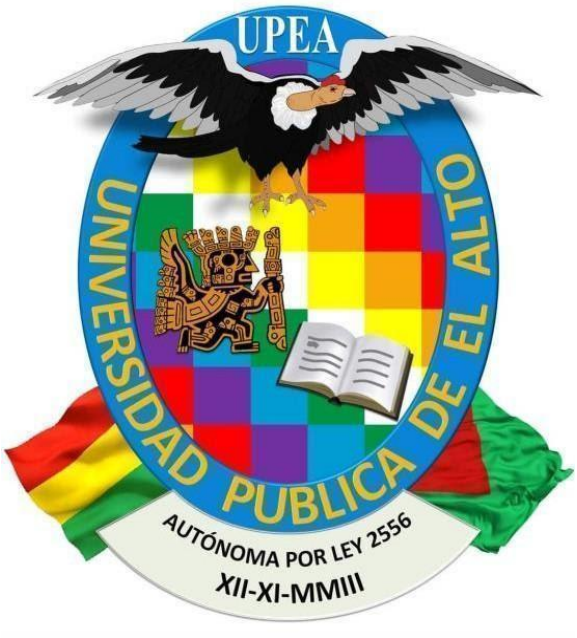
UNIVERSIDAD PÚBLICA DE EL ALTO

CARRERA INGENIERÍA DE SISTEMAS



TESIS DE GRADO

"MODELO DE PROYECCIÓN DE ÍNDICE DE CONTAMINACIÓN DE BASURA EN LA CIUDAD DE EL ALTO APLICANDO LA MINERÍA DE DATOS"

Para Optar al Título de Licenciatura en Ingeniería de Sistemas
MENCIÓN: INFORMÁTICA Y COMUNICACIONES

Postulante: Maribel Huchani Silvestre
Tutor Metodológico: M. Sc. Ing. Enrique Flores Baltazar
Tutor Revisor: Ing. Ramiro Kantuta Limachi
Tutor Especialista: Ing. William Roque Roque

EL ALTO - BOLIVIA

2022

Dedicatoria

A Dios por siempre estar a mi lado guiándome en mi camino por brindarme salud, bendición y fortaleza.

A mi padre Carlos Huchani por su apoyo incondicional, por su comprensión en todo el transcurso de mi preparación profesional.

A mis Sobrinos Eydan y Adair, por la motivación y alegrías que día a día me dan.

Maribel Huchani Silvestre

Agradecimiento

Agradezco a Dios por guiar mi camino, brindarme Salud, y por la bendición que día a día me regala.

A mi Padre Carlos Huchani, por todo su apoyo, cariño, y comprensión que día a día me entrega, por el sacrificio en brindarme educación y velar mi salud.

A mi tutor Ing. William Roque por compartir sus conocimientos y el apoyo que fueron importantes para la culminación de este trabajo..

A mi tutor Ing. Ramiro Kantuta por su apoyo en mi camino profesional y colaboración en el cumplimiento de mi meta.

A mi tutor Ing. Enrique Baltazar, por la guía y orientación para culminar el presente trabajo.

ÍNDICE GENERAL

CAPITULO I página

CAPITULO II

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

RESUMEN

El tema de la contaminación de la basura en la ciudad de El Alto es una preocupación constante ante los problemas de salud que pone en riesgo a la población en general debido a que carece de un eficiente tratamiento de los Residuos Sólidos.

En respuesta a esta situación en el presente trabajo de investigación tiene como objetivo principal “Desarrollar un Modelo de Proyección de índice de contaminación de la basura para la Ciudad de El Alto basado en las técnicas de la Minería de Datos, esto para la toma de decisiones en el resguardo del medio ambiente y la sociedad en general de la Ciudad de El Alto”, para lograr el objetivo se realizó el análisis de la aplicación de técnicas de minería de datos para identificar patrones de comportamiento con el fin de proyectar el índice de crecimiento. Los experimentos se realizaron con los registros de la Dirección Integral de Gestión de Residuos de la Ciudad de El Alto y del Instituto Nacional de Estadística.

Para la implementación se utilizó la metodología CRISP-DM que estructura el proceso de minería de datos y la metodología KDD (Knowledge Discovery in Databases) que interactúan entre ellas de forma iterativa, se aplicaron algoritmos de clasificación y algoritmos de series de tiempo.

El presente trabajo contribuye a la toma de decisiones a las autoridades encargadas del medio ambiente para optimizar el tratamiento de los residuos en bien de la población alteña.

Palabras clave: CRISP-DM, Residuos, Contaminación, Minería de Datos, Basura.

ABSTRACT

The issue of garbage contamination in the city of El Alto is a constant concern due to the health problems that put the general population at risk because of the lack of efficient treatment of solid waste.

In response to this situation, the main objective of this research work is "To develop a model for the projection of the garbage contamination index for the city of El Alto based on data mining techniques for decision making in order to protect the environment and society in general in the city of El Alto". To achieve this objective, the analysis of the application of data mining techniques to identify patterns of behavior in order to project the growth index was carried out. The experiments were carried out with the records of the Integral Waste Management Directorate of the City of El Alto and the National Institute of Statistics.

The CRISP-DM methodology that structures the data mining process and the KDD (Knowledge Discovery in Databases) methodology that interacts between them in an iterative way were used for the implementation, classification algorithms and time series algorithms were applied.

The present work contributes to the decision-making process of the authorities in charge of the environment in order to optimize the waste treatment for the benefit of the population of El Alto.

Key words: CRISP-DM, Waste, Pollution, Data Mining, Garbage.

GLOSARIO DE ABREVIACIONES

KDD: Knowledge Discovery in Databases (Descubrimiento de Conocimiento en bases de datos)

CRISP-DM: Cross Industry Standard Process for Data Mining (Modelo de Proceso Estándar para Minería de Datos)

WEKA: Waikato Environment Knowledge Analysis (Entorno de Análisis De Conocimiento de la Universidad de Waikato)

ANN: Red Neuronal Artificial.

INE: Instituto Nacional de Estadistica.

IBM: International Business Machines Corporation.

CHAID: Chi Squared Automatic Interaction Detector.

DM: Minería de datos.

DW: Data WereHouse

DGIR: Dirección general de Integración de Residuos.

OpenUP: Proceso Unificado Abierto.

WORA: Write Once and Run Anywhere.

SQuaRE: System and Software Quality Requirements and Evaluation.

CAPÍTULO I

Marco Preliminar

1 MARCO PRELIMINAR

El presente capitulo tiene como principal objetivo dar a conocer el presente trabajo de investigación “Modelo de proyección de índice de contaminación de la basura de la Ciudad de El Alto” dando a conocer la problemática, el objetivo, las herramientas que se aplicaran en el transcurso del desarrollo.

1.1 INTRODUCCIÓN

Hoy en día la Minería de Datos es utilizada en diferentes campos de la ciencia, ya que está automatiza el proceso de encontrar información predecible en grandes bases de datos. La Minería de Datos hoy en día se ha convertido en una práctica esencial para mantener ventaja competitiva en todas las fases del ciclo de vida del cliente y más aún en la toma de decisiones.

Los residuos se han manejado desde una perspectiva lineal; producir, consumir y desechar, esto ha originado diversos problemas ambientales afectando la salud y los servicios de recolección no son suficientes. La basura es considerada uno de los mayores problemas ambientales de nuestra sociedad. La población y el consumo va en constante crecimiento, y por ende la basura; pero el espacio no y además su tratamiento no es el adecuado.

En la Ciudad de El Alto existe altos niveles de contaminación según datos del INE (2021), la urbe alteña generó 257894 toneladas de basura, y un promedio de recojo de 600 toneladas por día, causando el deterioro del medio ambiente.

Con el Modelo de Proyección se pretende realizar predicciones del índice de crecimiento de la contaminación de la basura mediante el uso de la Minería de Datos con enfoque predictivo, el método científico, la metodología CRIS-DM y KDD, como herramientas para su desarrollo se usará el lenguaje de programación Java y la metodología Open Up.

La presente investigación pretende desarrollar un modelo de Proyección de la contaminación de la basura de la Ciudad de El Alto para la toma de decisiones de las autoridades encargadas del área de Medio Ambiente de la Ciudad de El Alto con respecto a la acumulación de la basura.

1.2 ANTECEDENTES

1.2.1 local

- ✓ Hidalgo, (2020), presenta su tema: “MODELO DE PREDICCIÓN BASADO EN MINERÍA DE DATOS SOBRE ÍNDICES DE DESERCIÓN DE ALUMNOS” en la Universidad Pública de El Alto.

  Tiene como objetivo desarrollar un modelo de predicción en base a la Minería de Datos y los factores socioeconómicos, sobre el índice de deserción de alumnos de la Universidad. Como resultado se obtiene que el modelo de Minería de Datos muestra la aplicación de algoritmos óptimos en la predicción, teniendo en cuenta la identificación de los factores de deserción universitaria.

- ✓ Apaza, (2020), presenta el tema “MODELO DE PREDICCIÓN SOBRE EL ÍNDICE DE CRECIMIENTO DEL CÁNCER DE MAMA EN LAS MUJERES DE EDADES ENTRE 20 A 40 AÑOS DE LA CIUDAD DE LA PAZ, BASADO EN MINERÍA DE DATOS” en la Universidad Pública de El Alto.

  El presente trabajo de investigación tiene como finalidad realizar un Modelo de Predicción del índice de crecimiento del cáncer de mama en las mujeres de edades entre 20 a 40 años de la Ciudad de La Paz, basado en Minería de Datos, para la toma de decisiones por un intervalo de 5 años con el propósito de evitar futuras muertes.

1.2.2 Nacional

- ✓ Valenzuela, (2004), presenta su tema: "HERRAMIENTA DE APOYO PARA DIAGNÓSTICO PRECOZ DE CÁNCER DE MAMA BASADO EN TÉCNICAS DE MINERÍA DE DATOS” en la Universidad Mayor de San Simón.

Propone una herramienta que permita el uso de algoritmos de explotación de datos aplicados al procesamiento de imágenes médicas de mamografías digitales. También indica que dicha herramienta cuente con una interfaz gráfica adaptable e intuitiva para los diferentes usuarios.

- ✓ Hernández, (2018), presenta el tema: “PROYECCIÓN DE LOS EFECTOS DE DESECHOS TECNOLÓGICOS EN EL MEDIO AMBIENTE, CASO BOTADERO MUNICIPAL ZONA VILLA INGENIO CIUDAD DE EL ALTO" en la Universidad Mayor de San Andrés.

  Tiene por objetivo realizar un estudio que permitirá obtener, aproximaciones de la cantidad de desechos tecnológicos generados con relación a la cantidad de habitantes en cada municipio durante ciertos intervalos de tiempo, dichos datos y variables podrán ser manejados en un “prototipo de simulación" que proyecte a futuro cuáles serán las consecuencias y se puedan adoptar medidas para prevenir desastres ecológicos que afecten la vida en nuestras Ciudades y su medio ambiente.

1.2.3 Internacional

- ✓ López, (2013). presenta su tema: “MINERÍA DE DATOS COMO SOPORTE EN EL DIAGNÓSTICO Y TRATAMIENTO DEL CÁNCER DE MAMA.” En el Centro de Investigación Científica y de Educación Superior de Ensenada, México.

  El objetivo del presente trabajo de tesis es aplicar el proceso de Minería de Datos en repositorios relacionados con el cáncer de mama con la finalidad de descubrir conocimiento útil que apoye al proceso de diagnóstico y tratamiento del cáncer de mama en sus diferentes etapas. Utilizo la herramienta de Minería de Datos juntamente al proceso de KDD.

- ✓ Villalba, (2019), presenta su tema: “PREDICCIÓN DE LA CALIDAD DEL AIRE DE MADRID MEDIANTE MODELOS SUPERVISADOS” en la Universidad: Universidad Oberta de Catalunya.

El objetivo de este trabajo es utilizar diferentes algoritmos de Minería de Datos para crear un modelo que nos permita realizar una predicción de la calidad del aire de Madrid de manera precisa y con una mayor antelación, haciendo uso con diferentes técnicas y algoritmos de Minería de Datos (MLP, LSTM, CNN y SVM).

- ✓ Ortega, (2018), presenta el tema: "IMPACTO DE LA APLICACIÓN DE ALGORITMOS DE MINERÍA DE DATOS EN VARIABLES DE CONTAMINACIÓN DEL AIRE". En la Universidad Católica de Valparaíso.

  El proyecto desarrolla modelos agrícolas, entregando evidencias sobre su utilidad, para brindar información de vital importancia para los agricultores, ya que puede contribuir al desarrollo exitoso de esta actividad económica. Y de esta manera evitar la disminución de cultivos en Chile. Para el desarrollo del Modelo de Proyección se utilizó la metodología KDD, CRISP-DM, Minería de Datos y la herramienta WEKA.

1.3 PLANTEAMIENTO DEL PROBLEMA

Según el informe del Banco Mundial (2018), el mundo genera 2010 millones de toneladas de residuos sólidos municipales anualmente. Toda esta basura generada, transmite enfermedades, aumentando las infecciones respiratorias por causa de la quema.

El Banco Mundial (2012) afirma que mientras más desarrollado es un país, sus patrones de consumo aumentan la generación de residuos sólidos. Asimismo, existe una correlación positiva entre el nivel de ingreso per cápita y la generación de basura, es decir, a mayor ingreso, mayor generación de basura (Andersen et al., 2016).

Actualmente la gestión de la basura se centra principalmente en la eliminación de los mismos a través de basurales, rellenos sanitarios y en algunos casos, de incineradores. No se tiene en cuenta la necesidad de reducir el consumo de materias primas y de energía, y se plantean serios riesgos para el medio ambiente y la salud de las personas a medida que van pasando los días.

Según los datos de INE (2012), aproximadamente el 42% de los hogares en Bolivia, eliminan su basura mediante formas alternativas ya que no cuentan con servicios de recolección de la misma o basureros públicos. Las formas alternativas más utilizadas son el incendio (23%), botar a la calle o a algún terreno baldío (7%) y botar al río (7%). Sin embargo, a pesar de una cobertura del recojo de basura sea del 100%, el problema de la basura continuaría, ya que Bolivia apenas recicla el 4% de las 5400 toneladas de basura que genera al día.

En la Ciudad de El Alto, se registra el recojo de 600 toneladas de basura por día, esta cifra aumenta debido a la actividad comercial de la feria 16 de Julio. La secretaria municipal de Agua, Saneamiento, Gestión Ambiental y Residuos informo que el relleno sanitario de Villa ingenio con el que cuenta la Ciudad de El Alto ya cumplió su vida útil, su estado actual no abastece la cantidad de basura que se genera en la Ciudad de El Alto, y a pesar de su situación la empresa Trebol sigue con el descargo de la basura en el botadero (Ramos, 2018).

1.3.1 Problema Principal

La acumulación de basura es un problema cada vez más grave, que afecta a la salud pública y al medio ambiente de la Ciudad de El Alto, la enorme demanda de bienes de consumo aumenta a su vez la cantidad de residuos, ya que el principal botadero Villa Ingenio con el que se cuenta actualmente es un foco de infección, afectando la salud de la población, es necesario realizar un estudio para que las autoridades responsables tomen decisiones para reducir los impactos sociales y ambientales que causa la basura.

1.3.2 Problemas Secundarios

- ✓ La falta de aplicación de nuevas técnicas y diseño que puedan proyectar el índice de crecimiento de la contaminación de basura en la Ciudad de El Alto.
- ✓ Falta de información adecuada para la toma de decisiones de las autoridades competentes.
- ✓ Información sobre la contaminación de basura no representadas ni interpretadas.

1.3.3 Formulación del Problema

¿De qué manera ayudaría un Modelo de Proyección del índice de la contaminación de la basura en la Ciudad de El Alto, basado en Minería de Datos, para la toma de decisiones?

1.4 OBJETIVOS

1.4.1 Objetivo General

Desarrollar un Modelo de Proyección de índice de contaminación de la basura para la Ciudad de El Alto basado en las técnicas de la Minería de Datos, esto para la toma de decisiones en el resguardo del medio ambiente y la sociedad en general de la Ciudad de El Alto.

1.4.2 Objetivo Especifico

- ✓ Recolección y análisis de los datos de la contaminación de la basura de la Ciudad de El Alto.
- ✓ Analizar y seleccionar las técnicas y métodos de la Minería de Datos que sean útiles para el Modelo de Proyección.
- ✓ Aplicar la metodología CRIPS-DM y KDD para el manejo de la información
- ✓ Diseñar un Modelo de Proyección del Índice de crecimiento de la basura aplicando pruebas de Algoritmo de Minería de Datos seleccionados.
- ✓ Interpretar el resultado obtenido del Modelo de Proyección de la basura generada en la Ciudad de El Alto.

1.5 HIPÓTESIS

1.5.1 Hipótesis General

Aplicando las técnicas de la Minería de Datos y el área de la Ingeniería de Software, se desarrollará el Modelo de Proyección de índice de la contaminación de la basura en la Ciudad de El Alto, esto con una eficiencia al 95% para coadyuvar en la

toma de decisiones a las Autoridades competentes que tienen como visión la planificación como pilar fundamental en las políticas de salubridad de la urbe alteña.

1.5.2 Hipótesis Alterna

Aplicando las técnicas de la Minería de Datos y el área de la Ingeniería de Software, se desarrollará el Modelo de Proyección de índice de la contaminación de la basura en la Ciudad de El Alto, esto con una eficiencia al 65% para coadyuvar en la toma de decisiones a las Autoridades competentes que tienen como visión controlar les futuras predicciones como pilar fundamental en las políticas de salubridad de la urbe alteña.

1.5.3 Hipótesis Nula

Aplicando las técnicas de la Minería de Datos y el área de la ingeniería de software, se desarrollará el Modelo de Proyección que información con datos erróneos de la contaminación de la Ciudad de El Alto.

1.5.4 Identificación de variables

- Variable dependiente: Contaminación de la basura en la Ciudad de El Alto
- Variable independiente: Modelo de Proyección
- Variable interviniente: Minería de Datos

1.6 OPERACIONALIZACIÓN DE VARIABLES

Tabla 1.1

Operacionalización de Variables

VARIABLES	DEFINICIÓN	DIMENSIONES	INDICADORES	HERRAMIENTAS
Variable independiente: Contaminación de la basura en la Ciudad de El Alto	Es la que implica daños al suelo, agua, aire y la salud por la acumulación de residuos que ya no son necesarios. (Beltran Prieto, 2020)	Análisis de información Grado de contaminación	Datos históricos obtenidos para el análisis Relevamiento de datos	Reportes validados y analizados
Variable dependiente: Modelo de Proyección	Es un mecanismo que permite el análisis de datos estadísticos para poder predecir lo que va ocurrir en el futuro. (Arimetrics, 2020)	Técnicas de Minería de Datos Algoritmos	Prototipo desarrollado Referencias del grado de contaminación Cálculo de eficiencia más cercano al 100%	Minería de Datos Observación del comportamiento del algoritmo con la información
Variable Interviniente: Minería de Datos	Es el proceso de hallar anomalías, patrones y correlaciones en grandes cantidades de datos para predecir resultados. (SAS, 2021)	• Algoritmos • Métodos • Técnicas • procesos	• Proyección • Patrones • Características	• Proceso KDD • Metodología CRISP-DM

Nota: *Cuadro descriptivo de las variables identificadas.*

1.7 JUSTIFICACIÓN

1.7.1 Científica

En el presente trabajo se busca aprovechar los beneficios de la Minería de Datos que se aplica para el hallazgo de anomalías, patrones y correlaciones en conjuntos de datos para predecir resultados.

El Modelo de Proyección realizara un análisis en el comportamiento de la tasa de crecimiento de la contaminación de basura de la Ciudad de El Alto, con el fin de apoyar a la toma de decisiones previniendo los impactos medioambientales.

1.7.2 Técnica

Para el desarrollo del Modelo de Proyección, se utilizará las siguientes herramientas: Lenguaje de programación Dart, IDE Visual Studio Code, Y las siguientes metodologías: Método científico, Metodología CRISP-DM (Cross Industry Standard Process for Data Mining), Metodología de KDD (Knowiedge Discovery in Databases), Minería de Datos técnica predictiva, llegando a obtener un modelo de proyección de índice de la contaminación que proporcionara datos en diferentes intervalos de tiempo coadyuvando a las autoridades pertinentes al tema.

1.7.3 Económica

La presente investigación tendrá un costo bajo, ya que será desarrollado en software libre y se usará herramientas con precios módicos y convenientes para su implementación, además el modelo coadyuvará a una mejor toma de decisiones para prevenir problemas futuros, beneficiará a las autoridades pertinentes y a las instituciones responsables minimizando los recursos económicos, teniendo conocimiento de los resultados mostrados por el modelo, estos datos les servirán a contrarrestar el deterioro del medio ambiente y efectos en la salud pública.

1.7.4 Social

La investigación beneficiará a las autoridades e instituciones pertinentes a tener un mejor seguimiento y control de la contaminación de la basura. También beneficiará a toda la población de la Ciudad de El Alto a tener una mejor calidad de vida lo que conducirá a un impacto positivo en la sociedad con la toma decisiones y acciones correctas de las autoridades

1.8 METODOLOGÍA

1.8.1 Método Científico

El método de investigación adoptado es un método científico cuantitativo, porque es un proceso de investigación, que consiste en un conjunto de tecnologías, procedimientos y mecanismos, estas tecnologías, procedimientos y mecanismos siguen ciertos pasos y reglas para ayudar a resolver problemas de conocidos a desconocidos (Collado & Lucio, 2014).

Fases de la investigación científica cuantitativa:

- ✓ **Fase 1: Idea**

  Se plantea estudiar la cantidad de basura que genera el ser humano y su crecimiento.

- ✓ **Fase 2: Planteamiento del problema**

  Según informes del Ministerio de Medio Ambiente y Agua, el 2016 Bolivia genero un aproximado de 2 millones de toneladas de residuos sólidos en el año, el equivalente a 5400 toneladas al día. Y según datos del INE (2017), más del 70% provenían de las 9 capitales y El Alto.

  (Gonzales, 2019)

- ✓ **Fase 3: Revisión de la literatura y desarrollo del marco teórico**

  Se realizo una revisión literaria referente al tema para ampliar los conocimientos sobre la investigación en el Capítulo ll del presente trabajo.

- ✓ **Fase 4: Visualización del alcance del estudio**

  Se abordará un estudio con alcance exploratorio.

- ✓ **Fase 5: Elaboración de hipótesis y definición de variables**

  Se tiene como hipótesis principal:

  “Aplicando las técnicas de la Minería de Datos y el área de la Ingeniería de Software, se desarrollará el Modelo de Proyección de índice de la

contaminación de la basura en la Ciudad de El Alto, esto con una eficiencia al 95% para coadyuvar en la toma de decisiones”

Y como variables:

Variable dependiente: Modelo de Proyección

Variable independiente: Contaminación de la basura en la Ciudad de El Alto

- ✓ **Fase 6: Desarrollo del diseño de investigación**

  Se realiza un diseño de investigación no experimentales.

- ✓ **Fase 7: Definición y selección de la muestra**

  Se tomo como muestra los datos de la Ciudad de El Alto para realizar los estudios necesarios

- ✓ **Fase 8: Recolección de los datos**

  Se utilizo el tipo de datos secundarios (recolectados por otros investigadores), el Análisis de contenido cuantitativo y la Observación.

- ✓ **Fase 9: Análisis de los datos**

  Se efectuará la Minería de Datos y todos los procesos pertenecientes para analizar los datos recolectados

- ✓ **Fase 10: Elaboración del reporte de resultados**

  Se presentará el modelo de proyección con la interpretación de los datos y los resultados de la investigación.

1.8.2 Metodología CRISP-DM

IBM, (2021) menciona:

El método CRISP-DM (Cross-Industry Standard Process for Data Mining) es un modelo de procesos jerárquicos que consta de un conjunto de tareas descritas en 4 niveles de abstracción, de lo general a lo específico. El modelo CRISP-DM cubre las diversas fases de un proyecto, sus respectivas tareas y las relaciones entre esas tareas. En este nivel de descripción, no es posible identificar todas las relaciones; las

relaciones pueden existir entre cualquier tarea, según los objetivos, el contexto y el interés del usuario en los datos.

1.8.3 Metodología KDD

Maimon y Rokech (2010), indican que:

La Minería de Datos en realidad es el núcleo de todo un proceso llamado Descubrimiento de Conocimiento en Base de Datos (Knowledge Discovery in Databases KDD) el cual es un proceso metodológico para encontrar un "modelo" válido, útil y comprensible para describir un patrón basado en información, como modelo lo entendemos como un intento de explicar la representación de ese patrón en los datos. Cabe mencionar que hablar de un "modelo" como una fórmula mágica no es decir que hay maestros en cualquier problema, sino todo lo contrario, porque existen muchas formas o algoritmos para satisfacer las necesidades, dependiendo de los objetivos del modelo. Investigación y datos recopilados.

Estos pasos se dividen en 9 que son:

- Abstracción del escenario.
- Selección de datos.
- Limpieza y preprocesamiento.
- Transformación de los datos.
- Elección de tareas de Minería de Datos.
- Elección del algoritmo.
- Aplicación del algoritmo.
- Evaluación e interpretación.
- Entendimiento del conocimiento.

1.8.4 Metodología de desarrollo OpenUP

IEBS (2015), indica:

Es un proceso unificado ligero y ágil que aplica métodos iterativos e incrementales dentro de un ciclo de vida estructurado y contiene un conjunto mínimo de prácticas que ayudan a los equipos a desarrollar software de manera más eficiente.

Es un modelo de desarrollo de software que forma parte de Eclipse Process Model Framework, desarrollado por Eclipse Foundation.

Características de Open UP

- Desarrollo incremental.
- Uso de casos de uso y escenarios.
- Manejo de riesgos.
- Diseño basado en la arquitectura.

1.8.5 Métricas de calidad

Las normas ISO (Organización Mundial de Normalización), son aquellas directrices que proporcionan especificaciones positivas, de ámbito mundial, a los bienes, servicios y sistemas de una corporación para garantizar la máxima finura y eficiencia en sus consecuencias y funcionamiento.

Tomando esto como base, la tendencia ISO/IEC 25000, también llamada SQuaRE, se encarga de recoger dentro de un mismo informe el conjunto de normas o consejos para asegurar el más excelente uso y residencia del programa de software de una empresa. ¿En qué casos se consigue este procedimiento? Sobre todo, en las agencias que amplían sus propios paquetes o programas para el control y la organización interna; de esta manera, se consolida la correcta creación y usabilidad de los mismos en los planteamientos internos (CTMA, 2021).

1.8.6 Estimación de costos

1.8.6.1 COCOMO II

El Modelo Constructivo de costos (COCOMO II), desarrollado por Barry Boehm en 1981. Es un modelo jerárquico de estimación de costos de software. Debido a deficiencias encontradas surge COCOMO II un modelo que permite el costo, el esfuerzo y el tiempo cuando se planifica una nueva actividad de desarrollo de software.

Por tanto, COCOMO II es un modelo que permite la estimación de costes, esfuerzos y tiempos a la hora de planificar nuevas actividades de desarrollo de software

El principal cálculo en el modelo COCOMO es el uso de la ecuación del esfuerzo para estimar el número de personas o de meses necesarios para desarrollar el proyecto. El resto de resultados del modelo se derivan de esta medida. (Gomez et al., s.f.)

1.9 HERRAMIENTAS

1.9.1 Hardware

1.9.1.1 PC Portátil

Es una máquina capaz de realizar una secuencia de operaciones mediante un programa, de tal manera, que se realice un procesamiento sobre un conjunto de datos de entrada, obteniéndose otro conjunto de datos de salida, por lo cual se mencionan lo más básico que tiene una portátil.

- Microprocesador: Intel Core i7
- Velocidad del procesador: 5MHz (Mega Hertz)
- Capacidad de la RAM: 16 GB de RAM
- Almacenamiento Interno: 256 SD de espacio

1.9.2 Software

1.9.2.1 Sistema operativo Windows 10

Windows 10 es un dispositivo operativo avanzado por el uso de Microsoft como una parte de la propia familia de Windows NT de las estructuras operativas. Se convirtió en oficialmente presentado en septiembre de 2014, observado por medio de una breve presentación de demostración en la conferencia Build 2014. Entró en la comprobación de la beta en octubre de 2014 y se lanzó al público en general el 29 de julio de 2015. (Microsoft, 2021)

1.9.2.2 Python

Se trata de un lenguaje de programación, multiparadigma y multinivel, con asistencia para la programación orientada a objetos, imperativa y propositiva. Con este tipo de lenguaje se pueden crear aplicaciones nativas e híbridas, y tiene una sintaxis

al alcance de personas con un nivel básico de "alfabetización" en lenguajes de programación.

Según una encuesta realizada a través de programadores en la comunidad del portal web Stack Overflow, el setenta y tres por ciento de los constructores consideran que Python es el lenguaje más requerido sobre las opciones que actualmente se pueden encontrar en el mercado. (Caminiti, 2021)

1.9.2.3 Java

Java es una plataforma de lenguaje de programación para ordenadores portátiles creada por Sun Microsystems en 1995. Ha evolucionado desde sus humildes comienzos hasta la electricidad de gran parte del mundo digital moderno, ya que es una plataforma fiable sobre la que se construyen muchas ofertas y paquetes. Los nuevos e innovadores productos y servicios virtuales diseñados para el futuro también se basan en Java (java, 2022).

1.9.2.4 Weka

Es una plataforma de software para el aprendizaje automático y la Minería de Datos escrito en Java y desarrollado en la Universidad de Waikato. Weka es un software libre distribuido bajo la licencia GNU-GPL.

Cuenta con un conjunto de herramientas y algoritmos intuitivos para el análisis de datos y el modelado predictivo, junto con una interfaz gráfica de usuario para acceder fácilmente a sus funciones. El lanzamiento inicial de Weka fue una interfaz para TCL/TK para modelar algoritmos implementados en otros lenguajes de programación, más unas utilidades para preprocesamiento de datos desarrolladas en C para hacer experimentos de aprendizaje automático. (EcuRed, s.f.)

1.9.2.5 Herramienta IDE

- **Visual Code**

Es un editor de código fuente desarrollado por Microsoft para Windows, Linux y macOS, incluye soporte de depuración, controles de Git en línea, resaltado de sintaxis, finalización de código inteligente, fragmentos y refactorización de código. También es

personalizable, por lo que los usuarios pueden cambiar los temas del editor, las teclas de acceso rápido y las preferencias. Es gratuito y de código abierto 12 aunque la descarga oficial está bajo software privativo e incluye características personalizadas por Microsoft: (EcuRed, s.f.).

1.9.2.6 Base de Datos MySQL

MySQL es el sistema de gestión de base de datos relacional hoy en día, ya que se basa totalmente en código abierto. Originalmente desarrollado a través de MySQL AB, se convirtió en recibido a través de Sun MicroSystems en 2008 y en flip comprado por Oracle Corporation en 2010, que ya poseía su propio motor InnoDB para MySQL.

MySQL es un sistema de gestión de bases de datos con una licencia doble. Por un lado, es de código abierto, por otro lado, tiene una versión empresarial controlada por el uso de Oracle. (Robleado, 2019)

1.10 LÍMITES Y ALCANCES

1.10.1 Limites

El modelo de índice de proyección será enfocado en la contaminación de la basura de la Ciudad de El Alto. Tomando datos estadísticos de la misma.

- Brindará información del crecimiento de la contaminación solo de la Ciudad de El Alto.
- Mostrará datos del índice de crecimiento en intervalos de tiempo.
- No mostrara índice de otro componente de contaminación.

1.10.2 Alcances

El presente Modelo de Proyección tendrá las siguientes fases:

- Selección de datos útiles y necesarios. Procesamientos y limpieza de datos. Transformación de datos para vincular a la Minería de Datos.
- Aplicación de Minería de Datos a través del proceso KDD.
- Interpretación y evaluación.

1.11 APORTES

El modelo de índice de proyección brindará información sobre el comportamiento del crecimiento de la contaminación de la basura en la Ciudad de El Alto, esto deberá determinar a las autoridades en la toma de decisiones en un futuro próximo, donde se podrá realizar predicciones de impactos medio ambientales y de salud pública que puede causar la basura y la contaminación, todo esto se podrá determinar con la Minería de Datos con enfoque predictivo.

CAPÍTULO II

Marco Teórico

2 MARCO TEÓRICO

El presente capitulo tiene como principal objetivo dar a conocer las definiciones necesarias para el entendimiento del presente trabajo de investigación donde se dará a conocer diferentes herramientas tecnológicas existentes para el manejo de la Minería de Datos y la problemática de los residuos sólidos en la ciudad de El Alto.

2.1 DATO

Los datos son la unidad semántica mínima, y corresponden a porciones de información número uno que no tienen sentido en sí mismas como ayuda para la toma de decisiones. También pueden considerarse como un conjunto discreto de valores, que no dicen nada sobre el porqué de las cosas y no están orientados a la acción.

Un número de teléfono o el nombre de alguien, por ejemplo, son datos que, sin propósito, aplicación o contexto, no sirven de base para ayudar a la toma de decisiones. Los datos pueden ser un grupo de registros almacenados en una región física que consiste en un trozo de papel, una herramienta electrónica (CD, DVD, disco difícil...), o los pensamientos de alguien. En este sentido, las tecnologías estadísticas han contribuido mucho a la recopilación de datos.

Como es de esperar, los registros pueden proceder de activos externos o internos a la empresa, y pueden ser de naturaleza objetiva o subjetiva, o de naturaleza cualitativa o cuantitativa, etc. (Davenport & Prusak, 1999)

2.2 INFORMACIÓN

La información puede definirse como un conjunto de datos procesados que tienen un significado (relevancia, motivo y contexto) y que, por tanto, son útiles para los responsables de la toma de decisiones al reducir su incertidumbre. Los datos pueden convertirse en estadísticas mediante el uso de la inclusión de costes:

- **Contextualización:** se sabe en qué contexto y por qué motivo se han generado.
- **Categorizar:** se reconocen los dispositivos de medición que ayudan a interpretarlos.
- **Calcular:** los hechos también pueden haber sido procesados matemática o estadísticamente.
- **Corrección:** se eliminan los errores e incoherencias de los datos.
- **Condensar:** los hechos también pueden resumirse de forma más concisa (agregación).

Por lo tanto, la estadística es el intercambio verbal de conocimientos o inteligencia, y es capaz de cambiar la forma en que el receptor percibe algo, impactando en sus juicios de valor y comportamientos. (Davenport & Prusak, 1999)

Información = Datos + Contexto (incluido el coste) + Utilidad (reducción de la incertidumbre)

2.3 CONOCIMIENTO

El conocimiento es una mezcla de disfrute, valores, información y datos que sirve de marco para incorporar nuevas revisiones y hechos, y es útil para el movimiento. Se origina y se aplica en la mente de los conocedores. En las agencias se encuentra con frecuencia no sólo en los archivos o almacenes de registros, sino también en los ejercicios, métodos, prácticas y normas de la organización. (Davenport & Prusak, 1999)

El conocimiento se deriva de los registros, al igual que los datos se derivan de las estadísticas. Para que la información se convierta en conocimiento es vital llevar a cabo acciones que incluyan:

- Comparación con otros factores.
- Predicción de consecuencias.
- Búsqueda de conexiones.
- Conversación con otros proveedores de conocimientos.

2.4 MINERÍA DE DATOS

2.4.1 Evolución histórica

El autor Beltran (2018) indica que los componentes de la Minería de Datos (DM) han existido durante muchos años en la investigación en áreas que incluyen la inteligencia artificial, la estadística o el aprendizaje automático, se puede decir que ahora estamos asistiendo al reconocimiento de la madurez de esas estrategias, que junto con el maravilloso desarrollo de los motores de bases de datos y las herramientas para la integración de la información justifica su introducción dentro de la esfera empresarial.

Las raíces del DM se remontan a los años 50. Los departamentos de TI solían preparar resúmenes de información, especialmente de carácter comercial, que se almacenaban en archivos centrales, para poder facilitar los cuadros de control. Nacieron los sistemas de gestión de datos, pero eran engorrosos, inflexibles y difíciles de leer para los usuarios que no eran de informáticos. Los años 60 vieron la entrega de estructuras de control de bases de datos, que sin embargo eran inflexibles y carecían de la capacidad de realizar consultas. Luego aparecieron los motores relacionales, que resolvieron esos problemas, aunque las revisiones eran muy agotadoras de preparar y depurar, y perdían relevancia debido a su bajo grado de actualización. Otro problema crítico fue la diversidad de bases de datos no incluidas instaladas con la ayuda de los departamentos únicos de una empresa. Nadie consideraba la utilidad del destino viable de un sistema interdependiente (Beltran, 2018).

El Data Warehouse (DW) vino a resolver este problema a finales de los años 80. La vida del DW ha estimulado la mejora de las técnicas de DM, en las que las obligaciones de evaluación se automatizan y cruzan un paso similar al permitir la extracción inductiva de información.

Tabla 2.1

Evolución de las tecnologías relacionadas con Data Mining

Etapa	Cuestión Planteada	Tecnologías	Características
Recolección de datos (A 60)	'Dime mis beneficios totales en los últimos 4 años	Ordenadores, cintas, discos.	Retrospectivo, datos estáticos.
Acceso a los datos (años 80)	Ventas en Cataluña durante las últimas Navidades	Bases de Datos Relacionales (SQL) ODBC	Retrospectivo, datos dinámicos a nivel de registro.
Data Warehouse y soporte a la toma de decisiones. {Anos 90	Ventas en Andalucía detalle por delegación y descender a nivel tienda.	(OLAP), bases de datos multidimensionales, data warehouse	Retrospective, obtención dinámica de datos a múltiples niveles.
Data Mining	Justifica la tendencia de venta en Castilla para el próximo año	Algoritmos avanzados, ordenadores, multiprocesadores, bases de datos masivas.	Prospectivo, obtención proactiva de información.

Nota: *Cuadro descriptivo de la evolución de las tecnologías relacionadas con Data Mining (Beltran, 2018).*

2.4.2 ¿Qué es la Minería de Datos?

El autor Perez (2014) define la minería de la información como conjunto de estrategias dirigidas a dar con el descubrimiento de la información. Esto implica el análisis de comportamientos, patrones, tendencias, asociaciones y diferentes características de la información. Hoy en día se dispone de grandes cantidades de datos y es mucho más necesario tener la capacidad de investigarlos de forma ordenada para poder investigarlos de forma ordenada con el fin de extraer de forma automatizada la Inteligencia contenida en ellos mediante técnicas especializadas apoyadas en el uso de equipos informáticos.

Por otro lado, en Witten et al., (2000) define a la Minería de Datos como un sistema de extracción de información útil y comprensible, antes desconocida, a partir de grandes cantidades de información almacenada en distintos formatos.

En otras palabras, el reto esencial de la Minería de Datos es encontrar modelos inteligibles a partir de datos. Para que esta manera sea efectiva necesita ser automatizada o semiautomatizada (asistida), y utilizando los patrones descubiertos debería ayudar a tomar decisiones más seguras para la organización.

2.4.3 Tipos de modelos

La Minería de Datos pretende analizar los datos para extraer conocimientos. Esto puede ser en forma de relaciones, patrones o reglas inferidas de los datos y previamente desconocidas, o en forma de una descripción más concisa. Estas relaciones o resúmenes constituyen el modelo de los datos analizados. Existen muchos enfoques únicos de representación de los modelos y cada uno de ellos determina el tipo de técnica que puede utilizarse para inferirlos (Hernandez et al., 2004).

En la práctica, los modelos pueden ser de dos tipos: predictivos y descriptivos. Los modelos predictivos pretenden estimar valores de destino o desconocidos de variables de interés, que llamamos variables objetivo o dependientes, utilizando otras variables o campos dentro de la base de datos, a los que llamamos variables independientes o predictivas. Por ejemplo, un modelo predictivo podría ser el que nos permite estimar la demanda de un nuevo producto como característica del gasto en publicidad y marketing (Hernandez et al., 2004).

Los modelos descriptivos, por su parte, identifican patrones que explican o resumen los datos, es decir, sirven para explorar las propiedades de los datos analizados, no para predecir nuevos hechos. Por ejemplo, una agencia de viajes desea conocer grupos de seres humanos con los mismos gustos, para organizar ofertas únicas para cada organización y poder remitirles estos datos; para ello, analiza los viajes que han realizado sus clientes e infiere una versión descriptiva que caracteriza a estos grupos (Hernandez et al., 2004).

