77 |
| | 5,000 o mas | 36 | 3,216 | 6" o más | 227,70 |
| | TOTAL | 522.498 | | | |

Nota: La EMAPAG EP, mensualmente muestra en su portal web la tarifa mensual del metro cubico respecto al rango de consumo para información del usuario con todos los valores por recaudación de impuestos que son variables.

La adopción de un sistema de gestión de la calidad (ISO 9001:2008) es una decisión estratégica para una organización que le puede ayudar a mejorar su desempeño global y proporcionar una base sólida para las iniciativas de desarrollo sostenible.

Los beneficios potenciales para una organización de implementar un sistema de gestión de la calidad basado en esta Norma Internacional son, ver anexo 1, para mayor información:

- La capacidad para proporcionar regularmente productos y servicios que satisfagan los requisitos del cliente y los legales y reglamentarios aplicables;

- Facilitar oportunidades de aumentar la satisfacción del cliente;
- Abordar los riesgos y oportunidades asociadas con su contexto y objetivos;
- La capacidad de demostrar la conformidad con requisitos del sistema de gestión de la calidad especificados.

Esta Norma Internacional puede ser utilizada por partes internas y externas. No es la intención de esta Norma Internacional presuponer la necesidad de:

- Uniformidad en la estructura de los distintos sistemas de gestión de la calidad;
- Alineación de la documentación a la estructura de los capítulos de esta Norma Internacional;
- Utilización de la terminología específica de esta Norma Internacional dentro de la organización.

Los requisitos del sistema de gestión de la calidad especificados en esta Norma Internacional son complementarios a los requisitos para los productos y servicios.

Esta Norma Internacional emplea el enfoque a procesos, que incorpora el ciclo Planificar-Hacer-Verificar-Actuar (PHVA) y el pensamiento basado en riesgos.

El enfoque a procesos permite a una organización planificar sus procesos y sus interacciones.

El ciclo PHVA permite a una organización asegurarse de que sus procesos cuenten con recursos y se gestionen adecuadamente, y que las oportunidades de mejora se determinen y se actúe en consecuencia.

El pensamiento basado en riesgos permite a una organización determinar los factores que podrían causar que sus procesos y su sistema de gestión de la calidad se desvíen de los resultados planificados, para poner en marcha controles preventivos para minimizar los efectos negativos y maximizar las oportunidades a medida que surjan.

El cumplimiento permanente de los requisitos y la consideración constante de las necesidades y expectativas futuras representa un desafío para las organizaciones en un entorno cada vez más dinámico y complejo. Para lograr estos objetivos, la

organización podría considerar necesario adoptar diversas formas de mejora además de la corrección y la mejora continua, tales como el cambio abrupto, la innovación y la reorganización.

Cabe recalcar que la EMAPAG EP, se encuentra en proceso de actualización de la norma ISO 9001:2015, auditoria que fue superada con éxito.

MARCO CONTEXTUAL

2.4 GLOBAL ¿DÓNDE SE VA A DESARROLLAR?

La EMAPAG, EP contribuye a la sociedad de la ciudad de Guayaquil con la mejora continua en las condiciones básicas de vida y salud de la población; controlando que INTERAGUA, ejecute sus responsabilidades de manera eficiente en los servicios de agua potable, alcantarillado sanitario y drenaje pluvial en el Cantón Guayaquil.

En la siguiente lista se destacan los objetivos de calidad de la EMAPAG EP, («Objetivos de la Calidad | Sitio Web de EMAPAG-EP Empresa Municipal de Agua Potable y Alcantarillado de Guayaquil», s. f.):

- Obtener la calificación “buena” y “excelente” en el 85% de las encuestas mensuales de satisfacción a los usuarios atendidos en recurso de apelación.
- Responder a los usuarios los recursos de apelación en treinta y cinco (35) días máximo en el 85% de los recursos de apelación atendidos mensualmente.
- Cumplir con el 90% mínimo de la capacitación programada anualmente para el personal de atención al usuario.
- Cumplir con el 90% mínimo del programa de mantenimiento anual preventivo de equipos de información y tecnológicos.
- Lograr que el nivel de servicio del Call center sea mínimo del 80% mensual.
- Cumplir en el 100% con los requerimientos legales aplicables a tercera instancia, recurso de apelación.
- Obtener la calificación “buena” y “excelente” en el 85% de las encuestas mensuales de satisfacción a los usuarios por las inspecciones ejecutadas.

La problemática a resolver involucraría el beneficio para la Jefatura de Atención al Usuario en la inversión del tiempo utilizado para la resolución de reclamos en tercera

instancia por alta facturación en casos o situaciones que realmente ameriten más tiempo y el uso razonable de los recursos disponibles dentro de la jefatura.

En la actualidad el número de reclamos por alta facturación que se encuentran en Recurso de Apelación-Tercera Instancia pueden ser atendidos mediante el modelo predictivo respecto a las similitudes encontradas posteriormente con las variables que encaminan esta investigación.

2.5 INFRAESTRUCTURA TECNOLÓGICA

El desarrollo y prestación de los servicios en Tecnologías de la Información y comunicaciones, ha tenido resultados exitosos gracias a las facilidades que otorga la plataforma tecnológica implementada en la institución; reflejando un mejor servicio prestado a la población de la ciudad de Guayaquil.

La Unidad de Tecnologías de la Información, con la ayuda de avanzadas herramientas tecnológicas de información y comunicación da apoyo a los procesos administrativos; garantizando la seguridad de la información, continuidad y seguridad en el servicio.

Servidor de Aplicaciones

Contiene los ejecutables y binarios de los aplicativos administrativos y financieros.

Servidor de Base de Datos

Base de Datos Oracle 12G, contiene toda la data de los aplicativos administrativos y financieros, así como los perfiles y roles de usuarios para el acceso a los mismos aplicativos.

CAPÍTULO III

METODOLOGÍA Y RESULTADOS DE LA INVESTIGACIÓN

3.1. METODOLOGÍA DE INVESTIGACIÓN

En este capítulo se describirá la metodología a utilizar y los resultados del respectivo levantamiento de información, de esta manera las entrevistas utilizadas como herramienta para el procesamiento de la información se extraerán los puntos más relevantes que servirán para darle un enfoque más claro a la investigación; concluyendo con un análisis de resultados de acuerdo a la información extraída en las entrevistas.

3.2. TIPO Y MÉTODO DE INVESTIGACIÓN

La investigación se la puede definir como una agrupación de procesos que son aplicados a la observación de un problema. *“Minería de Datos aplicada a La Detección de Patrones de Reclamos del Servicio de Agua Potable en la Ciudad de Guayaquil”*, tiene un enfoque mixto, es decir, cuantitativo, ya que uno de los objetivos de la investigación es encontrar un modelo predictivo en base a datos cuyos comportamientos se definen por patrones; y, cualitativo, ya que existe documentación observable para analizar; partiendo del resumen de datos obtenidos a lo largo del periodo investigativo.

Es además una investigación histórica ya que se parte de situaciones anteriores que, en el caso de la investigación, dependerá del estado de la resolución que demande cada reclamo procesado por la Jefatura para hallar relación con los distintos factores variables que pueden llegar a influir en el resultado.

La finalidad de la investigación es fundamental ya que el resultado será la producción de conocimientos nuevos a partir de simples datos y la aplicación de los mismos para efectuar cambios representativos en la entidad que sean favorables a largo plazo; ya que la obtención de datos en el tiempo ocurre más de una vez para determinar comportamiento o patrones a futuro para el análisis respecto a la toma de decisiones gerenciales.

La investigación es exploratoria en cuanto a profundidad, ya que se basa en el análisis específico sobre la minería de datos y su aplicación en el mundo de los negocios como herramienta tecnológica.

El método a aplicar y desarrollar en la investigación es Inductivo-Sintético, ya que el punto inicial de la investigación son los datos que generarían el resultado del estado de un reclamo y esto a su vez se determina por un patrón que será utilizado de manera general en la resolución de reclamos de usuarios; analizando globalmente los parámetros o comportamientos que debe cumplir el reclamo.

3.3. POBLACIÓN, MUESTRA Y MUESTREO

La investigación no está definida por población y muestra ya que se basa fundamentalmente en la revisión, diseño de vista, pruebas de datos históricos en base a documentación de la jefatura. Cabe recalcar que el uso de modelo predictivo a diseñar será utilizado en la Jefatura de Atención al Usuario, siendo beneficiaria directa la entidad.

3.4. TÉCNICAS E INSTRUMENTOS PARA RECOLECCIÓN DE INFORMACIÓN

La técnica de recolección de datos de la investigación a ejecutar es la entrevista, de manera individual a los tres miembros claves del proceso de reclamo en la EMAPAG-EP; utilizando como canal de comunicación, el cara a cara y correo electrónico, manejado libremente, en su desarrollo constaran menos de cinco preguntas de tipo abierta.

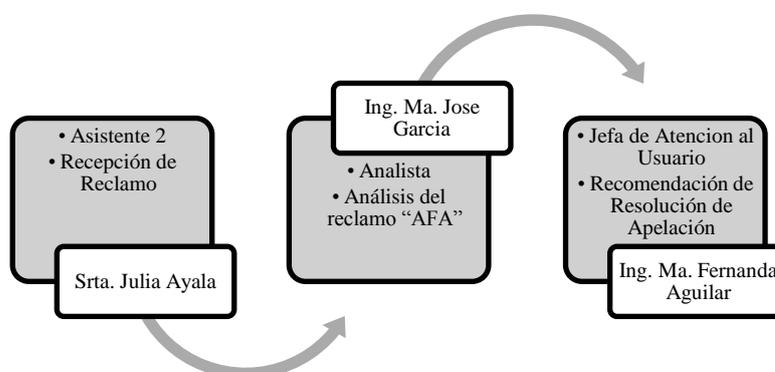


Figura 9. Proceso de reclamos en EMAPAG EP y sus miembros claves.

3.5. HERRAMIENTA PARA PROCESAMIENTO DE INFORMACIÓN

La herramienta utilizada como instrumento de recolección de datos es una guía de entrevista previamente informada a los miembros claves del proceso.

Tabla 4

Codificación para los entrevistados

| Entrevistado | Código |
|--|--------|
| Srta. Julia Ayala, asistente 2, recepta el reclamo | E1 |
| Ing. María José García, analista, analista del recurso de apelación | E2 |
| Ing. María Fernanda Aguilar, Jefa de Atención al Usuario, aprueba las resoluciones de los recursos de apelación. | E3 |

Nota: La codificación, facilita la manipulación de la información dentro de la investigación.

Guía de Entrevista

Preguntas para E1:

1. ¿Cómo se llama el procedimiento que realizas en la JAU?
2. Descripción breve acerca del procedimiento antes mencionado.
3. ¿Cómo te das cuenta que el reclamo que presenta el usuario no procede en la dirección?

Preguntas para E2:

1. ¿Cómo se llama el procedimiento que realizas en la JAU?
2. Descripción breve acerca del procedimiento antes mencionado.
3. ¿Cuáles son los tipos de resultados que se obtienen de las resoluciones que manejas diariamente?

4. Descripción breve acerca de cada caso antes mencionado. ¿Por qué el análisis da ese resultado?
5. Los casos negados, tienen un alto porcentaje de demanda en la Jefatura. ¿Cuál es tu opinión?

Preguntas para E3:

1. ¿Cuánto es el tiempo de duración en la atención de reclamos negados?
2. Opinión acerca de “Porcentaje de casos negados vs casos aprobados”
3. ¿Cuál es la cantidad de reclamos negados?
4. Descripción del proceso de Recurso de Apelación - Tercera instancia

Tabla 5

Matriz de la Guía de Entrevistas

| Preguntas Abiertas | E1 | E2 | E3 |
|--|---|---|---|
| <p>Pregunta#1:</p> <p>Descripción del Proceso de Reclamo en EMAPAG EP.</p> | <ol style="list-style-type: none"> 1. Ingreso de reclamo. 2. Análisis de reclamo. 3. Recomendación de reclamo. 4. Revisión de jurídico. 5. Aprobación de gerencia. | <ol style="list-style-type: none"> 1. Ingreso de reclamo. 2. Análisis de reclamo. 3. Recomendación de reclamo. 4. Revisión de jurídico. 5. Aprobación de gerencia. | <ol style="list-style-type: none"> 1. Ingreso de reclamo. 2. Análisis de reclamo. 3. Recomendación de reclamo. 4. Revisión de jurídico. 5. Aprobación de gerencia. |
| <p>Pregunta#2:</p> <p>Descripción del Procedimiento del Recurso de Apelación.</p> | | <p>Verificación de algún incumplimiento por parte de la Concesionaria al Reglamento de Servicios Internos, de ser el caso el reclamo inmediatamente sale favorable para el usuario.</p> | <p>El proceso de Recurso de Apelación – Tercera Instancia está regido bajo el Reglamento Interno de Manejo de los Servicios:</p> <p>Art. 89.- Recurso de apelación del usuario ante ECAPAG o reclamo de tercera instancia.</p> |
| <p>Pregunta#3:</p> <p>¿Cómo te das cuenta que el reclamo que presenta el usuario no procede?</p> | | <p>Cuando existe: Consumos regularizados. Verificando si hay IEC.</p> | <p>Cuando existe: Consumos y lecturas Cuanto es el promedio histórico de consumo del usuario Existencia de IEC “Incremento Extraordinario de Consumo”</p> |
| <p>Pregunta#4:</p> <p>¿Cuál es la cantidad de reclamos negados en la JAU?</p> | | <p>80%/100% al mes</p> | <p>Negados 444 de enero a noviembre del 2017.</p> |

Nota: Respuestas corta de las tres entrevistas realizadas a personal clave dentro del proceso de Recurso de Apelación en la EMAPA EP.

3.6. ANÁLISIS DE RESULTADOS

En la entrevista realizada al informante E1, se evidenció que esta persona desempeña un rol importante en la dirección, se lo puede considerar el primer filtro de decisión de un reclamo proveniente de INTERAGUA. Cabe recalcar que dos son las entidades que intervienen en estos procesos, cada una con una función específica, pero con el mismo objetivo que es el de brindar servicio a la comunidad, y para ello cada una solicita a sus usuarios distintos requisitos para el ingreso de su denuncia; en el caso de la EMAPAG EP se solicita lo siguiente: copia de cédula de ciudadanía, carta dirigida a la empresa y la resolución o informe en segunda instancia por parte de la Concesionaria. Uno de los puntos importantes mencionados por E1 fue la validación del reclamo presentado por el usuario, para lo cual considera relevante los siguientes parámetros:

1. ¿Cuándo un reclamo procede?

Cuando existe:

- Consumos no regularizados, se visualiza en el link de Interagua
- Proceso de crítica (Incremento Extraordinario de Consumo)
- Fuga interna
- Fuga externa

2. ¿Cuándo un reclamo no procede?

Cuando existe:

- Consumos regularizados, se visualiza en el link de Interagua (lectura actual – lectura anterior = consumo mensual m3)
- Verificando lo siguiente cuando el usuario insiste: “Sumatoria de los consumos de los últimos seis meses de acuerdo a lo que se visualiza por medio del link de Interagua, dividir esa cantidad para seis y finalmente multiplicar el valor por dos. De esta manera nos damos cuenta si el consumo del mes en reclamo pasa el 100% procede caso contrario no procede.

En la entrevista realizada al informante E2, se evidenció que su papel tiene un rol importante en la dirección y en el procedimiento de la respuesta al recurso de apelación, es decir, en la resolución del conflicto entre el usuario y la concesionaria.

La resolución es la respuesta al reclamo que el usuario interpuso en tercera instancia en EMAPAG EP, luego del ingreso del reclamo que lo realiza E1 en el ERCO, y con los respectivos requerimientos pasan a la fase de análisis conjuntamente con la información proporcionada por INTERAGUA referente al caso. La misión de E2 es la verificación de algún incumplimiento por parte de la Concesionaria al Reglamento de Servicios Internos, de ser el caso el reclamo inmediatamente sale favorable para el usuario.

Casos de AFA “Alta Facturación”, se observa:

1. Consumos y lecturas
2. Cuanto es el promedio histórico de consumo del usuario
3. Existencia de IEC “Incremento Extraordinario de Consumo”, es decir, que la lectura del mes en reclamo supere el 100% del promedio histórico de consumo del usuario. En este caso es aplicable el proceso de crítica

Proceso de crítica:

- Aviso del incremento del 100% de consumo
 - Inspección para la verificar lo que produce el IEC
 - Existencia de fuga, si es el caso el usuario tiene 28 días para su reparación inmediata.
4. INTERAGUA cuando aplica el proceso de crítica cobra mensualmente el promedio histórico.
 5. Prevención al usuario por parte de INTERAGUA.

Los tipos de resoluciones que maneja la Jefatura de Atención al Usuario son los siguientes:

1. **Revocar.**_ resolución favorable.
 - IEC
 - Fugas visibles o no visibles que INTERAGUA no se las haya prevenido al usuario, dando el plazo por reglamento para el arreglo respectivo.
2. **Confirmar.**_ resolución negada porque no existe incumplimiento por parte de la concesionaria.

- No existe alta facturación.
- Lecturas bien tomadas, consecutivas.
- Medidor en perfecto estado de funcionamiento.
- Fuga de responsabilidad del usuario.

“Consumos que el usuario reclama pero que son registrados por el medidor”

3. **Revocar parcialmente.**_ resolución parcialmente favorable (un periodo sí y otro no).
4. **Confirmar con conciliación.**_ reclamo negado por reglamento pero se le consigue al usuario una conciliación por disposición de fuga con INTERAGUA.

Cabe recalcar que en la ciudad de Guayaquil actualmente los casos negados tienen un alto porcentaje en la Jefatura de Atención al Usuario; generalmente las peticiones del usuario no proceden sea por desconocimiento de los usuarios respecto a los trámites y a la falta de un plan de concientización de la población con respecto al cuidado y uso adecuado del servicio de agua potable, mantenimiento de sus instalaciones internas, observar mensualmente los accesorios en los baños y tuberías viejas de sus establecimientos, etc. Mensualmente del 100%/80% pertenecen a los casos negados que ingresan a la Jefatura de Atención al Usuario.

En la entrevista realizada al informante E3, se evidenció que desempeña un rol importante en la dirección y la aprobación a la resolución al recurso de apelación. Para la atención de reclamos el Reglamento Interno de Manejo de los Servicios determina el término máximo de 35 días hábiles para la resolución administrativa en recurso de apelación. El porcentaje de los casos recibidos en la EMAPAG EP son negados el 56% enero a noviembre del 2017 y aprobados son el 43%, es notorio que existe desconocimiento del Reglamento Interno de Manejo de los Servicios, en el cual constan los derechos y obligaciones de los usuarios y de la Operadora. Los reclamos negados son superiores al 50% mensual, ya que generalmente no aplican debido a fugas de agua internas no visibles y por la falta de cultura de mantenimiento preventivo en las instalaciones intradomiciliarias de los usuarios.

Los reclamos negados incluyen la opción de una inspección gratis, interna y coordinada para el usuario con supervisores de EMAPAG EP y con la presencia de la

Operadora. Si el usuario corrige las causas del alto consumo se concilia con la Operadora un descuento.

Cantidad de reclamos negados:

1. Negados 444 de enero a noviembre del 2017
2. Aprobados 339 de enero a noviembre del 2017
3. Parcialmente aprobados 39 de enero a noviembre del 2017.

El informante E3, hizo referencia al proceso de Recurso de Apelación – Tercera Instancia está regido bajo el Reglamento Interno de Manejo de los Servicios, en referencia al **Art. 89.- Recurso de apelación del usuario ante ECAPAG o reclamo de tercera instancia**, menciona que toda decisión expresa recaída en relación a un reclamo de segunda instancia ejercitado, podrá ser objeto de apelación ante ECAPAG por parte del usuario afectado. Mientras se sustancia el recurso de apelación ante ECAPAG, el pronunciamiento que hubiere emitido la Empresa, quedará suspendido en sus efectos.

El ejercicio y tramitación de este derecho estará sujeto a las siguientes reglas:

a) Persona legitimada.- El recurso de apelación deberá ser solicitado por la misma persona que hubiere planteado la revisión o por quien ejerciere su representación mediante delegación escrita;

b) Formalidades.- El recurso será presentado por escrito en las oficinas de ECAPAG, dentro de un plazo de treinta (30) días de notificado el pronunciamiento objeto del recurso. El recurrente deberá formular una crítica precisa y detallada a la decisión cuestionada, no pudiendo ofrecer nuevos medios de prueba;

c) Remisión de antecedentes.- ECAPAG cursará copia de la apelación a la Empresa dentro de un (1) día hábil de recibida, y ésta deberá, en un término máximo de cinco (5) días hábiles enviar el expediente conformado por la Empresa con la atención del reclamo en primera y segunda instancia, además de la información esencial que requiera ECAPAG para la atención del recurso;

d) Sustanciación.- ECAPAG podrá realizar las diligencias de prueba que estime oportunas, incluso requiriendo a la Empresa documentación adicional o informes especiales relativos al reclamo, los que deberán ser cumplidos dentro de los cinco (5)

días hábiles de notificado el pedido, salvo cuando la producción de los informes implique actividades que requieran razonablemente mayor plazo, en cuyo caso ECAPAG fijará el término respectivo; asimismo el término máximo señalado para ECAPAG en el literal f) se extenderá por el mismo lapso;

e) Desistimiento por parte del usuario.- El desistimiento por parte del usuario se lo realizará ante ECAPAG durante la sustentación del reclamo de tercera instancia y podrá ser presentado o notificado hasta antes de que se emita resolución;

f) Resolución.- ECAPAG deberá emitir resolución dentro del término máximo de treinta y cinco (35) días hábiles de recibida la apelación, término dentro del cual deberá sustanciar la prueba que estime conveniente. Vencido este término, el recurrente tendrá el derecho a solicitar ante ECAPAG la aceptación de su reclamo.

Si ECAPAG incurre en silencio administrativo a causa de retrasos ocasionados por la Empresa, los efectos inherentes a dicho silencio correrán por cuenta de la Empresa. Si ECAPAG incurre en silencio administrativo por su responsabilidad, responderá ante la Empresa por los efectos inherentes a dicho silencio, debiendo reembolsarle a ésta un valor equivalente a aquel que la Empresa, a efectos de ejecutar la tácita resolución de ECAPAG, tuviere que gastar y/o que reconocer en cualquier forma a favor del usuario y que no hubiera debido gastar o reconocerle si la resolución de la apelación hubiera sido favorable a la Empresa. Los valores que en este caso correspondan a la Empresa se los pagará directamente ECAPAG.

La resolución de ECAPAG se limitará a confirmar lo actuado y/o resuelto por la Empresa, o a revocar o modificar total o parcialmente la decisión cuestionada; y,

g) Ejecución de la decisión. La decisión de la apelación será notificada a las partes por ECAPAG dentro de las cuarenta y ocho (48) horas de adoptada, y será cumplida por la Empresa o el recurrente, según corresponda. En lo relativo a la Empresa, esta obligación de cumplimiento será sin perjuicio al derecho descrito en el siguiente artículo.

Cabe recalcar que días posteriores a la realización del levantamiento de la información, el informante E3 mencionó que hubo un cambio en el Reglamento Interno específicamente al procedimiento de reclamo emitida el 25 de octubre del 2017 “Sentencia Corte Constitucional Reglamento Interno Manejo de Servicios”, de este

particular que ya ha sido comunicada por la Concesionaria el 07 de diciembre del 2017 mediante oficio al Ente Regulador; indicando la anulación de la segunda instancia en el procedimiento de reclamos entre INTERAGUA y EMAPAG EP, lo cual de ninguna manera afecta al desarrollo del modelo predictivo para la Jefatura de Atención al Usuario.

CAPÍTULO IV

PROPUESTA

4.1. INTRODUCCIÓN

La minería de datos a lo largo de la investigación ha sido identificada como mecanismo fundamental de la inteligencia de negocios, por su capacidad de transformar los datos; otorgando utilidad de información. Este procedimiento involucra cuatro aspectos importantes:

- Distribución._ definir una estructura general para el empleo en los datos nuevos.
- Asociación._ descubrimiento de grupos y estructuras.
- Estudio de patrones de agrupaciones._ nuevas variables y su vínculo.
- Retroceso._ localización y modelación de datos con el mínimo error.

Las técnicas de minería de datos, se clasifican en:

1. No Supervisado, variables de la misma jerarquía y condición, es decir, al mismo nivel.
 - a. Asociación (reglas de asociación)
 - b. Agrupamiento (Clustering)
2. Supervisado, variables de supervisión
 - a. Clasificación (árboles de decisión)
 - b. Predicción (árboles de regresión)

En la investigación con la transformación de los datos y obtención de la vista minable, se visualizarán variables útiles para la generación del modelo predictivo con una sola variable a pronosticar o predecir respecto a las variables independientes, es decir, el set de variables que nos permitirán construir el pronóstico o la predicción con el uso de algoritmos, que analicen los datos proporcionados por el levantamiento de la información, en busca de tendencias de tipo específicos.

El modelo de minería de datos que crea un algoritmo partiendo de los datos puede tomar diversas formas, como:

- Conjunto de colección de objetos que son similares entre sí, que detallan la relación de los datos de un conjunto de datos.
- Segmentación predictiva de un resultado, que especifica cómo le afectan los diferentes principios.
- Conjunto de reglas, que delimitan la agrupación de las variables independientes y las posibilidades de que los Recursos de Apelación se resuelvan en función de los parámetros primarios.

Los resultados a obtener estarán expuestos en patrones respecto a los diferentes algoritmos que se utilicen para predecir el resultado de los Recursos de Apelación que cumplan con las reglas obtenidas a lo largo de la investigación; generando el modelo predictivo para las especificaciones del problema e influir en la toma de decisiones gerenciales.

4.2. METODOLOGÍA

La metodología a utilizar será KDD (Knowledge Discovery in Databases) “Descubrimiento de conocimiento en base de datos”, ya que esta metodología dentro de sus etapas incluye la minería de datos como parte del procedimiento para la predicción de modelos según los objetivos del modelo de negocio. Dentro de los beneficios y razones por las que esta investigación toma la iniciativa por esta metodología es porque el KDD es un proceso centrado en el usuario, es decir, que permite una interacción, a modo de diálogo, entre la máquina y el usuario. De esta manera el modelo podrá ser guiado por las decisiones que toma el usuario o un agente inteligente. También posee una característica favorable relevante que hace atractivo su

implementación, asistencia al usuario en la selección de técnicas y herramientas para la obtención de los objetivos.