2.5 PROCESO DE EXTRACCIÓN DEL CONOCIMIENTO

La mayoría de las organizaciones mundiales producen más información en algunos días que una sola persona en toda su vida. Cada día se genera gigabytes de

datos sé que no es posible analizar ni supervisar mediante personas porque el auge de la producción de datos es exponencial. La producción automatizada de datos exige el desarrollo de nuevas técnicas mecanizadas para para la elección, el filtrado y la evaluación de los datos (Fayyad et al., 1996).

El proceso de extracción del conocimiento o KDD (Knowledge Data Discovery) surgió como consecuencia de diversas disciplinas: Inteligencia Artificial, bases de datos, estadísticas,

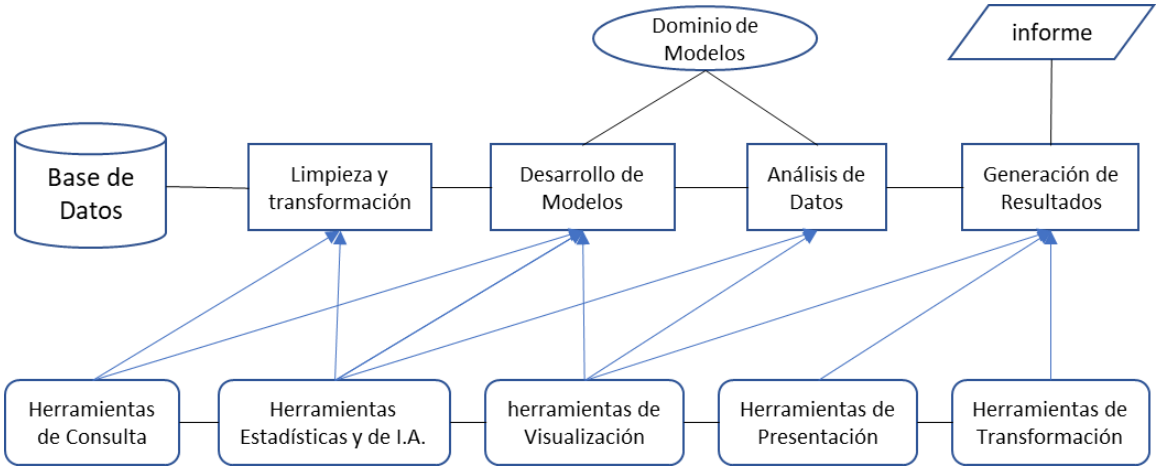
conocimiento de dispositivos, la visualización, la computación paralela y otras. La causa del KDD es el procesamiento de enormes bases de datos para que uno pueda descubrir la comprensión beneficiosa que se puede aplicar a la toma de decisiones.

El término minería de registros ha sido utilizado por muchos autores como sinónimo del método de descubrimiento de conocimientos, sin embargo, la Minería de Datos actualmente es un paso del proceso KDD, que corresponde únicamente a la etapa de descubrimiento de datos.

Formalizando, definiremos el sistema KDD porque el sistema no trivial de extraer información desconocida, implícita, antes desconocida y probablemente útil, de las estadísticas (Fayyad et al., 1996).

Figura 2.1

Fases de la metodología KDD



Nota: *Cuadro detallado de la metodología KDD (Perez & Santin, 2007).*

2.5.1 Fases del proceso de extracción de conocimiento

2.5.1.1 Recolección de datos

En la investigación por Beltran (2018), menciona que las primeras fases de KDD determinan que las fases sucesivas sean capaces de extraer conocimientos válidos y útiles partiendo de la información original. Generalmente, se determina la información que se desean investigar sobre un determinado dominio de la empresa:

- En bases de datos y diferentes fuentes muy diversas.
- Tanto internas como externas.
- Muchos de esos activos son los que se utilizan para los cuadros transaccionales.

El análisis posterior puede ser mucho más fácil si la fuente está unificada, accesible (interna) y desconectada del trabajo transaccional. Así que la siguiente técnica de Minería de Datos:

- Depende en gran medida de la fuente:
  - OLAP U OLTP.
  - Data warehouse o reproducción con el esquema auténtico.
  - ROLAP O MOLAP.
- Depende también de la forma de la persona:
  - Picapedreros (o granjeros): especialmente dedicados a hacer revisiones periódicas, ver la evolución de ciertos parámetros, seguir los valores atípicos, etc.
  - Exploradores: se encargan de localizar nuevos patrones considerables utilizando estrategias de minería estadística.
- Recogida de información externa. Además de los datos internos de la organización, los almacenes de registros pueden recopilar registros externos:
  - Demografía (censo), guía telefónica, psico geografía, gráficos de red, datos de diferentes grupos.

- datos compartidos en una empresa o región empresarial, organismos y asociaciones de expertos, catálogos, etc.
- datos resumida sobre zonas geográficas, distribución de la competencia, evolución económica, registros de calendario y clima, horarios de televisión, actividades deportivas, catástrofes.
- Bases de datos externas adquiridas a diferentes empresas.

2.5.1.2 Selección, limpieza y transformación de datos.

Beltrán (2018), indica que en esta fase se debe eliminar la mayor cantidad posible de datos erróneos o incoherentes (limpieza) e irrelevantes (cribado).

Métodos estadísticos casi exclusivamente.

- Histogramas (detección de datos anómalos).
- Selección de datos (muestreo, tanto vertical, eliminando atributos, como horizontal, eliminando tuplas).
- Redefinición de atributos (agrupación o separación).

Acciones en caso de información anómala (outliers):

- **Ignorar:** algunos algoritmos son robustos a la información anómala (por ejemplo, Timber).
- **Filtrar (eliminar o actualizar) la columna:** Solución extrema, pero de vez en cuando hay alguna otra columna dependiente con registros más agradables. Preferible a eliminar la columna es sustituirla por una columna discreta que anuncie si el coste se convierte en ordinario o atípico (por encima o por debajo).
- **Filtrar la fila:** Sesga claramente los datos, debido a que muchas veces las razones de las estadísticas erróneas están relacionadas con instancias o tipos únicos.
- **Sustituir el precio:** Por el precio "nulo" si el conjunto de reglas lo trata bien o por máximos o mínimos, dependiendo de dónde esté el valor atípico, o a través de las medias. A veces se puede predecir a partir de información diferente, el uso de cualquier técnica de ML.

- **Discretizar:** la remodelación de un valor no continuo en uno discreto (por ejemplo, muy alto, excesivo, medio, bajo, muy bajo) hace que los valores atípicos caigan en "muy alto" o "muy bajo" sin problemas esenciales.

Acciones para los registros que faltan (valores que faltan):

- **Ignorar:** algunos algoritmos son robustos a la falta de información (por ejemplo, los árboles).
- **Filtrar (eliminar o sustituir) la columna:** respuesta excesiva, pero en ocasiones puede haber cualquier otra columna dependiente con mejor información. Preferible a deshacerse de la columna, es reemplazarla por una columna booleana que anuncie si la tasa existió o no.
- **Filtrar la fila:** sesga claramente los datos, debido a que regularmente las causas de los datos que faltan están asociadas a casos o tipos únicos.
- **Sustituir la tasa por promedios.** A veces se puede esperar de otros registros, utilizando cualquier método de ML.
- **Segmentación:** las tuplas se segmentan por medio de los valores que deben tener. Se reciben diferentes modelos para cada segmento y luego se combinan.
- Modificación de la política de registros y espera hasta que los datos faltantes sean tenidos en cuenta.

Razones para la falta de valores:

A veces es crucial echar un vistazo a las razones que hay detrás de las estadísticas que faltan y actuar en consecuencia:

- Algunos valores que faltan explicitan rasgos relevantes: por ejemplo, un teléfono móvil que falta también puede constituir en muchos casos un deseo de que el hombre o la mujer en cuestión no sea molestado, o un último cambio de dirección.
- Valores inexistentes: muchos valores que faltan existen de hecho, pero otros no. Por ejemplo: el cliente que acaba de registrarse no tiene el consumo medio de los doce meses restantes.

- Registros incompletos: si los datos proceden de fuentes únicas, al combinarlos se suele hacer la unión y no la intersección de campos, por lo que muchos datos que faltan constituyen que esas tuplas proceden de una o varias fuentes diferentes al descanso.

2.5.1.3 Minería de Datos.

Características especiales de los datos:

Aparte del gran volumen, ¿por qué el aprendizaje del sistema y las estrategias estadísticas no son inmediatamente relevantes?

- Los datos residen en el disco. No se pueden escanear un par de veces.
- Algunas técnicas de muestreo no son compatibles con los algoritmos no incrementales.
- Dimensionalidad muy excesiva (muchos campos).
- Prueba positiva.
- Datos imperfectos

Aunque algunos se aplican de forma casi inmediata, la afición en los estudios de minería de la información está en adaptarlos.

Patrones a localizar:

- Una vez acumulada la información de interés, un explorador puede decidir qué tipo estilos desea averiguar.
- El tipo de conocimiento a extraer marcará absolutamente el enfoque de minería de información a utilizar.
- Dependiendo del tipo de conocimiento que se busque, se puede distinguir entre
    - Minería de información dirigida: se sabe realmente lo que se busca, normalmente se esperan datos positivos o instrucciones.
    - Minería de estadísticas no dirigida: ahora no se reconoce lo que se busca, se figura con los hechos.

En el primer caso, las propias estructuras de minería de registros son normalmente responsables de decidir el máximo conjunto apropiado de reglas entre las disponibles para un determinado tipo de patrón a buscar (Beltran, 2018).

2.5.1.4 Evaluación y validación.

El apartado anterior produce una o varias hipótesis de modelo. Para elegir y validar esos modelos, es necesario utilizar criterios de evaluación de la especulación. Por ejemplo:

1a Fase: Probar la precisión del modelo en un banco de ejemplos no sesgados de los únicos utilizados para examinar la versión. Se puede elegir el modelo agradable.

2a Fase: Se puede realizar un disfrute piloto con esa versión. Por ejemplo, si el modelo determinado se utilizara para esperar la reacción de los consumidores a un nuevo producto, se puede enviar un mailing a un subconjunto de clientes y evaluar la fiabilidad del modelo (Beltran, 2018).

2.5.1.5 Interpretación y difusión.

El despliegue del modelo es de vez en cuando trivial, pero de vez en cuando requiere una técnica de implementación o interpretación:

- El modelo puede requerir una implementación (por ejemplo, detección de tarjetas fraudulentas en tiempo real).
- La versión es descriptiva y requiere una interpretación (por ejemplo: una caracterización de las regiones geográficas en función de la distribución de las mercancías vendidas).
- El modelo puede tener muchos usuarios y desea ser difundido: el modelo también puede necesitar ser expresado de forma comprensible para ser distribuido dentro de la corporación (p.ej.: Las cervezas y los productos congelados se compran regularmente de forma colectiva y se ubican en estanterías remotas) (Beltran, 2018).

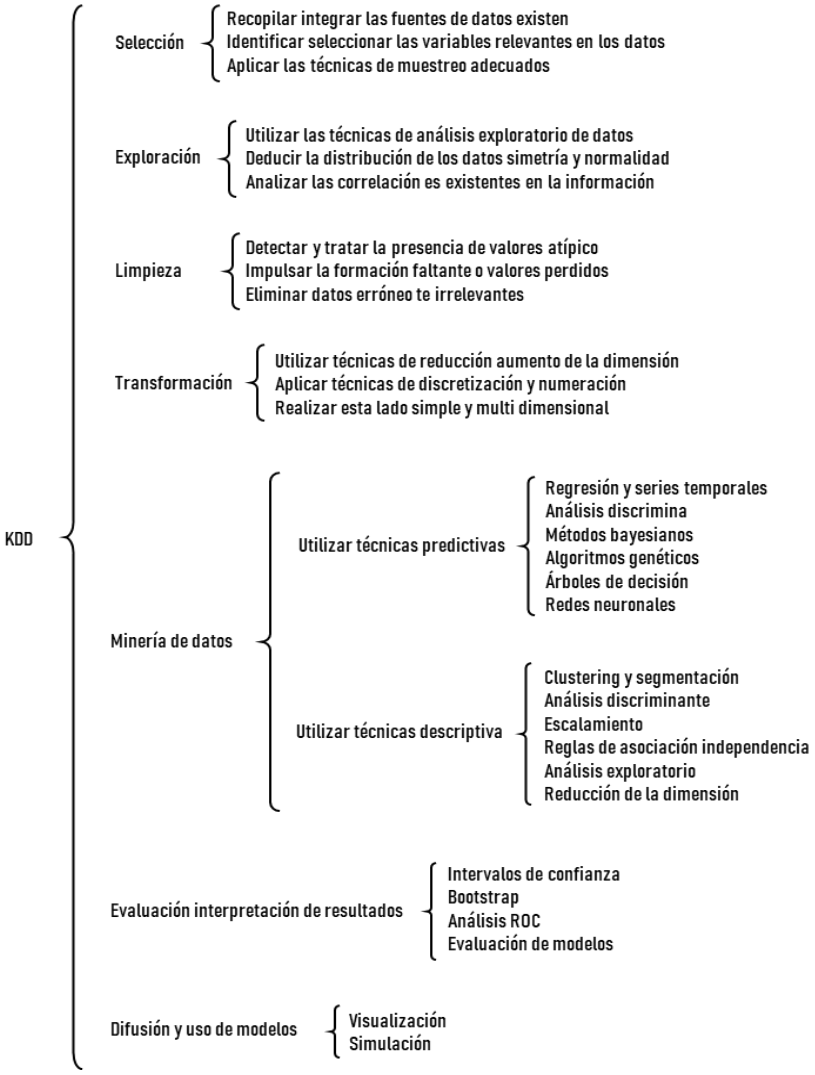
2.5.1.6 Actualización y seguimiento.

Los procesos conducen a la renovación:

- **Actualización:** Un modelo legítimo puede no ser válido: alternancia de contexto (financiero, oposición, recursos de datos, etc.).
- **Seguimiento:** Consiste en revalidar el modelo con una frecuencia positiva sobre la y nueva información, de manera que se localice si la versión requiere un reemplazo (Beltran, 2018).

Figura 2.2

Etapas descriptiva de la metodología KDD



Nota: Proceso de descubrimiento del conocimiento KDD (Perez & Santin, 2007)

2.6 MÉTODOS Y TÉCNICAS DE MINERÍA DE DATOS

2.6.1 Métodos

Beltran, (2018) indica que los métodos de Minería de Datos tienen como metas principales la predicción de registros desconocidos y la descripción de estilos.

Se pueden utilizar diferentes normas para clasificar los sistemas de Minería de Datos y en particular, las estructuras de dominio inductivo en computadoras:

- Dependiendo de la razón por la que se realiza el estudio, los sistemas pueden destacarse por: tipo, regresión, agrupación de conceptos, compactación, modelado de dependencias, detección de desviaciones, etc.
- Dependiendo de la tendencia con la que se aborde el problema, pueden destacarse líneas fundamentales de investigación o paradigmas: estructuras conexionistas (redes neuronales), estructuras evolutivas (algoritmos genéticos) y sistemas simbólicos.
- Según el lenguaje utilizado para simbolizar el conocimiento, distinguiremos: representaciones basadas en la lógica de predicados proposicional, representaciones basadas totalmente en la lógica de predicados de primer orden, representaciones basadas, representaciones mediante ejemplos y representaciones no simbólicas junto con redes neuronales.

2.6.1.1 Agrupamiento ("Clustering"):

También llamado Segmentación, este dispositivo permite identificar tipologías o negocios en los que los factores son iguales entre sí y diferentes a los de otros grupos. Para obtener las distintas tipologías o agencias existentes en una base de datos, estos equipos requieren, como entrada, información en el grupo a segmentar. Esta información corresponderá a los valores específicos, para cada elemento en un momento dado, de una serie de variables ("segmentación estática") o a través del comportamiento a lo largo de los años de cada uno de los elementos de la organización ("segmentación dinámica") (Beltran, 2018).

2.6.1.2 Asociación (" Association Pattern Discovery");

Establece las posibles relaciones o correlaciones entre acciones y ocasiones únicas supuestamente imparciales, siendo capaz de reconocer cómo la incidencia de una ocasión o movimiento puede resultar o generar la ocurrencia de otros.

Normalmente, este tipo de herramienta se basa principalmente en estrategias estadísticas que incluyen la correlación y la evaluación de la varianza. (Beltran, 2018)

2.6.1.3 Secuenciamiento ("Sequential Pattern Discovery"):

Permite percibir cómo, a lo largo de los años, la ocurrencia de un movimiento desencadena finalmente otros. Puede ser muy parecido al analizado anteriormente, aunque en este caso, el tiempo es una variable crucial y vital a proteger dentro de la información a analizar (Beltran, 2018).

2.6.1.4 Reconocimiento de Patrones ("Pattern Matching"):

Permiten la asociación de una información de entrada de signos con aquella que es más comparable y que podrían estar catalogadas dentro del sistema.

Estas herramientas son utilizadas por elementos que pueden ser tan habituales como un procesador de textos o un despertador. Los patrones pueden ser cualquier detalle de información que deseemos (Beltran, 2018).

2.6.1.5 Previsión ("Forecasting"):

La previsión establece el comportamiento del destino máximo en toda probabilidad basándose en la evolución pasada y presente.

Tiene su uso esencial dentro del tratamiento de Series Temporales y las estrategias asociadas tienen una gran madurez (Beltran, 2018).

2.6.1.6 Simulación:

La simulación puede describirse como la tecnología de más de una situación o posibilidad sujeta, normalmente, a directrices o esquemas para investigar la idoneidad y el comportamiento de una selección o prototipo en un marco de condiciones futuras

viables o para investigar todas las versiones o alternativas factibles a una selección o situación, y también para investigar todas las variaciones o alternativas factibles a una elección o situación. alternativas a una decisión o situación y se utiliza igualmente para el cálculo numérico (Beltran, 2010).

2.6.1.7 Optimización:

La optimización resuelve el problema de minimizar o maximizar una función que depende de una serie de variables, localizando los valores de éstas que cumplen la situación de máximo, normalmente beneficios, o mínimo, normalmente gastos (Beltran Martinez, 2018).

2.6.1.8 Clasificación ("Clasification", "Prediction" or "Scoring"):

La clasificación agrupa todas aquellas herramientas que nos permiten asignar un dato para pertenecer a un grupo o clase. Esto se instrumenta a través de la dependencia de la pertenencia a una clase de los valores de una secuencia de atributos o variables (Beltran, 2018).

2.6.2 Técnicas

La Minería de Datos ha llevado a una lenta sustitución del análisis de la información basado en la verificación por una técnica de descubrimiento de la comprensión para la evaluación de los datos. La distinción importante entre las dos es que, dentro de esta última, los datos se observan sin necesidad de formular una hipótesis de antemano. La aplicación automatizada de los algoritmos de Minería de Datos hace que sea factible sin problemas encontrar patrones dentro de los datos, por lo que este enfoque es mucho más eficiente que la evaluación impulsada por la verificación mientras se trata de explorar los hechos de grandes y particularmente complejos datos (Beltran, 2018).

En la tabla siguiente se muestra algunas de las técnicas de Minería de Datos en ambas categorías:

Tabla 2.2

Clasificación de las técnicas de Minería de Datos

SUPERVISADOS	NO SUPERVISADOS
Arboles de decisión	Detección de desviaciones
Inducción neuronal	Segmentación
Regresión	Agrupamiento
Series temporales	Reglas de asociación
	Patrones secuenciales

Nota: Técnicas de la Minería de Datos (Beltran, 2018)

La aplicación de los algoritmos de Minería de Datos requiere la realización de una serie de actividades previas encaminadas a preparar los datos de entrada debido a que, en muchas ocasiones dichos datos proceden de fuentes heterogéneas, no tienen el formato adecuado o contienen ruido. Por otra parte, es necesario interpretar y evaluar los resultados obtenidos (Martinez, 2018).

Tabla 2.3

Técnicas de la Minería de Datos

Métodos estadísticos	ANOVA
	Prueba ji cuadrado
	Análisis de componentes principales
	Análisis de clusters
	Análisis discriminante
	Regresión lineal
	Regresión logística
Arboles de decisión	CHAID
	CART
Reglas de asociación	
Redes neuronales	
Algoritmos genéticos	
otros	Lógica difusa
	Series temporales

Nota: Listado de Técnicas de la Minería de Datos (Martinez, 2018)

2.6.2.1 Métodos estadísticos:

Beltran, (2018) indica que la estadística es históricamente la técnica que se ha utilizado para el tratamiento de grandes volúmenes de datos numéricos, y nadie duda de su eficacia porque posee un conjunto totalmente masivo de fórmulas de análisis para cubrir el tratamiento de todo tipo de poblaciones y series de datos. Estos son algunos de los métodos estadísticos más utilizados:

- **ANOVA:** Análisis de la Varianza, contrasta si existen o no variaciones masivas entre las medidas de una o más variables continuas en diferentes grupos de población.
- **Jl cuadrada:** Contrasta la hipótesis de independencia entre variables.
- **Componentes principales:** Permite reducir la cantidad de variables observadas a una menor cantidad de variables ficticias, preservando la mayoría de las estadísticas sobre la mayoría de los hechos en la varianza de las variables.
- **Análisis de Clusters:** Permite categorizar una población en una amplia variedad de organismos determinados, basados en la de empresas, sobre la premisa de las similitudes y diferencias dentro de los perfiles actuales presentes entre los aditivos únicos de la población
- **Análisis discriminante:** Una técnica de clasificación de individuos en grupos que han sido previamente enganchados y que han sido previamente montados, y que permite localizar la regla de tipo de los de los factores de esas empresas, y en consecuencia tomar conciencia de cuáles pueden ser las variables que mejor definen la qué variables perfilan mejor la pertenencia a la organización.
- **Regresión lineal:** La técnica más fundamental de la Minería de Datos. Se implementa a través de la identificación de una variable base (y) y todas las variables insesgadas (X1, X2, ...). Se piensa que la conexión entre ellas es lineal. Todas las variables deben ser imparciales. El resultado final es la ecuación de la línea que mejor se adapta al conjunto de hechos y esta ecuación se interpreta o se utiliza para la predicción.

- **Regresión Logística:** Puede trabajar con variables discretas. También requiere que todas las variables sean lineales.

2.6.2.2 Métodos Basados en Arboles de Decisión

Son herramientas analíticas utilizadas para el descubrimiento de regulaciones y relaciones mediante la ruptura y subdivisión sistemática de la información contenida en el conjunto de registros. El árbol de elección se construye dividiendo el conjunto de registros en (CART) o más (CHAID) subconjuntos de observaciones basados totalmente en los valores tomados por las variables predictoras. Cada uno de estos subconjuntos se vuelve a dividir utilizando el mismo algoritmo (Beltran, 2018).

Así se mantiene hasta que no se observan variaciones considerables en el impacto de las variables predictoras de este tipo de empresas a costa de la variable de respuesta.

La raíz del árbol es el conjunto de hechos completo, los subconjuntos y subconjuntos forman las ramas del árbol. Un conjunto en el que se realiza una partición se denomina nodo.

El enfoque CHAID (Chi Squared Automatic Interaction Detector) es útil en situaciones en las que el objetivo es dividir una población en segmentos específicos basándose principalmente en algún criterio de decisión.

2.6.2.3 Reglas de Asociación

Derivan de un tipo de análisis que extrae información por coincidencias. Este análisis a veces llamado "cesta de la compra" permite descubrir correlaciones o co-ocurrencias en los

de la base de datos a analizar y se formaliza en la obtención de reglas de tipo; SI ... ENTONCES...

2.6.2.4 Redes Neuronales (“Neural Networks")

Las Redes Neuronales constituyen una técnica inspirada en los trabajos de investigación, iniciados en 1930, que pretendían modelar computacionalmente el

aprendizaje humano llevado a cabo a través de las neuronas en el cerebro (Martinez, 2018).

Las redes neuronales son una nueva forma de analizar la información con una diferencia fundamental con respecto a las técnicas tradicionales: son capaces de detectar y aprender patrones y características dentro de los datos. Se comportan de forma parecida a nuestro cerebro aprendiendo de la experiencia y el pasado y aplicando tal conocimiento a la resolución de problemas nuevos.

Una vez adiestradas las redes neuronales pueden hacer previsiones, clasificaciones y segmentación.

Las redes neuronales se construyen estructurando en una serie de niveles o capas compuesta por nodos o "neuronas". Poseen dos formas de aprendizaje derivadas del tipo de paradigma que usan: el supervisado y el no supervisado.

Son métodos de proceso numérico en paralelo que tratan de modelizar el funcionamiento del cerebro. La red asigna pesos al azar a cada variable independiente y determina si existe algún patrón predictivo en los datos. Una vez que encuentra un patrón la red lo optimiza reforzando los pesos de las variables y comparando con los datos del grupo de validación. Luego prosigue el proceso y aprende de los resultados una y otra vez. Finalmente, se puede aplicar el modelo aprendido a cualquier nuevo conjunto de datos de entrada. Pueden manejar datos continuos y discretos, lineales y no-lineales simultáneamente. El único inconveniente que presentan es que no genera una ecuación o modelo que explique el comportamiento del sistema, siendo muy difícil determinar la influencia de cada variable en el comportamiento global del sistema (Martinez, 2018).

2.6.2.5 Algoritmos Genéticos ("Genetic Algorithms")

Los Algoritmos Genéticos son otra técnica que debe su inspiración, de nuevo, a la Biología como las Redes Neuronales.

Estos algoritmos representan la modelización matemática de como los cromosomas en un marco evolucionista alcanzan la estructura y composición más óptima en aras de la supervivencia. Entendiendo la evolución como un proceso de

búsqueda y optimización de la adaptación de las especies que se plasma en mutaciones y cambios en los genes o cromosomas (Beltran, 2018).

Los Algoritmos Genéticos hacen uso de las técnicas biológicas de reproducción (mutación y cruce) para ser utilizadas en todo tipo de problemas de búsqueda y optimización.

Esta aproximación está enfocada a problemas de optimización. Se comienza con una población de partida y se va alterando y optimizando su composición para la solución de un problema particular mediante mecanismos tomados de la teoría de la evolución (introducir elementos aleatorios para la modificación de las variables o mutaciones). El material genético o información de los individuos puede ser transmitido a las siguientes generaciones, de diferentes formas que van optimizando el proceso. A través de la reproducción, los mejores segmentos perduran y su proporción crece de generación en generación. Al cabo de cierto número de iteraciones, la población estará constituida por buenas soluciones al problema de optimización.

Esta herramienta se usa en las primeras fases del Data Mining, para seleccionar las variables que luego se emplearán con otra técnica, como las redes de neuronas o la regresión logística

2.6.2.6 Lógica Difusa (“fuzzy logic")

La Lógica Difusa surge de la necesidad de modelizar la realidad de una forma más exacta evitando precisamente el determinismo o la exactitud. La Lógica permite el tratamiento probabilístico de la categorización de un colectivo.

La Lógica Difusa es aquella técnica que permite y trata la existencia de barreras difusas o suaves entre los distintos grupos en los que categorizamos un colectivo o entre los distintos elementos, factores o proporciones que concurren en una situación o solución (Martinez, 2018).

2.6.2.7 Series Temporales

Consisten en el estudio de una variable a través del tiempo para, a partir de ese conocimiento, y bajo el supuesto de que no van a producirse cambios estructurales,

poder realizar predicciones. Suelen basarse en un estudio de la serie en ciclos, tendencias y estacionalidades, que se diferencian por el ámbito de tiempo abarcado, para, por composición, obtener la serie original. Se pueden aplicar enfoques híbridos con los métodos anteriores, en los que la serie se puede explicar no sólo en función del tiempo sino como combinación de otras variables de entorno más estables y, por lo tanto, más fácilmente predecibles (Beltran, 2018).

2.6.2.8 Redes Bayesianas

Las redes bayesianas son una alternativa para Minería de Datos, la cual tiene varias ventajas:

- Permiten aprender sobre relaciones de dependencia y causalidad.
- Permiten combinar conocimiento con datos.
- Evitan el sobreajuste de los datos.
- Pueden manejar bases de datos incompletos.

El obtener una red bayesiana a partir de datos es un proceso de aprendizaje, el cual se divide, naturalmente, en dos aspectos:

1. Aprendizaje paramétrico: dada una estructura, obtener las probabilidades a priori y condicionales requeridas.
2. Aprendizaje estructural: obtener la estructura de la red Bayesiana, es decir, las relaciones de dependencia e independencia entre las variables involucradas.

Las técnicas de aprendizaje estructural dependen del tipo de estructura de red: árboles, poli árboles y redes multiconectadas. Otra alternativa es combinar conocimiento subjetivo del experto con aprendizaje. Para ello se parte de la estructura dada por el experto, la cual se valida y mejora utilizando datos estadísticos.

2.6.2.9 Inducción de Reglas

Las técnicas de Inducción de Reglas surgieron hace dos décadas y permiten la generación y contraste de árboles de decisión o reglas y patrones a partir de los datos de entrada.

Como información de entrada, tendremos un conjunto de casos donde se ha asociado una clasificación o evaluación a un conjunto de variables o atributos. Con tal información estas técnicas obtienen el árbol de decisión o conjunto de reglas que soportan la evaluación o clasificación.

En los casos en que la información de entrada posee algún tipo de "ruido" o defecto estas técnicas pueden habilitar métodos estadísticos de tipo probabilístico para generar, en estos casos, árboles de decisión podados o recortados.

2.6.2.10 Sistemas basados en el Conocimiento y Sistemas Expertos ("Knowledge Based Systems" & "Expert Systems")

Estos sistemas son un clásico de la inteligencia Artificial.

Estas técnicas permiten la formalización de árboles y reglas de decisión extraídas de la formalización del conocimiento de los expertos.

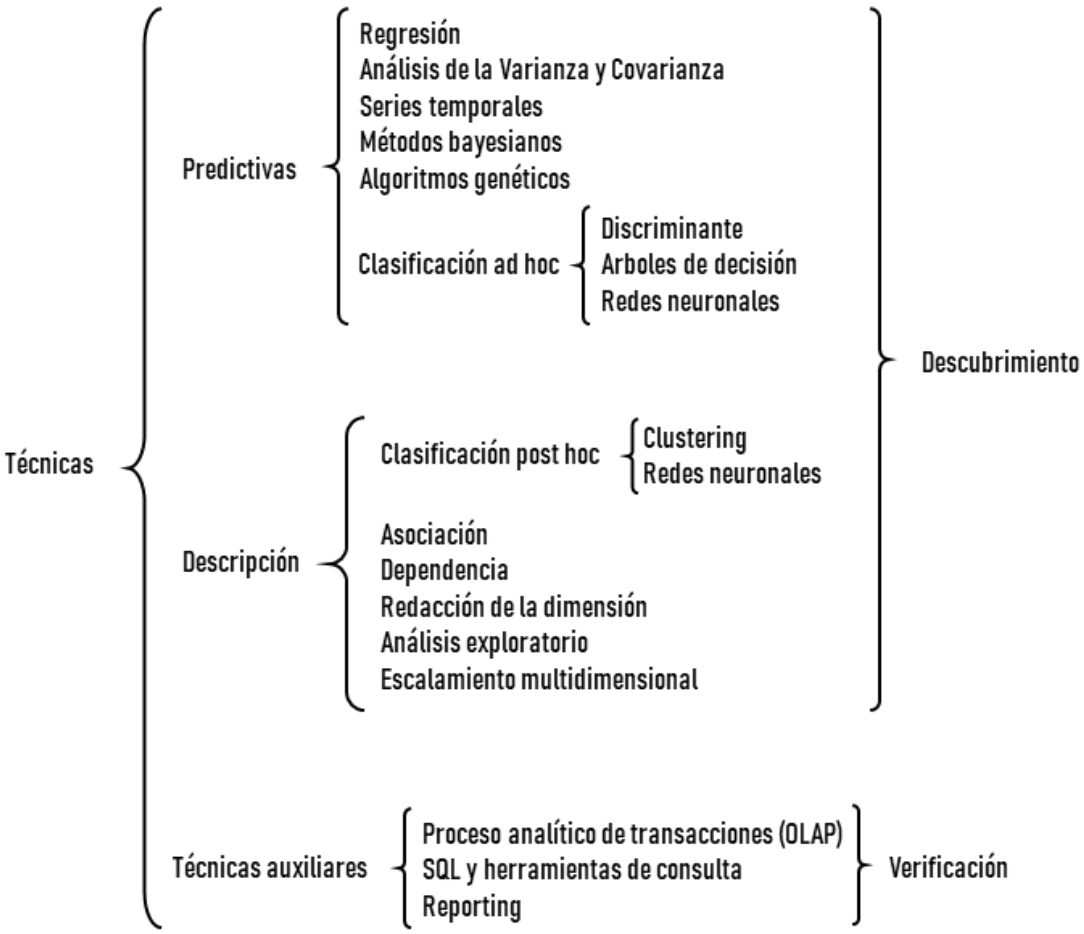
Poseen motores llamados "Motores de Inferencia" que se encargan de gestionar las distintas preguntas al ser realizadas de forma que el proceso de decisión sea lo más eficiente y rápido posible (Beltran, 2018).)

2.6.2.11 Algoritmos Matemáticos

Sin llegar a ser técnicas que den soporte a unas necesidades concretas como las anteriores, existe una amplia gama de algoritmos matemáticos que son especialmente útiles y eficaces en la resolución y tratamiento de problemas muy específicos y puntuales y que, normalmente, son incorporados en alguna de aquellas técnicas con el objeto de mejorarlas.

Figura 2.3

Descripción general de técnicas de Minería de Datos



Nota: Descripción general de técnicas de Minería de Datos (Perez & Santin, 2007)

2.7 HERRAMIENTAS DE LA MINERÍA DE DATOS

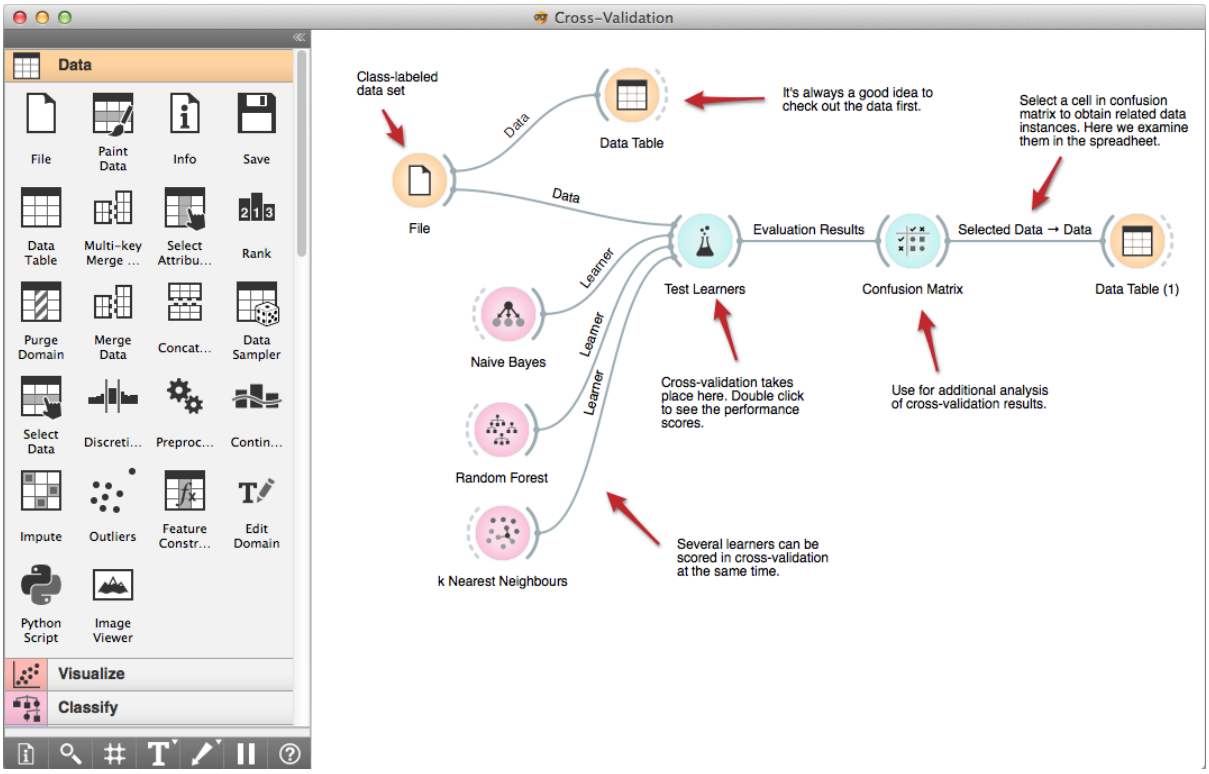
2.7.1 Orange data mining

De acuerdo a Predictive Analitics Today (2022):

Orange es una herramienta de análisis y visualización de información de código abierto. Orange está avanzado en el Laboratorio de Bioinformática de la Facultad de Informática y Ciencias de la Información de la Universidad de Liubliana, Eslovenia, junto con la comunidad de código abierto. La Minería de Datos se realiza mediante programación visual o scripts en Python.

Figura 2.4

Interfaz de Orange data Mining



Nota: Interfaz de trabajo de Orange Data Mining (PAT RESEARCH, 2022*).*

La herramienta tiene componentes para el aprendizaje automático, plugins para la bioinformática y la minería de textos y está llena de funciones para el análisis de datos. Orange es una biblioteca de Python. Los scripts de Python pueden ejecutarse en una ventana de terminal, en entornos integrados que incluyen PyCharm y Python Win, o en shells junto con iPython. Orange incluye una interfaz de lienzo en la que la persona coloca widgets y crea un flujo de trabajo de análisis de datos

Los widgets proporcionan funcionalidades como la lectura de información, la visualización de tablas de datos, la elección de funciones, la predicción de entrenamiento, la evaluación de los algoritmos de aprendizaje.

2.7.1.1 Características

- Código abierto
- Visualización interactiva de datos.

- Programación visual.
- Admite capacitación práctica e ilustraciones visuales.
- Complementos extiende la funcionalidad.

2.7.1.2 Beneficios

- Para todos: principiantes y profesionales.
- Ejecute análisis de datos simples y complejos.
- Cree gráficos hermosos e interesantes.
- Utilícelos en una clase de análisis de datos.
- Acceda a funciones externas para análisis avanzados.

2.7.2 R software Enviroment

De acuerdo a Predictive Analitics Today (2022):

R es un entorno de software libre para la computación estadística y las gráficas. Se compila y se ejecuta en un amplio tipo de sistemas UNIX, Windows y MacOS. R es un conjunto incorporado de instalaciones de programas de software para la manipulación de la información, el cálculo y la visualización gráfica. Algunas de las funciones abarcan una instalación eficiente de manejo y almacenamiento de datos, un conjunto de operadores para los cálculos en las matrices, especialmente las matrices, una enorme, coherente e incorporado colección de herramientas intermedios para el análisis de datos, las instalaciones gráficas para el análisis de la información y la visualización sin demora en el ordenador portátil o en papel, y un lenguaje de programación bien desarrollado, simple y potente que consiste en condicionales, bucles, definidos por el consumidor funciones recursivas, y las instalaciones de entrada y salida.

R puede ser en gran medida un automóvil para el desarrollo de nuevos métodos de análisis de información interactiva. Ha evolucionado rápidamente y se ha ampliado con una gran serie de paquetes. El lenguaje R es ampliamente utilizado entre los estadísticos y los mineros de hechos para ampliar el programa de software estadístico y la evaluación de datos.

2.7.2.1 Características

- Código abierto - Software gratuito.
- Proporciona una amplia variedad de estadísticas (modelado lineal y no lineal, pruebas estadísticas clásicas, análisis de series de tiempo, clasificación, agrupamiento) y técnicas gráficas.
- Facilidad de manejo y almacenamiento de datos efectivos.
- Conjunto de operadores para cálculos en matrices, en particular matrices.
- Colección grande, coherente e integrada de herramientas intermedias para el análisis de datos.
- Facilidades gráficas para el análisis y visualización de datos en pantalla o en papel.
- Lenguaje de programación bien desarrollado, simple y efectivo que incluye condicionales, bucles, funciones recursivas definidas e instalaciones de entrada y salida.

2.7.2.2 Beneficios

- Trae análisis a sus datos
- Se ejecuta en una amplia variedad de plataformas: UNIX, Windows, Macos.
- Software estadístico ampliamente utilizado.
- Fácil de aprender.
- Natural y expresivo.
- Capacidades reconocidas para visualizar datos.