Lo más relevantes dentro del proceso de KDD es lo siguiente:

1. Datos
2. Selección
3. Datos seleccionados
4. Pre-proceso
5. Datos Pre-procesados
6. Algoritmos de minería de datos
7. Modelo
8. Post-proceso; interpretación y validación

En la investigación se determinan relaciones entre los datos; aplicando la fase de modelamiento, donde los algoritmos inteligentes son aplicados con el objetivo de extraer patrones desconocidos inicialmente; tomando como origen, los datos.

4.3. SELECCIÓN DE LA HERRAMIENTA PARA IMPLEMENTAR LA METODOLOGIA

La herramienta para implementar la metodología es KNIME Analytics Platform, ya que es una plataforma utilizada para la innovación respecto a los datos y el descubrimiento de su potencial. KNIME, es una plataforma a código abierto, en la cual se pueden implementar y probar todos los algoritmos principales que se aplicaran en la investigación. Mediante la *Tabla 5* se pone en conocimiento la existencia de las diferentes herramientas que trabajan a partir de datos.

Tabla 5

Herramientas de Minería de Datos

| Minería de Datos | | | | | | |
|---|--|----------------|-------------------|---------------------|---|--------------------|
| | Nombre | Tipo/Sw | Plataforma | Arquitectura | Algoritmos | Tipo/Modelo |
| Herramientas de Minería de Datos | Orange Data Mining FruitFul&Fun | Libre | Multiplataforma | | | Predictivo |
| | RapidMiner | Libre | Windows, Linux | Cliente/Servidor | Clustering, árboles de decisión, redes neuronales | Predictivo |
| | Weka | Libre | Windows, Linux | | Clustering, Regresión | Predictivo |
| | JHepWork | Libre | | | | Predictivo |
| | KNIME | Libre | Multiplataforma | | Algoritmos segmentación, árboles de decisión, redes neuronales, SVM | Predictivo |

Nota: Las herramientas de minería de datos, ayudan en el descubrimiento de patrones; extrayendo conocimiento de los datos para alcanzar la excelencia mediante la estrategia empresarial.

4.4. DESARROLLO DE LA METODOLOGIA

El desarrollo de la metodología, trae consigo la participación del usuario para la toma de decisiones. El proceso contiene el siguiente detalle:

1. Selección

La selección de los datos, respecto a los objetivos de esta investigación ha llevado a la creación de un conjunto de datos objetivo, es decir, la investigación estará centrada en dos bases de datos principales:

- Recurso de apelación
- Inspecciones

Las bases de datos provienen del sistema ERCO, el cual es operado por la Jefatura de Atención al Usuario, el módulo de Atención al Cliente. La tabla de recurso de apelación contiene los siguientes campos:

Tipo_Reclamo, Aspecto, Subaspecto, Cuenta_Contrato, Identificación Cliente, Dirección, Teléfono, Respuesta, Motivo, Periodo, Fecha_Ingreso Resolución, Respuesta, Fecha_Resolucion, Secuencia, Acción Nombre_Accion

The screenshot shows a software window titled 'Atención al cliente - [Ingreso de Oficios]'. The interface includes a header with the ERCO logo and user information: 'Modulo: ATENCION AL CLIENTE', 'Programa: ACF0020', 'Concesionario: EMAPAG-EP', 'Fecha: 01-02-2010 15:38:03', and 'Usuario: MVERA'. The main area is divided into several sections:

- Datos del reclamo:** Includes fields for 'Tipo de Reclamo' and 'Medio de Promoción'.
- Datos del propietario del contrato:** Fields for 'Contrato', 'Cédula', 'Nombres', 'Apellidos', 'Dirección', 'Teléfono', and 'E-mail'.
- Datos de la persona que presenta el reclamo:** Fields for 'Parentesco', 'Cédula', 'Nombre', 'Dirección', 'Teléfono', and 'Celular'.
- Datos adicionales del reclamo:** Fields for 'Trámite', 'Resolución', 'Fecha Resolución', 'Aspecto', 'Subaspecto', 'Motivo', 'Hro. Queja', 'Fecha', 'Hora Provi', 'Respuesta', 'Periodo', 'F. Ingreso', 'Vcto', and 'Comentario'.
- Acciones Emapag:** A table with columns 'Acción', 'Usuario', 'Fecha Ingreso', and 'Estado'.

A small dialog box titled 'Calificador - Oper - Ecuinfo v1.0.15' is open in the bottom right, showing a dropdown menu with 'Recursos de Apelación' selected.

Figura 11. Módulo de Atención al Usuario, Ingreso de Recurso de Apelación.

La tabla de inspecciones contiene los siguientes campos:

Cuenta, Año, Id_Inspeccion, Secuencia, Fecha, Tipo_Inspeccion, Id_Inspector, Obs_Predio, Obs_Medidor, Obs_Pruexa, Obs_Anacon, Obs_Revvis, Obs_Rescon, Obs_Resgeo, Observa, Orden_Rubro, Desc_Rubro, Id_Concepto, Orden_Concepto, Concepto, Valor.

Formulario: 2017 10098 Reclamo: 9979617 Tipo: CONJUNTO Estado: Pendiente

Datos del Usuario
 Cuenta: 256506 Nombre: HECTOR RICHARD ORTIZ Teléfono: 2241784
 Sector: Dirección: SAMANES II MZ 220 VILLA Cod. Catastral:

Datos del Predio
 TIPO DE CONSTRUCCION
 UNIDADES HABITACIONALES
 UNIDADES NO HABITACIONA
 NUMERO DE FAMILIAS

Prueba de Exactitud al medidor
 MARCA
 SERIE
 CLASE
 PRESTION DE APP EN LA RED

| Lectura Inicial | Lectura Final | Volumen Pedido | Lectura Inicial | Lectura Final | Volumen Pedido | Tiempo Ensayo | Caudal Ensayo | Error Relativo | Error Med. Piloto | Error Med. Usuario |
|-----------------|---------------|----------------|-----------------|---------------|----------------|---------------|---------------|----------------|-------------------|--------------------|
| | | | | | | | | | | |
| | | | | | | | | | | |

Medidor cumple con la norma Sí No

Figura 12. Dirección de Regulación Técnica, Ingreso de Inspecciones.

2. Pre-procesamiento/limpieza

El Pre-procesamiento/limpieza, ayuda en la importancia de los datos obtenidos de la etapa de selección. En la investigación se han aplicado procedimientos para el empleo de datos forasteros, es decir, repetitivos y blanco; sumándole la interacción con la Jefa de Atención al Usuario para afinación del modelo.

Respecto a los valores, que no se encuentran en el rango de valores esperados; son visualizados en los datos por errores humanos, cambios en el sistema, información no disponible en un periodo de tiempo determinado.

3. Transformación/reducción

En la transformación/ reducción de datos, el modelo se acoge en la representación de los datos dependiendo del objetivo del proceso en general. Se utilizaron métodos de transformación de los datos para obtener la vista minable para la ejecución de las herramientas y los diferentes algoritmos estudiados.

Los métodos de transformación, tienen como objetivo reducir horizontal o verticalmente los datos. En la investigación se aplicó la reducción vertical; eliminando registros duplicados, insignificantes con respecto a la problemática.

La data recibida, se clasifico en campos de acuerdo al requerimiento de la Jefa de Atención al Usuario que guardan relevancia con el procedimiento analítico de los diferentes casos de Recursos de Apelación.

En la investigación se determina que la despersonalización de la información es relevante en el estudio, que coincide con la eliminación o sustitución de los campos que asocie lo estudiado, es decir, de manera específica y particular. Para encontrar patrones respecto a los reclamos de Tercera Instancia de receptados en la Jefatura, se utilizaron los atributos sector, estado_medidor, cambio_medidr, fuga, IEC, reparo_fuga, fuga_resp_IA, incumplimiento y para la búsqueda de patrones en relación a la minería de datos se utiliza respuesta_emapag de acuerdo a los estados de resoluciones manejados con frecuencia en la Jefatura.

Tabla 6

La vista minable generada en KNIME

| Atributo | Tipo | Valores posibles |
|-----------------|-------------|--|
| Sector | Nominal | Norte (N), Sur (S), Este (E), Oeste(O), Noroeste (NO), Noreste (NE), Suroeste (SE), Suroeste (SO). |
| Estado_medidor | Categórico | Bueno, Malo |
| Cambio_medidor | Nominal | Si (S), No (N) |
| Fuga | Categórico | Si (S), No (N) |
| IEC | Nominal | IES, No (N) |
| Reparo_fuga | Categórico | Si (S), No (N) |

| | | |
|------------------|------------|--|
| Fuga_resp_IA | Nominal | Si (S), No (N) |
| Incumplimiento | Catagórico | Si (S), No (N) |
| Respuesta_emapag | Nominal | Confirmar, Revocar, RP (Revocar Parcialmente) |

Nota: Diccionario de datos de la vista minable.

| | A | B | C | D | E | F | G | H | I | J |
|----|----------|--------|------------------|----------------|----------------|------|-----|-------------|--------------|----------------|
| 1 | CONTRATO | SECTOR | RESPUESTA_EMAPAG | ESTADO_MEDIDOR | CAMBIO_MEDIDOR | FUGA | IEC | REPARO_FUGA | FUGA_RESP_IA | INCUMPLIMIENTO |
| 2 | 8962688 | NO | CONFIRMAR | Bueno | N | N | N | N | N | N |
| 3 | 6592902 | NO | CONFIRMAR | Bueno | N | S | N | N | N | N |
| 4 | 5498990 | NO | CONFIRMAR | Bueno | N | N | N | N | N | N |
| 5 | 0029944 | NO | CONFIRMAR | Bueno | N | N | N | N | N | N |
| 6 | 5516122 | NO | REVOCAR | Malo | S | N | N | N | N | N |
| 7 | 5268970 | NO | CONFIRMAR | Bueno | N | N | N | N | N | N |
| 8 | 6653917 | NO | REVOCAR | Malo | S | N | N | N | N | N |
| 9 | 4467400 | S | REVOCAR | Bueno | N | S | IEC | S | N | N |
| 10 | 8595647 | NO | CONFIRMAR | Bueno | N | N | N | N | N | N |
| 11 | 6209578 | S | REVOCAR | Bueno | N | N | N | S | S | S |
| 12 | 5325695 | NO | CONFIRMAR | Bueno | N | N | N | N | N | N |
| 13 | 2116481 | S | CONFIRMAR | Bueno | N | N | N | N | N | N |
| 14 | 2622914 | NO | CONFIRMAR | Bueno | N | S | N | N | N | N |
| 15 | 1050203 | NO | REVOCAR | Malo | S | N | N | N | N | N |
| 16 | 7564860 | NO | CONFIRMAR | Bueno | N | N | N | N | N | N |
| 17 | 0099422 | NO | REVOCAR | Bueno | N | S | N | S | N | N |
| 18 | 6664326 | NO | CONFIRMAR | Bueno | N | N | N | N | N | N |
| 19 | 2010013 | S | CONFIRMAR | Bueno | N | N | N | N | N | N |
| 20 | 0000000 | S | CONFIRMAR | Bueno | N | N | N | N | N | N |

Figura 13. Campos seleccionados.

4. Minería de datos

En esta etapa, el objetivo principal es el descubrimiento de patrones. Como lo menciona (Aranda Robles & Sotolongo, 2013, p. 3), las técnicas de minería de datos constituyen un enfoque conceptual y, habitualmente, son implementadas por varios algoritmos. Estas pueden clasificarse, según su utilidad, en técnicas de clasificación, de predicción, de asociación o de agrupamiento (clustering), es decir, se crean modelos descriptivos y predictivos, que guardan relación con los objetivos a cumplir dentro de la investigación, como la generación de un modelo predictivo que permita establecer patrones para la toma de decisiones en la entidad.

En la investigación se usaran las siguientes técnicas:

- Las **técnicas de predicción**, permiten obtener pronósticos de comportamientos futuros a partir de los datos recopilados, de ahí que se apliquen frecuentemente.

- Las **técnicas de agrupamiento**, concentran datos dentro de un número de clases preestablecidas o no, partiendo de criterios de distancia o similitud, de manera que las clases sean similares entre sí y distintas de las otras clases. Su utilización ha proporcionado significativos resultados en lo que respecta a los clasificadores o reconocedores de patrones, como en el modelado de sistemas.
- Las **técnicas de reglas de asociación**, permiten establecer las posibles relaciones correlaciones entre distintas acciones o sucesos aparentemente independientes; pudiendo reconocer como la ocurrencia de un suceso o acción puede inducir o generar la aparición de otros.
- Las **técnicas de clasificación**, definen unas series de clases, en que se pueden agrupar los diferentes casos. Dentro de este grupo se encuentran las técnicas de árboles de decisión y reglas de inducción. (Aranda Robles & Sotolongo, 2013, p. 3)

5. Interpretación/evaluación

En esta etapa el conocimiento previamente descubierto, nos expone patrones o reglas del modelo predictivo que guardan relación con lo que se analiza diariamente en la Jefatura, sustentando los resultados con el modelo de negocio; por medio de la interpretación y evaluación llegar al establecimiento de un patrón para el modelo predictivo. Los resultados serán visualizados mediante la herramienta y la aplicación de los algoritmos correspondientes, donde lo más cercano a la realidad serán los resultados obtenidos y su variación de acuerdo al porcentaje de éxito o error de los datos, los cuales otorgaran conocimiento a esta investigación.

4.5. ELECCION DEL ALGORITMO, GENERACION DE REGLAS Y DESCRIPCION DEL MODELO PREDICTIVO

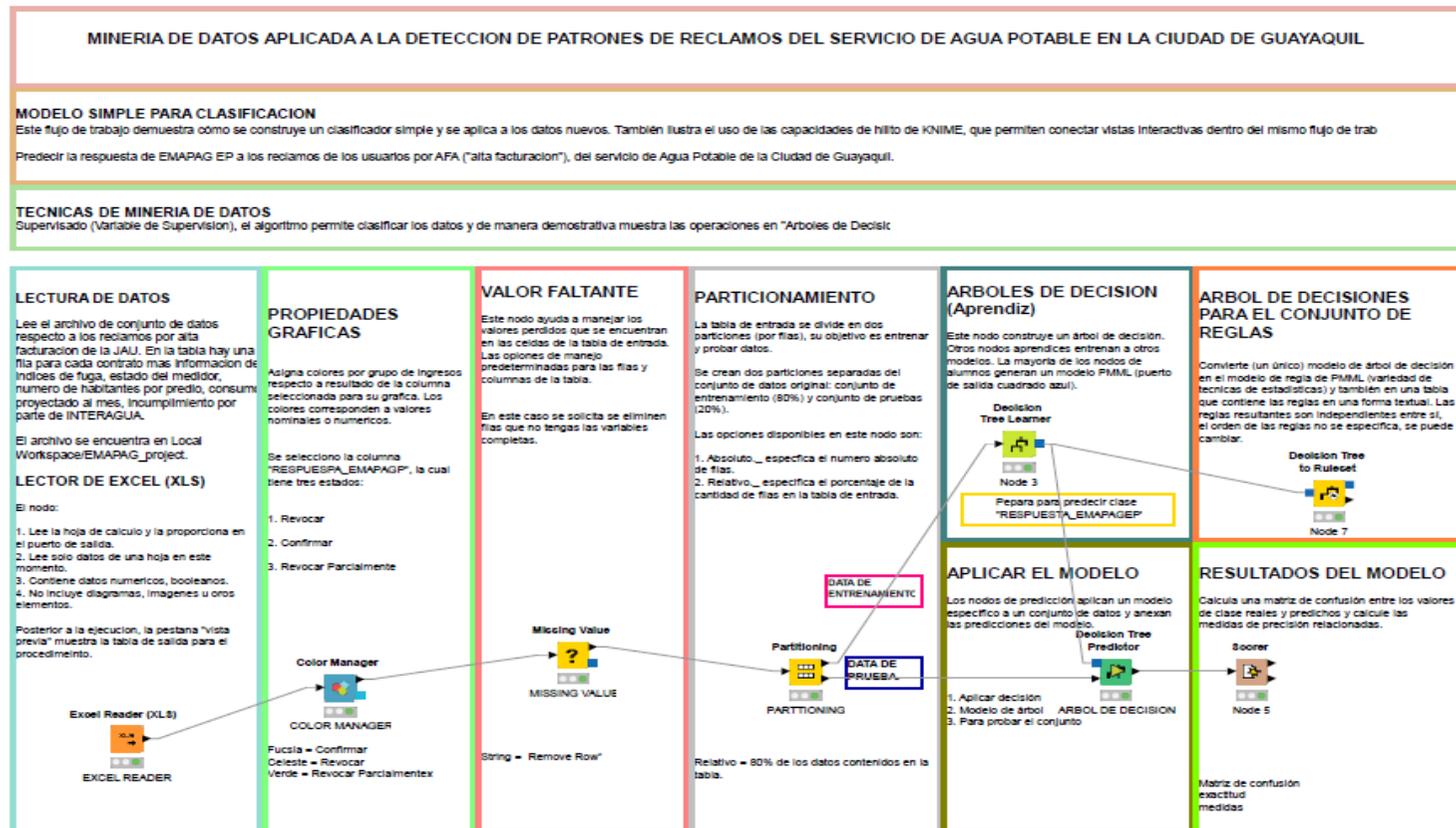


Figura 14. Flujo de Trabajo respecto al Algoritmo de Árbol de Decisión. Autoría propia.

MINERIA DE DATOS APLICADA A LA DETECCION DE PATRONES DE RECLAMOS DEL SERVICIO DE AGUA POTABLE EN LA CIUDAD DE GUAYAQUIL

CLASIFICADOR BAYESIANO SIMPLE Y RAPIDO

Este flujo de trabajo demuestra cómo se construye un clasificador simple y se aplica a los datos nuevos. También ilustra el uso de las capacidades de hilo de KNIME, que permiten conectar vistas interactivas dentro del mismo flujo de trabajo.

Predice la respuesta de EMAPAG EP a los reclamos de los usuarios por AFA ("alta facturación"), del servicio de Agua Potable de la Ciudad de Guayaquil.

TECNICAS DE MINERIA DE DATOS

Supervisado (Variable de Supervision), el algoritmo permite clasificar los datos y de manera demostrativa muestra las operaciones en "Navie Bayes".

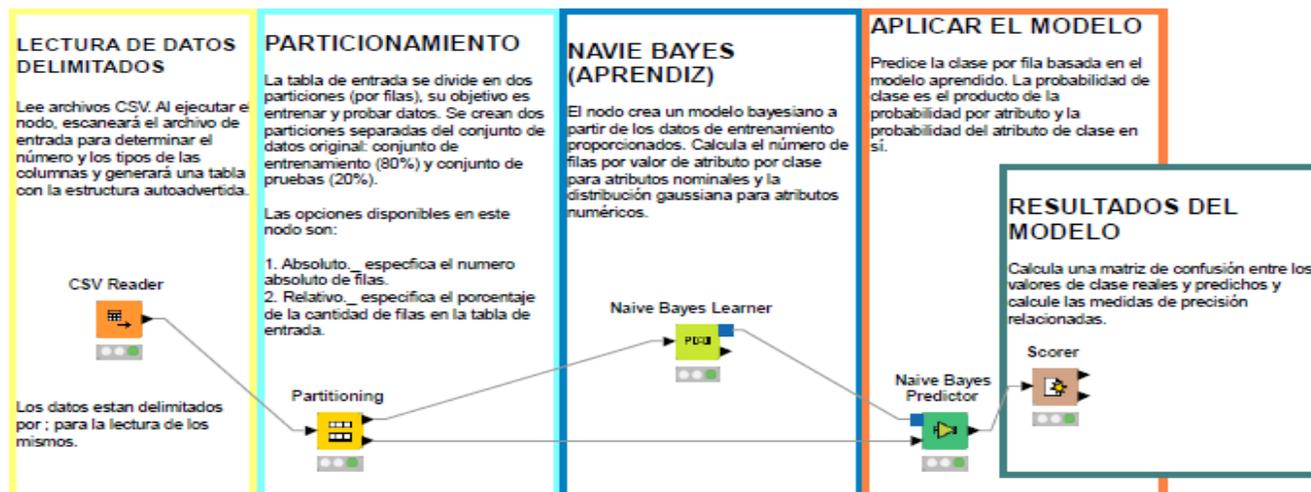


Figura 15. Flujo de Trabajo respecto al Algoritmo de Naive Bayes. Autoría propia.

MINERIA DE DATOS APLICADA A LA DETECCIÓN DE PATRONES DE RECLAMOS DEL SERVICIO DE AGUA POTABLE EN LA CIUDAD DE GUAYAQUIL

MODELO DE CLASIFICACION

Este flujo de trabajo demuestra cómo se construye un clasificador y se aplica a los datos nuevos. También ilustra el uso de las capacidades de hilo de KNIME, que permiten conectar vistas interactivas dentro del mismo flujo de trabajo.

Predecir la respuesta de EMAPAG EP a los reclamos de los usuarios por AFA ("alta facturación"), del servicio de Agua Potable de la Ciudad de Guayaquil.

TECNICAS DE MINERIA DE DATOS

WEKA (Waikato Environment for Knowledge Analysis) es una herramienta que permite la experimentación de análisis de datos mediante la aplicación, análisis y evaluación de las técnicas más relevantes de análisis de datos, principalmente las provenientes del aprendizaje automático, sobre cualquier conjunto de datos del usuario.

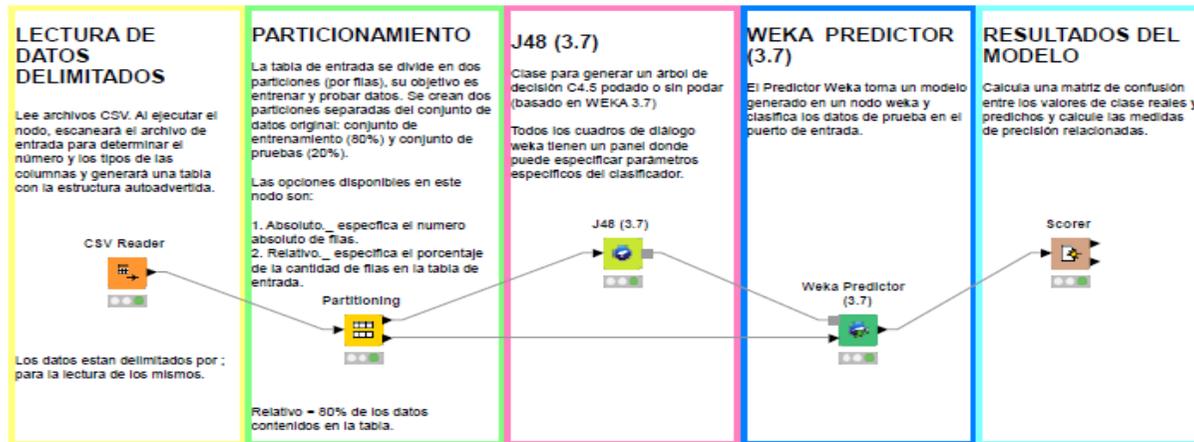


Figura 16. Flujo de Trabajo respecto al Algoritmo de WEKA j48. Autoría propia.

4.6. ALGORITMOS EN MINERIA DE DATOS

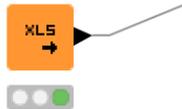
4.7. ARBOLES DE DECISION

Un árbol de decisión es un conjunto de condiciones organizadas en una estructura jerárquica, de tal manera que la decisión final a tomar se puede determinar siguiendo las condiciones que se cumplen desde la raíz del árbol hasta alguna de sus hojas. Una de las grandes ventajas de los árboles de decisión es que, en su forma más general, las opciones posibles a partir de una determinada condición son excluyentes. Esto permite analizar una situación y, siguiendo el árbol de decisión apropiadamente, llegar a una sola acción o decisión a tomar.

4.7.1. APLICACIÓN DEL ALGORITMO EN LA HERRAMIENTA

FLUJO DE TRABAJO – WORKFLOW

Excel Reader (XLS)



EXCEL READER

Lee el archivo de conjunto de datos respecto a los reclamos por alta facturación de la JAU. En la tabla hay una fila para cada contrato más información de índices de fuga, estado del medidor, número de habitantes por predio, consumo

proyectado al mes, incumplimiento por parte de INTERAGUA.

El archivo se encuentra en Local Workspace/EMAPAG_project.

LECTOR DE EXCEL (XLS)

El nodo:

1. Lee la hoja de cálculo y la proporciona en el puerto de salida.
2. Lee solo datos de una hoja en este momento.
3. Contiene datos numéricos, booleanos.
4. No incluye diagramas, imágenes u otros elementos.