2.7.3 WEKA

Es la herramienta que principalmente en base a las referencias consultadas es propicio para la modelización de esta investigación, en consonancia con Predictive Analytics Today (2018)

Weka es un conjunto de algoritmos de aprendizaje para tareas de Minería de Datos. Los algoritmos pueden ser implementados directamente a un conjunto de datos

o invocados desde su propio código Java. Las funciones de Weka consisten en el aprendizaje de dispositivos, la Minería de Datos, el preprocesamiento, la clasificación, la regresión, la agrupación, las pautas de asociación, la selección de atributos, los experimentos, el flujo de trabajo y la visualización. Weka está escrito en Java, desarrollado en la Universidad de Waikato, Nueva Zelanda. Todas las técnicas de Weka se basan principalmente en la suposición de que las estadísticas se tienen como un archivo plano único o relación, en el que cada factor de información se define a través de una cantidad de flujo de atributos. Weka ofrece la admisión a las bases de datos SQL el uso de Java Database Connectivity y puede manera el resultado final de nuevo mediante el uso de una consulta de base de datos. No es capaz de la Minería de Datos multi-relacionales (Today, 2022).

La interfaz de usuario predominante de Weka es el Explorador, la capacidad idéntica también se puede acceder a través de la interfaz de knowledge. También existe el Experimentador, que permite la comparación sistemática del rendimiento predictivo de la máquina de Weka ganando conocimiento de los algoritmos en una colección de conjuntos de información. La interfaz del Explorador cuenta con varios paneles que ofrecen el derecho de entrada a los componentes principales de la mesa de trabajo, que incluye el panel de preprocesamiento que permite la importación de datos, el panel de clase permite al consumidor utilizar algoritmos de tipo y regresión, el panel de cómplices presenta el acceso a la regla de afiliación de las personas sin experiencia, el panel de cluster ofrece el acceso a las técnicas de Clustering, el panel de selección de atributos presenta algoritmos para conocer los atributos más predictivos en un conjunto de datos. (Today, 2022)

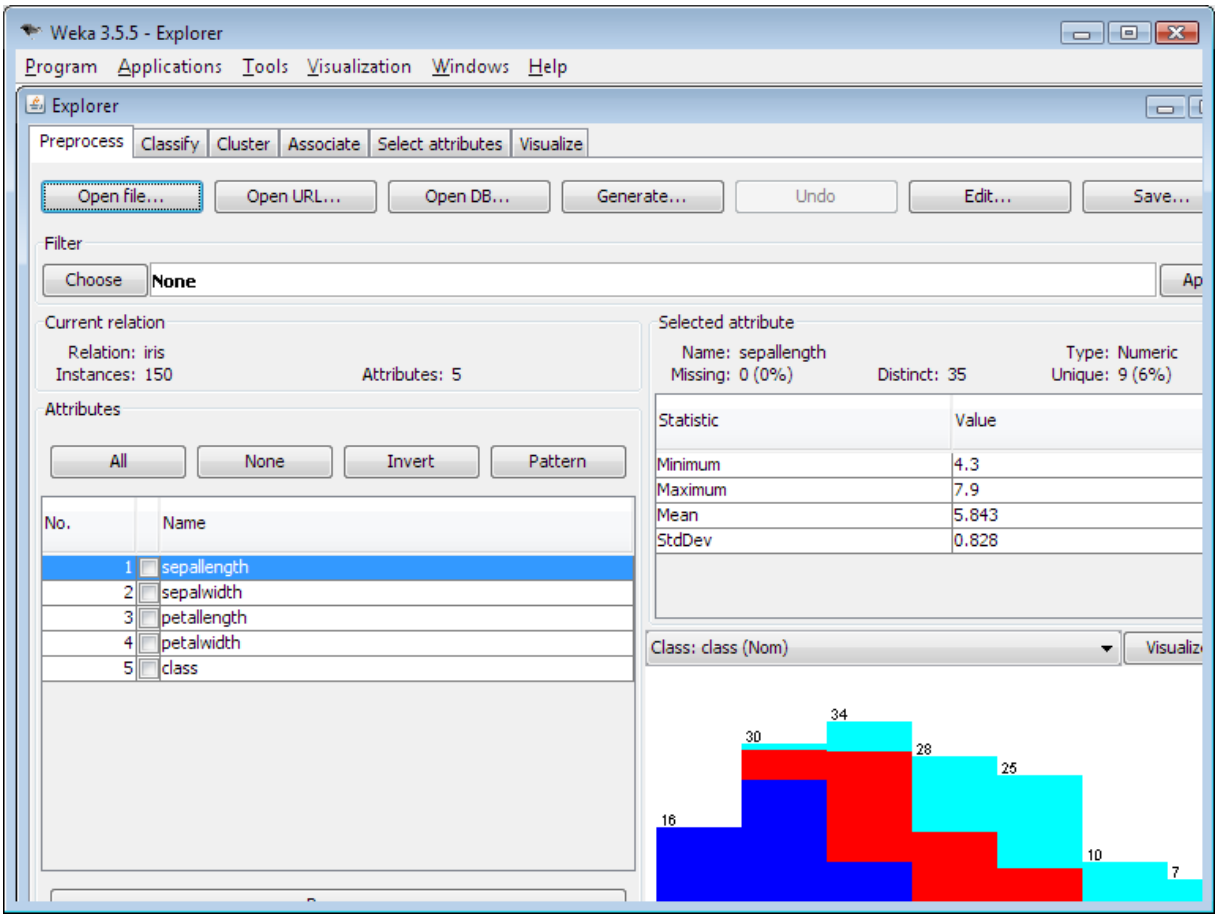
Weka proporciona un conjunto completo de engranajes de preprocesamiento de datos, ganando conocimiento de los algoritmos y estrategias de evaluación, interfaces gráficas de los consumidores y un entorno para la evaluación de los algoritmos de maestría. Los datos pueden ser importados desde un registro en diversos formatos, que incluyen ARFF, CSV, C4.5, binario. Los datos también pueden ser examinados desde una URL o desde una base de datos SQL (utilizando JDBC). Las herramientas de preprocesamiento en WEKA se conocen como "filtros" y los filtros

están disponibles para la discretización, normalización, remuestreo, selección de atributos, transformación y combinación de características.

Los esquemas de aprendizaje aplicados son arboles de decisión y listas, clasificadores totalmente basados en instancias, máquinas de vectores de ayuda, perceptrones multicapa, regresión logística, redes de Bayes. Las metas del clasificador incluidos son los códigos de salida bagging, boosting, stacking, corrección de errores, masterización ponderada regionalmente. Los esquemas aplicados son ok-Means, EM, Cobweb, X-manner, FarthestFirst. Los clusters pueden visualizarse y compararse con los clusters "verdaderos" Apriori puede calcular todas las regulaciones que tienen una guía mínima dada y superan una autocreencia dada. En Weka, los activos de hechos, clasificadores, etc. son granos y pueden ser conectados gráficamente.

Figura 2.5

Interfaz de Weka



Nota: Interfaz de la herramienta de Weka (Waikato, 2022).

2.7.3.1 Características

- Procesamiento de datos
- Clasificación de datos
- Regresión de datos
- Agrupación de datos
- Reglas de asociación de datos
- Visualización de datos

2.7.3.2 Beneficios

- portátil
- De uso gratuito.
- Fácil de usar.
- Adaptado para crear nuevas formas de diseños de aprendizaje automático.
- Contiene herramientas con múltiples usos.
- Cursos en línea gratuitos disponibles.
- Profesores altamente educados, capacitados y comprometidos.
- Libros y publicaciones extremadamente ingeniosos disponibles.
- Últimas tendencias en inteligencia artificial.

2.7.4 RapidMiner

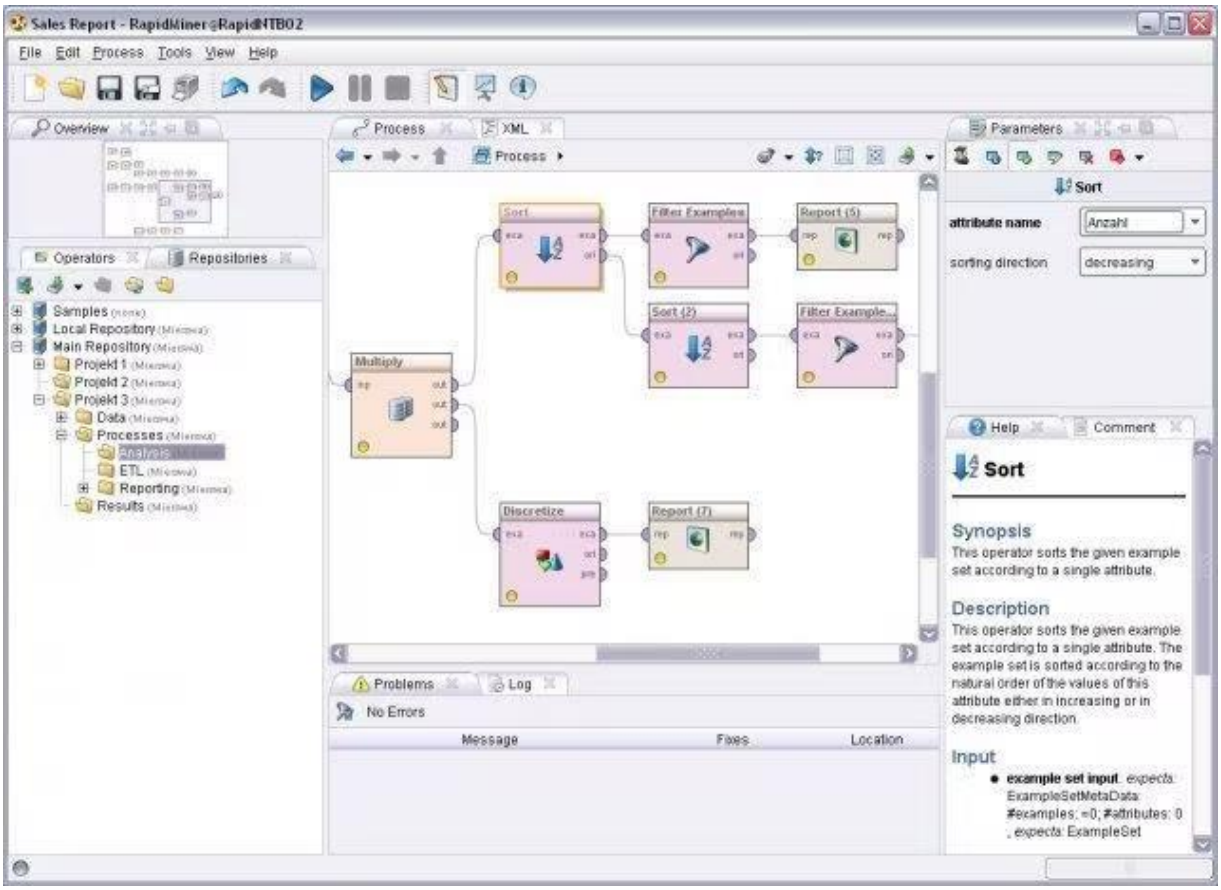
Predictive Analytics Today (2019), indica:

RapidMiner Studio ofrece una gran cantidad de funcionalidades para acelerar y optimizar la exploración de las tareas, la fusión y la limpieza de los datos, reduciendo el tiempo dedicado a la importación y la disputa de sus datos. RapidMiner ofrece un entorno incorporado para el aprendizaje de los datos, el estudio de la máquina, la ganancia de conocimiento profundo, la minería de texto y el análisis predictivo. Se utiliza para la empresa y los paquetes de negocios, además de para los estudios, la enseñanza, la escolarización, la creación de prototipos rápidos y el desarrollo de alerta, y apoya todos los pasos de la manera de aprendizaje de gadget, que consiste en la

instrucción de los hechos, la visualización de los resultados, la validación del modelo y la optimización. Cientos de sistema de adquisición de conocimientos, análisis de texto, algoritmos de modelado predictivo, la automatización y el proceso de manipular las características le ayudan a construir más alto modas más rápido que nunca. RapidMiner Studio (estadísticas: 10.000), RapidMiner Server (2 GB de RAM) y RapidMiner Radoop (restringido a una sola persona) están disponibles en la versión de inicio con obstáculos.

Figura 2.6

Interfaz de RapidMiner



Nota: Interfaz gráfica del área de trabajo de RapidMiner (RapidMiner, 2022)

2.7.4.1 Características

- La multitud de algoritmos de clasificación y regresión facilitan el aprendizaje supervisado.

- La amplia gama de algoritmos de agrupación, similitud y segmentación admiten el aprendizaje no supervisado.
- La integración perfecta de los scripts Ry Python en los flujos de trabajo proporciona una mayor extensibilidad.
- Capacidades de modelado y algoritmos de aprendizaje automático.

2.7.4.2 Beneficios

- Conéctese a cualquier fuente de datos, cualquier formato, a cualquier escala.
- Descubra rápidamente patrones o problemas de calidad de datos.
- Cree el conjunto de datos óptimo para el análisis predictivo.
- Limpie de manera experta los datos para algoritmos avanzados.

2.8 MÉTODO CIENTÍFICO

El método científico es una técnica que nos permite llegar a conocimientos que pueden considerarse válidos desde el punto de vista de la ciencia. (Collado & Lucio, 2014)

El método científico cumple dos características fundamentales:

- **Falsabilidad:** Las pautas legales o teorías adquiridas a partir de esta técnica pueden ser reevaluadas, es decir, es lejos una proposición que, con el paso de los años, posiblemente con más pruebas, puede observarse que es defectuosa.
- **Reproductividad:** Se puede replicar de nuevo, y a través de otra persona, obteniendo el mismo resultado. Piense en un experimento que cuando se repite en instancias únicas y por medio de investigadores únicos, si se completa de la misma manera, tiene que causar el mismo final.

El método científico consiste, en una manera de tecnificar una realidad, y es el resultado de un procedimiento imparcial de las creencias del investigador. Incluso a lo largo de los años, la información médica se perfecciona y se intenta descubrir cómo

funciona el ámbito, basándose siempre totalmente en la prueba y la observación (Collado & Lucio, 2014).

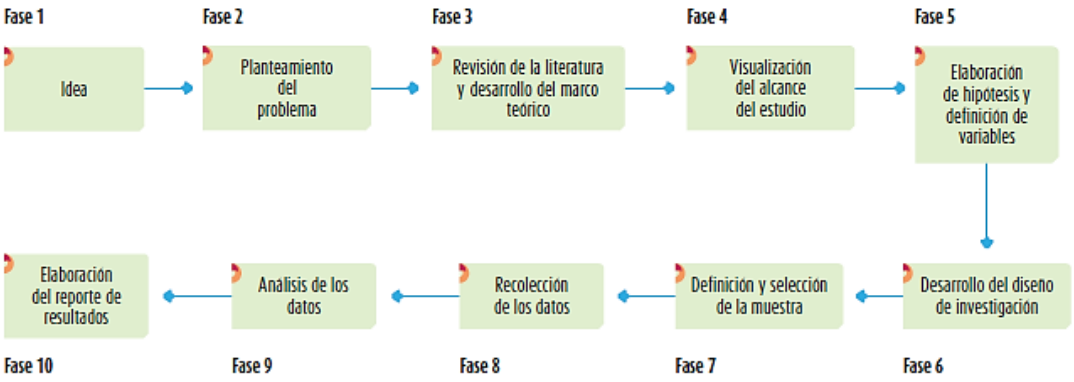
2.8.1 Características

Entre los rasgos que perfilan este enfoque, podemos destacar los siguientes:

- Es una metodología diseñada para obtener nuevos conocimientos.
- Consiste en el comentario sistemático, la medición, la experimentación y los componentes, el análisis y el cambio de hipótesis.
- Asimismo, las 2 características fundamentales de este método son la falibilidad y la reproducibilidad.
- En este sentido, la reproducibilidad porque se puede replicar en todo momento y por medio de algún otro personaje, obteniendo el resultado idéntico.
- Por otro lado, la falsabilidad debido a que las leyes o teorías recibidas desde este enfoque pueden ser reevaluadas.
- El enfoque científico reúne las prácticas conocidas por la comunidad clínica como legítimas para mostrar y confirmar nuevas teorías.
- Las políticas de la técnica científica minimizan, como vemos, el impacto de la subjetividad del científico en su mirada. De este modo, se refuerza la validez de los resultados y, por tanto, de la nueva comprensión.

Figura 2.7

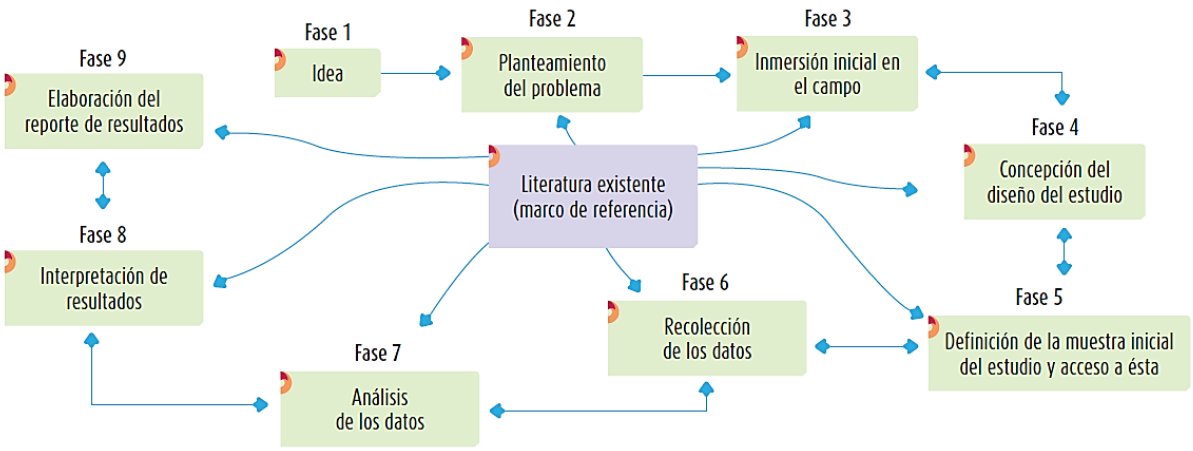
Fases del método científico



Nota: Pasos del proceso del método científico (Collado & Lucio, 2014)

Figura 2.8

Fases del método Cualitativo



Nota: Pasos del proceso del método científico cualitativo (Collado & Lucio, 2014)

Tabla 2.4

Cuadro comparativo del método cuantitativo y cualitativo

Características cuantitativas	Procesos fundamentales del proceso general de investigación	Características cualitativas
▪ Direcciona el proceso ▪ Justifica el planteamiento y la necesidad del estudio	Revisión de la literatura	▪ Contextualiza el proceso ▪ Justifica el planteamiento y la necesidad del estudio
▪ Generalmente predeterminadas, se prueban	Hipótesis	▪ Generalmente emergentes
▪ Preestablecidos, se implementan “al pie de la letra”	Diseños	▪ Emergentes, se implantan de acuerdo con el contexto y circunstancias
▪ El tamaño depende de qué tan grande sea la población (un número representativo de casos). Se determina a partir de fórmulas y estimaciones de probabilidad	Selección de la muestra	▪ El tamaño depende de que comprendamos el fenómeno bajo estudio (casos suficientes). La muestra se determina de acuerdo al contexto y necesidades
▪ Instrumentos predeterminados ▪ Antes de proceder al análisis se recaban todos los datos	Recolección de los datos	▪ Los instrumentos se van afinando ▪ Los datos emergen paulatinamente

▪ Los datos encajan en categorías predeterminadas ▪ Análisis estadístico ▪ Descripción de tendencias, contraste de grupos o relación entre variables ▪ Comparación de resultados con predicciones y estudios previos	Análisis de los datos	▪ Los datos generan categorías ▪ Análisis temático ▪ Descripción, análisis y desarrollo de temas ▪ Significado profundo de los resultados
▪ Distribuciones de variables, coeficientes, tablas y figuras que relacionan variables, así como modelos matemáticos y estadísticos	Presentación de resultados	▪ Categorías, temas y patrones; tablas y figuras que asocian categorías, materiales simbólicos y modelos
▪ Estándar ▪ Objetivo y sin tendencias	Reporte de resultados	▪ Emergente y flexible ▪ Reflexivo y con aceptación de tendencias

Nota: Cuadro comparativo del método científico en su enfoque cualitativo y cuantitativo (Collado & Lucio, 2014)

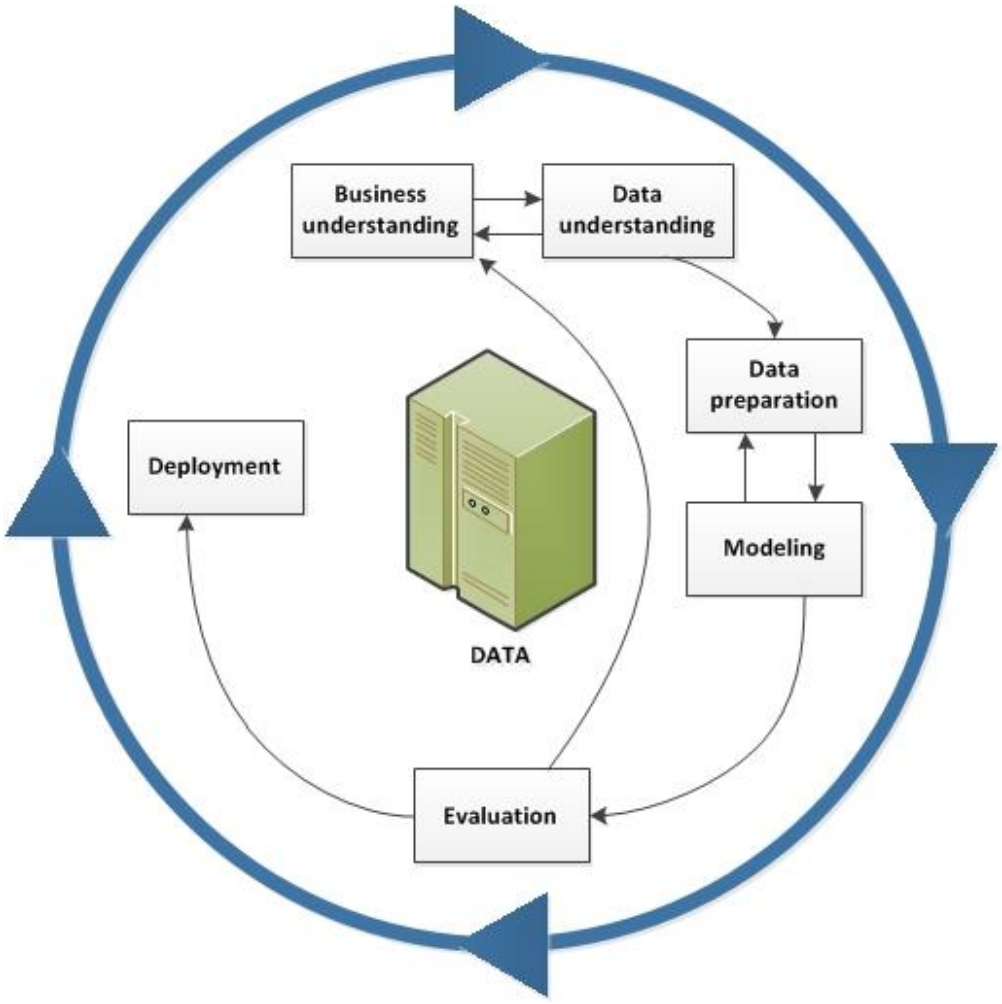
2.9 METODOLOGÍA CRISP-DM

En la página web IBM, (2012) describe que la metodología CRISP-DM, que significa Cross-Industry Standard Process for Data Mining, es un método probado para orientar sus trabajos de Minería de Datos.

- Como metodología, incluye descripciones de las fases normales de un proyecto, las tareas necesarias en cada fase y una explicación de las relaciones entre las tareas.
- Como modelo de proceso, CRISP-DM ofrece un resumen del ciclo vital de Minería de Datos.

Figura 2.9

Ciclo de vida de la Minería de Datos



Nota: Proceso del ciclo de vida de la Minería de Datos (IBM Docs, 2021)

El ciclo vital del modelo contiene seis fases con flechas que indican las dependencias más importantes y frecuentes entre fases. La secuencia de las fases no es estricta. De hecho, la mayoría de los proyectos avanzan y retroceden entre fases si es necesario.

El modelo de CRISP-DM es flexible y se pueden personalizar fácilmente. Por ejemplo, si su organización intenta detectar actividades de blanqueo de dinero, es probable que necesite realizar una criba de grandes cantidades de datos sin un objetivo de modelado específico. En lugar de realizar el modelado, su trabajo se centrará en explorar y visualizar datos para descubrir patrones sospechosos en datos

financieros. CRISP-DM permite crear un modelo de Minería de Datos que se adapte a sus necesidades concretas. (IBM, 2021)

En tal situación, las fases de modelado, evaluación y despliegue pueden ser menos relevantes que las fases de preparación y comprensión de datos. Sin embargo, es muy importante considerar algunas cuestiones que surgen durante fases posteriores para la planificación a largo plazo y objetivos futuros de Minería de Datos. (IBM, 2021)

2.9.1 Fases de la metodología CRISP-DM

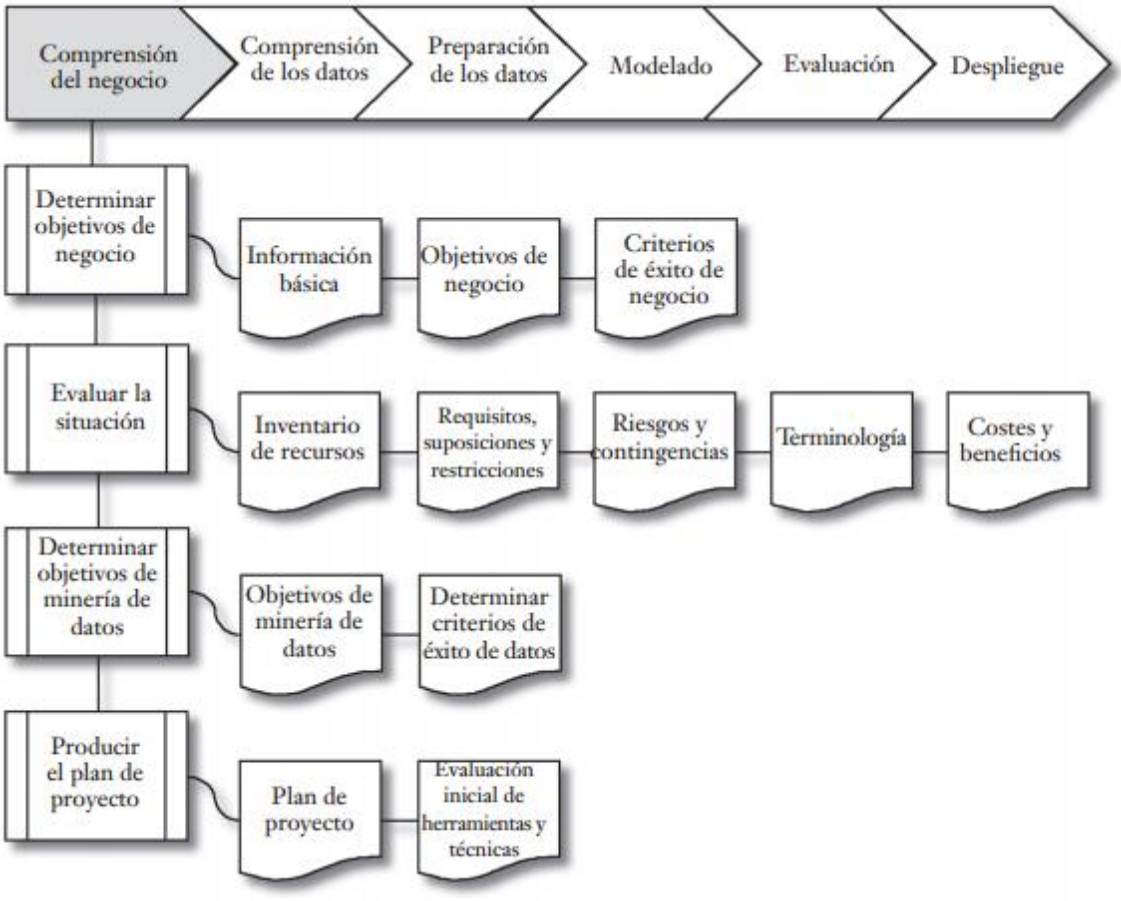
2.9.1.1 Comprensión del negocio

"Esta fase inicial se enfoca en la comprensión de los objetivos de proyecto. Después se convierte este conocimiento de los datos en la definición de un problema de minería. De datos y en un plan preliminar diseñado para alcanzar los objetivos." (BizMetriks, 2013)

- Establecimiento de los objetivos del negocio (Contexto inicial, objetivos, criterios de éxito).
- Evaluación de la situación (Inventario de recursos, requerimientos, supuestos, terminologías propias del negocio).
- Establecimiento de los objetivos de la Minería de Datos (objetivos y criterios de éxito).
- Generación del plan del proyecto (plan, herramientas, equipo y técnicas).

Figura 2.10

Fase comprensión del negocio



Nota: Etapas de la fase de Comprensión del negocio de la Metodología CRIPS-DM, (BizMetriks, 2013)

2.9.1.2 Comprensión de los Datos

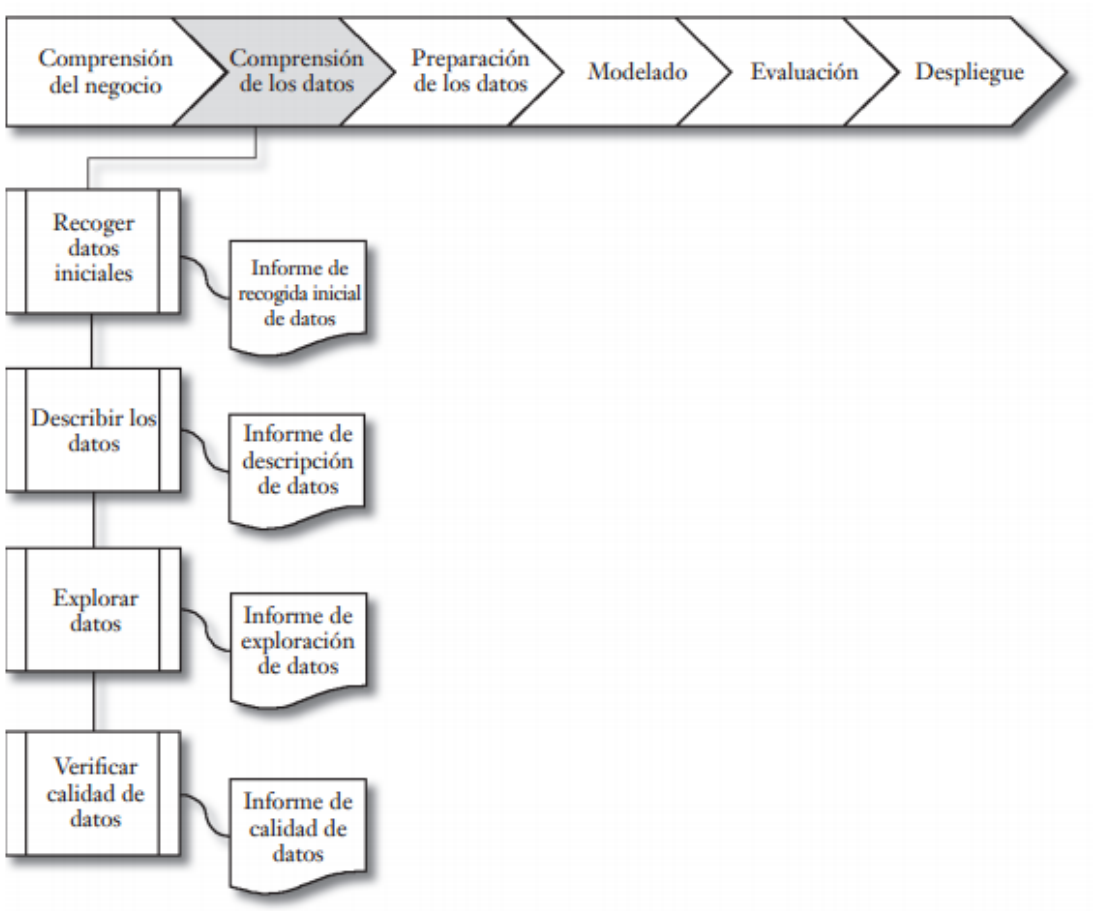
La fase de comprensión de los datos inicia con la recolección de datos y prosigue con diferentes actividades que permitan familiarizarse con los datos, la identificación de los problemas de la calidad de los datos, realizar el descubrimiento de los primeros conocimientos de los datos y/o detectar los subconjuntos interesantes para la elaboración de hipótesis en relación a la información oculta (BizMetriks, 2013).

- Recopilación inicial de datos
- Descripción de los datos

- Exploración de los datos
- Verificación de calidad de datos

Figura 2.11

Fase comprensión de datos



Nota: : Etapas de la fase de comprensión de datos de la metodología CRIPS-DM, (BizMetriks, 2013)

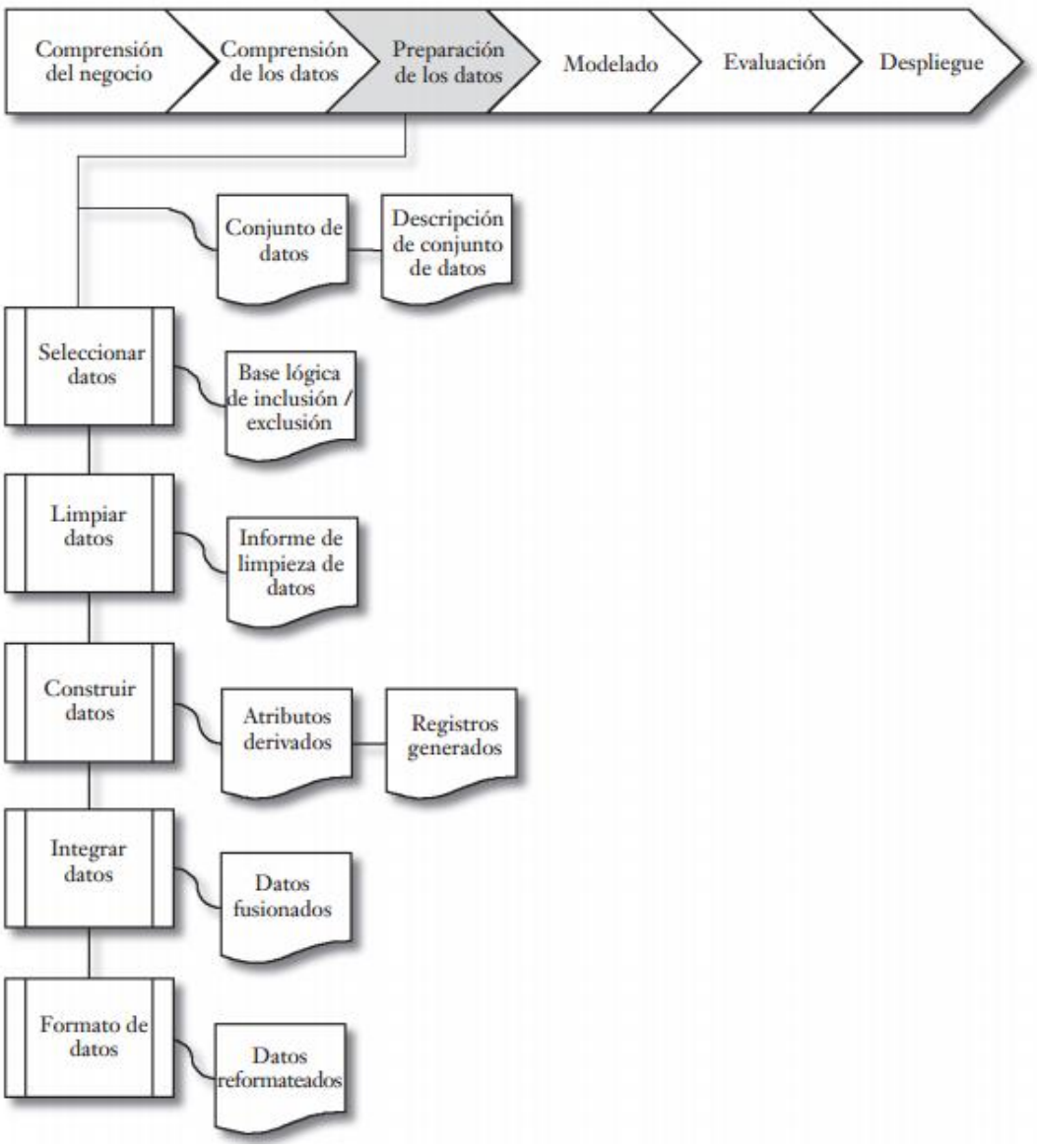
2.9.1.3 Preparación de los Datos

La fase de preparación de datos cubre todas las actividades necesarias para la construcción de conjuntos de datos finales (datos que serán introducidas en las herramientas de modelado] desde los datos en bruto iniciales. Las tareas de la preparación de datos posiblemente se realizarán varias veces y sin seguir ningún orden prescrito. Las tareas incluyen a la selección de tablas, registros y atributos, así como la transformación y limpieza de datos para las herramientas de modelado (BizMetriks, 2013).

- Selección de los datos
- Limpieza de datos
- Construcción de datos
- Integración de datos
- Formateo de datos

Figura 2.12

Fase Preparación de los Datos



Nota: : Etapas de la fase de preparación de datos de la metodología CRIPS-DM, (BizMetrika, 2013)

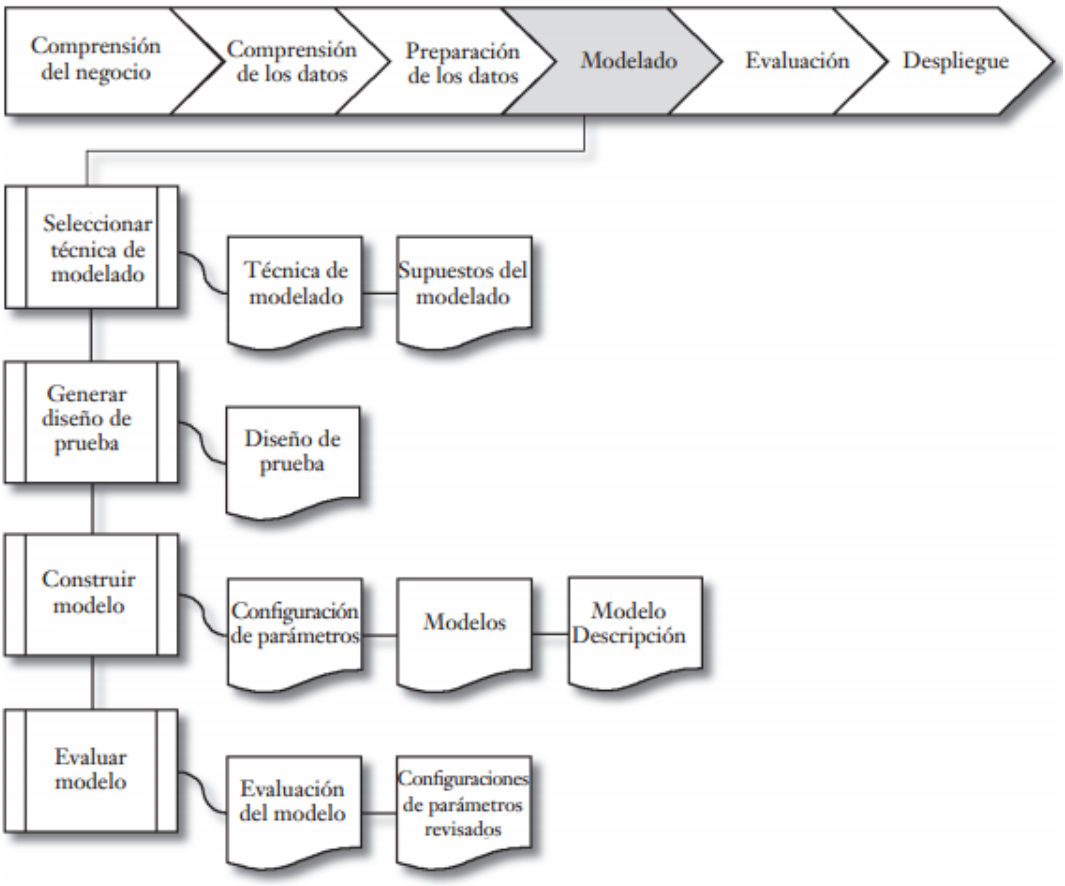
2.9.1.4 Obtención de los Modelos

En esta fase se realiza la selección y aplicación de varias técnicas de modelado, y sus diferentes parámetros se calibran en sus valores óptimos. Por lo general existen varias técnicas para el mismo tipo de problema de Minería de Datos. Algunas técnicas tienen requisitos precisos sobre el formato de los datos. Por lo tanto, también es esencial para regresar a la fase de preparación de datos. (BizMetriks, 2013)

- Selección de la/s técnicas de modelado
- Diseño de la evaluación
- Construcción del modelo
- Evaluación del modelo

Figura 2.13

Fase de modelado



Nota: metodología CRIPS-DM, (BizMetriks, 2013)

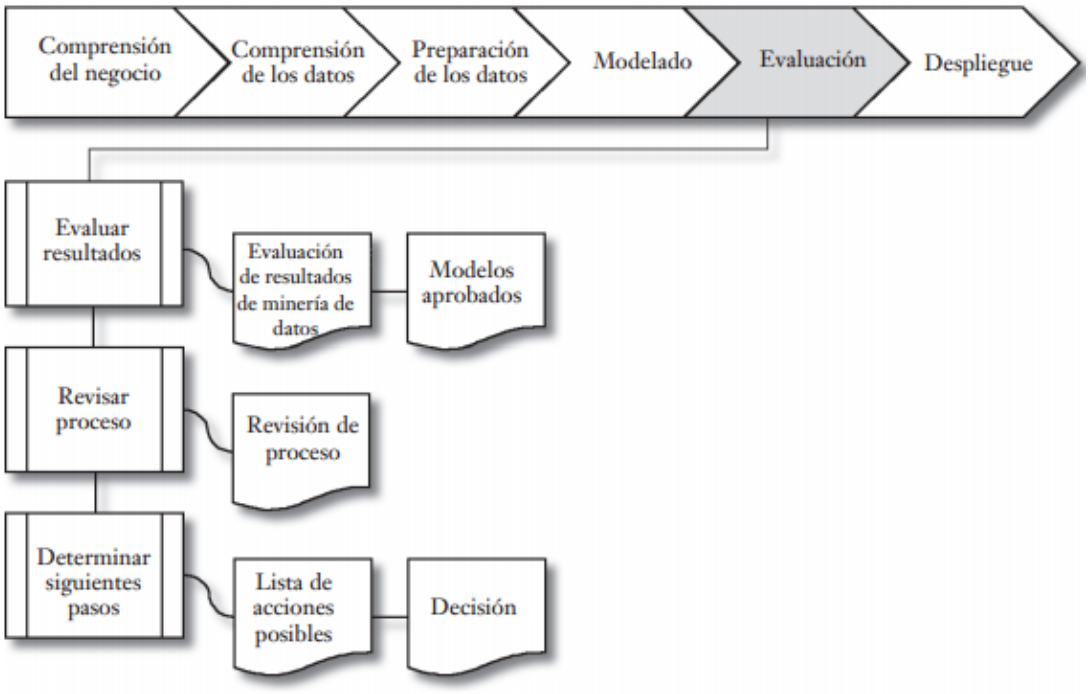
2.9.1.5 Evaluación de los Modelos

En esta etapa del proyecto se ha creado un modelo (o modelos) que parecen tener una alta calidad desde un ángulo del análisis de datos. Antes de proceder al desarrollo final del modelo, es importante evaluar el modelo y revisar los pasos realizados para crearlo, con la finalidad de asegurar que el modelo logra de forma correcta los objetivos de negocio. Un objetivo clave es la determinación de la existencia de algún problema de negocio importante que no se haya considerado suficientemente. Al finalizar esta fase se debe llegar a una decisión sobre el uso de los resultados de la Minería de Datos (BizMetriks, 2013).