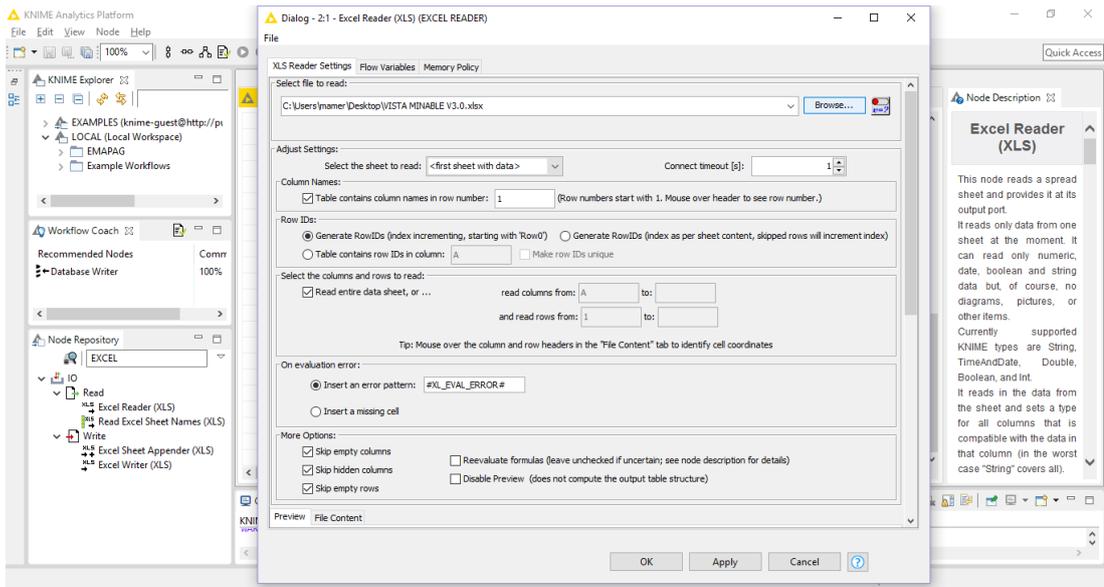


Figura 17. Entorno de Trabajo KNIME. Excel Reader (nodo)

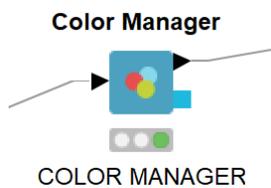
Output table - 2-1 - Excel Reader (XLS) (EXCEL READER)

Table "VISTA MINABLE V3.0.xlsx [16]js1" - Rows: 1679

| Row ID | \$ CLIENT... | \$ SECTOR | \$ RESPU... | \$ ESTAD... | \$ CAMBL... | \$ FUGA | \$ RESULT... | \$ REPAR... | \$ FUGA R... | \$ INCOM... |
|--------|--------------|-----------|--------------|-------------|-------------|---------|--------------|-------------|--------------|-------------|
| Row0 | 8962688 | NO | CONFIRMAR | Bueno | N | N | NO IEC | N | N | N |
| Row1 | 8592902 | NO | CONFIRMAR | Bueno | N | S | NO IEC | N | N | N |
| Row2 | 5488990 | NO | CONFIRMAR | Bueno | N | N | NO IEC | N | N | N |
| Row3 | 0029944 | NO | CONFIRMAR | Bueno | N | N | NO IEC | N | N | N |
| Row4 | 5516122 | NO | REVOCAR | Malo | S | N | NO IEC | N | N | N |
| Row5 | 5268970 | NO | CONFIRMAR | Bueno | N | N | NO IEC | N | N | N |
| Row6 | 8653917 | NO | REVOCAR | Malo | S | N | NO IEC | N | N | N |
| Row7 | 4467400 | S | REVOCAR | Bueno | N | S | IEC | S | N | N |
| Row8 | 8595647 | NO | CONFIRMAR | Bueno | N | N | NO IEC | N | N | N |
| Row9 | 6209578 | S | REVOCAR | Bueno | N | N | NO IEC | S | S | S |
| Row10 | 5325695 | NO | CONFIRMAR | Bueno | N | N | NO IEC | N | N | N |
| Row11 | 2116481 | S | CONFIRMAR | Bueno | N | N | NO IEC | N | N | N |
| Row12 | 2622914 | NO | CONFIRMAR | Bueno | N | S | NO IEC | N | N | N |
| Row13 | 1050203 | NO | REVOCAR | Malo | S | N | NO IEC | N | N | N |
| Row14 | 7564860 | NO | CONFIRMAR | Bueno | N | N | NO IEC | N | N | N |
| Row15 | 0099422 | NO | REVOCAR | Bueno | N | S | NO IEC | S | N | N |
| Row16 | 8664326 | NO | CONFIRMAR | Bueno | N | N | NO IEC | N | N | N |
| Row17 | 2010013 | S | CONFIRMAR | Bueno | N | N | NO IEC | N | N | N |
| Row18 | 3576375 | SE | REVOCAR P... | Bueno | N | S | NO IEC | P | S | N |
| Row19 | 5005157 | S | CONFIRMAR | Bueno | N | N | NO IEC | N | N | N |
| Row20 | 5009126 | NO | CONFIRMAR | Bueno | N | S | NO IEC | N | N | N |
| Row21 | 0002650 | S | REVOCAR | Malo | S | N | NO IEC | N | N | N |
| Row22 | 4920161 | NO | CONFIRMAR | Bueno | N | N | NO IEC | N | N | N |
| Row23 | 4946839 | NO | CONFIRMAR | Bueno | N | N | NO IEC | N | N | N |
| Row24 | 6086512 | NO | CONFIRMAR | Bueno | N | S | NO IEC | N | N | N |
| Row25 | 2601152 | S | REVOCAR | Bueno | N | S | NO IEC | S | N | N |
| Row26 | 5366132 | NO | REVOCAR | Bueno | N | N | IEC | S | S | S |
| Row27 | 6122145 | SO | CONFIRMAR | Bueno | N | N | NO IEC | N | N | N |
| Row28 | 7536696 | NO | REVOCAR | Bueno | N | PS | NO IEC | S | S | S |
| Row29 | 6388660 | NO | CONFIRMAR | Bueno | N | N | NO IEC | N | N | N |
| Row30 | 2541100 | SO | REVOCAR | Bueno | N | S | NO IEC | S | N | N |
| Row31 | 2999744 | S | CONFIRMAR | Bueno | N | N | NO IEC | N | N | N |
| Row32 | 7503540 | S | CONFIRMAR | Bueno | N | S | NO IEC | N | N | N |
| Row33 | 7009328 | NO | CONFIRMAR | Bueno | N | N | NO IEC | N | N | N |
| Row34 | 3047232 | S | CONFIRMAR | Bueno | N | N | NO IEC | N | N | N |
| Row35 | 0209009 | S | CONFIRMAR | Bueno | N | N | NO IEC | N | N | N |
| Row36 | 8028270 | NO | CONFIRMAR | Bueno | N | N | NO IEC | N | N | N |
| Row37 | 4401500 | NO | CONFIRMAR | Bueno | N | N | NO IEC | N | N | N |
| Row38 | 3512354 | SO | CONFIRMAR | Bueno | N | S | NO IEC | N | N | N |

Figura 18. Entorno de Trabajo KNIME. Posterior a la ejecución, la pestaña "vista previa" muestra la tabla de salida para el procedimiento.

PROPIEDADES GRAFICAS



Asigna colores por grupo de ingresos respecto a resultado de la columna seleccionada para su gráfica. Los colores corresponden a valores nominales o numéricos.

Se seleccionó la columna "RESPUESTA_EMAPAGP", la cual tiene tres estados:

1. Revocar (Celeste)
2. Confirmar (Fucsia)
3. Revocar Parcialmente (Azul)

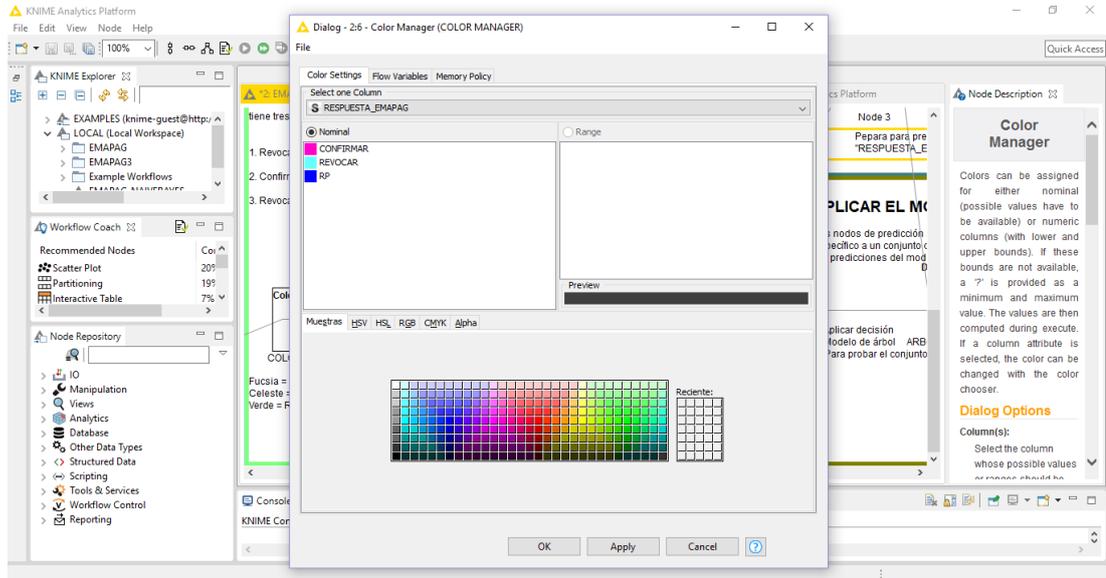
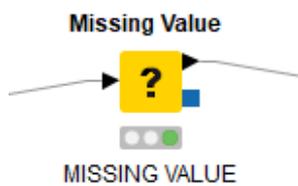


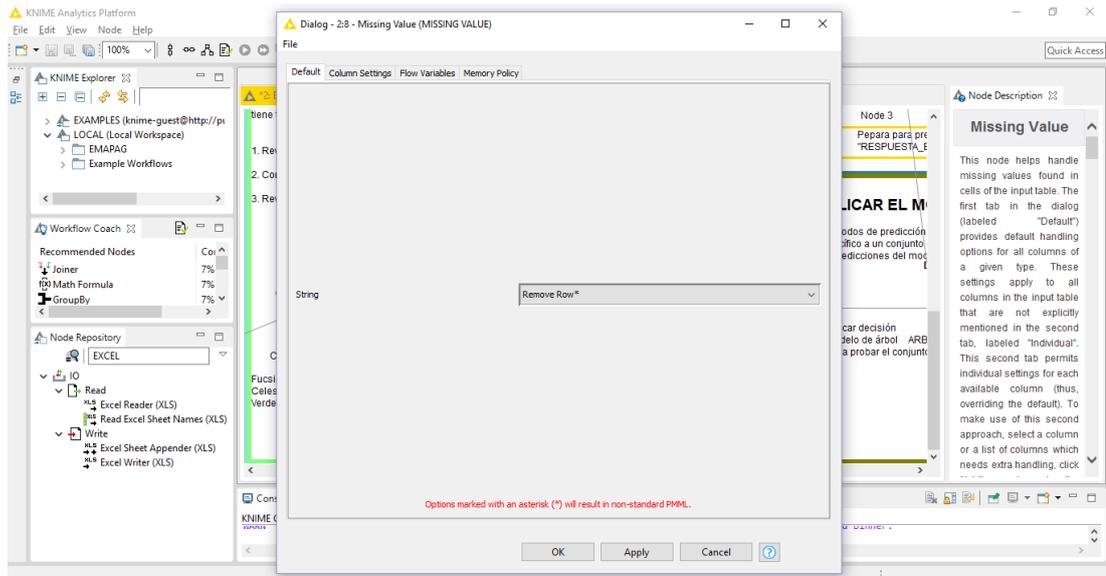
Figura 19. Entorno de Trabajo KNIME. *Color Manager (nodo)*

VALOR FALTANTE



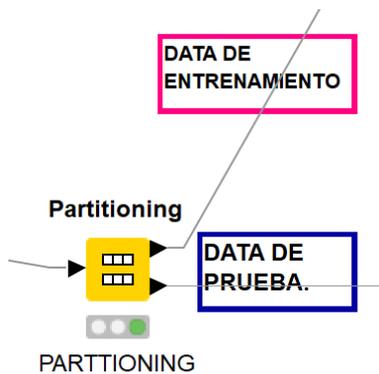
Este nodo ayuda a manejar los valores perdidos que se encuentran en las celdas de la tabla de entrada. Las opciones de manejo predeterminadas para las filas y columnas de la tabla.

En este caso se solicita se eliminen filas que no tengas las variables completas.



*Figura 20. Entorno de Trabajo KNIME. Missing Value (nodo); String = Remove Row**

PARTICIONAMIENTO



La tabla de entrada se divide en dos particiones (por filas), su objetivo es entrenar y probar datos. Se crean dos particiones separadas del conjunto de datos original: conjunto de entrenamiento (80%) y conjunto de pruebas (20%).

Las opciones disponibles en este nodo son:

1. Absoluto._ especifica el número absoluto de filas.
2. Relativo._ especifica el porcentaje de la cantidad de filas en la tabla de entrada.

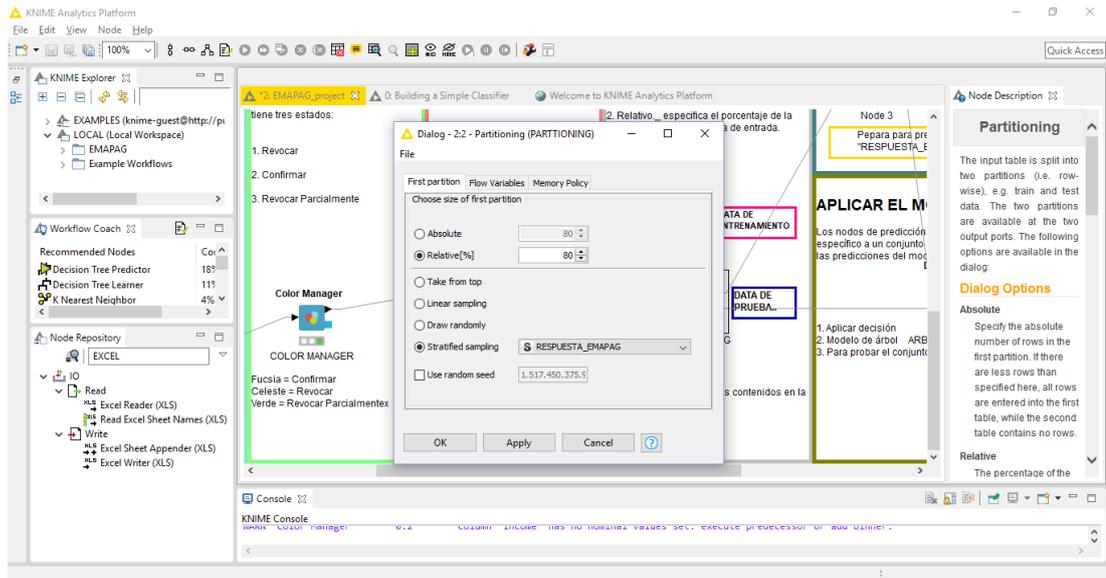
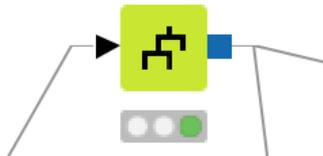


Figura 21. Entorno de Trabajo KNIME. Partitioning (nodo)
Relativo = 80% de los datos contenidos en la tabla.

ARBOLES DE DECISION (APRENDIZ)

Decision Tree Learner

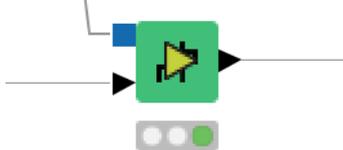


Este nodo construye un árbol de decisión. Otros nodos aprendices entrenan a otros modelos. La mayoría de los nodos de alumnos generan un modelo PMML (puerto de salida cuadrado azul).

Prepara para predecir clase "RESPUESTA_EMAPAGEP"

APLICAR EL MODELO

Decision Tree Predictor

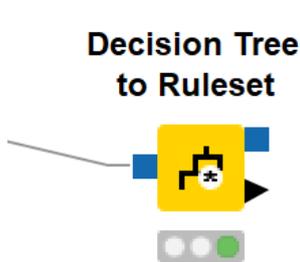


ARBOL DE DECISION

Los nodos de predicción aplican un modelo específico a un conjunto de datos y anexan las predicciones del modelo.

1. Aplicar decisión
2. Modelo de árbol
3. Para probar el conjunto

ARBOL DE DECISIONES PARA EL CONJUNTO DE REGLAS



Convierte (un único) modelo de árbol de decisión en el modelo de regla de PMML (variedad de técnicas de estadísticas) y también en una tabla que contiene las reglas en una forma textual. Las reglas resultantes son independientes entre sí, el orden de las reglas no se especifica, se puede cambiar.

Rules table - 27 - Decision Tree to Ruleset

| Row ID | Rule | Record ... | D Number... |
|--------|--|------------|-------------|
| Row1 | $\$INCUMPLIMENTOS = "N"$ AND $\$ESTADO_MEDIDORS = "Bueno"$ AND $\$REPARO_FUGAS = "N" \Rightarrow "CONFIRMAR"$ | 685 | 685 |
| Row2 | $\$INCUMPLIMENTOS = "S"$ AND $\$ESTADO_MEDIDORS = "Bueno"$ AND $\$REPARO_FUGAS = "N" \Rightarrow "RP"$ | 24 | 24 |
| Row3 | $\$ESTADO_MEDIDORS = "Malo"$ AND $\$REPARO_FUGAS = "N" \Rightarrow "REVOCAR"$ | 28 | 27 |
| Row4 | $\$ESTADO_MEDIDORS = "Bueno"$ AND $\$REPARO_FUGAS = "N" \Rightarrow "CONFIRMAR"$ | 2 | 2 |
| Row5 | $\$FUGA_RESP_IAS = "N"$ AND $\$INCUMPLIMENTOS = "N"$ AND $\$REPARO_FUGAS = "S" \Rightarrow "REVOCAR"$ | 107 | 107 |
| Row6 | $\$FUGAS = "N"$ AND $\$FUGA_RESP_IAS = "S"$ AND $\$INCUMPLIMENTOS = "N"$ AND $\$REPARO_FUGAS = "S" \Rightarrow "..."$ | 304 | 289 |
| Row7 | $\$CAMBIO_MEDIDORS = "N"$ AND $\$FUGAS = "S"$ AND $\$FUGA_RESP_IAS = "S"$ AND $\$INCUMPLIMENTOS = "N"$ A... | 8 | 8 |
| Row8 | $\$CAMBIO_MEDIDORS = "S"$ AND $\$FUGAS = "S"$ AND $\$FUGA_RESP_IAS = "S"$ AND $\$INCUMPLIMENTOS = "N"$ A... | 2 | 2 |
| Row9 | $\$FUGAS = "S"$ AND $\$FUGA_RESP_IAS = "S"$ AND $\$INCUMPLIMENTOS = "N"$ AND $\$REPARO_FUGAS = "S" \Rightarrow "..."$ | 4 | 4 |
| Row10 | $\$INCUMPLIMENTOS = "S"$ AND $\$REPARO_FUGAS = "S" \Rightarrow "REVOCAR"$ | 156 | 156 |

Decision Tree to Ruleset

Converts (a single) decision tree model to PMML RuleSet model and also to a table containing the rules in a textual form. The resulting rules are independent of each other, the order of rules is not specified, can be changed. Missing value strategies are ignored, it will always evaluate to missing value.

Dialog Options

Split rules to condition and outcome columns

When checked, two columns will be created for the rules, Condition and Outcome otherwise

Figura 22. Entorno de Trabajo KNIME. Generador de Reglas (nodo)

RESULTADOS DEL MODELO

Calcula una matriz de confusión entre los valores de clase reales y predichos y calcula las medidas de precisión relacionadas.

REGLAS O PATRONES DEL ALGORITMO DE ARBOL DE DECISION:

Mediante este algoritmo, se obtuvieron 10 reglas que contienen las variables independientes que tienen valor relevante en la investigación.

Detalle de las reglas:

Cuando se produce un incumplimiento por parte de la Concesionaria, el usuario tendrá a favor así sea que el mismo haya tenido la culpa. Cuando no existe incumplimiento, esto deberá ser demostrado por la Concesionaria mediante un expediente, todo esté funcionando de manera correcta el resultado no ser favorable ya que una de las variables analizadas es el medidor, el cual rige su medición en pruebas de consumo que por cualquier eventualidad Interagua realiza y sin indicios de alguna fuga.

Cuando el medidor se encuentra sin anomalías, y no evidencian fugas el resultado será negativo para el usuario. Caso contrario y sin fuga el caso es favorable por la indisposición del medidor y este a su vez es reemplazado.

Cuando existe fuga de responsabilidad de Interagua, y el usuario repara; y de ser una fuga interna de responsabilidad del usuario y reparada el caso será favorable.

Los parámetros relevantes dentro del proceso de Recurso de Apelación son: sector, estado del medidor, cambio de medidor, fuga, IEC, usuario repara fuga, fuga responsabilidad de Interagua e incumplimiento por parte de la Concesionaria.

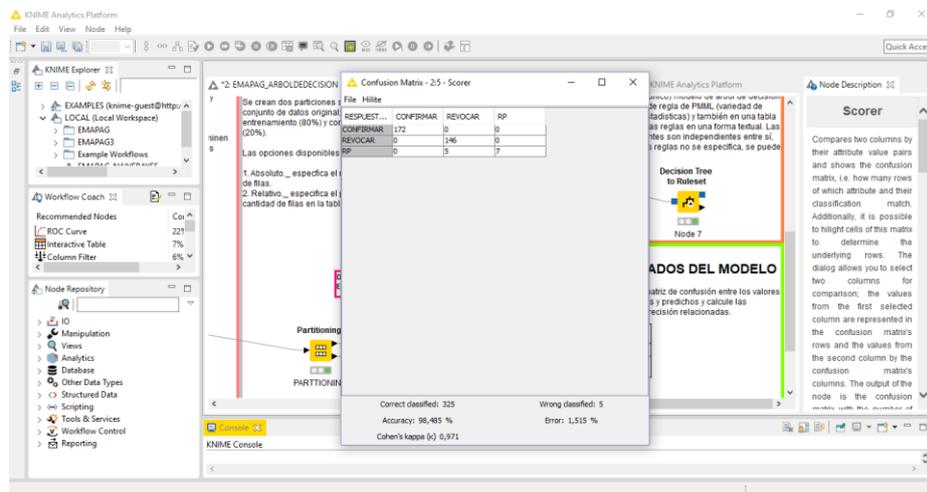


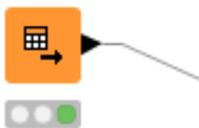
Figura 23. Entorno de Trabajo KNIME. Matriz de Confusión (nodo)

4.8. WEKA J48

El mencionado algoritmo es un implementación libre en java C4.5, que utiliza el concepto de entropía de la información para la selección de variables que mejor clasifiquen a la variable (clase) estudiada. Permite predecir la variable de destino de un nuevo registro de conjunto de datos.

4.8.1. APLICACIÓN DEL ALGORITMO EN LA HERRAMIENTA

CSV Reader



Lee archivos CSV. Al ejecutar el nodo, escaneará el archivo de entrada para determinar el número y los tipos de las columnas y generará una tabla con la estructura auto advertida.

Los datos están delimitados por “punto y coma ;” para la lectura de los mismos.

J48 (3.7)



Clase para generar un árbol de decisión C4.5 podado o sin podar (basado en WEKA 3.7). Todos los cuadros de diálogo WEKA tienen un panel donde puede especificar parámetros específicos del clasificador.

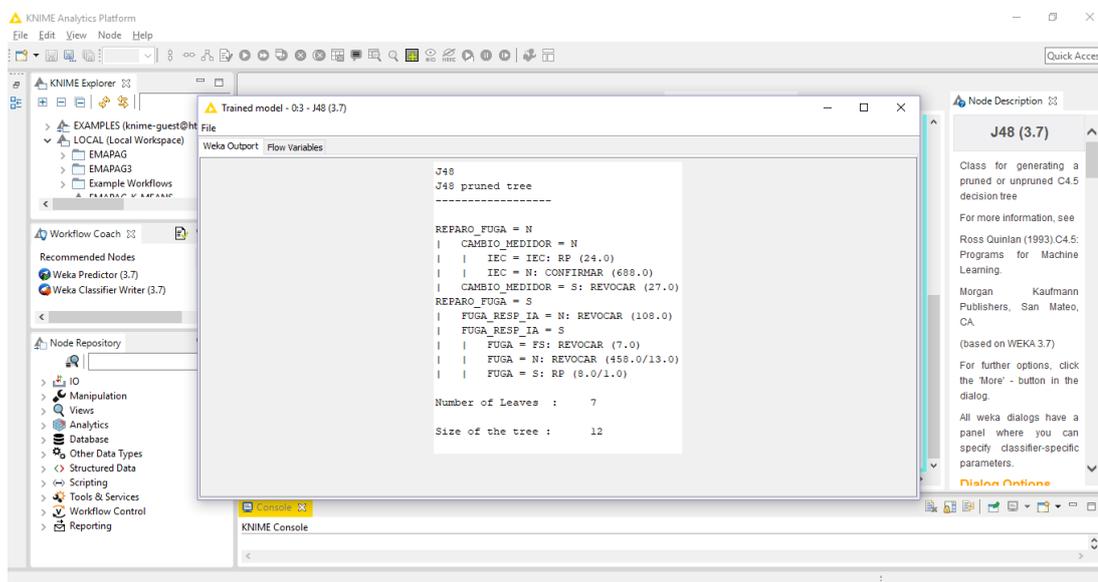
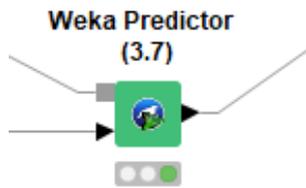


Figura 24. Entorno de Trabajo KNIME. Weka J48 Output (nodo)



El Predictor Weka toma un modelo generado en un nodo weka y clasifica los datos de prueba en el puerto de entrada.

RESULTADOS DEL MODELO

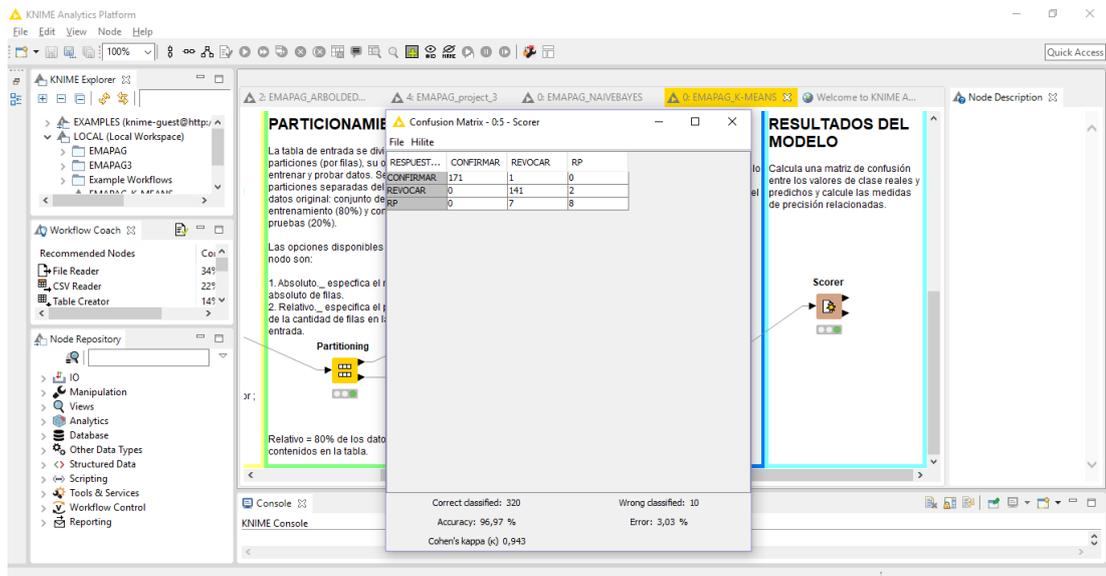


Figura 25. Entorno de Trabajo KNIME. Weka Matriz de Confusión (nodo)

REGLAS O PATRONES DEL ALGORITMO WEKA J48:

Mediante este algoritmo se obtuvo un árbol, elaborado bajo reglas con el aporte de las variables independientes del modelo.