- Evaluación de resultados
- Revisión el proceso
- Establecimiento de los siguientes pasos o acciones

Figura 2.14

Fase de evaluación



Nota: : Etapas de la fase Evaluación de la metodología CRIPS-DM, (BizMetriks, 2013)

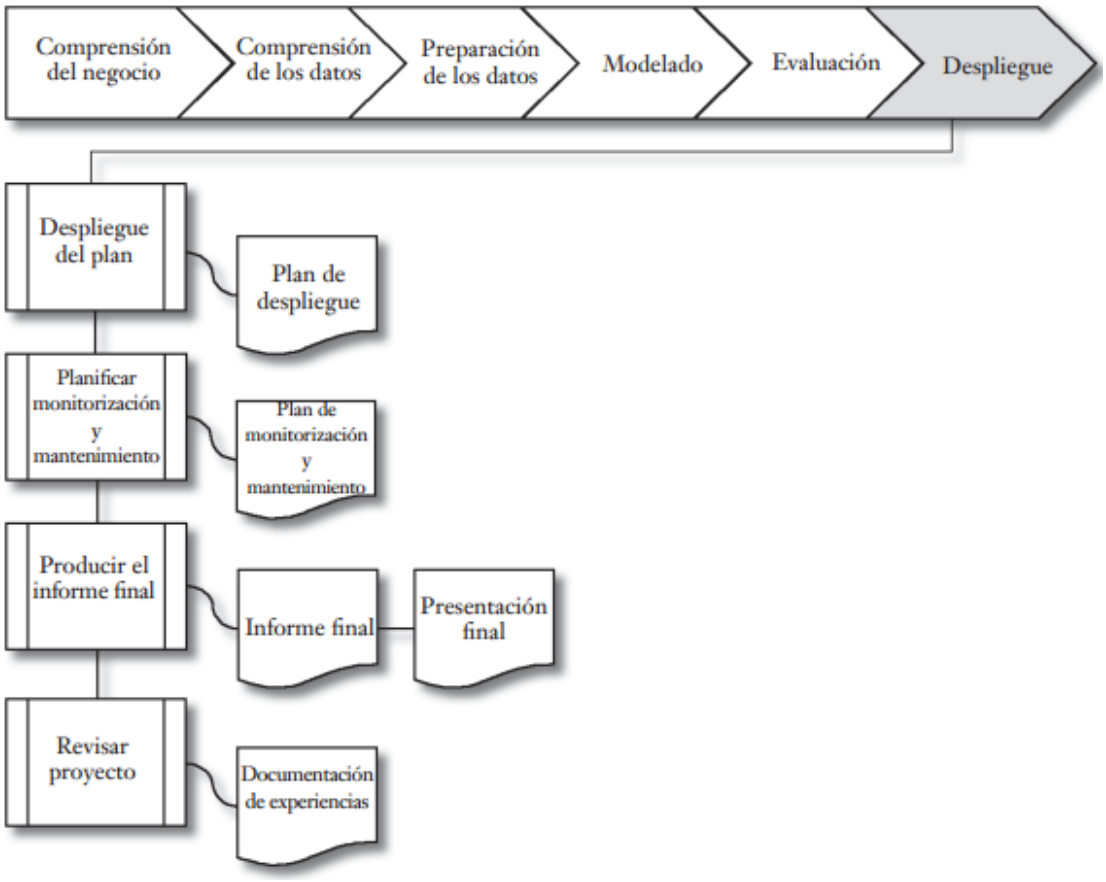
2.9.1.6 Implementación

La creación del modelo por lo general no es el final del proyecto. Aunque el objetivo del modelo sea aumentar el conocimiento que se tiene de los datos, el conocimiento adquirido deberá organizarse y presentarse de una forma que el cliente pueda utilizarlo. A menudo implica el uso de modelos “vivos" en los procesos de decisión de una organización – por ejemplo, la personalización en tiempo real de páginas Web o la puntuación repetida de bases de datos de marketing. dependiendo de los requisitos, la fase del despliegue puede ser tan fácil como generar un informe o tan compleja como la implementación de un proceso repetible de Minería de Datos en toda la empresa. En muchos casos, es el cliente, no el analista de datos, quién lleva a cabo los pasos del despliegue. No obstante, aunque el analista realice las acciones de despliegue, es importante que el cliente sepa anticipadamente las acciones que deben llevarse a cabo para poder hacer un uso real de los modelos creados (BizMetriks, 2013).

- Planificación de la implementación
- Planificación del monitoreo y mantenimiento
- Generación de informe final
- Revisión del proyecto

Figura 2.15

Fase de despliegue



Nota: Etapas de la fase de despliegue de la metodología CRIPS-DM, (BizMetriks, 2013)

2.10 INGENIERÍA DE SOFTWARE

Pressman, (2016) define al software informático como el producto que los programadores profesionales construyen y luego mantienen durante un largo periodo de tiempo. Incluye programas que se ejecutan en un ordenador de cualquier tamaño y arquitectura, contenidos que se presentan a medida que los programas informáticos se ejecutan, e información descriptiva en formatos impresos y virtuales que abarcan prácticamente cualquier medio electrónico. La ingeniería del software consiste en un proceso, un conjunto de métodos (prácticas) y un conjunto de herramientas que permiten a los profesionales producir software informático de alta calidad.

¿Quién realiza la ingeniería de software?

Los ingenieros de software desarrollan y mantienen programas informáticos, y prácticamente todo el mundo industrializado los utiliza, ya sea directa o indirectamente.

¿Por qué es importante?

El software es importante porque afecta a casi todos los aspectos de nuestras vidas y ha invadido nuestro comercio, nuestra cultura y nuestras actividades diarias. La ingeniería del software es importante porque nos permite construir sistemas complejos en un tiempo razonable y con alta calidad.

¿Cuáles son los pasos?

Los programas informáticos se construyen de la misma manera que cualquier producto de éxito, con la aplicación de un proceso ágil y adaptativo para obtener un resultado de alta calidad que satisfaga las necesidades de las personas que van a utilizar el producto. En estos pasos se aplica el enfoque de la ingeniería del software.

¿Cuál es el producto final?

Desde el punto de vista de un ingeniero de software, el producto final es el conjunto de programas, contenidos (datos) y otros productos acabados que conforman el software informático. Pero desde la perspectiva del usuario, el producto final es la información resultante que, de alguna manera, mejora el mundo en el que vive.

2.11 METODOLOGÍA OPEN UP

OpenUP (Proceso Unificado Abierto) se trata de un procedimiento versionable y extensible, orientado a la gestión y mejora de proyectos de software basado principalmente en el desarrollo iterativo, ágil e incremental, apropiado para proyectos pequeños y de bajos recursos; y es aplicable a un amplio conjunto de estructuras y paquetes de desarrollo.

Sin embargo, OpenUP es completo en el sentido de que manifiesta completamente el método de construcción de una máquina. Para hacer frente a los deseos que no están cubiertos en su material de contenido OpenUP es extensible para

su uso como una base sobre la que añadir o adaptar a otro material de proceso como se desee. (Medina, 2014)

2.11.1 Proceso iterativo

- Mínimo: Sólo incluye el contenido técnico fundamental.
- Completo: Puede manifestarse como una técnica completa para construir un aparato.
- Extensible: Puede utilizarse como base para añadir o adaptar mayores estrategias.

2.11.2 Características de OpenUP

- Desarrollo incremental
- Uso de instancias de uso y escenarios.
- Control de riesgos.
- Diseño basado principalmente en la arquitectura.

2.11.3 Principios de OpenUP

- Colaborar para sincronizar los pasatiempos y proporcional el conocimiento. Este precepto promueve las prácticas que fomentan un entorno de equipo saludable, facilitan la colaboración y aumentan la información compartida del proyecto.
- Equilibrar las prioridades para maximizar la ventaja obtenida mediante el uso de las partes interesadas en la asignación. Este precepto promueve prácticas que permiten a las partes interesadas del proyecto ampliar una respuesta que maximiza los beneficios obtenidos con la ayuda de las partes interesadas y cumple con los requisitos y las limitaciones del desafío.
- Centrarse en la estructura desde el principio para reducir el riesgo y preparar el desarrollo. (Medina, 2014)

- Mejora evolutiva para las observaciones y la mejora continua. Este principio promueve las prácticas que permiten a los grupos de desarrollo obtener observaciones tempranas y continuas de las partes interesadas en la asignación, lo que les permite demostrar a los clientes el aumento progresivo de la capacidad.

2.11.4 Ciclo de vida

2.11.4.1 Iteración de la fase inicial.

En este segmento se tienen en cuenta las necesidades de todos los participantes en el proyecto y se traducen en objetivos de asignación. Se describen el alcance, los límites, los criterios de atracción, los casos de uso vitales, una estimación inicial de los honorarios y un esquema de planificación para la empresa.

Objetivos:

- Entender qué hay que construir.
- Identificar la funcionalidad clave.
- Determinar como mínimo una solución viable.
- Comprender los honorarios, el calendario y los riesgos de la misión.

2.11.4.2 Iteración de la fase de elaboración.

En esta sección se terminan las obligaciones de evaluación del área y definición de la arquitectura del dispositivo. Se debe elaborar un plan de tareas, organizando las necesidades fuertes y la arquitectura. Al final del segmento debe haber una definición limpia y única de los casos de uso, los actores, la arquitectura del sistema y un prototipo ejecutable.

Objetivos:

- Obtener una comprensión más definida de las necesidades.
- Diseñar, implementar y validar la línea de base arquitectónica.
- Mitigar los riesgos y obtener estimaciones de precios y calendarios más precisos.

2.11.4.3 Iteración de la fase de construcción.

En esta fase se descubren, prueban e integran todos los aditivos y funcionalidades de la máquina que se van a aplicar. Las consecuencias adquiridas dentro de la forma de incrementos ejecutables deben avanzar lo más rápido posible sin descuidar la excelencia de lo que se ha evolucionado.

Objetivos.

- Aumentar de forma iterativa un producto completo que pueda ser transferido a la comunidad de usuarios.
- Minimizar los costes de mejora y alcanzar un cierto nivel de paralelismo.

2.11.4.4 Iteración de la fase de transición.

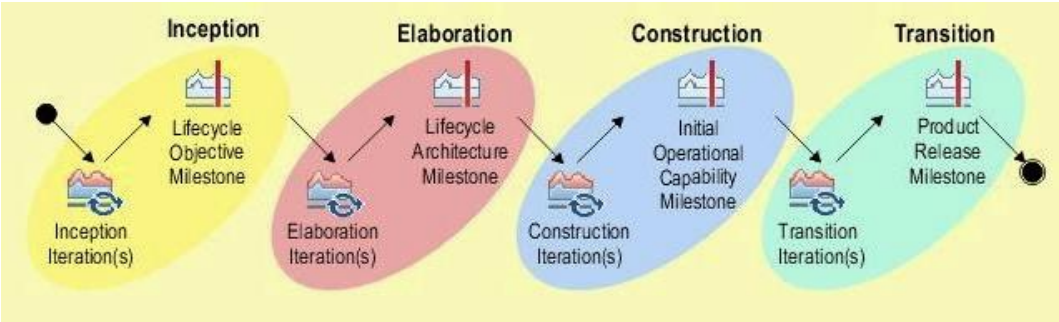
Esta sección corresponde a la creación del producto a la comunidad de consumidores, cuando el producto está adecuadamente maduro. La fase de transición se compone de los subniveles de pruebas beta, pilotaje y formación presencial de los mantenedores del gadget. Dependiendo de la respuesta obtenida con la ayuda de los clientes puede ser importante realizar modificaciones dentro de las últimas entregas o poner en marcha algunas funcionalidades más solicitadas por la mayoría de las personas.

Objetivos.

- Realizar pruebas beta para decidir si se han cumplido las expectativas de los consumidores.
- Llegar a un acuerdo con las partes interesadas de que el producto está terminado.
- Mejorar el rendimiento general futuro mediante el análisis retrospectivo de la empresa.

Figura 2.16

Fases de la metodología de Open Up



Nota: Recopilado de la pagina web metodologia OpenUp por Medina (2014)

2.11.5 Beneficios de utilizar OpenUP

- Al ser adecuado para tareas pequeñas y de uso ocasional de recursos, disminuye la oportunidad de fracaso en las tareas pequeñas y aumentará la probabilidad de logro.
- Permite descubrir errores tempranos mediante un ciclo iterativo.
- Evita la elaboración de documentación, diagramas e iteraciones inútiles que requiere la técnica RUP.
- Al ser un método ágil, tiene un método centrado en el cliente e iteraciones cortas.

Ventajas

- Es un método ágil
- Se puede adaptar a diferentes procesos.

Desventajas

- A veces omite contenidos que pueden ser de interés dentro del emprendimiento.
- Se prevé que cubra una gran variedad de necesidades de las iniciativas de desarrollo en un periodo de tiempo muy rápido.
- Al ser un método de bajo formalismo, existe la posibilidad, si no se tiene cuidado, de que la misión pierda el rumbo debido a la desorganización. (Medina, 2014)

2.12 CONTAMINACIÓN DE RESIDUOS SÓLIDOS

Uno de los principales problemas que enfrenta el medio ambiente y la sociedad humana es la contaminación causada por la basura. Solo aquellos residuos sólidos que se depositan en los vertederos, para bien o para mal, suelen identificarse como basura. Sin embargo, la basura es un concepto más amplio que tiene graves consecuencias para el medio ambiente y la salud de los seres humanos, los animales y las plantas (Arriol, 2019).

2.12.1 Causas

La principal causa de la contaminación por residuos es la ineficiente o inexistente gestión de residuos. Hay que tener en cuenta que no es el material en sí mismo el que genera los residuos, sino la forma en que se gestiona o no. De esta forma, si tomamos como ejemplo un papel, puede ser basura o puede ser materia prima, dependiendo de cómo se gestione, cuando queda inservible y oficialmente se convierte en residuo. De esta forma, si el papel se deposita en el medio ambiente, se convierte en basura, contaminando el medio ambiente en tanto se descompone. Por el contrario, si el mismo papel se almacena en un contenedor de reciclaje y se gestiona adecuadamente, no se convierte en basura, se convierte en materia prima. En otras palabras, no es la naturaleza de los materiales lo que determina si los residuos se convierten en basura, sino cómo se gestionan.

Nuevamente, entre estas razones, también cabe mencionar que el consumismo actual juega un papel protagónico. No porque el consumismo signifique necesariamente basura, sino porque cuanto mayor sea el consumismo, más basura, y cuanta más basura, más probable es que se gestione mal. En otras palabras, el consumismo ayuda a que la gestión de los residuos sea inadecuada y, por tanto, pueda considerarse una causa indirecta de la contaminación por basura. Sin embargo, es necesario aclarar que es la gestión de los residuos (buenos o malos) la que genera residuos. (Arriol, 2019)

2.12.2 Efectos

La principal consecuencia de la contaminación por basura implica la degradación de la salud biológica. Recuerde que la basura libera sustancias tóxicas al medio ambiente y se esparce por el suelo, el agua y el aire. Cuando estas sustancias tóxicas entran en contacto con los seres vivos, ya sean personas, animales o plantas, pueden afectar negativamente a su salud.

Además, hay que recordar que, desde un punto de vista estético, la contaminación por basura tiene un impacto negativo muy importante en el medio ambiente, ya que destruye el paisaje (tanto natural como urbano), lo que también se considera uno de los principales problemas causados por la basura, hoy contaminar. (Arriol, 2019)

2.13 CONTAMINACIÓN DE BASURA EN LA CIUDAD DE EL ALTO

En la Ciudad de El Alto existen altos niveles de contaminación ambiental causado por la basura que la población bota a la calle. Cada día se produce un promedio de 450 toneladas de residuos sólidos, según la dirección de Medio Ambiente de la Alcaldía. Este hecho, dicen las autoridades, se constituye en el principal factor de deterioro medioambiental. (eabolivia, 2018)

El titular de la Dirección de Medio Ambiente de la alcaldía alteña, Eduardo Quille, explicó que lamentablemente algunos sectores de la población no tienen educación Ciudadana y echan sus residuos en las calles, plazas y avenidas sin tener en cuenta que este hecho contamina tanto la tierra como el aire. (eabolivia, 2018)

Quille precisó que alrededor del 45 por ciento de la basura que se produce en El Alto es recogida de las calles debido a vecinos que no respetan los horarios en los que pasan los carros basureros y “sacan su basura a cualquier hora, pero también los transeúntes botan los residuos en las calles y no utilizan los papeleros que se instalan en lugares como la Ceja", explicó.

“La basura es lo que más preocupa a las autoridades municipales y, por ello, hemos implementado programas de concientización con las organizaciones sociales; sabemos que es un trabajo a largo plazo, pero queremos que la población esté

consciente de que si bota basura en la calle está aportando para que la vida en este planeta se acabe pronto", advirtió la autoridad. (eabolivia, 2018).

Como parte de las medidas preventivas y de educación Ciudadana, la Alcaldía declaró La Semana del Medioambiente, con la realización de varias actividades, como la feria educativa que se realizó ayer en inmediaciones de la Alcaldía Quemada, la cual se repetirá esta semana en varios sectores de la Ciudad.

Quille añadió que la Dirección de Medio Ambiente desarrolla proyectos de reciclaje para enseñar a la población a aprovechar desechos que pueden ser reutilizados y disminuir los residuos que no son asimilables por los suelos. (eabolivia, 2018)

2.13.1 Factores de contaminación

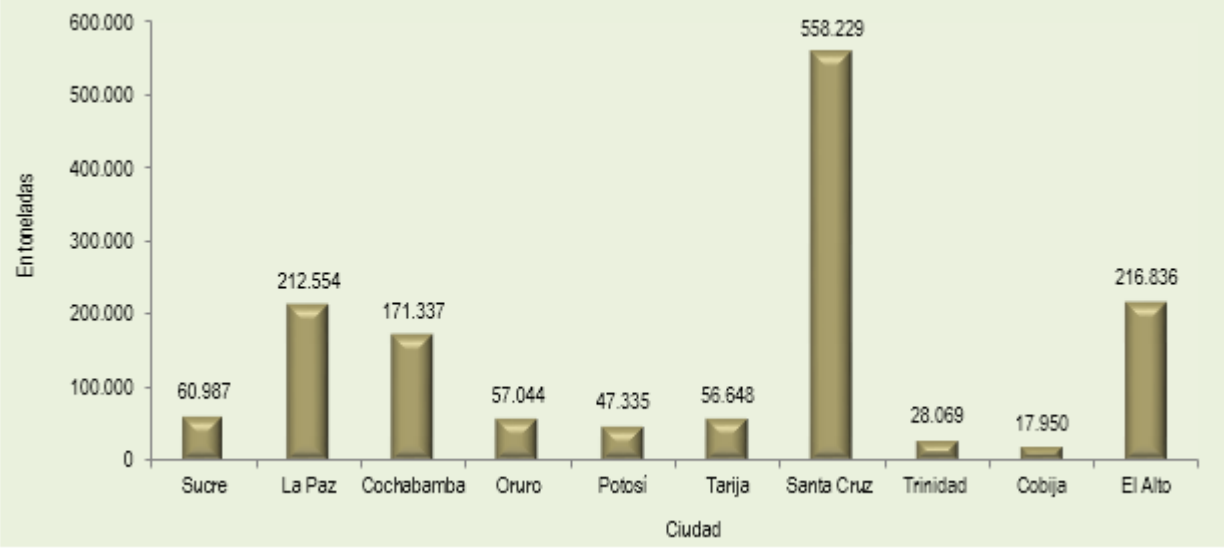
Juana Acarapi, responsable del proyecto Programa Urbano, impulsado por Red Hábitat, explicó que además de la basura existen otros cinco factores de contaminación en la urbe alteña, que son preocupantes y sobre los cuales esa organización está trabajando.

"Podemos identificar todos estos problemas en La Ceja, tenemos la contaminación acústica, por la música con alto volumen de los negocios y bares; la contaminación visual, en la alta cantidad de afiches, carteles y básicamente el cableado eléctrico que en la Ceja es extremo", señaló.

La profesional también citó la falta de espacios y áreas verdes, la contaminación del aire por la emanación de vapores contaminantes producidos por orines, heces fecales y desechos, además del monóxido de carbono.

Figura 2.17

Recolección de Basura 2006 – 2016



Nota: Residuos sólidos en Ciudades capitales y El Alto (INE, 2017)

2.13.2 Datos sobre basura

Cantidad: La urbe alteña produce un promedio de 450 toneladas de basura. Esta cantidad puede subir a 900 toneladas.

Porcentajes: El 60 por ciento de la basura corresponde a residuos sólidos.

Botadero: El botadero de basura de Villa Ingenio está a punto de cumplir su vida útil

(Gabriela, Cambio)

2.14 HERRAMIENTAS

2.14.1 Python

Se trata de un lenguaje de programación, multiparadigma y multinivel, con asistencia para la programación orientada a objetos, imperativa y propositiva. Con este tipo de lenguaje se pueden crear aplicaciones nativas e híbridas, y tiene una sintaxis al alcance de personas con un nivel básico de "alfabetización" en lenguajes de programación.

Según una encuesta realizada a través de programadores en la comunidad del portal web Stack Overflow, el setenta y tres por ciento de los constructores consideran que Python es el lenguaje más requerido sobre las opciones que actualmente se pueden encontrar en el mercado. (Caminiti, 2021).

Se trata de un lenguaje de programación de código abierto y de razón estándar, y es muy libre, por lo que no es necesario pagar una licencia para aplicarlo. Es interpretado, es decir, no está compilado, lo que significa un mayor tiempo de ejecución en comparación con los paquetes desarrollados con lenguajes compilados.

Sin embargo, algunos especialistas aseguran que el problema de la tasa no es un inconveniente porque la banda diferencial es mínima y las iniciativas de desarrollo de programas de software que se pueden llevar a cabo actualmente están orientadas a la nube, lo que le da una capacidad de computación desembolsada de primer nivel a un costo menor. (Caminiti, 2021)

¿Para qué sirve Python?

Python es una era vital en el disfrute y los sistemas de medios sociales, orientado a la obtención de dispositivos y el desarrollo de algoritmos de asesoramiento, por lo que las aplicaciones que consisten en Instagram, Pinterest, Dropbox, Facebook, Spotify y Netflix tienen este lenguaje en su mejora, ya que permite las tareas de programación con el objetivo de procesar grandes cantidades de registros y la obtención de datos atesorados. (Caminiti, 2021)

Además, es viable abordar los siguientes tipos de iniciativas con el lenguaje:

- Aplicaciones web.
- Tecnología de datos.
- Dominio automático.
- Evaluación y automatización de datos.
- Inteligencia artificial.

Características de Python

- Es un buen lenguaje para las personas que desean iniciarse en el mundo de la programación, especialmente por sus múltiples campos de aplicación.

- Frameworks y entornos incluidos para la mejora ágil y ecológica de las aplicaciones de Internet.
- Es uno de los lenguajes de programación más utilizados en el ámbito educativo y científico.
- Es interpretado y no compilado, siendo la depuración (debugging) más rápida.
- Se puede utilizar programación orientada a objetos, establecida o útil.
- Aplica el código de suministro, permitiendo la creación de grandes aplicaciones.
- Proporciona sistemas de registros dinámicos.
- Tiene una implementación de recolección de basura automática para un mejor control de la reminiscencia.
- Se puede integrar con los lenguajes C, C++, COM, ActiveX, CORBA y Java.

8 motivos para examinar Python

- Es el lenguaje apropiado para empezar a programar. Con una combinación de motivación, ganas de estudiar e información técnica se puede ser capaz de captar esta herramienta y por ello ser capaz de ampliar tus propias aplicaciones sin ninguna molestia.
- Destaca la claridad y nitidez del código porque es sencillo escribir instrucciones positivas o realizar estrategias específicas con el lenguaje.
- Es multiplataforma, esto significa que usted será capaz de crear programas dentro del sistema de ejecución que tiene que pinturas con: Windows, Linux, IOS, siendo muy portable.
- Dado que Python funciona a través de un intérprete, lo convierte en un lenguaje fácilmente portable, por lo que serás capaz de ejecutar programas en distintas plataformas.
- Para aplicar en Python no es obligatorio utilizar un IDE (Entorno de Desarrollo Integrado) o un editor de código, incluso es posible programar el uso del bloc de notas del PC; es decir, está en la posición número 2 dentro

del índice TIOBE 2021 en Minería de Datos e IA. Esta empresa se encarga de estudiar millones de cepas de código para determinar los lenguajes más utilizados a nivel internacional.

- Si estás interesado en hacer crecer interfaces gráficas de manera sencilla, Python es también la alternativa recomendada porque tiene una librería predefinida para permitirte crear botones, controles de lista, contenedores, tablas y todo lo que quieras para tener una UI dinámica.
- Cuenta con el framework Django que es específico y realmente útil para diseñar y desarrollar paquetes de red el uso de este lenguaje de programación. Este framework permite simplificar la mejora de funcionalidades que incluyen: login, control de usuarios, permanencia y protección de datos, además de la creación de módulos para crear, examinar, reemplazar y eliminar datos rápidamente con scaffolding.
- Python es el más elegido para la ciencia de los registros, en particular el Aprendizaje Automático, un campo que abarca el auto-estudio de las computadoras, es decir, esta sub-rama de la computación busca que los gadgets o sistemas informáticos sistematizados puedan analizar y generar nuevos conocimientos a través de la interacción con otras estructuras.

¿Dónde podemos encontrar Python?

Al ser un lenguaje multiplataforma, es viable desarrollar paquetes en sistemas operativos exclusivos. La sencillez y la potencia del lenguaje para controlar arquitecturas y tecnologías excepcionales, unidas a su rendimiento en el procesamiento de hechos, hacen de Python un lenguaje popular para las empresas de todo el mundo. A continuación, podemos especificar sus campos de utilidad predominantes:

- Data Analytics y Big records.
- Minería de Datos
- Ciencia de datos
- Inteligencia Artificial (IA)

- Blockchain
- Dominio de la máquina
- Juegos e imágenes en 3D

2.14.2 Java

Java es un lenguaje de programación muy utilizado para codificar programas en red. Ha sido una opción popular entre los desarrolladores durante más de un tiempo, con miles y miles de paquetes Java en uso hoy en día. Java es un lenguaje multiplataforma, orientado a objetos y centrado en la web, que puede utilizarse como plataforma por derecho propio. Es un lenguaje de programación rápido, fácil y fiable para codificar todo, desde programas para móviles y software de organización hasta grandes programas de información y tecnología de servidores. (aws, 2022)

a) ¿Para qué se utiliza el lenguaje de programación Java?

Dado que Java es un lenguaje flexible y de uso libre, crea programas de software localizados y dispensados. Algunos usos comunes de Java abarcan:

1. **Desarrollo de juegos en línea**

   Muchos videojuegos, además de los juegos para móviles y portátiles, se crean con Java. Incluso los videojuegos actuales que integran tecnología avanzada, junto con el conocimiento de la máquina o la verdad digital, se crean con la generación de Java.

2. **Computación en la nube**

   Java es frecuentemente conocido como WORA (Write Once and Run Anywhere), lo que lo hace ideal para aplicaciones descentralizadas basadas en la nube. Los proveedores de la nube eligen el lenguaje Java para ejecutar programas en una amplia gama de sistemas subyacentes.

3. **Información macro**

   Java se utiliza para motores de procesamiento de registros que pueden trabajar con unidades de registros complejas y grandes cantidades de estadísticas en tiempo real.

4. **Inteligencia artificial**

   Java es una fuente inagotable de bibliotecas de aprendizaje automático. Su equilibrio y velocidad lo hacen ideal para el desarrollo de programas de inteligencia artificial que incluyen el procesamiento del lenguaje natural y el estudio profundo.

5. **Internet de las cosas**

   Java se ha utilizado para programar sensores y hardware en dispositivos periféricos que pueden conectarse independientemente a Internet.

b) ¿Por qué es Java una preferencia tan famosa entre los desarrolladores de programas de software hoy en día?

Java es famoso porque está diseñado para ser limpio de aplicar. Algunos motivos por los que los desarrolladores eligen Java en lugar de otros lenguajes de programación son:

- **Fuentes de dominio de alta calidad.**

  Java ha existido durante mucho tiempo, por lo que hay un montón de recursos de aprendizaje disponibles para los nuevos programadores. La documentación detallada, los libros completos y los cursos ayudan a los desarrolladores a lo largo de la curva de adquisición de conocimientos. Además, los novatos pueden empezar a escribir código en Core Java antes de pasar a Advanced Java.

- **Funciones y bibliotecas incorporadas**

  Al utilizar Java, los desarrolladores no quieren escribir cada nueva función desde cero. En su lugar, Java proporciona una rica atmósfera de funciones incorporadas y bibliotecas para el crecimiento de una expansión de paquetes.

- **Guía de red activa**

  Java tiene muchos usuarios enérgicos y una comunidad que puede ayudar a los constructores cuando se enfrentan a situaciones exigentes de

codificación. Asimismo, el software de la plataforma Java se mantiene regularmente y se actualiza.

- **Equipo de mejora altamente excepcional**

  Java ofrece numerosos equipos para ayudar a la modificación automatizada, la depuración, las pruebas, el despliegue y la gestión alternativa. Estas herramientas hacen que la programación en Java sea más rápida y rentable.

- **Independiente de la plataforma**

  El código Java puede ejecutarse en cualquier plataforma subyacente, incluida Windows, Linux, iOS o Android, sin necesidad de reescribirlo. Esto lo hace especialmente eficaz en el entorno actual, en el que queremos ejecutar aplicaciones en un par de gadgets.

- **Seguridad**

  Los usuarios pueden descargar código Java no fiable a través de una red y ejecutarlo en un entorno seguro en el que no puede causar ningún daño. El código no fiable no puede infectar el dispositivo anfitrión con una epidemia ni puede estudiar o escribir documentos del disco duro. Los niveles de seguridad y las normas en Java también son relativamente configurables.

c) ¿Cómo funciona Java?

Todos los lenguajes de programación son un método para hablar con las máquinas. El hardware del sistema responde más eficazmente a la comunicación digital. Los lenguajes de programación de alto nivel, entre los que se encuentra Java, actúan como un puente entre el lenguaje humano y el lenguaje del hardware. Para utilizar Java, un desarrollador debe entender dos cosas:

1. El lenguaje Java y las API.

Se trata de la comunicación frontal entre el desarrollador y la plataforma Java.

2. La máquina virtual Java

Es la comunicación de fondo entre la plataforma Java y el hardware subyacente. A continuación, veremos cada uno de estos aspectos con más detalle.

2.14.3 Weka

Es una plataforma de software para el aprendizaje automático y la Minería de Datos escrito en Java y desarrollado en la Universidad de Waikato. Weka es un software libre distribuido bajo la licencia GNU-GPL.

Cuenta con un conjunto de herramientas y algoritmos intuitivos para el análisis de datos y el modelado predictivo, junto con una interfaz gráfica de usuario para acceder fácilmente a sus funciones. El lanzamiento inicial de Weka fue una interfaz para TCL/TK para modelar algoritmos implementados en otros lenguajes de programación, más unas utilidades para preprocesamiento de datos desarrolladas en C para hacer experimentos de aprendizaje automático. (EcuRed, s.f.)

2.15 MÉTRICAS DE CALIDAD

2.15.1 ISO 25010

El modelo de calidad representa la piedra angular en torno a la cual se establece el sistema para la evaluación de la calidad del producto. En este modelo se determinan las características de calidad que se van a tener en cuenta a la hora de evaluar las propiedades de un producto software determinado. La ISO/IEC 25000, denominada SQuaRE (System and Software Quality Requirements and Evaluation), es una familia propia de normas orientadas al desarrollo de un marco común para la comparación de la calidad de los productos de software. Está dirigida a las empresas de programas de software, independientemente de su tamaño o cantidad. Así como a empresas que crean internamente sus propias herramientas de programas de software para desarrollar su negocio (Cava & Lisset, 2018).

Tabla 2.5

Modelo de calidad de software

Seguridad	• Confidenciabilidad • Integridad • No repudio • Autenticidad • Responsabilidad
Adecuación Funcional	• Completitud functional • Corrección functional • Pertinencia funcional
Eficiencia de desempeño	• Comportamiento temporal • Utilización de recursos • Capacidad
Portabilidad	• Adaptabilidad • Facilidad de Instalación • Capacidad de ser reemplazado
Fiabilidad	• Madurez • Disponibilidad • Tolerancia de fallos • Capacidad de recuperación
Compatibilidad	• Coexistencia • Interoperabilidad
Mantenimiento	• Modularidad • Responsabilidad • Analizabilidad • Capacidad de ser modificado • Capacidad de ser probado
Usabilidad	• Inteligibilidad • Aprendizaje • Operabilidad • Proteccion frente a errores de usuario • Estética • Accesibilidad

Nota: ISO 25010, 2021

2.16 EVALUACIÓN DE COSTOS COCOMO II

El Modelo Constructivo de Costos (o COCOMO, por su acrónimo del inglés COnstructive COst MOdel) es un modelo matemático de base empírica utilizado

para estimación de costos de software. Incluye tres submodelos, cada uno ofrece un nivel de detalle y aproximación, cada vez mayor, a medida que avanza el proceso de desarrollo del software: básico, intermedio y detallado.

Este modelo fue desarrollado por Barry W. Boehm a finales de los años 70 y comienzos de los 80, exponiéndolo detalladamente en su libro "Software Engineering Economics" (Prentice-Hall, 1981).

2.16.1 Características generales

Pertenece a la categoría de modelos estimadores basados en estimaciones matemáticas. Está orientado a la magnitud del producto final, midiendo el "tamaño" del proyecto, en función de la cantidad de líneas de código, principalmente.

2.16.2 Modelos de estimación

Las ecuaciones que se utilizan en los tres modelos son:

- $E = a(Kl)^b * m(X)$, en persona-mes
- $Tdev = c(E)^d$, en meses
- $P = {}^{E}/_{Tdev}$, en personas

donde:

- ***E*** es el esfuerzo requerido por el proyecto, en persona-mes
- ***Tdev*** es el tiempo requerido por el proyecto, en meses
- ***P*** es el número de personas requerido por el proyecto
- ***a***, ***b***, ***c*** y ***d*** son constantes con valores definidos en una tabla, según cada submodelo
- ***Kl*** es la cantidad de líneas de código, en miles.
- ***m(X)*** Es un multiplicador que depende de 15 atributos.

A la vez, cada submodelo también se divide en **modos** que representan el tipo de proyecto, y puede ser:

- ***Modo orgánico***: un pequeño grupo de programadores experimentados desarrollan software en un entorno familiar. El tamaño del software varía desde unos pocos miles de líneas (tamaño pequeño) a unas decenas de miles (medio).
- ***Modo semilibre*** o ***semiencajado***: corresponde a un esquema intermedio entre el orgánico y el rígido; el grupo de desarrollo puede incluir una mezcla de personas experimentadas y no experimentadas.
- ***Modo rígido*** o ***empotrado***: el proyecto tiene fuertes restricciones, que pueden estar relacionadas con la funcionalidad y/o pueden ser técnicas. El problema a resolver es único y es difícil basarse en la experiencia, puesto que puede no haberla.

2.16.3 Modelo Básico

Se utiliza para obtener una primera aproximación rápida del esfuerzo, y hace uso de la siguiente tabla de constantes para calcular distintos aspectos de costes:

Tabla 2.6

Modelo Básico

MODO	a	b	c	d
Orgánico	2.40	1.05	2.50	0.38
Semi - Orgánico	3.00	1.12	2.50	0.35
Empotrado	3.60	1.20	2.50	0.33

Nota: (Gómez et al., 2014)

Estos valores son para las fórmulas:

- Personas necesarias por mes para llevar adelante el proyecto (**MM**) = $a*(Kl^b)$
- Tiempo de desarrollo del proyecto (**TDEV**) = $c*(MM^d)$
- Personas necesarias para realizar el proyecto (**CosteH**) = MM/TDEV
- Costo total del proyecto (**CosteM**) = CosteH * *Salario medio entre los programadores y analistas*.

Se puede observar que a medida que aumenta la complejidad del proyecto (modo), las constantes aumentan de 2.4 a 3.6, que corresponde a un incremento del esfuerzo del personal. Hay que utilizar con mucho cuidado el modelo básico puesto que se obvian muchas características del entorno. (Gómez et al., 2014)

2.16.4 Modelo intermedio

Este añade al modelo básico quince modificadores opcionales para tener en cuenta en el entorno de trabajo, incrementando así la precisión de la estimación.

Para este ajuste, al resultado de la fórmula general se lo multiplica por el coeficiente surgido de aplicar los atributos que se decidan utilizar.

Los valores de las constantes a reemplazar en la fórmula son:

Tabla 2.7

Modelo Intermedio

MODO	a	b
Orgánico	3.20	1.05
Semi - Orgánico	3.00	1.12
Empotrado	2.80	1.20

Nota: (Gómez et al., 2014)

Se puede observar que los exponentes son los mismos que los del modelo básico, confirmando el papel que representa el tamaño; mientras que los coeficientes de los modos orgánico y rígido han cambiado, para mantener el equilibrio alrededor del semilibre con respecto al efecto multiplicador de los atributos de coste.