Detalle:

Cuando el usuario no repara la fuga dentro del plazo establecido por reglamento de 28 días a partir de la detección de la fuga; contando con el medidor de consumo en buen estado sin que registre incremento de consumo, el resultado será negativo para el cliente. Caso contrario de producirse un cambio de medidor como consecuencia habrá incremento de consumo la resolución será favorable para el usuario.

Cuando usuario repara fuga, dentro del plazo autorizado por reglamento y no existe fuga de responsabilidad de la concesionaria, tendrá un resultado favorable para el usuario. También lo será cuando la concesionaria es responsable de fuga.

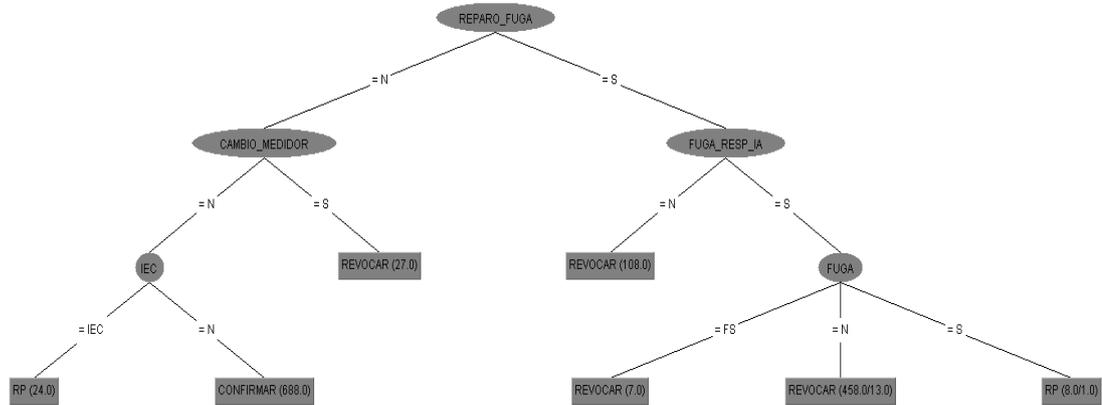


Figura 26. Entorno de Trabajo KNIME. Weka Resultados

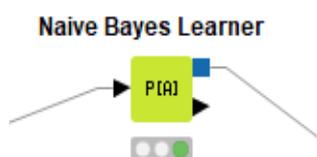
4.9. NAIVE BAYES

Método basado en la teoría de la probabilidad, usa frecuencias para calcular predicciones sobre nuevos casos. Es una técnicas tanto predictiva como descriptiva. A pesar de ser simple, ha sido desarrollada con éxito, produciendo buenos resultados en sus aplicaciones.

4.9.1. APLICACIÓN DEL ALGORITMO EN LA HERRAMIENTA

Este algoritmo para leer la data utiliza el mismo nodo del algoritmo WEKA “CSV Reader”.

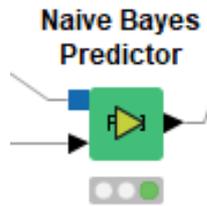
NAVIE BAYES (APRENDIZ)



El nodo crea un modelo bayesiano a partir de los datos de entrenamiento proporcionados. Calcula el número de filas por valor de atributo por clase

para atributos nominales y la distribución gaussiana para atributos numéricos.

APLICAR EL MODELO



Predice la clase por fila basada en el modelo aprendido. La probabilidad de clase es el producto de la probabilidad por atributo y la probabilidad del atributo de clase en sí.

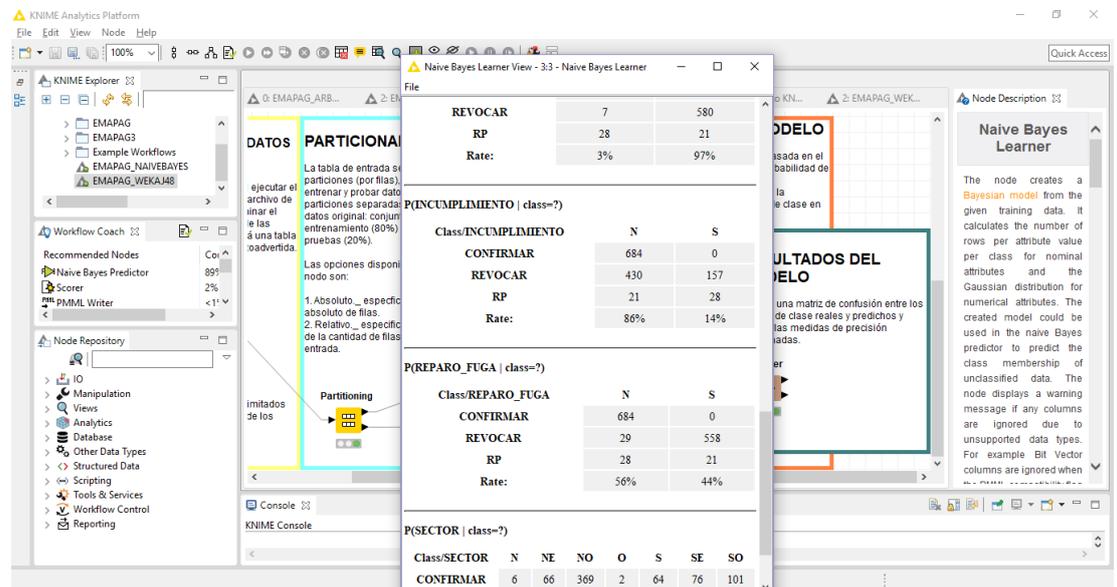


Figura 27. Entorno de Trabajo KNIME. Naive Bayes Resultados

REGLAS O PATRONES DEL ALGORITMO NAIVE BAYES:

Mediante este algoritmo se obtuvieron varios cuadros analíticos de las variables independientes del modelo.

Detalle:

| P(CAMBIO_MEDIDOR class=?) | | N | S |
|-----------------------------|--|-----|----|
| Class/CAMBIO_MEDIDOR | | | |
| CONFIRMAR | | 683 | 1 |
| REVOCAR | | 555 | 32 |
| RP | | 49 | 0 |
| Rate: | | 98% | 2% |

| P(FUGA class=?) | | FS | N | S |
|-------------------|--|----|-----|-----|
| Class/FUGA | | | | |
| CONFIRMAR | | 0 | 517 | 167 |
| REVOCAR | | 8 | 471 | 108 |
| RP | | 0 | 41 | 8 |
| Rate: | | 1% | 78% | 21% |

| P(FUGA_RESP_IA class=?) | | N | S |
|---------------------------|--|-----|-----|
| Class/FUGA_RESP_IA | | | |
| CONFIRMAR | | 684 | 0 |
| REVOCAR | | 131 | 456 |
| RP | | 0 | 49 |
| Rate: | | 62% | 38% |

| P(IEC class=?) | | IEC | N |
|------------------|--|-----|-----|
| Class/IEC | | | |
| CONFIRMAR | | 0 | 684 |
| REVOCAR | | 7 | 580 |
| RP | | 28 | 21 |
| Rate: | | 3% | 97% |

| P(INCUMPLIMIENTO class=?) | | N | S |
|-----------------------------|--|-----|-----|
| Class/INCUMPLIMIENTO | | | |
| CONFIRMAR | | 684 | 0 |
| REVOCAR | | 430 | 157 |
| RP | | 21 | 28 |
| Rate: | | 86% | 14% |

| P(REPARO_FUGA class=?) | | N | S |
|--------------------------|--|-----|-----|
| Class/REPARO_FUGA | | | |
| CONFIRMAR | | 684 | 0 |
| REVOCAR | | 29 | 558 |
| RP | | 28 | 21 |
| Rate: | | 56% | 44% |

| P(SECTOR class=?) | | N | NE | NO | O | S | SE | SO |
|---------------------|--|----|----|-----|----|-----|-----|-----|
| Class/SECTOR | | | | | | | | |
| CONFIRMAR | | 6 | 66 | 369 | 2 | 64 | 76 | 101 |
| REVOCAR | | 6 | 53 | 300 | 2 | 79 | 61 | 86 |
| RP | | 0 | 4 | 36 | 0 | 3 | 2 | 4 |
| Rate: | | 1% | 9% | 53% | 0% | 11% | 11% | 14% |

Figura 28. Entorno de Trabajo KNIME. Naive Bayes Resultados del análisis a cada variable independiente del modelo de negocio.

Se han obtenidos los siguientes resultados:

1. Los cambios de medidor no son frecuentes, ya que su reemplazo se da porque no pasa la prueba de exactitud, por manipulación y vida útil (8 a 10 años).
2. Que el mayor porcentaje en fugas es de responsabilidad de los clientes, es decir, fugas internas.

3. El incumplimiento al reglamento por parte de la Concesionaria es una variable importante para la toma de decisiones del analista. Esto se mide de acuerdo al expediente que envía Interagua para tener más conocimiento respecto al caso de estudio.
4. El usuario no es un agente responsable según la estadísticas en los casos que su predio tenga fuga.
5. los sectores que presentan reclamos de tercera instancia por AFA “alta facturación”, están en la estadísticas presentes en este orden: NO, SO, S, SE, NE, N , O.

4.10. GENERACION DE RESULTADOS PARA EL MODELO PREDICTIVO

Tabla 7

Resultados respecto a los algoritmos utilizados.

| Detalle/Algoritmo | Algoritmo |
|--|-------------------|
| Registros clasificados = 325 (80% data) Registros inválidos = 5 Éxito= 98,485% Error=1,515% | ARBOL DE DECISION |
| Registros clasificados = 321 (80% data) Registros inválidos = 9 Éxito= 97,273% Error=2,727% | NAIVE BAYES |
| Registros clasificados = 320 (80% data) Registros inválidos = 10 Éxito= 96,97% Error=3,03% | WEKA j48 |

Nota: Los resultados de las pruebas realizadas a los datos de la EMAPAGEP, respecto a las resoluciones de los reclamos de los usuarios que utilizan el servicio de agua potable de la Ciudad de Guayaquil.

Los patrones en esta investigación están en función de las diferentes soluciones que pueden acogerse los diferentes reclamos de los usuarios de la Ciudad de Guayaquil.

Como se conoce los estados son:

1. Confirmar = Negado
2. Revocar = Aprobado
3. Revocar Parcialmente = Aprobado parcialmente

El modelo predictivo o patrones:

CONFIRMAR = Todas las variables "OK" → *Reclamo no procede*

= Usuario repare sus fugas internas de acuerdo al plazo establecido en el reglamento.

REVOCAR = Medidor en mal estado + Usuario repara fuga dentro del plazo establecido en el reglamento.

= Usuario repara fuga interna.

= Cambio del medidor + Reparación de fuga (ambas entidades)

= Cuando existe algún incumplimiento al reglamento por parte de la Concesionaria.

REVOCAR PARCIALMENTE = Incumplimiento de la Concesionaria +
Reparación de fuga de ambas partes.

CONCLUSIONES Y RECOMENDACIONES

CONCLUSIONES:

El proceso de Tercera Instancia-Recurso de Apelación, en su etapa de análisis la mayoría de los casos son reclamos que no proceden y esto da como consecuencia que la resolución no sea favorable para el usuario, es decir, negado el reclamo. E2, utiliza su recurso más importante para dar respuesta a usuarios, el tiempo, el cual puede ser utilizado para reclamos especiales donde se observen otros parámetros de gran relevancia dentro del procedimiento, no para invertirlo en casos que no procedan y cumplan un mismo patrón de comportamiento.

Aplicando algoritmos de transformación se obtuvo la variable de clasificación y las variables que están alrededor de la misma, siendo parte del diccionario de la vista minable e influencia relevante para el proceso de los diferentes tipos de reclamos que un usuario puede ingresar a la entidad, fue ejecutada en la herramienta KNIME, opera en modalidad open source, contiene varias técnicas y algoritmos de KDD donde los resultados fueron el comportamiento de los casos según los tres estados que maneja la entidad para emitir resoluciones a usuarios. Se ejecutó la despersonalización de la información, es decir, no se utilizó información de fácil asociamiento a los clientes de la EMAPAG, se eliminó o sustituyó lo antes mencionado.

Los resultados obtenidos han sido evaluados por los algoritmos árboles de decisión (éxito=98,485%, error=1,515%), naive bayes (éxito=97,273%, error=2,727%) y weka (éxito=96,97%, error=3,03%) dando como resultado que mediante el uso de patrones, reglas, árboles de decisión se obtuvo un modelo.

De la predicción obtenida de las técnicas se generaron siete reglas que giran en torno a la variable de clasificación y las variables del diccionario de la vista minable. Por lo tanto se cuenta con tres soluciones a los distintos problemas (Confirmar, Revocar y Revocar parcialmente) y en general se observaron que entre las reglas, árboles de decisión y cuadros estadísticos siete tienen en común el conocimiento obtenido por los datos del modelo de negocio, es decir, el modelo quedó de la siguiente forma: Todas

las variables “OK” (Reclamo no procede), Usuario repare sus fugas internas de acuerdo al plazo establecido en el reglamento dando como resultado una resolución negada a usuario. Medidor en mal estado y Usuario repara fuga dentro del plazo establecido en el reglamento. Usuario repara fuga interna. Cambio del medidor y Reparación de fuga (ambas entidades), Cuando existe algún incumplimiento al reglamento por parte de la Concesionaria dando como resultado una resolución aprobada a usuario. Incumplimiento de la Concesionaria, Reparación de fuga de ambas partes dando como resultado una resolución parcialmente aprobada a usuario.

RECOMENDACIONES:

En la investigación se ha evidenciado que los reclamos que ingresan con mayor frecuencia a la EMAPAG EP dan como resultado “Confirmar = Negado”, es decir, el tiempo invertido en casos no procedentes para la jefatura lleva al recurso humano, la analista perder el tiempo, factor relevante dentro del procedimiento de “Recurso de Apelación”, por lo tanto sería importante que a nivel administrativo se sea más exigente al momento de estandarizar los reclamos que procedan para análisis; utilizando el primer filtro de reclamo.

Realizar publicidad informativa que imparta conocimiento a la sociedad que utiliza el servicio de Agua Potable en la Ciudad de Guayaquil, con la explicación de los reclamos no procedentes y de las variables relevantes que entrarían a análisis, de esta manera se crearía conocimiento a partir de los resultados del modelo.

Implementación de un Front-end del caso de estudio anteriormente mencionado para la efectividad de decisión cuando el usuario presente su queja en la entidad, esto sería de gran ayuda para los casos negados, que no proceden y el recurso tiempo sería invertido por parte de la analista de reclamo en casos que ameriten realmente el procedimiento analítico.

REFERENCIAS BIBLIOGRAFICAS

Aisemberg, L. (2013, enero 15). Qué es el análisis predictivo. Recuperado 18 de noviembre de 2017, a partir de <http://www.evaluandosoftware.com/que-es-el-analisis-predictivo-2/>

Alvarado Pérez, J. C., Caicedo Zambrano, S. J., Hernández Arteaga, I., Hidalgo Troya, A., & Timarán Pereira, S. R. (2016). *Descubrimiento de patrones de desempeño académico* (Universidad Cooperativa de Colombia). Bogota, Colombia: Universidad Cooperativa de Colombia. Recuperado a partir de <http://ediciones.ucc.edu.co/index.php/ucc/catalog/book/36>

Aranda Robles, Y., & Sotolongo, A. R. (2013). Integración de los algoritmos de minería de datos 1R, PRISM e ID3 a PostgreSQL. *JISTEM - Journal of Information Systems and Technology Management*, 10(2), 389-406. <https://doi.org/10.4301/S1807-17752013000200012>

Big data, el siguiente paso en la analítica de datos. (s. f.). Recuperado 18 de noviembre de 2017, a partir de <http://searchdatacenter.techtarget.com/es/cronica/Big-data-el-siguiente-paso-en-la-analitica-de-datos>

diegocond. (2017, mayo 22). Vista Minable. Definición. Recuperado 26 de noviembre de 2017, a partir de <https://datosmineriainformacion.wordpress.com/2017/05/22/vista-minable-definicion/>

El descubrimiento de Patrones: clave para la gestión de clientes. (2016, septiembre 1).

Recuperado 17 de noviembre de 2017, a partir de <https://www.questionpro.com/blog/es/patrones-para-la-gestion-de-clientes/>

Garcia, A. (2017). Minería de datos: Modelos predictivos con RapidMiner [Minería de datos: Modelos predictivos con RapidMiner]. Recuperado 19 de noviembre de 2017, a partir de <http://www.tradingsys.org/mineria-de-datos-modelos-predictivos-con-rapidminer>

Gil, Q., Ledis, N., & Valencia, C. A. (2012). KDD application process in the context of bibliomining: Elogim case. *Revista Interamericana de Bibliotecología*, 35(1), 97-108.

Los desafíos del marketing en la era del big data. (2016, enero 1), 30.

Merodio, J. (2014, octubre 8). Los modelos predictivos en el nuevo marketing. Recuperado 18 de noviembre de 2017, a partir de <https://www.juanmerodio.com/2014/modelos-predictivos-nuevo-marketing/>

Metodología CRISP-DM - Parte 1. (2016, mayo 11). Recuperado 26 de noviembre de 2017, a partir de <http://smartbasegroup.com/metodologia-crisp-dm-parte-i/>

Modelo de madurez en administracion de la informacion. (s. f.). Recuperado 18 de noviembre de 2017, a partir de <http://cdn.ttgtmedia.com/rms/madurezenIM2.jpg>

Objetivos de la Calidad | Sitio Web de EMAPAG-EP Empresa Municipal de Agua Potable y Alcantarillado de Guayaquil. (s. f.). Recuperado 14 de enero de 2018, a partir de <http://www.emapag-ep.gob.ec/emapag/objetivos-de-la-calidad/>

- Pacheco Aguila, Y., & Fernández Segredo, Y. (2015). Aplicación de técnicas de descubrimiento de conocimientos en el proceso de caracterización estudiantil. *Ciencias de la Información*, 46(3). Recuperado a partir de <http://www.redalyc.org/resumen.oa?id=181443340004>
- Rodríguez León, C., Lorenzo, G., & Matilde, M. (2016). ADECUACIÓN A METODOLOGÍA DE MINERÍA DE DATOS PARA APLICAR A PROBLEMAS NO SUPERVISADOS TIPO ATRIBUTO-VALOR. *Revista Universidad y Sociedad*, 8(4), 43-53.
- Roman Carrillo, J., & Virseda Benito, F. (s. f.). Minería de datos y aplicaciones, 8.
- Tercera Instancia – EMAPAG | Sitio Web de EMAPAG-EP Empresa Municipal de Agua Potable y Alcantarillado de Guayaquil. (2017, junio 12). Recuperado 18 de noviembre de 2017, a partir de <http://www.emapag-ep.gob.ec/emapag/tercera-instancia-emapag/>
- TI. (2017). *Empresa Municipal de Agua Potable y Alcantarillado de Guayaquil, EP*. Empresa Municipal de Agua Potable y Alcantarillado de Guayaquil, EP. Recuperado a partir de <http://www.emapag-ep.gob.ec/emapag/>
- Troche Clavijo, A. (2014). APLICACIÓN DE LA MINERÍA DE DATOS SOBRE BASES DE DATOS TRANSACCIONALES. *Fides et Ratio - Revista de Difusión cultural y científica de la Universidad La Salle en Bolivia*, 7(7), 58-66.
- Universidad Nacional Autónoma de México. (s. f.). Reconocimiento de patrones. Recuperado 19 de noviembre de 2017, a partir de <http://redyseguridad.fi-p.unam.mx/proyectos/biometria/basesteoricas/reconocimiento.html>

Veolia. (2017). *Interagua Cía. Ltda.* Interagua Cía. Ltda. Recuperado a partir de https://www.interagua.com.ec/user/login?destination=emapag/datos-referenciales&autologout_timeout=1

ANEXOS

Anexo 1

Reglamento de la Estructura Tarifaria para la Prestación de los Servicios de Agua Potable y Alcantarillado para el Cantón Guayaquil para el Segundo Quinquenio de la Concesión.



Presidencia
de la República
del Ecuador



Plan Nacional
de Ciencia, Tecnología,
Innovación y Saberes



SENESCYT

Secretaría Nacional de Educación Superior,
Ciencia, Tecnología e Innovación

DECLARACIÓN Y AUTORIZACIÓN

Yo, **Vera Letamendi María Mercedes**, con C.C: # **093108634-2** autor/a del trabajo de titulación: **Minería de datos aplicada a la detección de patrones de reclamos del servicio de agua potable en la ciudad de Guayaquil** previo a la obtención del título de **Ingeniera en Sistemas Computacionales** en la Universidad Católica de Santiago de Guayaquil.

1.- Declaro tener pleno conocimiento de la obligación que tienen las instituciones de educación superior, de conformidad con el Artículo 144 de la Ley Orgánica de Educación Superior, de entregar a la SENESCYT en formato digital una copia del referido trabajo de titulación para que sea integrado al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor.

2.- Autorizo a la SENESCYT a tener una copia del referido trabajo de titulación, con el propósito de generar un repositorio que democratice la información, respetando las políticas de propiedad intelectual vigentes.

Guayaquil, 08 de marzo de 2018

Vera Letamendi María Mercedes

Nombre: **Vera Letamendi, María Mercedes**

C.C: **093108634-2**

REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA

FICHA DE REGISTRO DE TESIS/TRABAJO DE TITULACIÓN

| | | | |
|-----------------------------------|---|------------------------|----|
| TEMA Y SUBTEMA: | Minería de datos aplicada a la detección de patrones de reclamos del servicio de agua potable en la ciudad de Guayaquil | | |
| AUTOR(ES) | Vera Letamendi María Mercedes | | |
| REVISOR(ES)/TUTOR(ES) | Ing. Galo Enrique Cornejo Gómez, Mgs. | | |
| INSTITUCIÓN: | Universidad Católica de Santiago de Guayaquil | | |
| FACULTAD: | Facultad de Ingeniería | | |
| CARRERA: | Ingeniería en Sistemas Computacionales | | |
| TÍTULO OBTENIDO: | Ingeniera en Sistemas Computacionales | | |
| FECHA DE PUBLICACIÓN: | 08 de marzo de 2018 | No. DE PÁGINAS: | 84 |
| ÁREAS TEMÁTICAS: | Metodología de la Investigación, Inteligencia Artificial, Ingeniería | | |
| PALABRAS CLAVES/ KEYWORDS: | Minería de datos, metodología KDD, base de datos, patrones, resoluciones, proceso. | | |

El presente trabajo denominado “**Minería de datos aplicada a la detección de patrones de reclamos del servicio de agua potable en la ciudad de Guayaquil**” realiza un diagnóstico a la Jefatura de Atención al Usuario de la EMAPAG EP, respecto al proceso de Recurso de Apelación – Tercera Instancia con la finalidad de la aplicación de la metodología KDD (“Descubrimiento de Conocimiento en Base de Datos”) para descubrir un modelo predictivo válido, útil y entendible que especifiquen patrones de acuerdo a la información; utilizando la herramienta KNIME como plataforma tecnológica dentro de la investigación y la ejecución de los algoritmos de clasificación fundamentados en inteligencia artificial como: árbol de decisión, naive bayes y weka j48 que satisfacen las necesidades dependiendo de los objetivos del caso de estudio en base a los datos proporcionados por la entidad.

El problema a resolver va de la mano con el cumplimiento del objetivo principal de la entidad el servicio a la comunidad de la Ciudad de Guayaquil; supervisando a la operadora Interagua Cía. Ltda., en el desempeño de sus funciones como ente privado a los sectores que utilizan el servicio de Agua Potable y Alcantarillado de la ciudad, la EMAPAG EP, busca la satisfacción de los usuarios y a través de este objetivo tener un crecimiento relevante en el sector público. Dentro del proceso de Atención al Usuario, una de las variables más relevantes es el tiempo que la analista de reclamos emplea para dar solución a los diferentes casos presentados por los usuarios. Cabe recalcar que estos datos son la materia prima que se utilizan en la metodología KDD.



El estudio de esta investigación es para determinar patrones respecto a los estados de las resoluciones que se maneja en la jefatura como: Revocar, Confirmar y Revocar Parcialmente, resultados que dependerán de las variables relevantes del proceso; realizándose un análisis de los datos.

| | | |
|--|---|---|
| ADJUNTO PDF: | <input checked="" type="checkbox"/> SI | <input type="checkbox"/> NO |
| CONTACTO CON AUTOR/ES: | Teléfono: +593-4-2174888 | E-mail: ma.mercedes_vl@hotmail.com |
| CONTACTO CON LA INSTITUCIÓN (COORDINADOR DEL PROCESO UTE):: | Nombre: Yanza Montalván, Ángela Olivia, Ing., Mgs. | |
| | Teléfono: +(593) 098-3035702 | |
| | E-mail: angela.yanza@cu.ucsg.edu.ec | |
| SECCIÓN PARA USO DE BIBLIOTECA | | |
| Nº. DE REGISTRO (en base a datos): | | |
| Nº. DE CLASIFICACIÓN: | | |
| DIRECCIÓN URL (tesis en la web): | | |

Guayaquil, 08 de marzo de 2018

Ma. Mercedes Vera Letamendi

Nombre: **Vera Letamendi, María Mercedes**

C.C: **093108634-2**