Atributos

Cada atributo se cuantifica para un entorno de proyecto. La escala es ***muy bajo*** - ***bajo*** - ***nominal*** - ***alto*** - ***muy alto*** - ***extremadamente alto***. Dependiendo de la calificación de cada atributo, se asigna un valor para usar de multiplicador en la fórmula (por ejemplo, si para un proyecto el atributo *DATA* es calificado como *muy alto*, el resultado de la fórmula debe ser multiplicado por 1000).

El significado de los atributos es el siguiente, según su tipo:

- De software
  - **RELY**: garantía de funcionamiento requerida al software. Indica las posibles consecuencias para el usuario en el caso de que existan defectos en el producto. Va desde la sola inconveniencia de corregir un fallo (*muy bajo*) hasta la posible pérdida de vidas humanas (*extremadamente alto*, software de alta criticidad).
  - **DATA**: tamaño de la base de datos en relación con el tamaño del programa. El valor del modificador se define por la relación: D/K, donde D corresponde al tamaño de la base de datos en bytes y K es el tamaño del programa en cantidad de líneas de código.
  - **CPLX**: representa la complejidad del producto.
- De hardware
  - **TIME**: limitaciones en el porcentaje del uso de la CPU.
  - **STOR**: limitaciones en el porcentaje del uso de la memoria.
  - **VIRT**: volatilidad de la máquina virtual.
  - **TURN**: tiempo de respuesta requerido.
- De personal
  - **ACAP**: calificación de los analistas.
  - **AEXP**: experiencia del personal en aplicaciones similares.
  - **PCAP**: calificación de los programadores.
  - **VEXP**: experiencia del personal en la máquina virtual.
  - **LEXP**: experiencia en el lenguaje de programación a usar.
- De proyecto
  - **MODP**: uso de prácticas modernas de programación.
  - **TOOL**: uso de herramientas de desarrollo de software.
  - **SCED**: limitaciones en el cumplimiento de la planificación.

El valor de cada atributo, de acuerdo a su calificación, se muestra en la siguiente tabla:

Tabla 2.8

Tabla de Estimación

Atributos	Valor					
	Muy bajo	**Bajo**	**Nominal**	**Alto**	**Muy alto**	**Extra alto**
Atributos de software						
Fiabilidad	0,75	0,88	1,00	1,15	1,40	
Tamaño de Base de datos		0,94	1,00	1,08	1,16	
Complejidad	0,70	0,85	1,00	1,15	1,30	1,65
Atributos de hardware						
Restricciones de tiempo de ejecución			1,00	1,11	1,30	1,66
Restricciones de memoria virtual			1,00	1,06	1,21	1,56
Volatilidad de la máquina virtual		0,87	1,00	1,15	1,30	
Tiempo de respuesta		0,87	1,00	1,07	1,15	
Atributos de personal						
Capacidad de análisis	1,46	1,19	1,00	0,86	0,71	
Experiencia en la aplicación	1,29	1,13	1,00	0,91	0,82	
Calidad de los programadores	1,42	1,17	1,00	0,86	0,70	
Experiencia en la máquina virtual	1,21	1,10	1,00	0,90		
Experiencia en el lenguaje	1,14	1,07	1,00	0,95		
Atributos del proyecto						
Técnicas actualizadas de programación	1,24	1,10	1,00	0,91	0,82	
Utilización de herramientas de software	1,24	1,10	1,00	0,91	0,83	
Restricciones de tiempo de desarrollo	1,22	1,08	1,00	1,04	1,10	

Nota: (Gómez et al., 2014)

2.16.5 Modelo Detallado

Presenta principalmente dos mejoras respecto al anterior:

- Los factores correspondientes a los atributos son sensibles o dependientes de la fase sobre la que se realizan las estimaciones. Aspectos tales como la experiencia en la aplicación, utilización de herramientas de software, etc., tienen mayor influencia en unas fases que en otras, y además van variando de una etapa a otra.
- Establece una jerarquía de tres niveles de productos, de forma que los aspectos que representan gran variación a bajo nivel, se consideran a nivel módulo, los que representan pocas variaciones, a nivel de subsistema; y los restantes son considerados a nivel sistema. (Gómez et al., 2014)

CAPÍTULO III

Diseño Metodológico

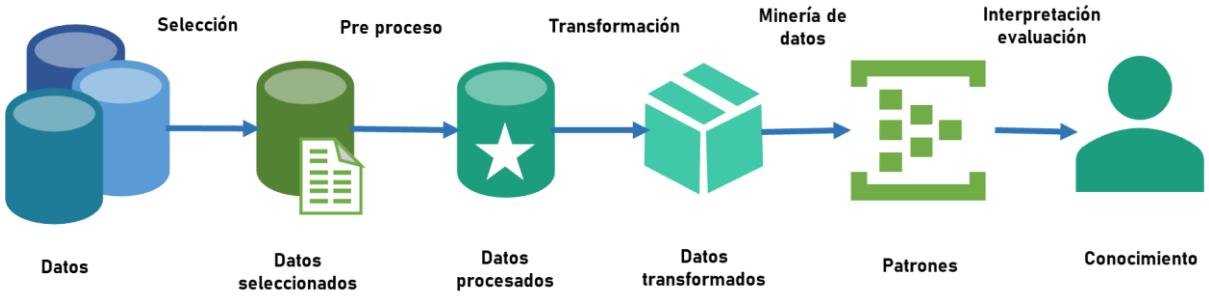
3 MARCO APLICATIVO

El presente capitulo tiene como principal objetivo dar a conocer la aplicación de las metodologías y técnicas mencionados en el capítulo I, llegando a construir el modelo de proyección de índice de contaminación de la basura de la Ciudad de El Alto y poder lograr el objetivo planteado anteriormente.

3.1 APLICACIÓN DE LA METODOLOGÍA CRISP-DM

Figura 3.1

Esquema de solución de la Minería de Datos



Nota: Proceso de Knowledge Discovery in Databases, recopilación de IBM (2022).

3.1.1 FASE I: Comprensión del negocio

El Alto es una Ciudad y municipio de Bolivia ubicado en la provincia Pedro Domingo Murillo del departamento de La Paz, al oeste del país en la meseta altiplánica. El Alto es la Ciudad más poblada del departamento y la segunda Ciudad más poblada de Bolivia.

En la Ciudad se tiene dificultades en el tratamiento de residuos sólidos este problema es evidente debido a un crecimiento de la mancha urbana a consecuencia de un considerable crecimiento poblacional. El tratamiento de los residuos sólidos es una competencia municipal, así lo señala la Constitución y las leyes, sin embargo, más

allá de los preceptos netamente jurídicos se deben analizar las causas y efectos que surten en el ámbito social y ambiental.

La Minería de Datos busca un nuevo conocimiento a partir de los datos utilizados en la presente investigación para lograr encontrar el índice de crecimiento de la contaminación de la basura.

3.1.2 FASE II: Comprensión de los datos

3.1.2.1 Recolección de datos iniciales

Los datos utilizados en la presente investigación son datos de la cantidad de recolección de basura de la Ciudad de El Alto, el cual esta representados por toneladas mensuales, anuales, agrupadas, estos datos fueron solicitados a la secretaria municipal de agua, saneamiento, gestión ambiental y riesgos del gobierno autónomo municipal de El Alto y del instituto nacional de estadística de Bolivia.

3.1.2.2 Datos recolectados

- **Datos de la recolección de residuos sólidos por Ciudades, según año y mes, 2005 – 2022**

  El siguiente cuadro representa los datos recopilados de la cantidad de residuos sólidos recolectadas en a nivel nacional de Bolivia, descritas por departamentos desde el año 2005 al 2022.

Tabla 3.1

Datos históricos 2005 al 2022

PERIODO	Sucre	La Paz	Cochabamba	Oruro	Potosí	Tarija	Santa Cruz	Trinidad	Cobija	El Alto
2005	**34.168**	**157.526**	**115.260**	**34.769**	**19.008**	**26.967**	**310.389**	**17.639**	**757**	**69.169**
Enero		15.288	10.841	3.197	1.730	2.348	27.807	1.352	64	6.214
Febrero		13.247	9.286	3.191	1.423	2.175	23.976	1.353	62	5.852
Marzo		14.327	10.422	3.024	1.707	2.375	25.784	1.384	64	6.075
Abril		13.328	8.923	3.121	1.560	2.289	25.007	1.459	65	6.117
Mayo		12.810	9.260	2.929	1.563	2.173	25.000	1.394	61	5.217
Junio		11.130	10.297	2.635	1.537	2.119	26.174	1.524	63	5.044
Julio		12.716	9.022	2.226	1.550	2.143	24.820	1.551	64	6.205
Agosto		12.302	9.943	2.631	1.572	2.234	25.353	1.631	62	6.150
Septiembre		12.105	9.263	2.784	1.556	1.973	23.694	1.331	63	6.245
Octubre		12.508	9.807	2.722	1.505	2.171	25.428	1.518	65	5.459
Noviembre		13.018	9.233	3.016	1.538	2.396	26.690	1.574	62	6.156
Diciembre		14.748	8.963	3.294	1.768	2.573	30.655	1.568	62	4.435
2006	**35.534**	**169.666**	**114.467**	**37.845**	**20.555**	**28.886**	**315.881**	**22.413**	**686**	**104.798**
2007	**38.801**	**168.205**	**117.473**	**38.794**	**33.488**	**30.143**	**329.337**	**20.803**	**938**	**109.830**
2008	**40.354**	**164.849**	**122.013**	**38.631**	**37.405**	**36.630**	**328.232**	**18.817**	**1.018**	**126.013**
2009	**41.316**	**168.285**	**125.182**	**42.810**	**37.287**	**40.464**	**381.681**	**20.381**	**n.d.**	**138.539**
2010	**44.965**	**177.817**	**131.866**	**44.277**	**58.670**	**47.709**	**359.826**	**24.264**	**7.794**	**143.296**
2011	**48.842**	**177.629**	**136.428**	**44.473**	**50.459**	**51.764**	**363.808**	**22.113**	**9.452**	**153.712**
2012	**54.041**	**181.267**	**140.233**	**47.996**	**53.914**	**53.794**	**376.507**	**26.424**	**3.756**	**161.785**
2013	**54.047**	**186.378**	**166.849**	**49.389**	**49.918**	**49.668**	**400.928**	**24.290**	**n.d.**	**185.627**
2014	**54.209**	**187.650**	**178.034**	**55.855**	**65.076**	**47.001**	**430.103**	**24.322**	**n.d.**	**191.853**
2015	**56.575**	**206.308**	**177.517**	**53.710**	**62.949**	**53.459**	**488.737**	**19.805**	**n.d.**	**200.315**
2016	**60.987**	**212.554**	**171.337**	**57.044**	**47.335**	**56.648**	**558.229**	**28.069**	**17.950**	**216.836**
2017	**57.199**	**224.433**	**192.008**	**60.512**	**47.717**	**59.060**	**599.853**	**19.553**	**16.511**	**245.038**
2018	**60.652**	**236.370**	**212.196**	**61.492**	**45.810**	**62.881**	**636.928**	**26.100**	**17.042**	**257.257**
2019	**22.461**	**230.674**	**199.123**	**64.781**	**43.460**	**64.317**	**672.341**	**30.642**	**18.359**	**254.780**
2020	**39.505**	**234.939**	**192.140**	**62.923**	**45.300**	**68.121**	**659.547**	**36.978**	**19.497**	**262.353**
2021(p)	**60.035**	**241.199**	**228.311**	**72.680**	**44.130**	**75.554**	**565.348**	**31.626**	**22.648**	**262.709**
2022(p)	**39.859**	**141.429**	**139.738**	**46.694**	**26.270**	**44.758**	**308.329**	**20.728**	**7.008**	**152.901**

Nota: Cuadro estadístico de recolección de residuos sólidos (INE, 2022).

- **Datos de la recolección de residuos Sólidos por tipo de procedencia y según año, mes desde 2010 al 2020**

Tabla 3.2

Datos históricos según la procedencia 2005 al 2022

PERIODO	Domiciliaria	Áreas Públicas	Mercados	Hospital	Otros[(1)]
2005	**596.485**	**66.605**	**48.836**	**5.940**	**33.620**
Enero	54.661	6.420	4.284	557	2.917
Febrero	47.904	5.329	3.946	493	2.893
Marzo	51.484	5.894	4.120	538	3.126
Abril	48.773	5.365	4.103	493	3.135
Mayo	48.331	4.931	4.182	493	2.470
Junio	48.531	5.355	4.040	441	2.157
Julio	48.116	5.062	3.841	482	2.796
Agosto	49.981	5.057	3.835	495	2.509
Septiembre	47.191	5.051	3.578	479	2.715
Octubre	48.001	5.424	3.905	473	3.380
Noviembre	50.041	6.352	4.090	481	2.718
Diciembre	53.469	6.365	4.911	517	2.804
2006	**658.017**	**61.665**	**52.784**	**6.146**	**36.585**
2007	**663.592**	**65.275**	**52.967**	**7.029**	**60.150**
2008	**711.283**	**63.008**	**52.848**	**7.276**	**39.193**
2009	**734.973**	**62.070**	**59.064**	**8.791**	**89.730**
2010	**757.988**	**70.469**	**73.263**	**9.642**	**84.158**
2011	**782.164**	**57.273**	**68.428**	**11.267**	**90.708**
2012	**875.353**	**39.287**	**73.516**	**12.104**	**45.414**
2013	**942.884**	**32.866**	**77.953**	**11.959**	**47.385**
2014	**983.238**	**32.937**	**96.181**	**16.893**	**50.644**
2015	**1.090.406**	**33.165**	**128.476**	**12.477**	**54.851**
2016	**1.185.712**	**28.854**	**139.799**	**11.311**	**61.312**
2017	**1.246.981**	**29.909**	**161.815**	**10.026**	**73.153**
2018	**1.363.874**	**27.461**	**158.757**	**9.802**	**56.835**
2019	**1.333.568**	**23.511**	**156.389**	**8.210**	**79.260**
2020	**1.405.321**	**21.865**	**135.411**	**9.824**	**48.883**
2021[(p)]	**1.362.106**	**25.289**	**146.597**	**10.564**	**59.684**
2022[(p)]	**777.855**	**15.408**	**93.544**	**5.511**	**35.397**

Nota: Cuadro estadístico de recolección de residuos sólidos por tipo de procedencia (INE, 2022)

- **Recolección de residuos sólidos recolectados, por Ciudades, según tipo de procedencia, 2010 - 2020**

Tabla 3.3

Datos históricos según la procedencia por Ciudades 2010 al 2022

TIPO DE PROCEDENCIA	SUCRE	LA PAZ	COCHABAMBA	ORURO	POTOSÍ	TARIJA	SANTA CRUZ	TRINIDAD	COBIJA	EL ALTO
2010	**n.d.**	**177.817**	**131.866**	**44.277**	**58.670**	**47.709**	**359.826**	**24.264**	**7.794**	**143.296**
Domiciliarios		144.290	113.628	38.744	32.150	34.473	233.866	18.531	3.128	139.178
Áreas Públicas		10.504	7.788	2.951	11.680	8.821	23.737	4.194	794	n.d.
Mercados		10.643	9.556	2.435	9.795	3.378	33.645	1.451	2.360	n.d.
Establecimientos de salud		3.269	894	147	3.365	309	852	88	594	124
Otros[(1)]		9.112	n.d.	n.d.	1.680	728	67.726	n.d.	918	3.993
2011	**n.d.**	**177.629**	**136.428**	**44.473**	**50.459**	**51.764**	**363.808**	**22.113**	**9.452**	**153.712**
2012	**n.d.**	**181.267**	**140.233**	**47.996**	**53.914**	**53.794**	**376.507**	**26.424**	**3.756**	**161.785**
2013	**n.d.**	**186.378**	**166.849**	**49.389**	**49.918**	**49.668**	**400.928**	**24.290**	**n.d.**	**185.627**
2014	**n.d.**	**187.650**	**178.034**	**55.855**	**65.076**	**47.001**	**430.103**	**24.322**	**n.d.**	**191.853**
2015	**56.575**	**206.308**	**177.517**	**53.710**	**62.949**	**53.459**	**488.737**	**19.805**	**n.d.**	**200.315**
2016	**60.987**	**212.554**	**171.337**	**57.044**	**47.335**	**56.648**	**558.229**	**28.069**	**17.950**	**216.836**
2017	**57.199**	**224.433**	**192.008**	**60.512**	**47.717**	**59.060**	**599.853**	**19.553**	**16.511**	**245.038**
2018	**60.652**	**236.370**	**212.196**	**61.492**	**45.810**	**62.881**	**636.928**	**26.100**	**17.042**	**257.257**
2019	**22.461**	**230.674**	**199.123**	**64.781**	**43.460**	**64.317**	**672.341**	**30.642**	**18.359**	**254.780**
2020[(p)]	**39.505**	**234.939**	**192.140**	**62.923**	**45.300**	**67.800**	**659.547**	**36.978**	**19.497**	**262.353**

Nota: Cuadro estadístico de recolección de residuos sólidos según la procedencia por ciudades (INE, 2022).

- **Recolección de residuos Sólidos en la Ciudad de El Alto por distritos**

Tabla 3.4

Datos histórico de la ciudad de El Alto por distritos 2014 al 2021

GESTIÓN 2014															
Periodo	**Distrito 1**	**Distrito 2**	**Distrito 3**	**Distrito 4**	**Distrito 5**	**Distrito 6**	**Distrito 7**	**Distrito 8**	**Distrito 9**	**Distrito 10**	**Distrito 11**	**Distrito 12**	**Distrito 13**	**Distrito 14**	**Varios Distritos**
ene	3675,44	2300,83	3716,16	2818,67	2322,44	2730,82	933,47	1813,76	0,00	0,00	0,00	1222,01	0,00	1273,02	0,00
feb	3632,43	1954,25	3223,55	2439,13	2242,16	2536,13	863,29	1919,27	0,00	9,74	0,00	1004,97	0,00	1162,66	0,00
mar	3939,43	2111,23	3394,43	2768,14	2362,96	2715,05	857,27	1950,04	6,04	0,00	0,00	1033,40	0,00	1205,09	0,00
abr	3650,71	2041,49	3195,85	2619,23	2124,53	2519,77	853,77	1783,89	0,00	5,99	0,00	1005,53	6,54	1135,80	0,00
may	3667,62	2018,84	3094,04	2680,16	2204,97	2539,57	956,08	1751,09	0,00	0,00	0,00	983,97	0,00	1152,08	0,00
jun	3429,56	2059,20	3098,21	2647,14	2201,39	2504,08	895,52	1581,83	5,50	0,00	0,00	1102,15	0,00	1129,26	0,00
jul	3474,52	2137,85	3134,79	2534,98	2208,99	2543,33	968,14	1689,49	0,00	8,07	0,00	1100,14	0,00	1165,27	0,00
ago	3377,54	1956,68	2897,00	2440,57	2070,85	2475,22	891,33	1748,73	6,22	19,03	0,00	1016,69	0,00	1084,34	600,22
sep	3338,85	1916,74	2888,65	2393,22	2040,95	2477,54	912,15	1763,28	2,78	0,00	0,00	1011,55	0,00	1080,96	0,00
oct	3593,70	1969,73	2973,89	2649,79	2090,62	2481,17	905,08	1808,48	0,00	0,00	0,00	1104,14	0,00	1101,35	0,00
nov	3660,30	2025,60	3080,23	2600,98	2250,12	2587,08	891,71	1858,25	0,00	0,00	0,00	1056,24	3,62	1160,23	0,00
dic	4243,52	2138,65	3369,10	3052,72	2697,95	2953,38	929,41	2147,27	0,00	0,00	0,00	1159,44	0,00	1397,78	0,00
2015	**37239,65**	**18151,13**	**29903,88**	**22119,09**	**19189,93**	**30902,31**	**4799,36**	**14918,88**	**0**	**0**	**0**	**6202,25**	**0**	**7806,96**	**0**
2016	**38613,23**	**21606,63**	**33033,62**	**24628,99**	**21196,51**	**31091,13**	**5574,44**	**17946,1**	**0**	**0**	**0**	**7715,72**	**0**	**8667,18**	**600,22**
2017	**40466,98**	**24447,6**	**34083,33**	**25001,54**	**22871,56**	**31804,75**	**6280,26**	**19533,05**	**0**	**0**	**0**	**9355,84**	**0**	**10058,33**	**0**
2018	**43294,11**	**24898,86**	**37109,71**	**27236,59**	**25386,54**	**34627,19**	**7959,98**	**21621,09**	**0**	**0**	**0**	**11369,65**	**0**	**13192,2**	**0**
2019	**45211,31**	**23856,26**	**38244,79**	**28017,92**	**25699,33**	**34648,97**	**9703,7**	**21219,98**	**33,41**	**290,8**	**0**	**12881,97**	**0**	**12663,28**	**0**
2020	**40981,56**	**26512,2**	**39313,85**	**31267,18**	**26391,46**	**31620,47**	**9783,46**	**21402,85**	**8,15**	**39,71**	**7,93**	**12854,62**	**4,71**	**13416**	**0**
2021	**43683,62**	**24631,09**	**38065,9**	**31644,73**	**26817,93**	**31063,14**	**10857,22**	**21815,38**	**20,54**	**42,83**	**0**	**12800,23**	**10,16**	**14047,84**	**0**

Nota: Cuadro estadístico de recolección de residuos sólidos de la Ciudad de El Alto (DGIR, 2021)

- **Distribución de forma de eliminación de residuos sólidos**

Tabla 3.5

Porcentaje de forma de eliminación de residuos sólidos

Tipo	Porcentaje
Depositado en basurero publico	13.4
En terrenos baldíos	7
La queman	23
Servicio de recolección	44.2
Ríos	7
Entierran	4
Otros	1.6
total	**100%**

Nota**:** Cuadro de forma de eliminacion de residuos solidos en porcentajes del total de residuos generados (IISEC, 2019)

- **Diagnóstico de la gestión de residuos sólidos (2010)**

Tabla 3.6

Porcentaje de tipo de residuos sólido.

Tipo	Porcentaje
Orgánica	55.2%
Reciclable	22.1%
No aprovechable	27.7%
Total	**100%**

Nota: Cuadro de diagnóstico del tipo de residuos sólidos (IISEC, 2019).

3.1.3 FASE III: Preparación de los datos

En esta fase se realiza la preparación de los datos juntamente con las primeras cuatro fases de la metodología KDD.

3.1.3.1 Selección de los datos

Se realizo la selección de los datos que servirán para realizar el estudio del caso, omitiendo datos que no son relevantes para el análisis.

Tabla 3.7

Selección de datos para el en Weka

CAMPO	MUESTRA
Id_clasificacion	7
Tipo_basura	Vidrios
Reciclable	SI
Orgánico	NO
Aprovechable	SI
Contaminante	NO

Nota: Datos estructurados para la base de datos

Tabla 3.8

Selección de datos históricos

CAMPO	MUESTRA
Id_registro	7
Gestión	2018
Cantidad_basura	3939.43
Id_distrito	Distrito 4

Nota: Datos históricos estructurados para la predicción.

3.1.3.2 Limpieza de datos

Se realizo una adecuación a los datos para que estos sean introducidos a la base de datos, corrigiendo la sintaxis generando un ID y un tipo de datos en los diferentes campos.

Tabla 3.9

Fragmento de datos normalizados

Id	Tipo_basura	Reciclable	Orgánico	Aprovechable	Contaminante
1	Orgánico	1	1	1	0
2	Otros	0	0	0	0
3	organices	1	1	1	0
4	Otros	0	0	1	1
5	orgánico	1	1	1	0
6	orgánico	1	1	1	0
7	orgánico	0	1	1	0
8	orgánico	1	1	1	0
9	Especiales	0	1	0	0
10	Especiales	0	1	0	0
11	orgánico	1	1	1	0
12	Plásticos	0	1	1	0
13	Especiales	0	1	0	0
14	Polilamin	0	1	1	1
15	orgánico	1	1	1	0
16	orgánico	1	1	1	0
17	Especiales	0	1	0	0
18	orgánico	1	1	1	0
19	orgánico	1	1	0	0
20	Especiales	0	1	0	0
21	Especiales	0	1	0	0
22	Especiales	0	1	0	0
23	Vidrios	0	0	1	1
24	Especiales	0	1	0	0
25	Polilamin	0	1	1	1
26	Vidrios	0	0	1	1
27	Peligrosos	0	0	0	1

Nota: Datos Normalizados para el entrenamiento en algoritmos de clasificación.

Tabla 3.10

Fragmento de datos históricos normalizados

id	Gestión	Cantidad_residuos
1	2014	253488.16
2	2015	191238.59
3	2016	210678.99
4	2017	223973.60
5	2018	246795.28
6	2019	252475.78
7	2020	253604.14
8	2021	253482.60

Nota: Datos normalizados para el ingreso a los algoritmos de minería de datos

3.1.4 FASE IV: Modelado

En esta fase se procede a la creación del modelo, la selección de técnicas y los algoritmos de Minería de Datos adecuados.

3.1.4.1 Proceso KDD fase Minería de Datos

Para lograr el modelado se aplicará la fase cinco del proceso de descubrimiento del conocimiento de datos (KDD), donde se realiza la selección de técnicas y algoritmos para el modelo de proyección de índice de contaminación de la basura.

3.1.4.2 Selección de Técnicas y algoritmos de Minería de Datos

La herramienta Weka tiene integrado diferentes técnicas y algoritmos de Minería de Datos, el cual se realizó el adecuado entrenamiento y la elección de técnicas y algoritmos más eficientes.

Tabla 3.11

Técnicas y algoritmos seleccionados

Técnicas	Algoritmos
Árbol de decisión	REPTree RandomTree J48
Series de tiempo	Arima LSTM Prophet

Nota: Algoritmos y técnicas elegidas para el entrenamiento en Weka.

3.1.4.3 Estructura de datos de entrada

La siguiente estructura representa un fragmento de los datos recolectados del tipo de residuo contaminante y no contaminante que será ingresado a la herramienta Weka es cual debe estar con extensión .arff, se realizó la estructuración de los datos.

```
@relation clasificación

@attribute tipo_basura
{Organicos,Plasticos,Papeles,Metales,Vidrios,Polilamin,Telas,Especiales,Peli
grosos,Otros}

@attribute reciclable {1,0}

@attribute orgánico {1,0}

@attribute aprovechable {1,0}

@attribute contamina {1,0}

@data


Orgánicos, 1, 1, 1, 0

Otros, 0, 0, 0, 0

Orgánicos, 1, 1, 1, 0

Otros, 0, 0, 1, 1

Organicos, 1, 1, 1, 0

Organicos, 1, 1, 1, 0

Organicos, 0, 1, 1, 0

Organicos, 1, 1, 1, 0

Especiales, 0, 1, 0, 0
```

```
Especiales, 0, 1, 0, 0
Organicos, 1, 1, 1, 0
Plasticos, 0, 1, 1, 0
Especiales, 0, 1, 0, 0
Polilamin, 0, 1, 1, 1
Organicos, 1, 1, 1, 0
Organicos, 1, 1, 1, 0
Especiales, 0, 1, 0, 0
Organicos, 1, 1, 1, 0
Organicos, 1, 1, 0, 0
Especiales, 0, 1, 0, 0
Especiales, 0, 1, 0, 0
Especiales, 0, 1, 0, 0
Vidrios, 0, 0, 1, 1
```

La siguiente estructura representa el total de casos de la recolección de residuos sólidos en la Ciudad de El Alto de la gestión 2014 al 2021 en estructura de archivo .arff.

```
@relation PROYECTO

@attribute año integer
@attribute peso numeric
@data

2014, 256100.83
2015, 191233.44
2016, 210673.77
2017, 223903.24
2018, 246695.92
2019, 252471.72
2020, 253604.15
2021, 255500.61
```

3.1.4.4 Entrenamiento de Weka

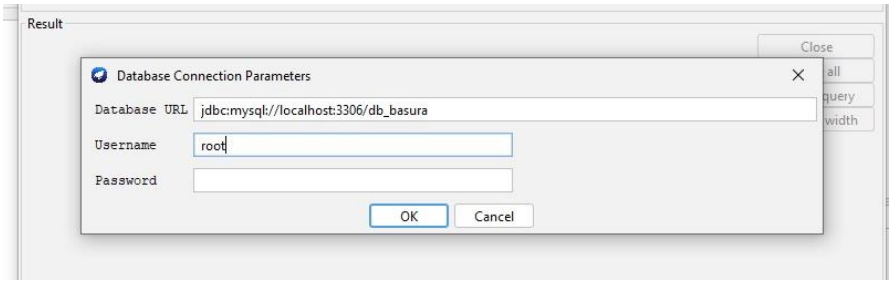
Para lograr un entrenamiento adecuado se vio más factible el uso de una base de datos para el envió de los datos que serán necesarios a Weka a través de consultas, de esta manera el algoritmo podrá leer los datos sin inconvenientes y se podrá realizar el entrenamiento.

- **Conexión de base de datos**

  Se realizo la conexión de la base de datos, introduciendo la URL, el usuario y el password.

  Figura 3.2

  Conexión de base de datos



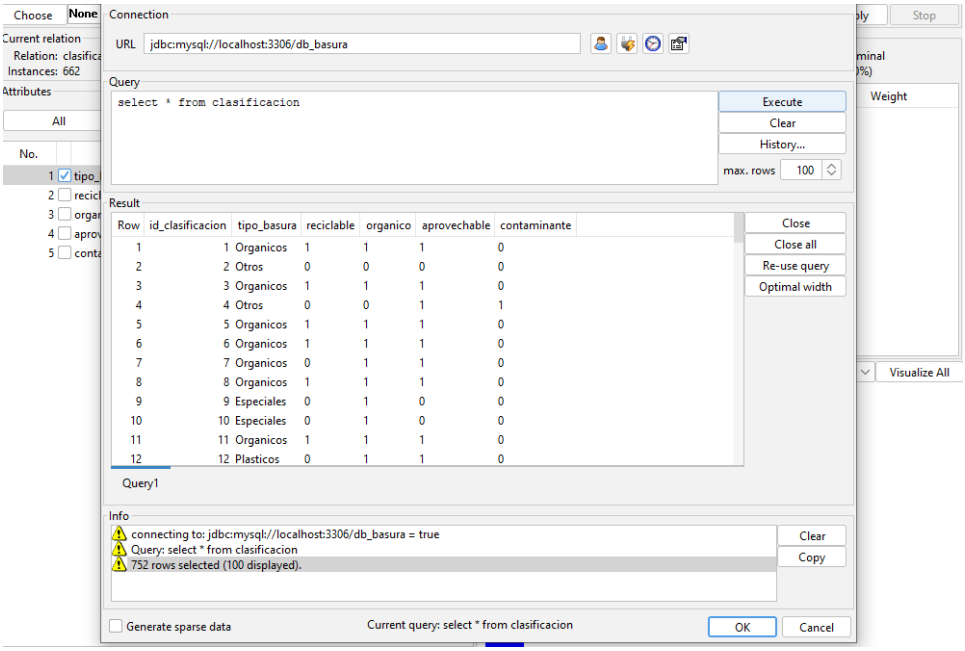
  Nota: Inicio de sesión en la base de datos

- **Selección de datos**

  Una vez conectada la base de datos en Weka, se procedió a realizar consultas para la selección de datos que se usara en el entrenamiento de los algoritmos.

Figura 3.3

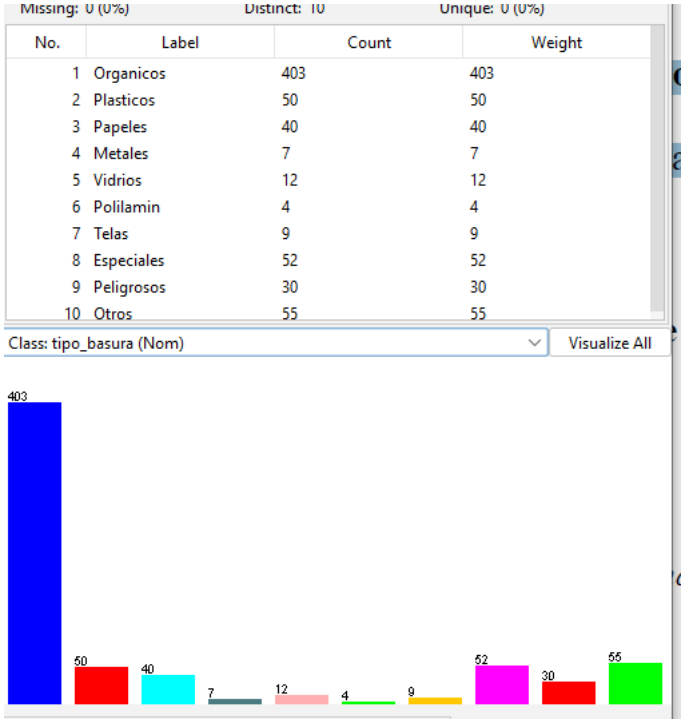
Selección de datos para el entrenamiento



Nota: Selección de datos para el entrenamiento en los algoritmos.

Figura 3.4

Tipo de residuos sólidos contaminantes y no contaminante



Nota: Gráfica descriptiva de los datos ingresados a Weka

a) Aplicación de algoritmo REPTree

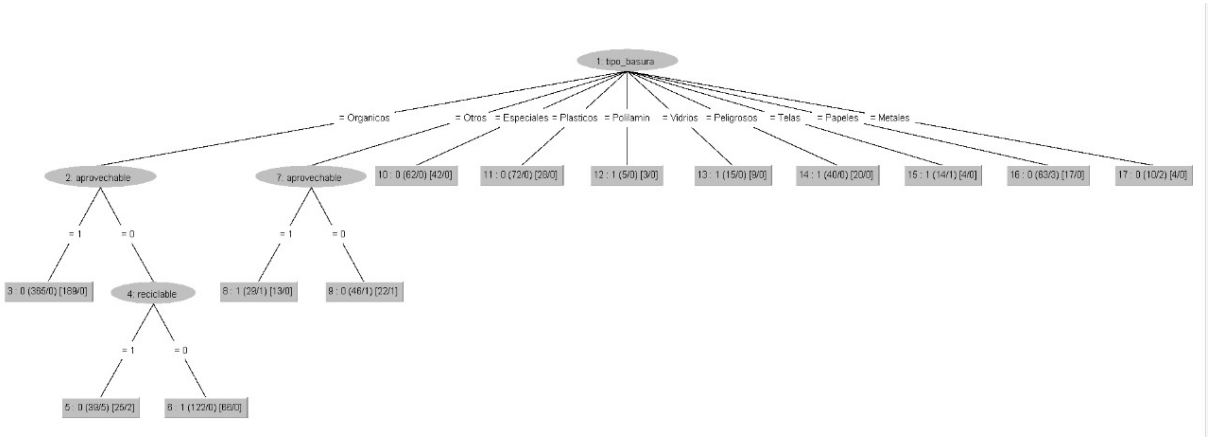
```
 === Run information ===

Scheme:       weka.classifiers.trees.REPTree -M 2 -V 0.001 -N 3 -S 1 -L
-1 -I 0.0
Relation:     QueryResult
Instances:    1324
Attributes:   6
              id_clasificacion
              tipo_basura
              reciclable
              organico
              aprovechable
              contaminante
Test mode:    10-fold cross-validation

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


REPTree
============

tipo_basura = Organicos
|   aprovechable = 1 : 0 (365/0) [189/0]
|   aprovechable = 0
|   |   reciclable = 1 : 0 (39/5) [25/2]
|   |   reciclable = 0 : 1 (122/0) [66/0]
tipo_basura = Otros
|   aprovechable = 1 : 1 (29/1) [13/0]
|   aprovechable = 0 : 0 (46/1) [22/1]
tipo_basura = Especiales : 0 (62/0) [42/0]
tipo_basura = Plasticos : 0 (72/0) [28/0]
tipo_basura = Polilamin : 1 (5/0) [3/0]
tipo_basura = Vidrios : 1 (15/0) [9/0]
tipo_basura = Peligrosos : 1 (40/0) [20/0]
tipo_basura = Telas : 1 (14/1) [4/0]
tipo_basura = Papeles : 0 (63/3) [17/0]
tipo_basura = Metales : 0 (10/2) [4/0]

Size of the tree : 17

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        1308               98.7915 %
Incorrectly Classified Instances        16                1.2085 %
Kappa statistic                          0.9688
Mean absolute error                      0.0213
Root mean squared error                  0.1091
Relative absolute error                  5.4585 %
Root relative squared error             24.6912 %
Total Number of Instances             1324
```

```
=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0,997    0,037    0,987      0,997    0,992      0,969    0,990     0,994     0
               0,963    0,003    0,991      0,963    0,977      0,969    0,990     0,978     1
Weighted Avg.  0,988    0,028    0,988      0,988    0,988      0,969    0,990     0,990

=== Confusion Matrix ===

   a   b   <-- classified as
 969   3 |   a = 0
  13 339 |   b = 1
```

En los resultados obtenidos del algoritmo REPTree se observa una matriz de confusión donde se el resultado indica el tipo de residuo solido contaminante y no contaminante, la matriz de confusión indica que 339 tipos de residuos sólidos son contaminantes, 969 no contaminantes donde 16 indica que tiende a ser contaminante y no contaminante.

Figura 3.5

Árbol generado por Weka



Nota: Proceso del algoritmo REPTree.

b) Aplicación del algoritmo RandomTree

```
=== Run information ===

Scheme:       weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1
Relation:     QueryResult
Instances:    1324
Attributes:   6
              id_clasificacion
              tipo_basura
              reciclable
              organico
```

```
              aprovechable
              contaminante
Test mode:    10-fold cross-validation

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


RandomTree
==========

reciclable = 1
|   tipo_basura = Organicos
|   |   id_clasificacion < 539.5
|   |   |   aprovechable = 1 : 0 (159/0)
|   |   |   aprovechable = 0
|   |   |   |   id_clasificacion < 307 : 0 (15/0)
|   |   |   |   id_clasificacion >= 307
|   |   |   |   |   id_clasificacion < 519
|   |   |   |   |   |   organico = 1
|   |   |   |   |   |   |   id_clasificacion < 485 : 0 (9/0)
|   |   |   |   |   |   |   id_clasificacion >= 485
|   |   |   |   |   |   |   |   id_clasificacion < 503.5 : 1 (1/0)
|   |   |   |   |   |   |   |   id_clasificacion >= 503.5 : 0 (1/0)
|   |   |   |   |   |   organico = 0
|   |   |   |   |   |   |   id_clasificacion < 406.5 : 1 (4/0)
|   |   |   |   |   |   |   id_clasificacion >= 406.5 : 0 (3/0)
|   |   |   |   |   id_clasificacion >= 519 : 1 (2/0)
|   |   id_clasificacion >= 539.5 : 0 (283/0)
|   tipo_basura = Otros
|   |   id_clasificacion < 419 : 0 (1/0)
|   |   id_clasificacion >= 419 : 1 (2/0)
|   tipo_basura = Especiales : 0 (1/0)
|   tipo_basura = Plasticos : 0 (11/0)
|   tipo_basura = Polilamin : 0 (0/0)
|   tipo_basura = Vidrios : 0 (0/0)
|   tipo_basura = Peligrosos : 1 (1/0)
|   tipo_basura = Telas : 1 (17/0)
|   tipo_basura = Papeles
|   |   aprovechable = 1
|   |   |   id_clasificacion < 531 : 0 (30/0)
|   |   |   id_clasificacion >= 531
|   |   |   |   id_clasificacion < 556.5 : 1 (1/0)
|   |   |   |   id_clasificacion >= 556.5
|   |   |   |   |   id_clasificacion < 1193 : 0 (39/0)
|   |   |   |   |   id_clasificacion >= 1193
|   |   |   |   |   |   id_clasificacion < 1218.5 : 1 (1/0)
|   |   |   |   |   |   id_clasificacion >= 1218.5 : 0 (8/0)
|   |   aprovechable = 0 : 1 (1/0)
```

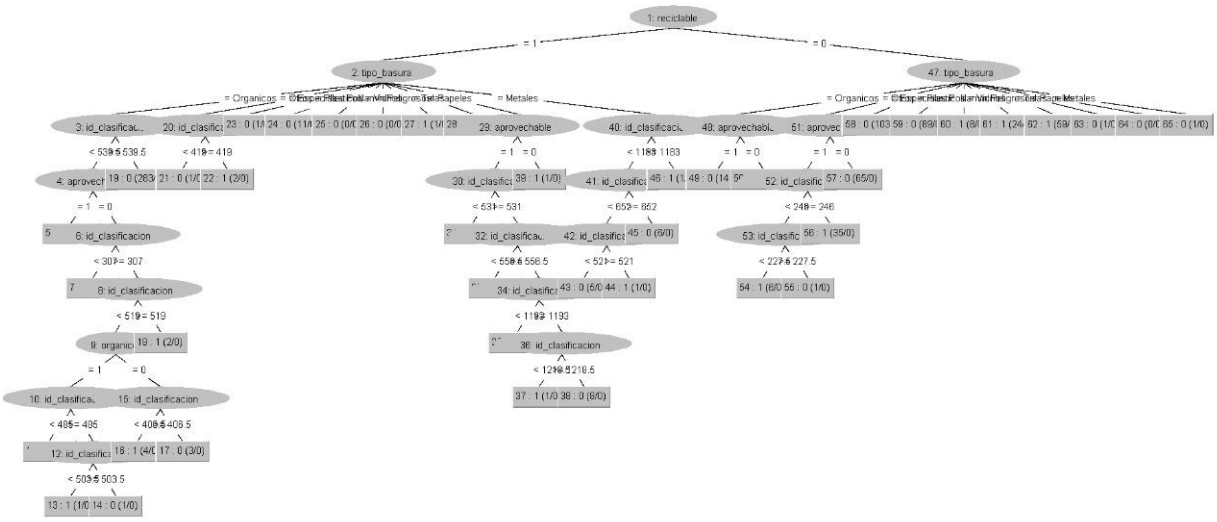
```
|   tipo_basura = Metales
|   |   id_clasificacion < 1183
|   |   |   id_clasificacion < 652
|   |   |   |   id_clasificacion < 521 : 0 (5/0)
|   |   |   |   id_clasificacion >= 521 : 1 (1/0)
|   |   |   id_clasificacion >= 652 : 0 (6/0)
|   |   id_clasificacion >= 1183 : 1 (1/0)
reciclable = 0
|   tipo_basura = Organicos
|   |   aprovechable = 1 : 0 (141/0)
|   |   aprovechable = 0 : 1 (188/0)
|   tipo_basura = Otros
|   |   aprovechable = 1
|   |   |   id_clasificacion < 246
|   |   |   |   id_clasificacion < 227.5 : 1 (6/0)
|   |   |   |   id_clasificacion >= 227.5 : 0 (1/0)
|   |   |   id_clasificacion >= 246 : 1 (35/0)
|   |   aprovechable = 0 : 0 (65/0)
|   tipo_basura = Especiales : 0 (103/0)
|   tipo_basura = Plasticos : 0 (89/0)
|   tipo_basura = Polilamin : 1 (8/0)
|   tipo_basura = Vidrios : 1 (24/0)
|   tipo_basura = Peligrosos : 1 (59/0)
|   tipo_basura = Telas : 0 (1/0)
|   tipo_basura = Papeles : 0 (0/0)
|   tipo_basura = Metales : 0 (1/0)

Size of the tree : 65

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        1299               98.1118 %
Incorrectly Classified Instances        25                1.8882 %
Kappa statistic                          0.9517
Mean absolute error                      0.019
Root mean squared error                  0.1375
Relative absolute error                  4.8699 %
Root relative squared error             31.1238 %
Total Number of Instances             1324

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0,987    0,034    0,988      0,987    0,987      0,952    0,976     0,984     0
               0,966    0,013    0,963      0,966    0,965      0,952    0,976     0,939     1
Weighted Avg.  0,981    0,029    0,981      0,981    0,981      0,952    0,976     0,972
```

```
=== Confusion Matrix ===

   a   b   <-- classified as
 959  13 |   a = 0
  12 340 |   b = 1
```

En los resultados obtenidos del algoritmo RandomTree se observa una matriz de confusión donde se indica el tipo de residuo solido contaminante y no contaminante, la matriz indica que 340 tipos de residuos sólidos son contaminantes, 959 son no contaminantes y 25 residuos sólidos tiendes a ser no contaminantes y contaminante.

Figura 3.6

Árbol generado por RandomTree



Nota: Proceso del algoritmo RandomTree

c) Aplicación del algoritmo J48

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     QueryResult
Instances:    1324
Attributes:   6
              id_clasificacion
              tipo_basura
              reciclable
```

```
              organico
              aprovechable
              contaminante
Test mode:    10-fold cross-validation

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

J48 pruned tree
------------------

reciclable = 1
|   tipo_basura = Organicos
|   |   aprovechable = 1: 0 (413.0)
|   |   aprovechable = 0
|   |   |   organico = 1: 0 (56.0/2.0)
|   |   |   organico = 0
|   |   |   |   id_clasificacion <= 406: 1 (4.0)
|   |   |   |   id_clasificacion > 406: 0 (4.0/1.0)
|   tipo_basura = Otros: 1 (3.0/1.0)
|   tipo_basura = Especiales: 0 (1.0)
|   tipo_basura = Plasticos: 0 (11.0)
|   tipo_basura = Polilamin: 0 (0.0)
|   tipo_basura = Vidrios: 0 (0.0)
|   tipo_basura = Peligrosos: 1 (1.0)
|   tipo_basura = Telas: 1 (17.0)
|   tipo_basura = Papeles: 0 (80.0/3.0)
|   tipo_basura = Metales: 0 (13.0/2.0)
reciclable = 0
|   tipo_basura = Organicos
|   |   aprovechable = 1: 0 (141.0)
|   |   aprovechable = 0: 1 (188.0)
|   tipo_basura = Otros
|   |   aprovechable = 1: 1 (42.0/1.0)
|   |   aprovechable = 0: 0 (65.0)
|   tipo_basura = Especiales: 0 (103.0)
|   tipo_basura = Plasticos: 0 (89.0)
|   tipo_basura = Polilamin: 1 (8.0)
```

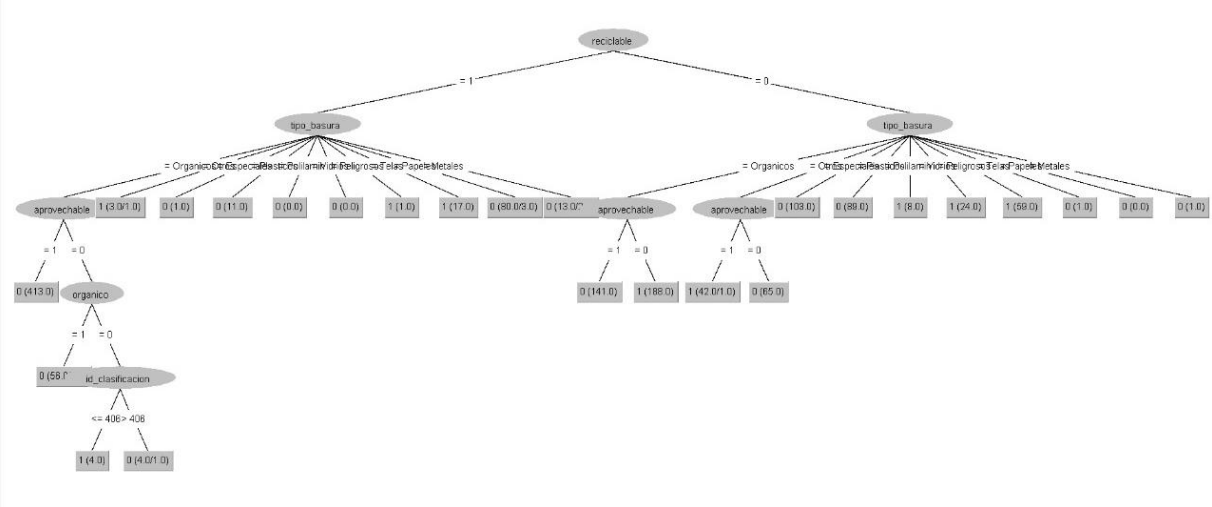
```
|   tipo_basura = Vidrios: 1 (24.0)
|   tipo_basura = Peligrosos: 1 (59.0)
|   tipo_basura = Telas: 0 (1.0)
|   tipo_basura = Papeles: 0 (0.0)
|   tipo_basura = Metales: 0 (1.0)

Number of Leaves  :    25

Size of the tree :    33
Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        1307               98.716  %
Incorrectly Classified Instances        17                1.284  %
Kappa statistic                          0.9668
Mean absolute error                      0.02
Root mean squared error                  0.1145
Relative absolute error                  5.1158 %
Root relative squared error             25.9211 %
Total Number of Instances             1324

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0,997    0,040    0,986      0,997    0,991      0,967    0,990     0,994     0
               0,960    0,003    0,991      0,960    0,975      0,967    0,990     0,977     1
Weighted Avg.  0,987    0,030    0,987      0,987    0,987      0,967    0,990     0,989

=== Confusion Matrix ===

   a   b   <-- classified as
 969   3 |   a = 0
  14 338 |   b = 1
```

En los resultados obtenidos del algoritmo J48 se observa una matriz de confusión donde se indica el tipo de residuo solido contaminante y no

contaminante, la matriz indica que 338 tipos de residuos sólidos son contaminantes, 969 son no contaminantes y 17 residuos sólidos tiendes a ser no contaminantes y contaminante.

Figura 3.7

Árbol generado por J48



Nota: Proceso del algoritmo J48

d) Aplicación del algoritmo PART

```
=== Run information ===

Scheme:       weka.classifiers.rules.PART -C 0.25 -M 2
Relation:     QueryResult
Instances:    1324
Attributes:   6
              id_clasificacion
              tipo_basura
              reciclable
              organico
              aprovechable
              contaminante
Test mode:    10-fold cross-validation

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

PART decision list
------------------

reciclable = 1 AND
tipo_basura = Organicos AND
aprovechable = 1: 0 (413.0)

tipo_basura = Especiales: 0 (104.0)
```

```
aprovechable = 1 AND
tipo_basura = Organicos: 0 (141.0)

tipo_basura = Organicos AND
reciclable = 0: 1 (188.0)

tipo_basura = Plasticos: 0 (100.0)

tipo_basura = Papeles: 0 (80.0/3.0)

tipo_basura = Peligrosos: 1 (60.0)

aprovechable = 0 AND
reciclable = 0: 0 (65.0)

tipo_basura = Organicos AND
organico = 1: 0 (56.0/2.0)

tipo_basura = Otros: 1 (45.0/2.0)

tipo_basura = Vidrios: 1 (24.0)

tipo_basura = Telas: 1 (18.0/1.0)

organico = 0 AND
tipo_basura = Metales: 0 (14.0/2.0)

tipo_basura = Polilamin: 1 (8.0)

id_clasificacion <= 406: 1 (4.0)

: 0 (4.0/1.0)

Number of Rules  :    16


Time taken to build model: 0.07 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        1306               98.6405 %
Incorrectly Classified Instances        18                1.3595 %
Kappa statistic                          0.9649
Mean absolute error                      0.0214
Root mean squared error                  0.1155
Relative absolute error                  5.4773 %
Root relative squared error             26.1469 %
Total Number of Instances             1324

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0,995    0,037    0,987      0,995    0,991      0,965    0,989     0,994     0
               0,963    0,005    0,985      0,963    0,974      0,965    0,989     0,976     1
Weighted Avg.  0,986    0,028    0,986      0,986    0,986      0,965    0,989     0,989
```

```
=== Confusion Matrix ===

   a   b   <-- classified as
 967   5 |   a = 0
  13 339 |   b = 1
```

En los resultados obtenidos del algoritmo PART se observa una matriz de confusión donde se indica el tipo de residuo solido contaminante y no contaminante, la matriz indica que 339 tipos de residuos sólidos son contaminantes, 967 son no contaminantes y 18 residuos sólidos tiendes a ser no contaminantes y contaminante.

e) Aplicación del algoritmo RandomForest

```
=== Run information ===

Scheme:       weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:     QueryResult
Instances:    1324
Attributes:   6
              id_clasificacion
              tipo_basura
              reciclable
              organico
              aprovechable
              contaminante
Test mode:    10-fold cross-validation

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

RandomForest
Bagging with 100 iterations and base learner
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities
Time taken to build model: 0.18 seconds

=== Stratified cross-validation ===

=== Summary ===
```

```
Correctly Classified Instances        1303               98.4139 %
Incorrectly Classified Instances        21                1.5861 %
Kappa statistic                          0.9593
Mean absolute error                      0.0179
Root mean squared error                  0.1076
Relative absolute error                  4.5827 %
Root relative squared error             24.3518 %
Total Number of Instances             1324

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0,991    0,034    0,988      0,991    0,989      0,959    0,995     0,997     0
               0,966    0,009    0,974      0,966    0,970      0,959    0,995     0,990     1
Weighted Avg.  0,984    0,027    0,984      0,984    0,984      0,959    0,995     0,995

=== Confusion Matrix ===

   a   b   <-- classified as
 963   9 |   a = 0
  12 340 |   b = 1
```

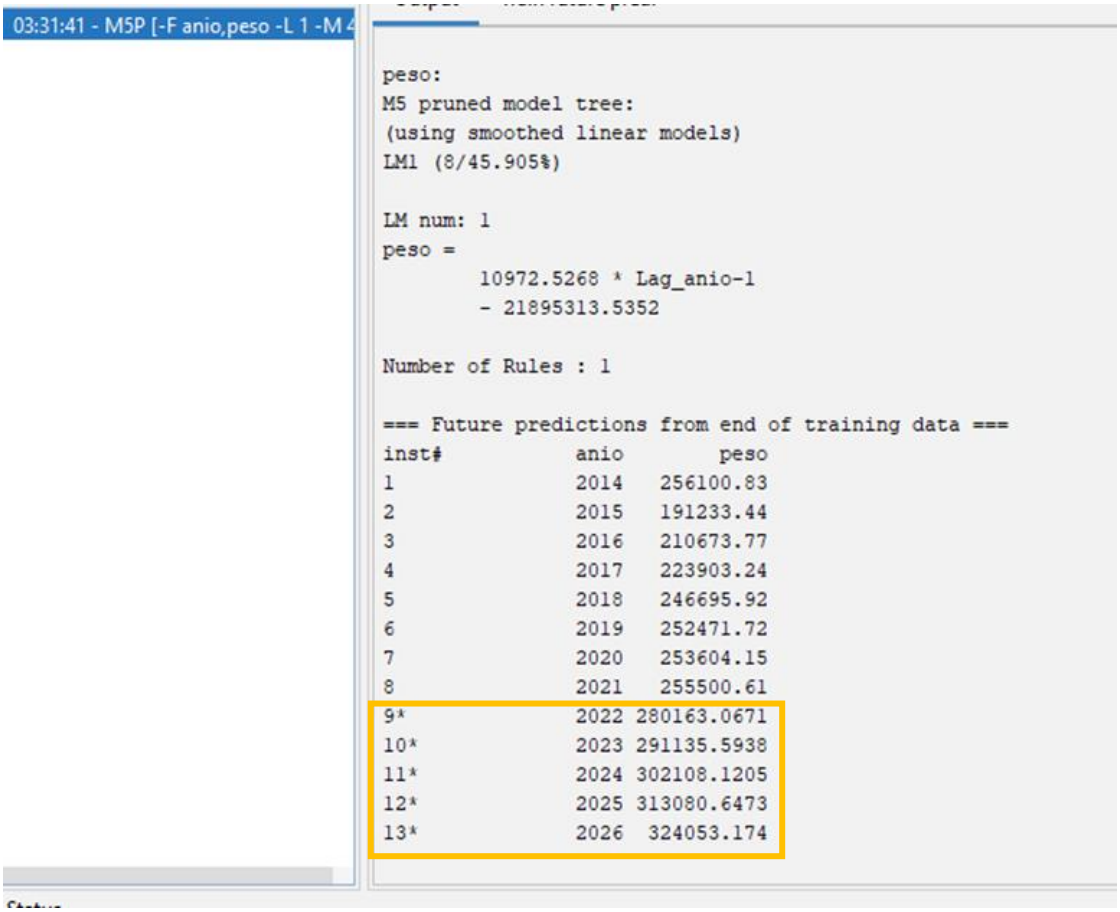
En los resultados obtenidos del algoritmo PART se observa una matriz de confusión donde se indica el tipo de residuo solido contaminante y no contaminante, la matriz indica que 340 tipos de residuos sólidos son contaminantes, 963 son no contaminantes y 21 residuos sólidos tiendes a ser no contaminantes y contaminante.

f) Aplicación del algoritmo MP5

Se aplico el algoritmo MP5 para lograr la predicción de residuos sólidos generados en la Ciudad de El Alto.

Figura 3.8

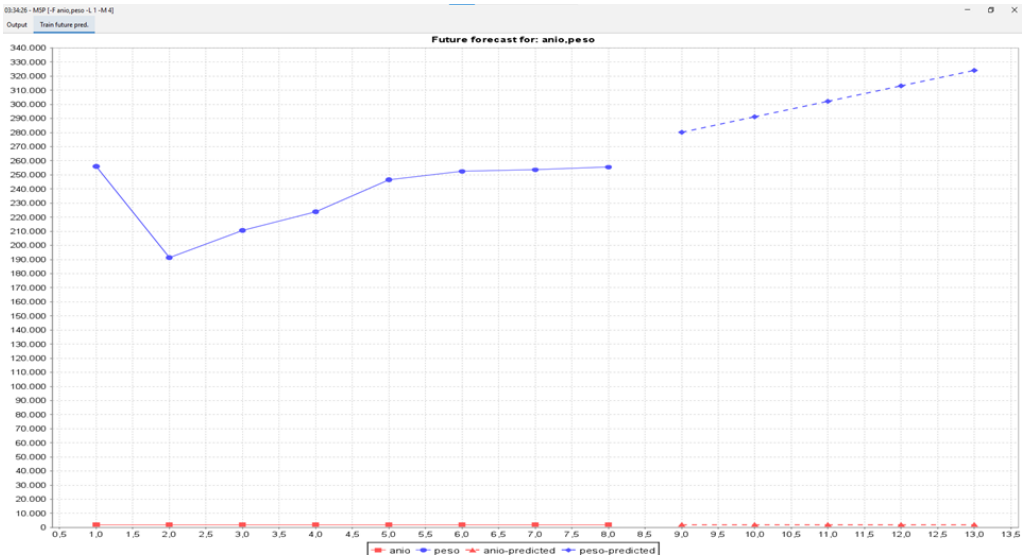
Resultados del algoritmo MP5



Se puede observar los resultados obtenidos con el algoritmo MP5 de los casos totales de la generación de residuos sólidos desde el año 2014 a 2026.

Figura 3.9

Gráfica de los resultados del algoritmo MP5



Nota: Resultados de Predicción según el algoritmo MP5

g) Aplicación del algoritmo RandomTree

=== Run information ===
Scheme:
RandomTree -K 0 -M 1.0 -V 0.001 -S 1

Lagged and derived variable options:
-F anio,peso -L 1 -M 4

Relation: PROYECTO
Instances: 8
Attributes: 2
anio
peso

Transformed training data:

Año
peso
ArtificialTimeIndex
Lag_anio-1
Lag_anio-2
Lag_anio-3
Lag_anio-4
Lag_peso-1
Lag_peso-2
Lag_peso-3
Lag_peso-4
ArtificialTimeIndex^2
ArtificialTimeIndex^3

ArtificialTimeIndex*Lag_anio-1
ArtificialTimeIndex*Lag_anio-2
ArtificialTimeIndex*Lag_anio-3
ArtificialTimeIndex*Lag_anio-4
ArtificialTimeIndex*Lag_peso-1
ArtificialTimeIndex*Lag_peso-2
ArtificialTimeIndex*Lag_peso-3
ArtificialTimeIndex*Lag_peso-4

anio:

RandomTree
==========

ArtificialTimeIndex < 4.5
| ArtificialTimeIndex < 2.5
| | ArtificialTimeIndex < 1.5 : 2014 (1/0)
| | ArtificialTimeIndex >= 1.5 : 2015 (1/0)
| ArtificialTimeIndex >= 2.5
| | ArtificialTimeIndex < 3.5 : 2016 (1/0)
| | ArtificialTimeIndex >= 3.5 : 2017 (1/0)
ArtificialTimeIndex >= 4.5
| ArtificialTimeIndex*Lag_anio-3 < 13107.5
| | ArtificialTimeIndex*Lag_anio-1 < 11096.5 : 2018 (1/0)
| | ArtificialTimeIndex*Lag_anio-1 >= 11096.5 : 2019 (1/0)
| ArtificialTimeIndex*Lag_anio-3 >= 13107.5
| | Lag_peso-2 < 249583.82 : 2020 (1/0)
| | Lag_peso-2 >= 249583.82 : 2021 (1/0)

Size of the tree : 15

peso:

RandomTree
==========
Lag_anio-1 < 2016.5
| ArtificialTimeIndex^3 < 4.5 : 256100.83 (0.43/0)
| ArtificialTimeIndex^3 >= 4.5
| | Lag_anio-1 < 2014.5 : 191233.44 (1/0)
| | Lag_anio-1 >= 2014.5
| | | ArtificialTimeIndex < 3.5 : 210673.77 (1/0)
| | | ArtificialTimeIndex >= 3.5 : 223903.24 (1/0)
Lag_anio-1 >= 2016.5
| Lag_anio-2 < 2016.5 : 247871.53 (1.14/9674473.82)
| Lag_anio-2 >= 2016.5
| | Lag_anio-3 < 2017.5
| | | ArtificialTimeIndex < 3.5 : 256100.83 (0.29/0)
| | | ArtificialTimeIndex >= 3.5 : 253037.93 (2/320599.43)
| | Lag_anio-3 >= 2017.5 : 255575.64 (1.14/39403.88)

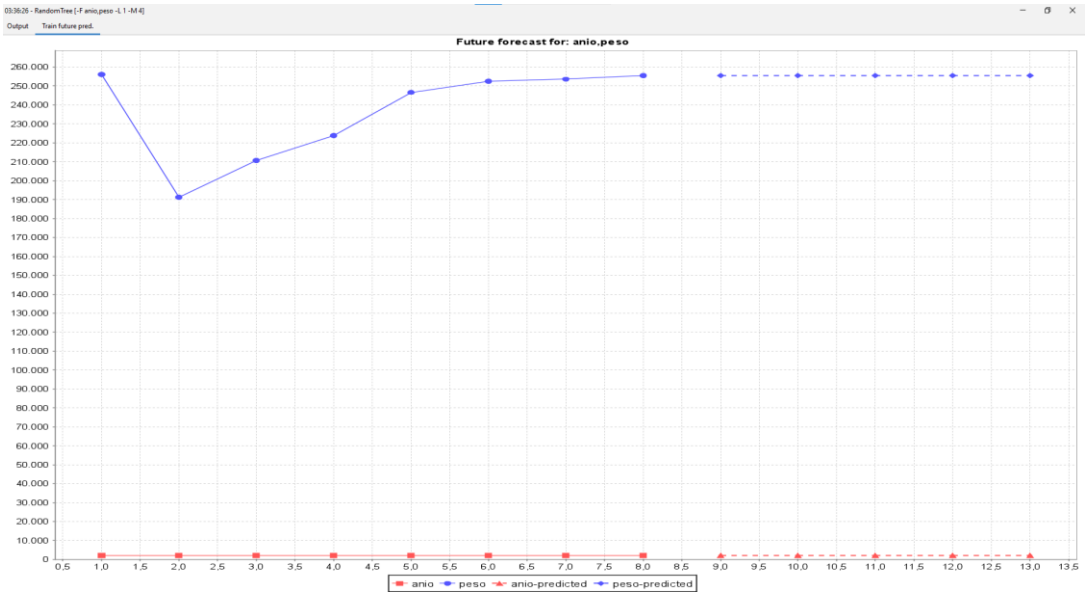
Size of the tree : 15
=== Future predictions from end of training data ===
inst# año peso

1	2014	256100.83
2	2015	191233.44
3	2016	210673.77
4	2017	223903.24
5	2018	246695.92
6	2019	252471.72
7	2020	253604.15
8	2021	255500.61
9*	2021	255575.6375
10*	2021	255575.6375
11*	2021	255575.6375
12*	2021	255575.6375
13*	2021	255575.6375

Se puede observar los resultados obtenidos con el algoritmo RamdomTree el cual no logro realizar una predicción efectiva dando como resultados la misma cantidad para cada año.

Figura 3.10

Gráfica generada por el algoritmo RandomTree



Nota: Resultados de Predicción según el algoritmo RandomTree

h) Aplicación del algoritmo MultilayerPerceptron

=== Future predictions from end of training data ===

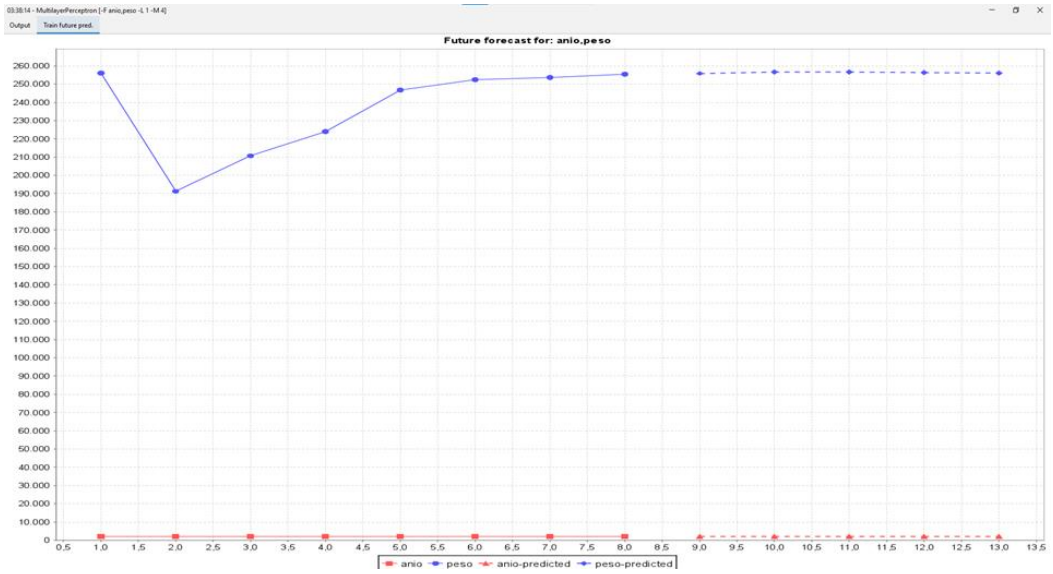
inst#	año	peso
1	2014	256100.83
2	2015	191233.44
3	2016	210673.77

4	2017	223903.24
5	2018	246695.92
6	2019	252471.72
7	2020	253604.15
8	2021	255500.61
9*	2022.0179	255841.9227
10*	2022.9429	256593.9288
11*	2023.768	256606.8734
12*	2024.4868	256280.3116
13*	2025.0752	255942.3114

Se puede observar los resultados obtenidos con el algoritmo MultilayerPerceptron de los casos totales de la generación de residuos sólidos desde el año 2014 al 2026.

Figura 3.11

Gráfica generada por el algoritmo MultilayerPerceptron



Nota: Resultados de Predicción según el algoritmo MultilayerPerceptron

3.1.5 FASE V: Evaluación

En esta fase se procede a mostrar los resultados de cada algoritmo entrenado en la anterior fase.

- **Evaluación algoritmo REPTree**

Figura 3.12

Resultados algoritmo REPTree

```
Correctly Classified Instances        1308               98.7915 %
Incorrectly Classified Instances        16                1.2085 %
Kappa statistic                          0.9688
Mean absolute error                      0.0213
Root mean squared error                  0.1091
Relative absolute error                  5.4585 %
Root relative squared error             24.6912 %
Total Number of Instances             1324

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,997    0,037    0,987      0,997    0,992      0,969    0,990     0,994     0
                 0,963    0,003    0,991      0,963    0,977      0,969    0,990     0,978     1
Weighted Avg.    0,988    0,028    0,988      0,988    0,988      0,969    0,990     0,990

=== Confusion Matrix ===

   a   b   <-- classified as
 969   3 |   a = 0
  13 339 |   b = 1
```

Nota: datos resultantes del entrenamiento del algoritmo REPTree

- **Evaluación algoritmo RandomTree**

Figura 3.13

Resultados algoritmo RandomTree

```
Correctly Classified Instances        1299               98.1118 %
Incorrectly Classified Instances        25                1.8882 %
Kappa statistic                          0.9517
Mean absolute error                      0.019
Root mean squared error                  0.1375
Relative absolute error                  4.8699 %
Root relative squared error             31.1238 %
Total Number of Instances             1324

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,987    0,034    0,988      0,987    0,987      0,952    0,976     0,984     0
                 0,966    0,013    0,963      0,966    0,965      0,952    0,976     0,939     1
Weighted Avg.    0,981    0,029    0,981      0,981    0,981      0,952    0,976     0,972

=== Confusion Matrix ===

   a   b   <-- classified as
 959  13 |   a = 0
  12 340 |   b = 1
```

Nota: datos resultantes del entrenamiento del algoritmo RandomTree

- **Evaluación algoritmo J48**

Figura 3.14

Resultados algoritmo J48

```
Correctly Classified Instances        1307               98.716  %
Incorrectly Classified Instances        17                1.284  %
Kappa statistic                          0.9668
Mean absolute error                      0.02
Root mean squared error                  0.1145
Relative absolute error                  5.1158 %
Root relative squared error             25.9211 %
Total Number of Instances             1324

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,997    0,040    0,986      0,997    0,991      0,967    0,990     0,994     0
                 0,960    0,003    0,991      0,960    0,975      0,967    0,990     0,977     1
Weighted Avg.    0,987    0,030    0,987      0,987    0,987      0,967    0,990     0,989

=== Confusion Matrix ===

   a   b   <-- classified as
 969   3 |   a = 0
  14 338 |   b = 1
```

Nota: datos resultantes del entrenamiento del algoritmo J48

- **Evaluación algoritmo PART**

Figura 3.15

Resultados del algoritmo PART

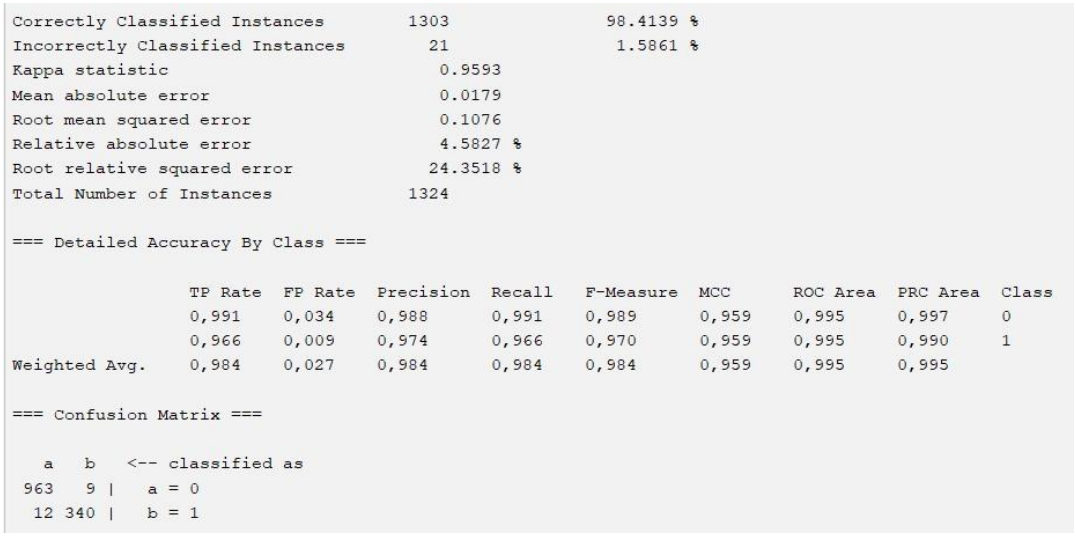
```
Correctly Classified Instances        1306               98.6405 %
Incorrectly Classified Instances        18                1.3595 %
Kappa statistic                          0.9649
Mean absolute error                      0.0214
Root mean squared error                  0.1155
Relative absolute error                  5.4773 %
Root relative squared error             26.1469 %
Total Number of Instances             1324

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,995    0,037    0,987      0,995    0,991      0,965    0,989     0,994     0
                 0,963    0,005    0,985      0,963    0,974      0,965    0,989     0,976     1
Weighted Avg.    0,986    0,028    0,986      0,986    0,986      0,965    0,989     0,989

=== Confusion Matrix ===

   a   b   <-- classified as
 967   5 |   a = 0
  13 339 |   b = 1
```

Nota: datos resultantes del entrenamiento del algoritmo PART

- **Evaluación algoritmo RandomForest**

Figura 3.16

Resultados algoritmo RandomForest



```
Correctly Classified Instances        1303               98.4139 %
Incorrectly Classified Instances        21                1.5861 %
Kappa statistic                          0.9593
Mean absolute error                      0.0179
Root mean squared error                  0.1076
Relative absolute error                  4.5827 %
Root relative squared error             24.3518 %
Total Number of Instances             1324

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,991    0,034    0,988      0,991    0,989      0,959    0,995     0,997     0
                 0,966    0,009    0,974      0,966    0,970      0,959    0,995     0,990     1
Weighted Avg.    0,984    0,027    0,984      0,984    0,984      0,959    0,995     0,995

=== Confusion Matrix ===

   a   b   <-- classified as
 963   9 |   a = 0
  12 340 |   b = 1
```

Nota: datos resultantes del entrenamiento del algoritmo RandomForest

Una vez aplicado los diferentes algoritmos se realiza la comparación de resultados de la siguiente manera:

Tabla 3.12

Comparación de resultados de algoritmos

Ecuación	Instancias clasificadas correctamente	Error de la Media Absoluta	Error Absoluto Relativo	Error cuadrático medio de raíz
REPTree	98.7915%	1.2085%	5.4585%	24.6912%
RandomTree	98.1118%	1.8882%	4.8699 %	31.1238%
J48	98.716%	1.2845%	5.1158%	25.9211%
PART	98.6405%	1.3595%	5.4773%	26.1469%
RandomForest	98.4139%	1.5861%	4.5827%	24.3518%

Nota: Cuadro comparativo de resultados del entrenamiento en los algoritmos

En la Tabla 3.12 se observa los resultados obtenidos del entrenamiento de los algoritmos REPTree, J48 y PART donde se consideran factibles para el Modelado por las instancias clasificadas correctamente que tienen un valor alto y el Error Absoluto más bajo.

3.2 APLICACIÓN DE LA METODOLOGÍA OPEN UP

En esta fase se aplica la metodología Open Up para el desarrollo del modelo de proyección de índice de contaminación de basura

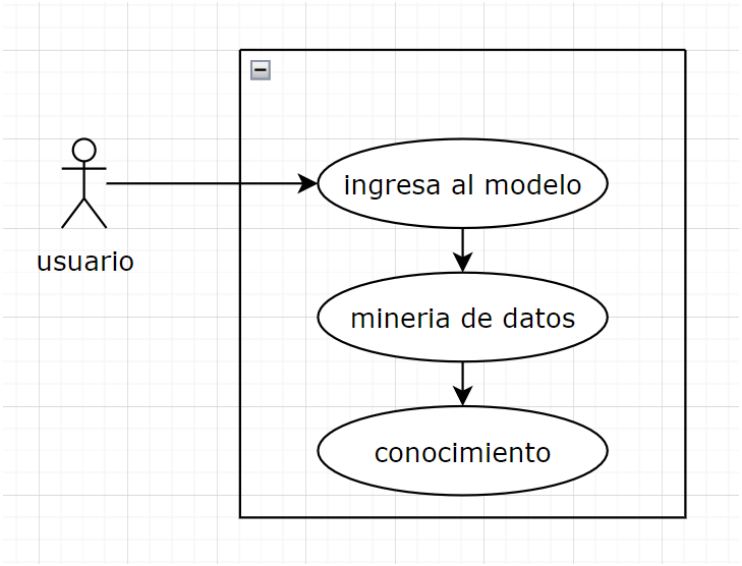
3.2.1 Fase de inicio

3.2.1.1 Descripción de actores

La identificación de los interesados o también llamados Stakeholders describe las responsabilidades y descripciones de cada interesado.

Figura 3.17

Caso de uso



Nota: Casos de uso general del modelo

Tabla 3.13

Descripción caso de uso

Nombre	Descripción
Usuario general	Es la persona que se encarga de visualizar los resultados del modelo de proyección Visualizara datos estadísticos Podrá manipular datos de la base de datos

Nota: Cuadro descriptivo del caso de uso general planteado

3.2.2 Fase de Elaboración

En esta fase se da inicio al desarrollo del prototipo propuesto.

3.2.2.1 Captura de requerimientos

A continuación, se realiza una descripción de los requerimientos funcionales del modelo de proyección de índice de contaminación de basura.

Tabla 3.14

Requerimientos del Modelo

Código	Requerimientos	Prioridad
R1	El modelo realizara una proyección del índice de crecimiento de contaminación de basura basado en la Minería de Datos	ALTA
R2	Realizar un entrenamiento predictivo del índice de crecimiento de la contaminación de basura.	ALTA
R3	Generar resultados del índice de crecimiento de la contaminación de basura por año.	ALTA
R4	Visualización de reportes estadísticos de los Resultados	ALTA

Nota: Requerimientos del Modelo planteado

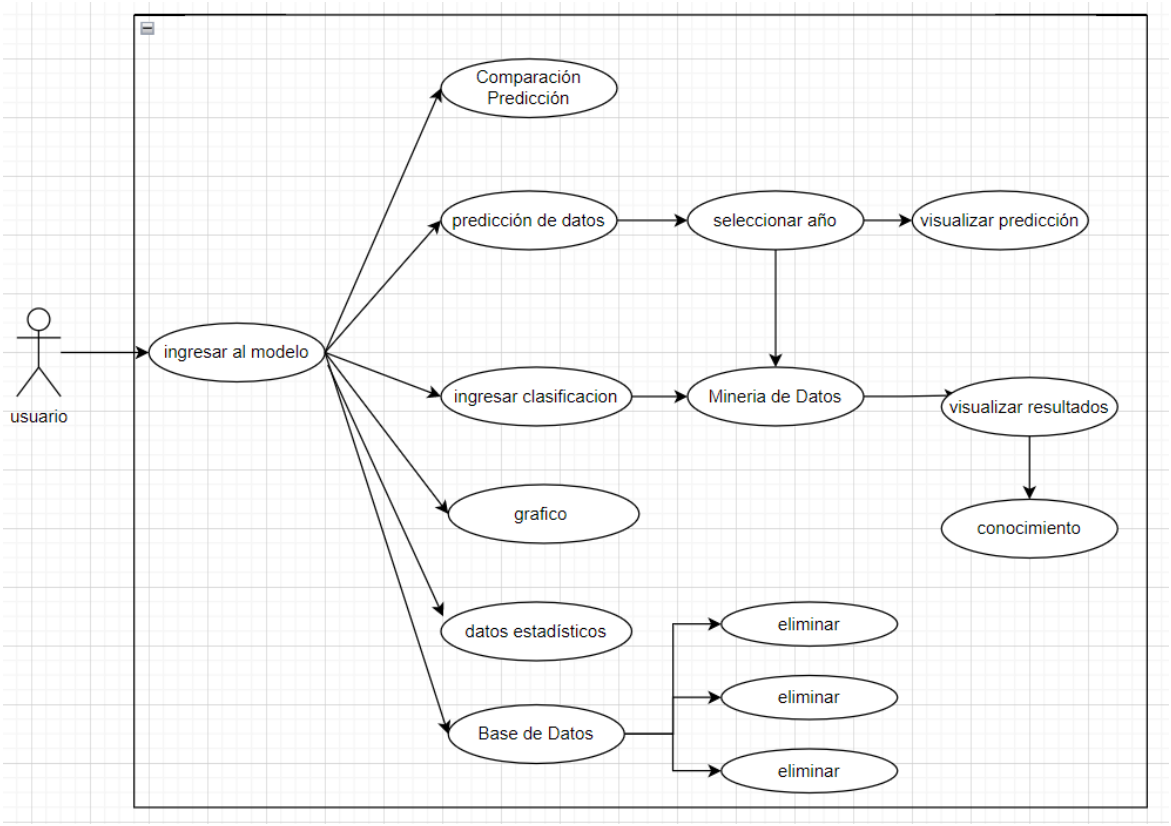
3.2.2.2 Análisis

a) Caso de uso general

Se diseña el caso de uso especifico donde se demuestra el proceso que tendrá el modelo.

Tabla 3.15

Caso de uso especifico



Nota: Descripción de Caso de uso específico del modelo

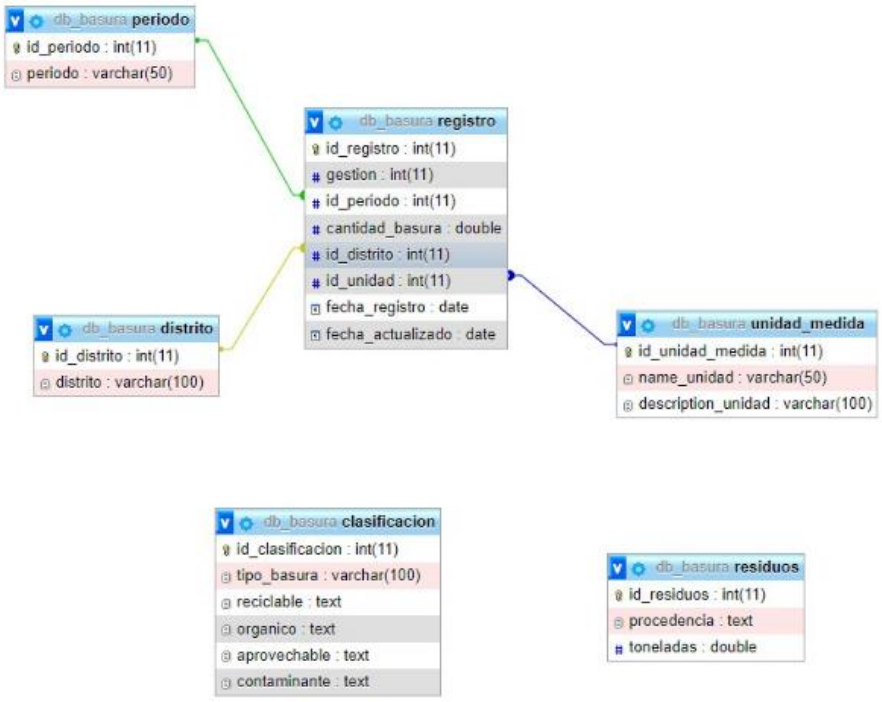
3.2.3 Fase de construcción

En esta fase se procedió al desarrollo del prototipo con la herramienta de NetBeans en lenguaje de programación Java.

3.2.3.1 Modelo físico del prototipo

Figura 3.18

Modelo Relacional



3.2.3.2 Diseño de interfases

a) Interfaz Home

Se desarrollo una interfaz inicial donde se muestra tres diferentes algoritmos de predicción en series de tiempo donde se elegirá el más a apropiado para realizar la predicción del índice de contaminación.

Figura 3.19

Interfaz home

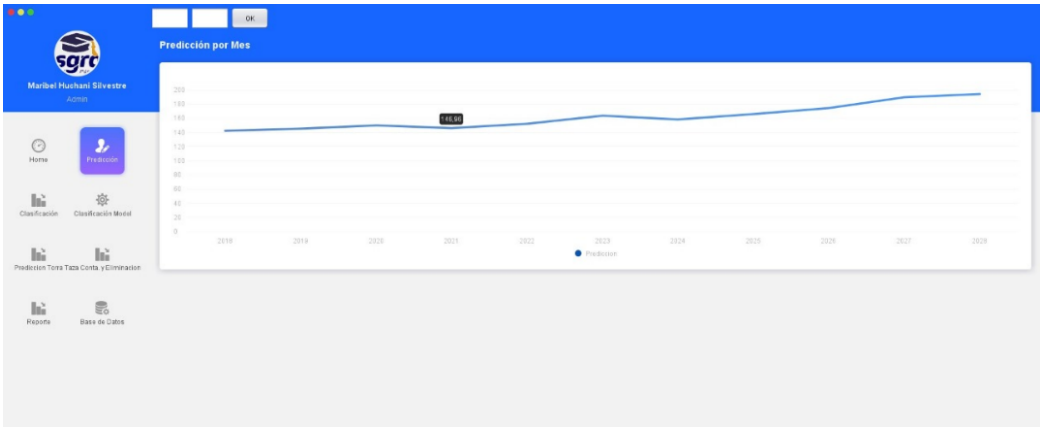
Nota: Interfaz gráfica de la pantalla de inicio del modelo

b) Interfaz de predicción

Se visualiza la proyección del índice de crecimiento de los residuos sólidos, donde el usuario podrá ingresar el año de inicio y el año final para lograr la predicción.

Figura 3.20

Interfaz predicción

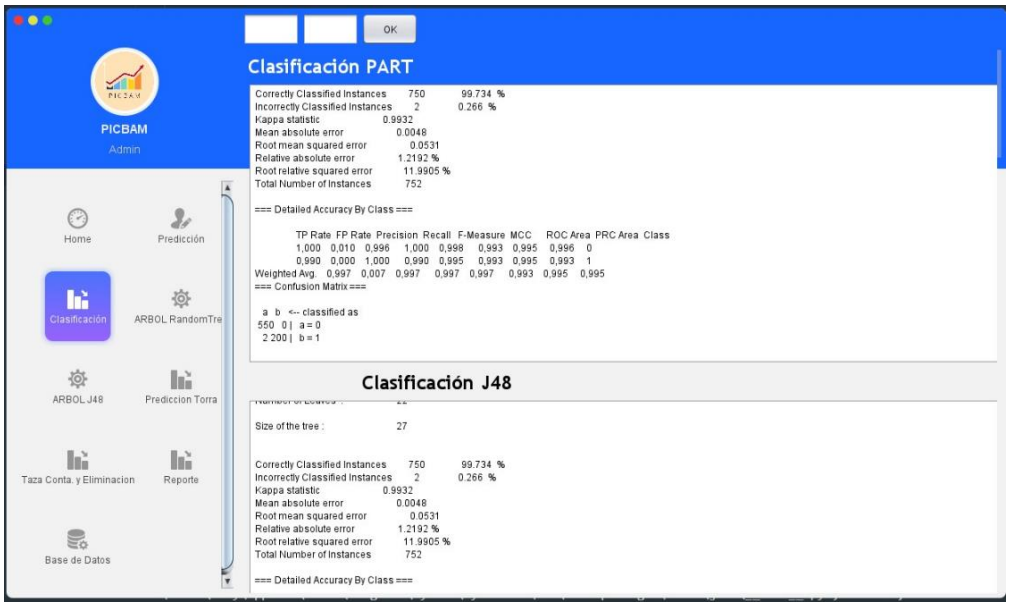


Nota: Interfaz gráfica con el Algoritmo de Series de tiempo seleccionado

c) Interfaz algoritmos de clasificación

Figura 3.21

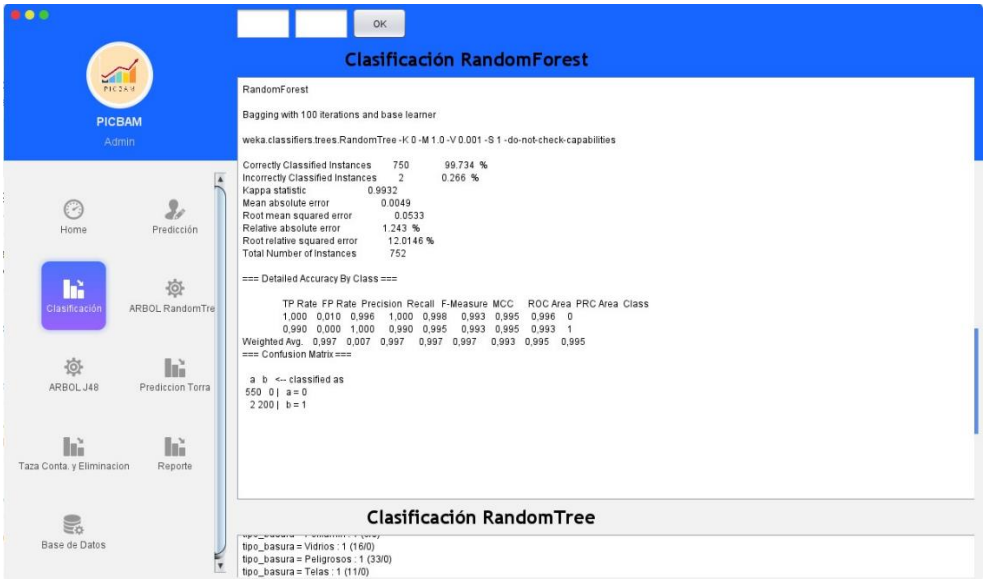
Interfaz algoritmos de clasificación PART



Nota: Interfaz gráfica de la resultante del entrenamiento del algoritmo PART

Figura 3.22

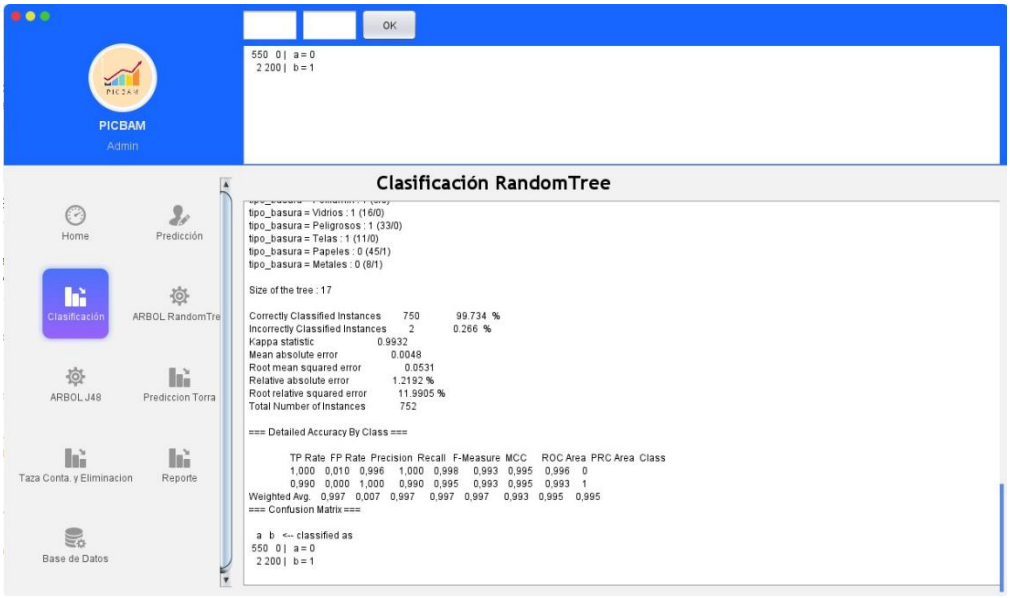
Interfaz algoritmo de clasificación J48



Nota: Interfaz gráfica de la resultante del entrenamiento del algoritmo J48

Figura 3.23

Interfaz algoritmo de clasificación RandomTree

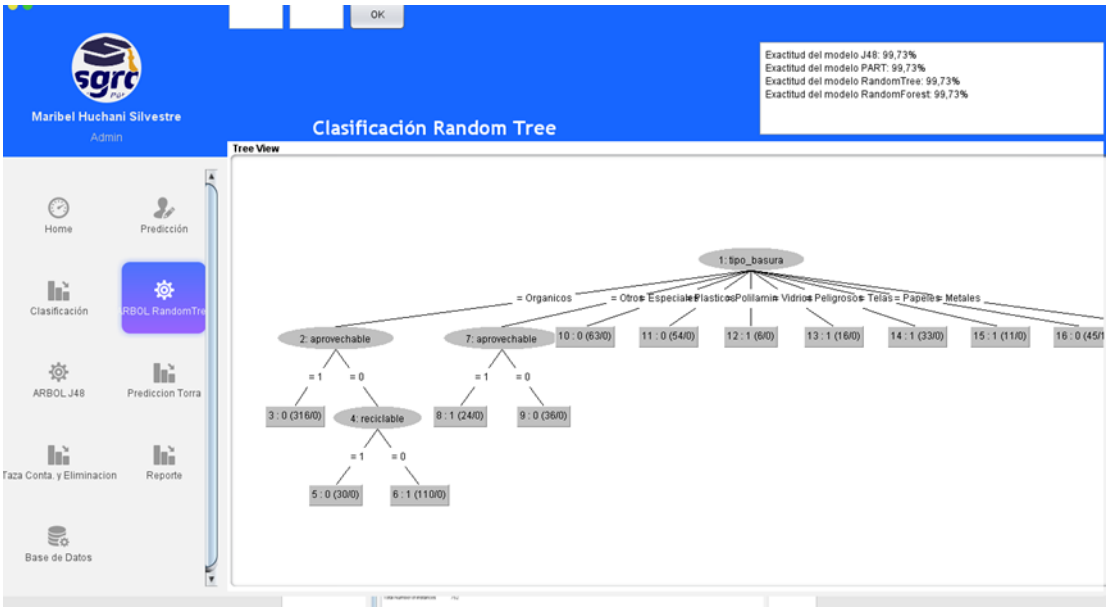


Nota: Interfaz gráfica de la resultante del entrenamiento del algoritmo RandomTree

d) Interfaz de diagrama de clasificación

Figura 3.24

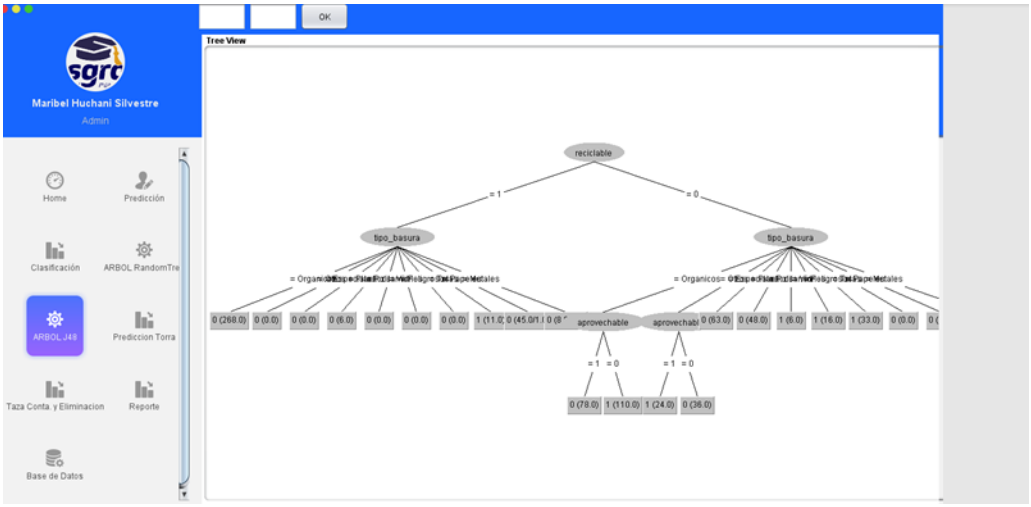
Interfaz generado por el algoritmo RandomTree



Nota: Interfaz gráfica del árbol generado por el algoritmo RandomTree

Figura 3.25

Interfaz generado por el algoritmo J48



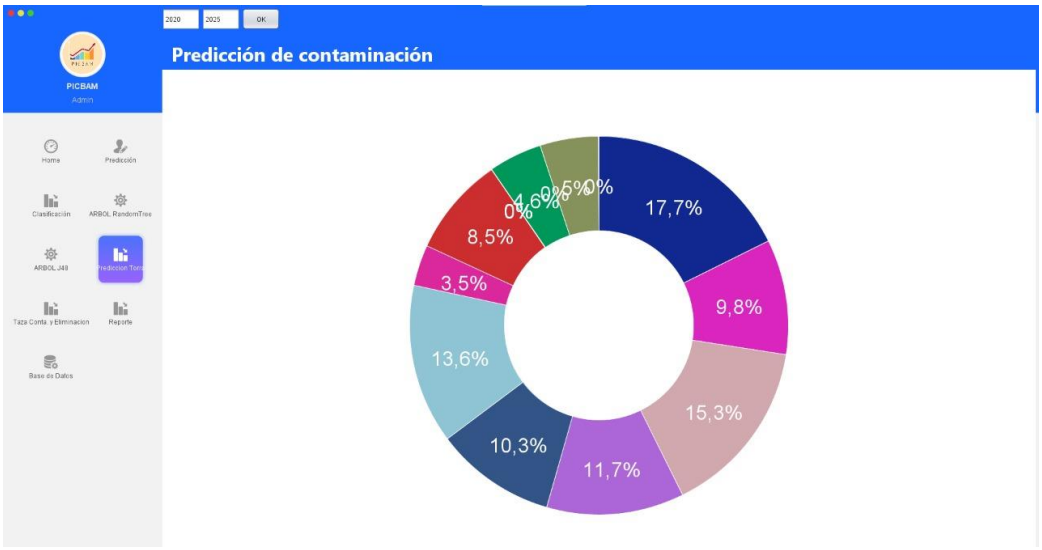
Nota: Interfaz gráfica del árbol generado por el algoritmo J48

e) Visualización de descripción general

se visualiza la cantidad de residuos sólidos generados por distrito de la Ciudad de El Alto.

Figura 3.26

Interfaz de porcentaje de generación de residuos sólidos



Nota: Interfaz gráfica del porcentaje del distrito que más residuos sólidos genera en la ciudad de El Alto

f) Visualización de datos generales

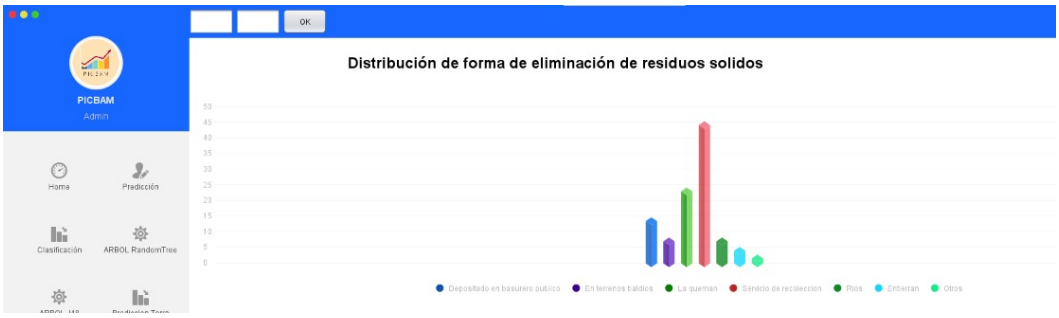
se visualizará el índice de forma en que las personas realizamos la eliminación de residuos sólidos en la Ciudad de El Alto.

De igual manera se visualiza el tipo de residuo solido más generado en la Ciudad de El Alto.

Y por último se visualiza el porcentaje de residuos sólidos aprovechables y no aprovechables.

Figura 3.27

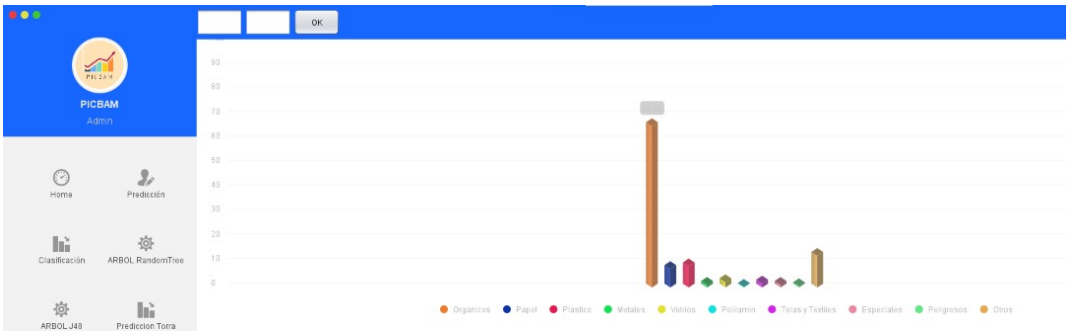
Interfaz de porcentaje de forma de eliminación de residuos solidos



Nota: Interfaz gráfica del porcentaje de la forma de eliminación de residuos sólidos que se genera en la ciudad de El Alto.

Figura 3.28

Interfaz de porcentajes de tipo de residuos solidos



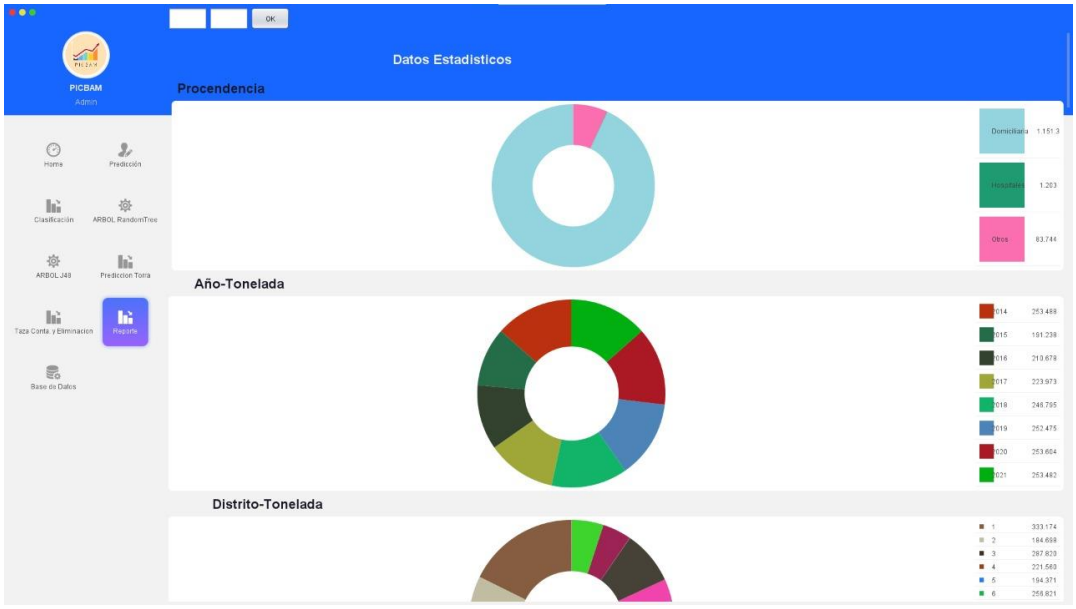
Nota: Interfaz gráfica del porcentaje del tipo de residuos sólidos que se genera en la ciudad de El Alto.

g) Datos actuales

Basado en los datos actuales se visualizará todos los datos existentes en porcentajes.

Figura 3.29

Interfaz datos estadísticos actuales



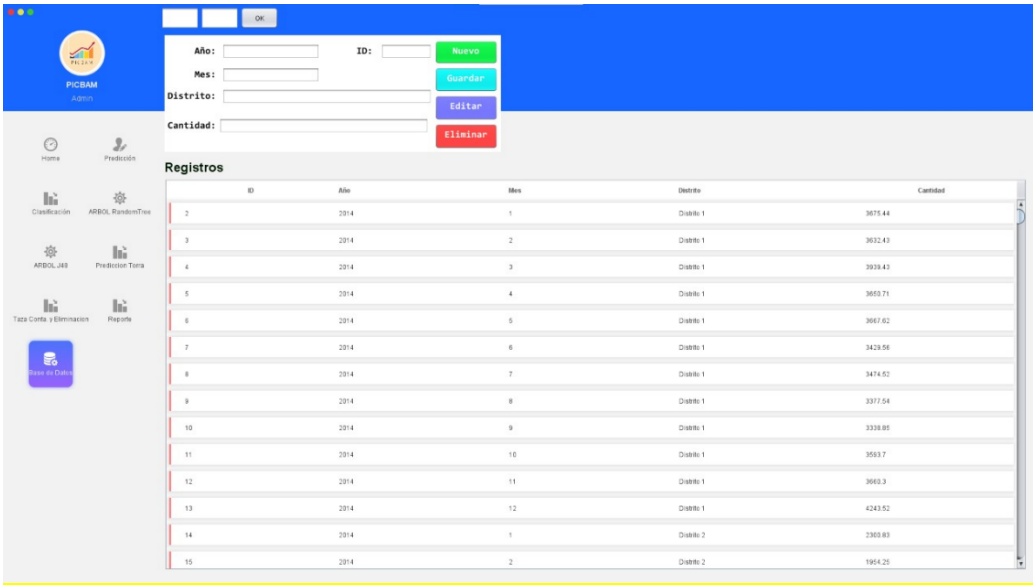
Nota: Interfaz gráfica de datos actuales de los residuos sólidos gestión 20222 de la ciudad de El Alto

h) Interfaz administrador

Se creo una base de datos donde se almacena todos los registros recuperados a través de informes para una adecuada administración del prototipo donde se tendrá las opciones de editar, eliminar y registrar nuevos datos.

Figura 3.30

Interfaz de la base de datos



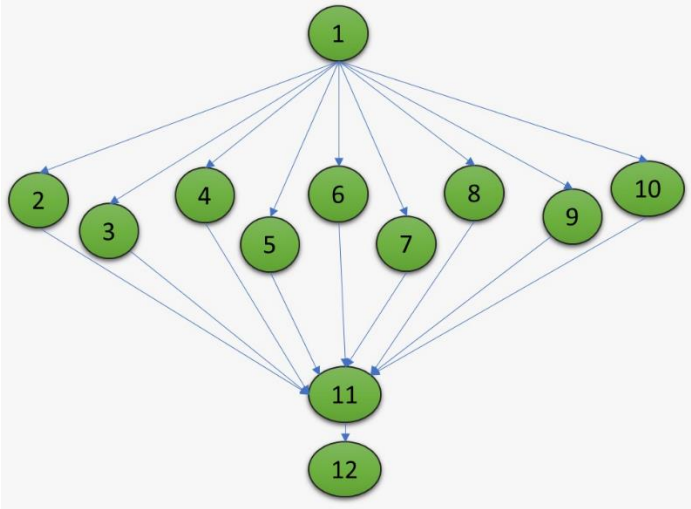
Nota: Interfaz gráfica donde se realiza la manipulación de datos con los que se trabajó para el entrenamiento de los algoritmos.

3.2.4 Fase de transición

3.2.4.1 Pruebas de caja negra

Figura 3.31

Nodos del Modelo



Nota: Nodos identificados en el modelo

1. INICIO
2. Pantalla Principal (Comparación de Predicción)
3. Predicción intervalo de año
4. Clasificación (PART, J48, RandomForest, RandomTree)
5. Árbol RandomTree
6. Árbol J48
7. Predicción Torta
8. Taza. Contaminación
9. Reportes
10. Base de datos (CRUD)
11. Fin ciclo
12. Fin modelo

3.3 MÉTRICAS DE CALIDAD

La familia ISO/IOE 25000 tiene como objetivo evaluar la calidad del producto software, es la evolución de las normas ISO/IEC 9126 e ISO/IEC 14598.

Para la medición de calidad se realizará a través de la métrica ISO/25010 el cual detalla como “el valor en el que un producto o sistema puede ser manejado por usuarios determinados para atender sus necesidades y lograr sus objetivos específicos con eficacia, eficiencia, libertad de riesgo y satisfacción en un ambiente propio de uso” (Organización Internacional para la Estandarización, 2011a).

3.3.1 Usabilidad

Se evalúa el esfuerzo necesario que se deberá invertir para usar el sistema en la siguiente tabla se menciona los valores.

Tabla 3.16

Cálculo de usabilidad

Nro.	PREGUNTAS	RESPUESTAS		% de SI
		SI	NO	
1	¿Aprendió a usar rápido el sistema?	9	1	90%
2	¿La vista de pantalla que vio fueron de su agrado?	10	0	100%
3	¿Las pantallas que vio fueron fáciles de comprender?	10	0	100%
4	¿El sistema responde rápido a sus solicitudes?	7	3	70%
5	¿El sistema le facilita el trabajo?	10	0	100%
6	¿El sistema reduce su tiempo de trabajo?	10	0	100%
7	¿Es fácil navegar por las distintas opciones?	10	0	100%
8	¿Las operaciones que se realizan no son complicadas?	10	0	100%
9	¿El sistema le proporciono las respuestas requeridas?	8	2	80%
10	¿El sistema no presento errores?	9	1	90%
Resultados de la usabilidad es de: 93%				

Nota: Resultado general de la Usabilidad.

Con el resultado obtenido se interpreta que se tiene 93% de calidad de uso

3.3.2 Confiabilidad

La confiabilidad del sistema define la probabilidad de operación libre de fallos en un entorno determinado y durante un tiempo específico. Para determinar la confiabilidad de un software especificamos desde el instante que empieza a funcionar es decir $t=0$, a partir de ese momento se realiza las observaciones pertinentes hasta un $t=n$.

$\boldsymbol{P(T \leq t)}$ **Probabilidad de fallas (termino en el cual el sistema trabaja sin fallas)**

$\boldsymbol{P(T \leq t) = 1 - F(t)}$**,** probabilidad de trabajo sin fallas (Tiempo en el cual no ocurren fallas en el sistema).

Para calcular la confiabilidad del sistema se debe tomar en cuenta el periodo de tiempo en el que se ejecuta el sistema, a partir de ello se irán obteniendo las muestras respectivas. $F(t) = f * e(-\mu * t)$

Donde:

f: Funcionalidad del sistema.

μ: Es la probabilidad de error que pude tener el sistema.

t: Tiempo de duración de gestión en el sistema.

Entonces, se considerará un periodo de 20 días como tiempo de prueba donde se define que de cada diez ejecuciones se presenta un fallo con el sistema.

Procedemos a realizar los cálculos respectivos:

$$F(t) = f * e(-\mu * t)$$

$$F(t) = 0{,}8962 * e(-110 * 20)$$

$$F(t) = 0{,}1212$$

$$F(t) = 12{,}12\ \%$$

Tomando en cuenta el resultado anterior y reemplazando el resultado en las fórmulas de probabilidades se tiene lo siguiente:

$$P(T \leq t) = F(t) \Rightarrow P(T \leq t) = 0{,}1212 = 12{,}12\%$$

$$P(T \leq t) = 1 - F(t) \Rightarrow P(T \leq t) = 1 - 0{,}1212$$

$$P(T \leq t) = 0{,}8788 = 87{,}89\%$$

Entonces, la confiabilidad del sistema es del 88% en un periodo de 20 días como tiempo de prueba.

3.3.3 Mantenibilidad

es la cualidad del software para determinar si debe ser modificar, corregir o mejorar. Para realizar este cálculo se utilizará el índice de madurez (IMS), para determinar la estabilidad del producto.

El índice de madurez se calcula con la siguiente formula:

$$IMS = \frac{M_t - (F_a + F_b + F_c)}{M_t}$$

Donde:

M_t: numero de módulos en la versión actual

F_a: numero de módulos en la versión actual que se han cambiado

F_b: numero de módulos en la versión actual que se han añadido

F_c: número de módulos en la versión actual que se han borrado en la versión actual

Los datos obtenidos son los siguientes:

Tabla 3.17

Estimación de valores

información	valor
M_t	8
F_a	0
F_b	1
F_c	0

Nota: Calculo general de IMS

Cálculo del IMS con los datos obtenidos:

$$IMS = \frac{M_t - (F_a + F_b + F_c)}{M_t}$$

$$IMS = \frac{8 - (0 + 1 + 0)}{8}$$

$$IMS = 87\%$$

Se concluye que el sistema tiene un índice de madurez del 87%, lo que indica que no se requiere mantenibilidad inmediata.

3.3.4 Eficiencia

Para pode obtener el cálculo de la eficiencia del sistema se consideró ponderar las características esenciales que el sistema desempeña.

Tabla 3.18

Evaluación de desempeño

Característica de desempeño	Ponderación
Rapidez de inicio	4
Rapidez de proceso	5
Proceso rápido de búsqueda	5
Fluidez	5
Disponibilidad	4

Nota: Cálculo general de la eficiencia.

En base a los datos de la anterior tabla se podría llegar a tener una idea de la eficiencia, para ello se utilizó la siguiente formula:

$$Eficiencia = \Sigma xi\ /n * 100/n$$

$$Eficiencia = 23/5 * 100/5$$

$$Eficiencia = 92\%$$

3.3.5 Calidad total

Es el resultado de la media de todos los cálculos realizados conforme a la norma ISO/IEC 25010 descritas de la siguiente manera:

Tabla 3.19

Cuadro de cálculo general

Atributos	Valor (%)
Usabilidad	93
Confiabilidad	88
Mantenibilidad	87
Eficiencia	92
Calidad global	**90%**

Nota: Cuadro general de resultados obtenidos

Se obtiene como calidad global del Modelo un 90% el cual indica que es aceptable.

3.4 EVALUACIÓN DE COSTOS

3.4.1 Adecuación funcional

Para esta parte se agrupará los atributos que calificaran si el producto de Software maneja las funciones para satisfacer las necesidades para las que fue diseñado.

Para este cálculo funcional se determinó las características de dominios de información de la siguiente manera:

- **Número de entradas de usuario:** es cuando el usuario proporciona distintos datos en el sistema.
- **Numero de salida de usuario:** son datos que ofrecen información al usuario
- **Número de peticiones:** se refieren a las consultas desde la base de datos que muestra información en el sistema
- **Numero de archivos:** se refieren a la base de datos o flujos de información requerido por el sistema
- **Numero de interfaces externos:** representa las interfaces externas con las que se conecta nuestro sistema.

Tabla 3.20

Cálculo de adecuación funcional

Nro.	PARÁMETROS DE MEDIDA	CANTIDAD
1	Número de entradas de usuario	6
2	Numero de salida de usuario	15
3	Número de peticiones	2
4	Numero de archivos	4
5	Numero de interfaces externos	0

Nota: Cuadro de resultados obtenidos del cálculo de la adecuación funcional

Una vez obtenida los parámetros de medida y cantidad se procede a calcular la cuenta total con el factor de ponderación media que se muestra en la siguiente tabla:

Tabla 3.21

Cálculo de cuenta total

Parámetros de medición	Cuenta total	Factor de ponderación	Valor obtenido
Número de entradas de usuario	5	3	15
Numero de salida de usuario	8	4	32
Número de peticiones	2	3	6
Numero de archivos	3	7	21
Numero de interfaces externos	0	5	0
TOTAL			74

Nota: Cuadro general del cálculo de la cuenta total

En la siguiente tabla se muestra el factor de ajuste de complejidad que están basadas en las respuestas de las siguientes preguntas formuladas:

Tabla 3.22

Cálculo factor de ajuste de complejidad

Nro.	Factores	0	1	2	3	4	5	*Fi*
1	¿Requiere el sistema copias de seguridad y de recuperación fiables?					X		4
2	¿Se requiere comunicación de datos?					X		4
3	¿Existen funciones de procesos distribuidos?				X			3
4	¿Es critico el rendimiento?					X		4
5	¿Sera ejecutado el sistema en un entorno operativo existente y fuertemente utilizado?					X		4

6	¿Requiere el sistema entrada de datos interactiva?		X			2
7	¿Se utilizaron los archivos maestros de forma iterativa?		X			2
8	¿Tiene facilidad operativa?			X		3
9	¿Son complejas las entradas, salidas y/o peticiones?		X			2
10	¿Es complejo el procesamiento interno?			X		3
11	¿Se ha diseñado el código para ser reutilizable?			X		3
12	¿Están incluidas en el diseño la conversación y la instalación?	X				2
13	¿Se ha diseñado el sistema para soportar diferentes instalaciones en diferentes organizaciones?	X				2
14	¿Se ha diseñado la aplicación para facilitar los cambios y para ser fácilmente utilizada por el usuario?				X	4
Factor ajuste de complejidad						42

Nota: cuadro general del cálculo del factor de ajuste de complejidad.

Para realizar el cálculo de punto de función se emplea la siguiente formula:

$$PF = CuentaTotal * \left(x + \min(Y) * \sum F_i\right)$$

Donde:

PF: medida de la adecuación funcional

Cuenta Total: es la sumatoria de numero de estradas, numero de salidas, número de peticiones, numero de archivos, y de numero de interfaces externas.

x: Confiabilidad del Proyecto, varía entre 1% o 100%

$\mathbf{min}(Y)$***:*** error mínimo aceptable de complejidad.

$\sum \boldsymbol{F_i}$**:** es el valor de ajuste de complejidad donde 1<=i=>14.

Reemplazando los valores obtenidos anteriormente se tendrá el siguiente resultado:

$$PF = Cuenta\ Total * \left[0.65 + (0.01 * \sum F_i)\right]$$

$$PF = 74 * (0.65 + 0.01 * 42)$$

$$PF = 74 * 1.07 = 74.18$$

Como resultado se obtuvo la funcionalidad de la aplicación móvil con el valor de 90% de funcionalidad el cual significa que le sistema responde de manera óptima a las funcionalidades requerida.

3.4.2 Aplicación de Cocomo II

Tabla 3.23

Factor LCD/PF de lenguaje de programación.

Lenguaje	Nivel	Factor LCD/PF
C	2.5	128
ANSI/basic	5	64
Java	6	53
PL/I	4	80
Visual Basic	7	46
ASP	9	36
PHP	11	29
Visual C++	9.5	34

Nota: Cuadro de factor LCD/PF

Se realizo el cálculo del costo del Modelo mediante las siguientes formulaciones:

$$E = a(KLDC)^b\ ;\ Personas - mes$$

$$Tdev = c(E)^d; Meses$$

$$P = E/Tdev; Personas$$

Dónde:

E: Esfuerzo requerido por el proyecto expresado en persona-mes.

D: Tiempo requerido por el proyecto expresado en meses.

P: Número de personas requeridas para el proyecto.

A, B, C y D: Constantes con valores definidos según cada sub-modelo.

KLDC: Cantidad de líneas de código distribuidas en miles

A la vez cada modelo se subdivide en tres modos:

Modo orgánico: Es un pequeño grupo de programadores experimentados desarrollando proyectos de software en un entorno familiar.

Modo semilibre: Corresponde a un esquema intermedio entre el modo orgánico y el rígido, el grupo de desarrollo puede incluir una mezcla de personas experimentadas y no experimentadas.

Modo rígido: El proyecto tiene fuertes restricciones, que pueden estar relacionadas con la funcionalidad y/o pueden ser técnicas.

A continuación, se describe las constantes de acuerdo a los modos mencionados anteriormente.

Tabla 3.24

Constantes a b c d COCOMO

Modo	A	B	C	D
Orgánico	2.4	1.05	2.5	0.38
Semilibre	3.0	1.12	2.5	0.35
Rígido	3.6	2.20	2.5	0.32

Nota: Constantes de Modos

3.4.2.1 Costos del sistema

Para la estimación de costos del sistema ha sido desarrollado bajo las KLDC (Kilo Líneas de Código) las que detallamos lo siguiente:

$$LDC = PFA * Factor\ LDC\ /PF$$

$$LDC = 74.18 * 53$$

$$LDC = 3431.54$$

$$KLDC = 3431.54\ /1000$$

$$KLDC = 3.43$$

Esfuerzo Requerido por el proyecto, en persona-mes

$$E = a(KLDC)^b\ ;\ Personas\ -\ mes$$

$$E\ =\ 2.4\ *\ (3.43)^{1.05}\ = 8\ \text{persona} - \text{mes}$$

Cálculo del tiempo requerido por el proyecto, en meses

$$Tdev\ =\ c(E)^d;\ Meses$$

$$Tdev =\ 2.5\ *\ (8)^{0.38}\ =\ 5\ \text{meses}$$

Para el cálculo del número de programadores para el desarrollo de software:

$$P\ = E/Tdev\ ;\ Personas$$

$$P\ = 8/5 =\ 2.33\ = 1\ personas$$

Estimando que el salario mínimo nacional es de Bs. 2250, esta cifra será tomada en cuenta para la siguiente estimación:

*Costo del software = Nro. de programadores * Salario de un programador.*

*Costo del Software por Persona = 1 * 2250 = 2250 Bs.*

*Costo total del Software Desarrollado = 2250 * 5= 11250 Bs.*

Tabla 3.25

Costo total del Modelo

Descripción	Estimación
Personas necesarias por mes para llevar adelante el proyecto	8 personas-mes
Tiempo de desarrollo del proyecto	5 meses
Personas necesarias para realizar el proyecto	1 persona
Costo total del proyecto	**Bs 11250**

Nota: Cuadro general del costo total del Modelo.

Lo que significa que el costo del sistema desarrollado por los 5 meses es bs 11250, equivalente en dólares a $ 1616.

CAPÍTULO IV

Pruebas y Resultados

4 PRUEBAS Y RESULTADOS

El presente capitulo tiene como principal objetivo dar a conocer las pruebas de Hipótesis del presente trabajo de investigación y su interpretación.

4.1 PRUEBAS AL MODELO

- **Hipótesis nula**

  H_0 : "Aplicando las técnicas de la Minería de Datos y el área de la ingeniería de software, se desarrollará el Modelo de Proyección que información con datos erróneos de la contaminación de la Ciudad de El Alto".

- **Hipótesis de la investigación**

  H_1 : "Aplicando las técnicas de la Minería de Datos y el área de la Ingeniería de Software, se desarrollará el Modelo de Proyección de índice de la contaminación de la basura en la Ciudad de El Alto, esto con una eficiencia al 95% para coadyuvar en la toma de decisiones a las Autoridades competentes que tienen como visión la planificación como pilar fundamental en las políticas de salubridad de la urbe alteña".

  Se realizó 84 pruebas para la primera muestra del Modelo de proyección, 71 pruebas fueron correctas y 13 pruebas fue incorrecta.

  Donde:

  n= Muestra

  P=proporción de la muestra

  α = nivel de confianza

  $$H_0 = P_0 >= 95\%$$

  $$H_1 = P_1 < 95\%$$

  $$\overline{P} = 0.845$$

  $$n = 84$$

  $$\alpha = 0.05$$

4.1.1 Aplicación de la Distribución Estándar Normal

Se formula de la siguiente manera:

$$z = \frac{(\overline{P} - P_0)}{\frac{\sqrt{P_0 * (1 - P_0)}}{n}}$$

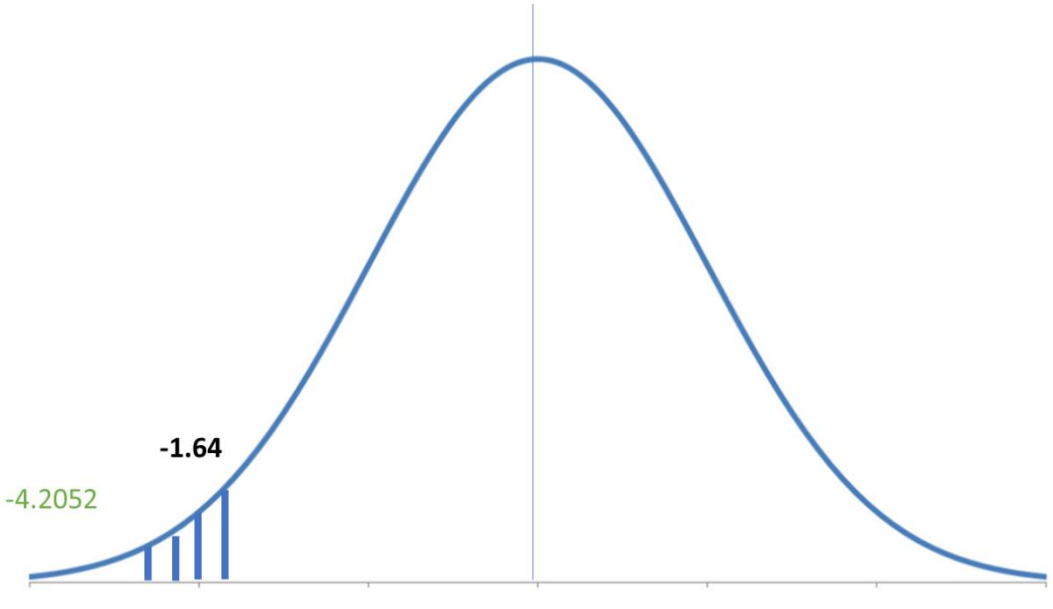
$$z = \frac{(0{,}845 - 0{,}95)}{\frac{\sqrt{0{,}95*}(1 - 0{,}95)}{84}} = \frac{(-0{,}05)}{\frac{\sqrt{0{,}95*}(0{,}05)}{84}}$$

$$Z = -4.2052$$

Zona critica izquierda -1.64

Figura 4.1

Distribución Estándar Normal



Nota: Resultado de la distribución normal que cae en la zona critica.

Se obtiene como resultado el cálculo z = -4.2052 de por lo que cae en la zona de crítica, por lo que se rechaza la Hipótesis Nula H_0, y se acepta la Hipótesis Alternativa H_1, con un intervalo de confianza de 95 %.

4.2 PRUEBAS AL MODELO

Se describe el margen de error que tiene los algoritmos de series de tiempo utilizados en el modelo.

Tabla 4.1

Margen de error de los modelos de serie de tiempo

Models	***Margen de Error***
ARIMA	0.015
LSTM	0.02
Prophet	0.03

Nota: Información del margen de error de cada Modelo

Figura 4.2

Resultados obtenidos con los algoritmos de series de tiempo.

Mes	Toneladas	ARIMA_Predictions	LSTM_Prediccion	Prophet_Predictions
2014-01-01	2234.06	2235.125572	2261.492198	2288.607421
2015-01-01	2252.38	2211.874581	2328.061867	2367.956464
2016-01-01	2306.78	2240.550001	2345.361995	2477.788100
2017-01-01	2339.73	2230.575985	2378.646503	2443.464148
2018-01-01	2467.95	2258.968486	2423.315472	2492.298393
2019-01-01	2424.75	2313.840778	2483.620628	2571.647436
2020-01-01	2536.04	2528.464361	2539.457555	2681.479072
2021-01-01	2584.82	2486.474718	2611.899732	2647.155120

Nota: Resultados obtenidos de cada modelo comparando con el original.

Figura 4.3

Gráfica de algoritmos de Series de tiempo

Nota: Comparación de los modelos de Predicción

Se puede observar que el algoritmo ARIMA tiene un menor porcentaje de margen de error y es el más eficiente para el uso, por el cual se decidió usar este algoritmo como base para la predicción de residuos sólidos.

CAPÍTULO V

Conclusiones y Recomendaciones

5 CONCLUSIONES Y RECOMENDACIONES

El presente capítulo tiene como principal objetivo describir las conclusiones que se llegó a obtener en el presente trabajo y las recomendaciones para futuras investigaciones y/o mejoras en el Modelo.

5.1 CONCLUSIONES

Una vez concluida el trabajo de investigación y haber realizado el desarrollo del modelo, se ha logrado el objetivo principal planteado anteriormente: “Desarrollar un Modelo de Proyección de índice de contaminación de la basura para la Ciudad de El Alto basado en las técnicas de la Minería de Datos, esto para la toma de decisiones en el resguardo del medio ambiente y la sociedad en general de la Ciudad de El Alto”. En el capítulo lll se demuestra el análisis de los datos, la aplicación de la minería de datos y el desarrollo del modelo de proyección PICBAM llegando a la totalidad del objetivo.

Con respecto a los objetivos específicos se cumplió de la siguiente manera:

- “Recolección y análisis de los datos de la contaminación de la basura de la Ciudad de El Alto”, se realizó la solicitud de datos a las instituciones correspondiente y se aplicó las primeras fases de la metodología CRISP-DM y KDD para su respectivo análisis y preparación.
- “Analizar y seleccionar las técnicas y métodos de la Minería de Datos que sean útiles para el Modelo de Proyección”, se procedió a realizar el entrenamiento de los datos en Weka donde se eligió a los algoritmos más eficientes.
- “Aplicar la metodología CRIPS-DM y KDD para el manejo de la información Diseñar un Modele de Proyección del Índice de crecimiento de la basura aplicando pruebas de Algoritmo de Minería de Datos seleccionados”, ambas metodologías se aplicaron en todo el proceso de la minería de datos descritas en el capítulo lll.
- “Interpretar el resultado obtenido del Modelo de Proyección de la basura generada en la Ciudad de El Alto”, se trabajó el desarrollo del Modelo con la

metodología Open Up con el ID NetBeans logrando interfaces comprensivas para la visualización de resultados e interpretación de los mismos.

Mediante el desarrollo de las metodologías desarrollados en el capítulo lll y realizando las respectivas pruebas al Modelo se dio solución al problema principal del presente trabajo de investigación:

"¿De qué manera ayudaría un Modelo de Proyección del índice de la contaminación de la basura en la Ciudad de El Alto, basado en Minería de Datos, para la toma de decisiones?"

Al desarrollar un Modelo de Proyección donde se visualiza datos futuros sobre la contaminación en la ciudad de El Alto, las autoridades competentes podrán tomar decisiones previniendo los riesgos y peligros de salud a los que se encuentran expuestos la población alteña.

5.2 RECOMENDACIONES

Se propone las siguientes recomendaciones, con el fin de mejorar el modelo de proyección:

- Registrar más datos históricos a la base de datos para una mejor proyección.
- Solicitar datos de registro de los residuos sólidos en diferentes instituciones, y complementar a la base de datos existente en el modelo.
- Desarrollar una segunda versión con más registros sobre los residuos sólidos de la Ciudad de El Alto, en un entorno web.
- Se recomienda estudiar los diferentes algoritmos de la Minería de Datos para una mejor interpretación de resultados.

BIBLIOGRAFÍA

Aggarwal, C. C. (2015). *Data Mining.* Springer International Publishing Switzerland.

Andersen, L., Del Granado, S., & Doyle, A. (2016). *Basura. El Abc Del Desarrollo En Bolivia.*

Arimetrics. (2020). *Arimetrics*. Https://Www.Arimetrics.Com/Glosario-Digital/Modelo-Predictivo

Aws. (2022). *Aws*. Https://Aws.Amazon.Com/Es/What-Is/Java/

Beltran Martinez, B. (S.F.). *Minería De Datos.*

Beltran Prieto, P. (2 de diciembre De 2020). *Medico Plus*. Https://Medicoplus.Com/Ciencia/Contaminacion-Basura

Bhatia, P. (2019). *Data Mining and Data Warehousing, Principles and Practical Techniques.* Cambridge University Press.

Caminiti, G. (31 De Agosto De 2021). *Coderhouse.* Https://Www.Coderhouse.Cl/Blog/Que-Es-Python?Utm_Term=&Utm_Campaign=9&Utm_Source=Google_Performance_Max&Utm_Medium=Cpc&Gclid=Cj0kcqjwpeaybhdxarisaezitbeyxbeygqbowpuw_Wrelqgvdwfx6gunq3yfjei-Qadz0qgyz4hgsymaamg6ealw_Wcb

Carrizosa, E. (2005). *Modelos Matematicos para la Mineria de Datos.*

Collado, D. C., & Lucio, D. M. (2014). *Metodologia de la Investigacion.* Mcgraw-Hill / Interamericana Editores, S.A. De C.V.

Ctma. (18 De Marzo De 2021). *Ctma*. Https://Ctmaconsultores.Com/Norma-Iso-25000/

Davenport, & Prusak. (1999). *Business Intelligence and Operation.*

Dean, J. (2014). *Big Data, Data Mining, and Machine Learning.* Sas Institute Inc.

Gómez, A., Del C.López, M., & Migani, S. (2014). *Cocomo un Modelo de Estimacion de Proyectos de Software.*

Gonzales Rocabado, A. (2019). *Iisec*. Https://Iisec.Ucb.Edu.Bo/Publicacion/La-Basura-Un-Problema-Creciente-En-Bolivia

Gonzales Rocabado, A. (2019). *Iisec Instituto de Investigaciones Socio - Economicas*. Https://Iisec.Ucb.Edu.Bo/Publicacion/La-Basura-Un-Problema-Creciente-En-Bolivia

Han, J., & Kamber, M. (2006). *Data Mining: Concepts And Techniques.* Morgan Kaufmann.

Ibm. (24 De Marzo De 2022). Https://Www.Ibm.Com/Docs/Es/Db2/11.1?Topic=Miner-Data-Mining-Process

Iebs, C. (20 De Mayo De 2015). *Metodologias Agiles.* Https://Comunidad.Iebschool.Com/Metodologiasagiles/General/Concepto-Metodologias-Agiles/

Ine. (2022). Www.Ine.Gob.Bo

Ine. (Septiembre De 2022). *Instituto Nacional de Estadistica*. Https://Www.Ine.Gob.Bo/

Iso 9001, N. (19 De Julio De 2022). *Norma Iso 9001*. Https://Www.Isotools.Org/Normas/Calidad/Iso-9001/#:~:Text=La%20iso%209001%20es%20una,Su%20tama%C3%B1o%20o%20actividad%20empresarial.

Java. (2022). *Java*.

Jones, H. (2019). *Minería de Datos Guía de Minería de Datos para Principiantes, Que Incluye Aplicaciones Para Negocios, Técnicas de Minería de Datos, Conceptos y Más.*

López, J. M., & Herrero, J. G. (2006). *Técnicas de Análisis de Datos Aplicaciones Prácticas Utilizando Microsoft Excel Y Weka.*

Luna, G. L. (S.F.). *Minería de Datos: Cómo Hallar Una Aguja En Un Pajar.*

Mailund, T., & Denmark, A. (2017). *Beginning Data Science In R, Data Analysis, Visualization, And Modelling For The Data Scientist.*

Masters, T. (2018). *Data Mining Algorithms In C++.* Timothy Masters.

Medina González, L. E. (23 De Febrero De 2014). *Open Up.* Http://Openup3.Blogspot.Com/2014/02/Metodologia-Open-Up.Html

Microsoft. (Octubre De 2021). *Microsoft.* Https://Www.Microsoft.Com/Windows10

Mitchell, T. M. (1997). *Machine Learning.*

Monjas, Y. B. (S.F.). *Minería de Datos.* Madrid, España.

Node.Js. (2022). *Node Js.* Https://Nodejs.Org/Es/About/

Orallo, J. H., Quintana, J. R., & Ramírez, C. F. (2005). *Intrroduccion a la Mineria de Datos.* Pearson Educacion S.A.

Ramos, C. (S.F.). *Mineria de Datos, Modelos Y Algoritmos Aplica la Conocimiento Al Analisis Predictivo.*

Ramos, G. (2018). *Eabolivia.*

Rapidminer. (2022). *Rapidminer*. Https://Rapidminer.Com/

Ratner, B. (2007). *Statistical And Machine-Learning Data Mining.*

Riquelme, J. C., Ruiz, R., & Gilbert, K. (2006). *Minería de Datos: Conceptos Y Tendencias.*

Rivera, I. W. (2006). *Minería de Datos: Herramienta De Apoyo En La Seleccion De Equipos De Proyectos Informaticos.* Cujae.

Robleado, A. (24 De Septiembre De 2019). *Open Webinars*. Https://Openwebinars.Net/Blog/Que-Es-Mysql/

Roig, J. G., Roma, J. C., Alfonso, J. M., & Quiles, R. C. (2017). *Minería de Datos: Modelos y Algoritmos.* Oberta Uoc Publishing, Sl.

Sas. (2021). *Sas*. Https://Www.Sas.Com/Es_Mx/Insights/Analytics/Data-Mining.Html

Secretaria Municipal De Agua, S. G. (2021). La Paz.

Suárez, R., & Amador, A. D. (2009). *Herramientas De Minería De Datos.*

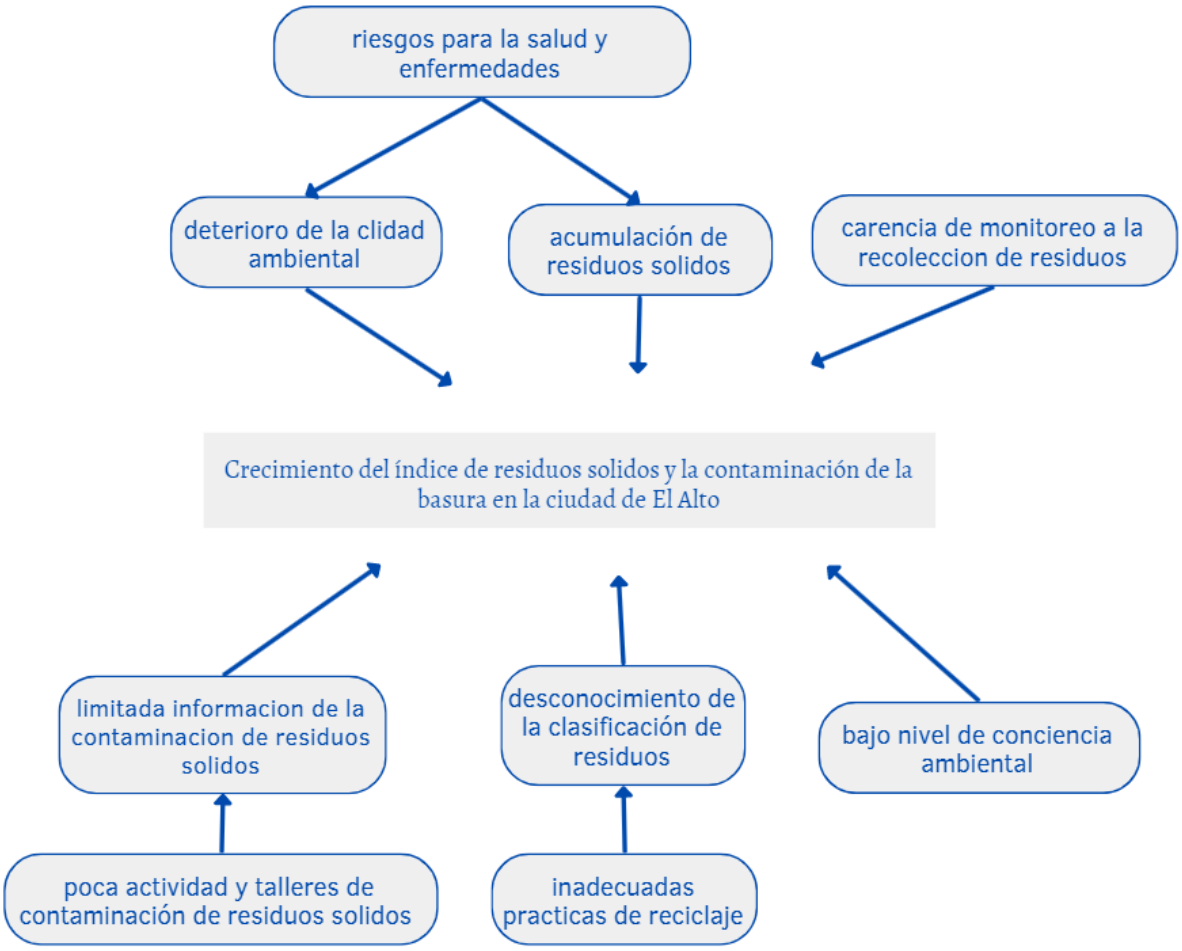
Tan, Y., & Shi, Y. (2016). *Data Mining And Big Data.* Board.

Waikato. (2022). *Waikato*. Https://Www.Cs.Waikato.Ac.Nz/Ml/Weka/

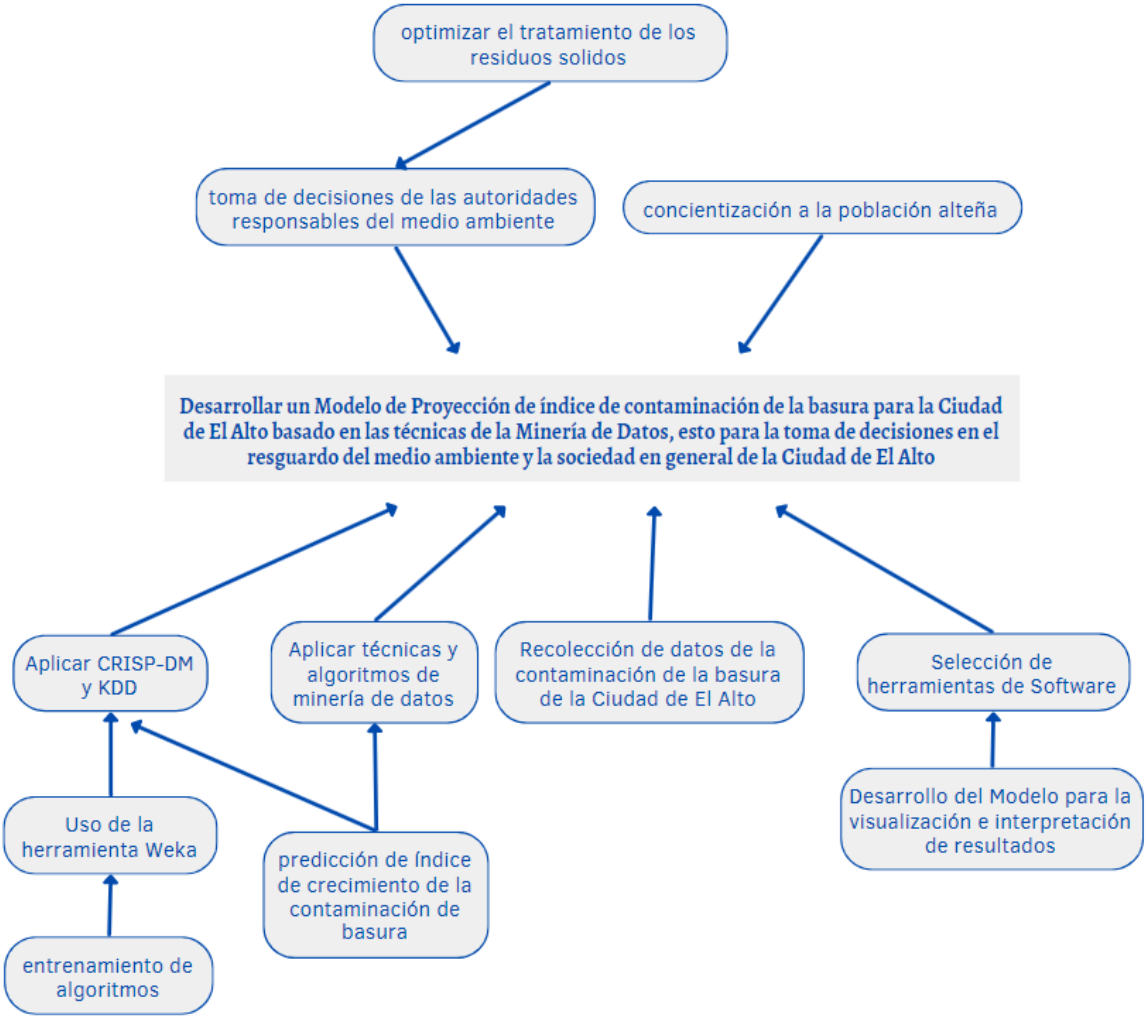
Witten, I. H., & Frank, E. (2005). *Data Mining Practical Machine Learning Tools And Techniques.* Acid-Free Paper.

ANEXOS

ANEXO 1. ÁRBOL DE PROBLEMAS



ANEXO 2. ÁRBOL DE OBJETIVOS



La Paz - El Alto Noviembre de 2022

Señor

M. Sc. Ing. David Carlos Mamani Quispe

DIRECTOR DE CARRERA
INGENIERÍA DE SISTEMAS

Presente.-

REF.: AVAL DE CONFORMIDAD

Distinguido ingeniero:

Mediante la presente tengo a bien comunicarle mi conformidad del trabajo de grado:

TITULO: MODELO DE PROYECCIÓN DE ÍNDICE DE CONTAMINACIÓN DE BASURA EN LA CIUDAD DE EL ALTO APLICANDO LA MINERÍA DE DATOS.

MODALIDAD: TESIS DE GRADO.

Univ.: Maribel Huchani Silvestre

Registro Universitario: 200012160

Cedula de identidad: 9965652 LP

Para su defensa pública y evaluación correspondiente a la materia de Taller de Grado II, de acuerdo al reglamento vigente de la carrera de Ingeniería de Sistemas de la Universidad Pública de El Alto.

Atentamente:

M. Sc. Ing. Enrique Flores Baltazar
TUTOR METODOLÓGICO
TALLER DE GRADO II

La Paz - El Alto Noviembre de 2022

Señor:

M. Sc. Ing. Enrique Flores Baltazar

TUTOR METODOLÓGICO
TALLER DE GRADO II

Presente.-

REF.: AVAL DE CONFORMIDAD

Distinguido tutor metodológico:

Mediante la presente tengo a bien comunicarle mi conformidad del trabajo de grado:

TITULO: MODELO DE PROYECCIÓN DE ÍNDICE DE CONTAMINACIÓN DE BASURA EN LA CIUDAD DE EL ALTO APLICANDO LA MINERÍA DE DATOS.

MODALIDAD: TESIS DE GRADO.

Univ.: Maribel Huchani Silvestre

Registro Universitario: 200012160

Cedula de identidad: 9965652 LP

Para su defensa pública y evaluación correspondiente a la materia de Taller de Grado II, de acuerdo al reglamento vigente de la carrera de Ingeniería de Sistemas de la Universidad Pública de El Alto.

Atentamente.

M. Sc. Ing. Ramiro Kantuta Limachi
TUTOR REVISOR

La Paz - El Alto Noviembre de 2022

Señor:

M. Sc. Ing. Enrique Flores Baltazar

TUTOR METODOLÓGICO
TALLER DE GRADO II

Presente.-

REF.: AVAL DE CONFORMIDAD

Distinguido tutor metodológico:

Mediante la presente tengo a bien comunicarle mi conformidad del trabajo de grado:

TITULO: MODELO DE PROYECCIÓN DE ÍNDICE DE CONTAMINACIÓN DE BASURA EN LA CIUDAD DE EL ALTO APLICANDO LA MINERÍA DE DATOS.

MODALIDAD: TESIS DE GRADO.

Univ.: Maribel Huchani Silvestre

Registro Universitario: 200012160

Cedula de identidad: 9965652 LP

Para su defensa pública y evaluación correspondiente a la materia de Taller de Grado II, de acuerdo al reglamento vigente de la carrera de Ingeniería de Sistemas de la Universidad Pública de El Alto.

Atentamente.

Ing. William Roque Roque
TUTOR ESPECIALISTA

Manual de Usuario

PICBAM

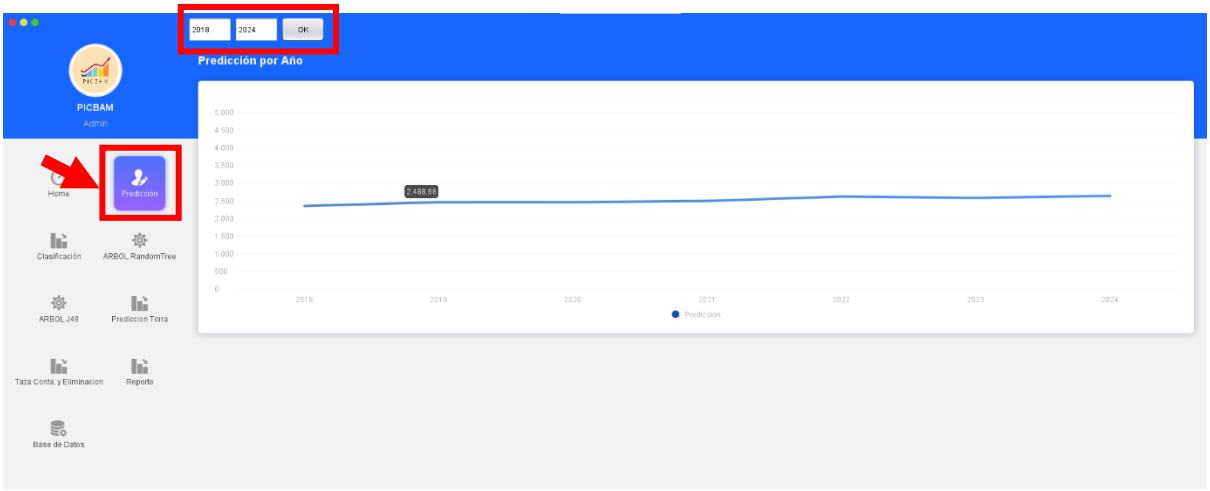
Modelo de Proyección del índice de contaminación de la basura de la Ciudad de El Alto

HOME

Se visualiza la cantidad de residuos sólidos generados del año 2014 hasta 2021 comparando con algoritmos de predicción en serie de tiempo (ARIMA, LSTM, PROPHET)

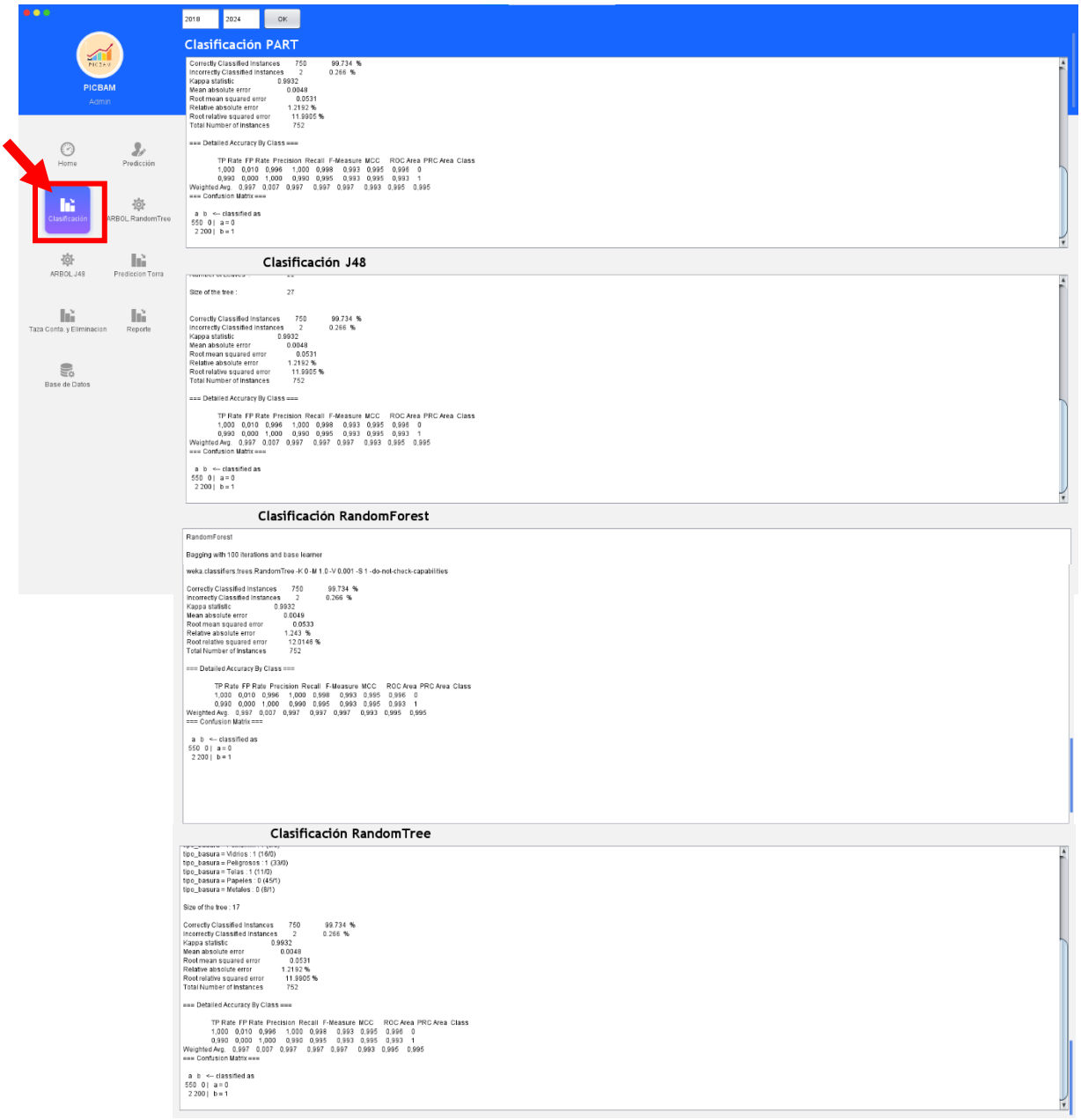
PREDICCIÓN

En esta opción se muestra la predicción hasta el año que desee, primero deberá ingresar parámetros el cuales son año-inicio/año-fin para realizar la muestra de la predicción.



CLASIFICACIÓN

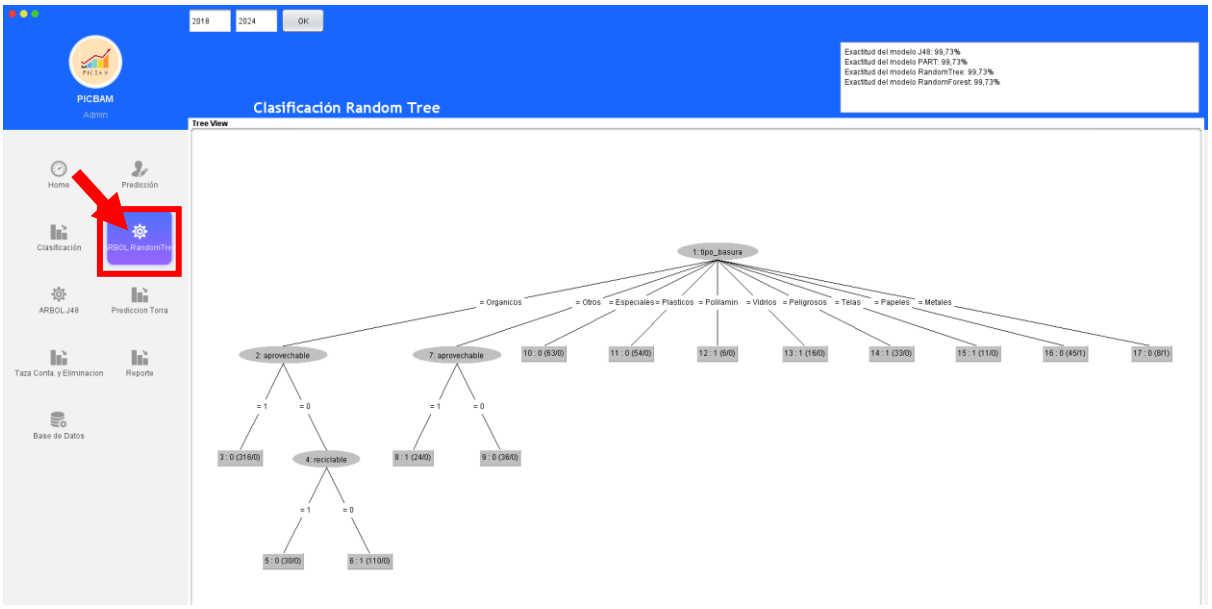
Muestra los diferentes tipos de clasificación con los algoritmos PART, J48, RandomForest y RandomTree para realizar la comparación de diferentes tipos de que determina el porcentaje de contaminación de los residuos sólidos.



GRÁFICAS GENERADAS DE LOS ALGORITMOS

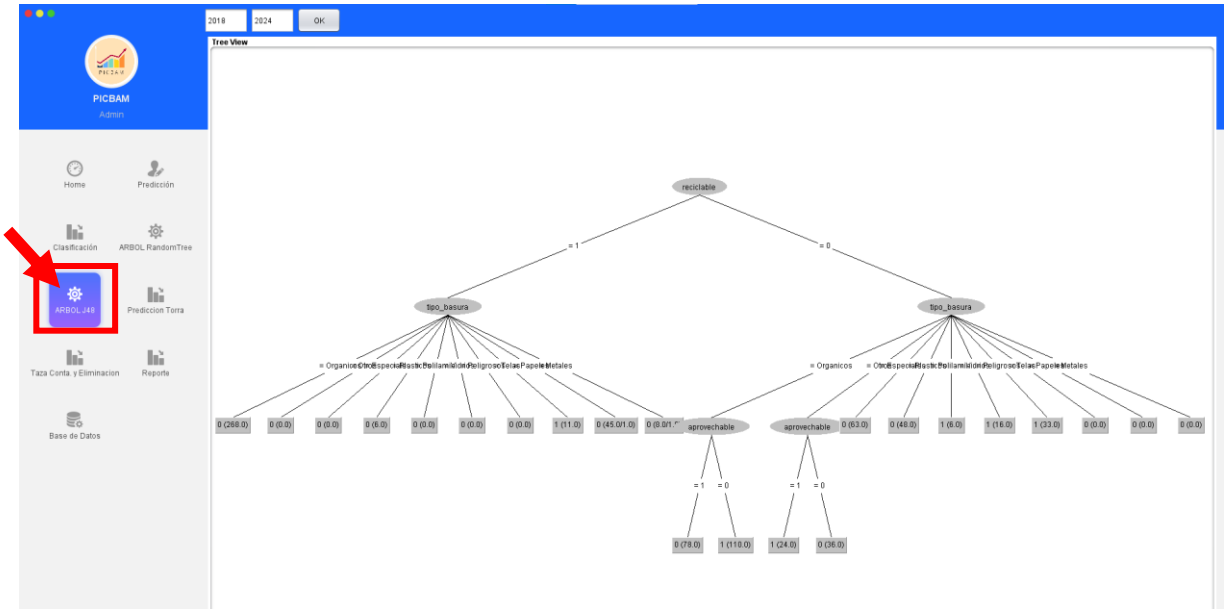
ÁRBOL RandomTree

Visualiza el árbol generado por el algoritmo Random Tree



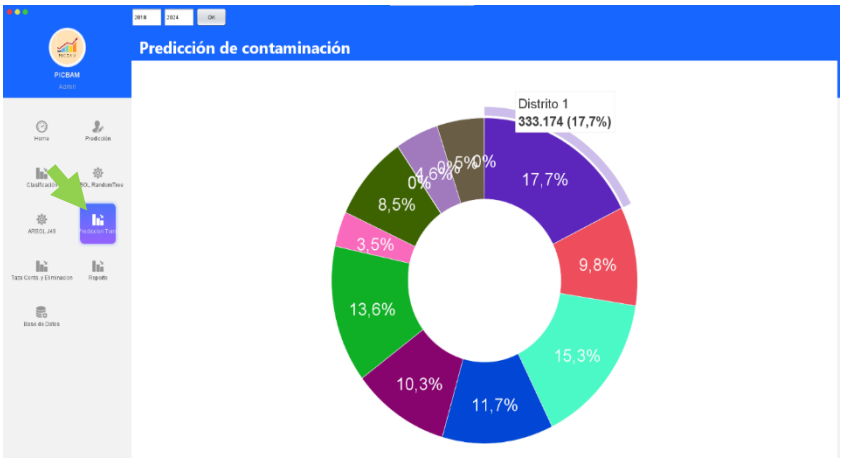
ÁRBOL J48

Visualiza el árbol generado por el algoritmo J48.



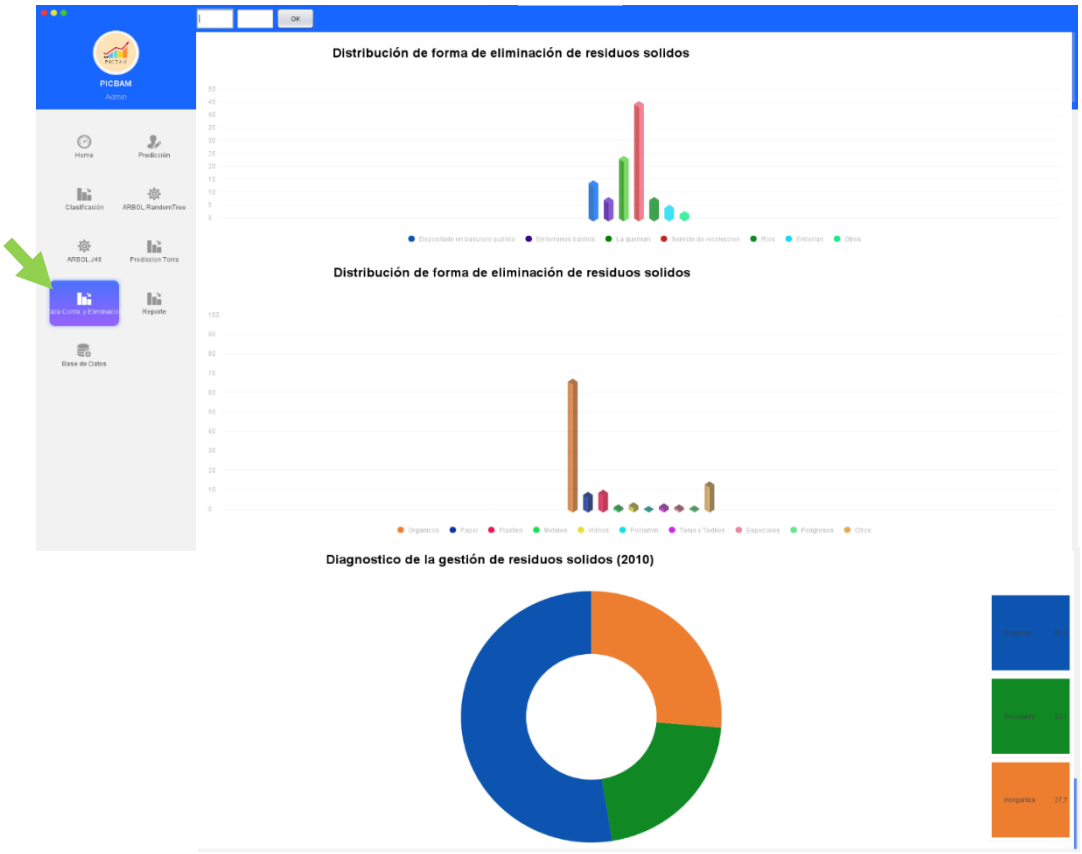
GRÁFICAS

En la siguiente opción se visualiza la cantidad de basura generado por distritos de la Ciudad de El Alto indicando el distrito que más genera residuos y el que más contamina en este caso el **Distrito 1** genera el 17,7%



Taza Contaminación y Eliminación

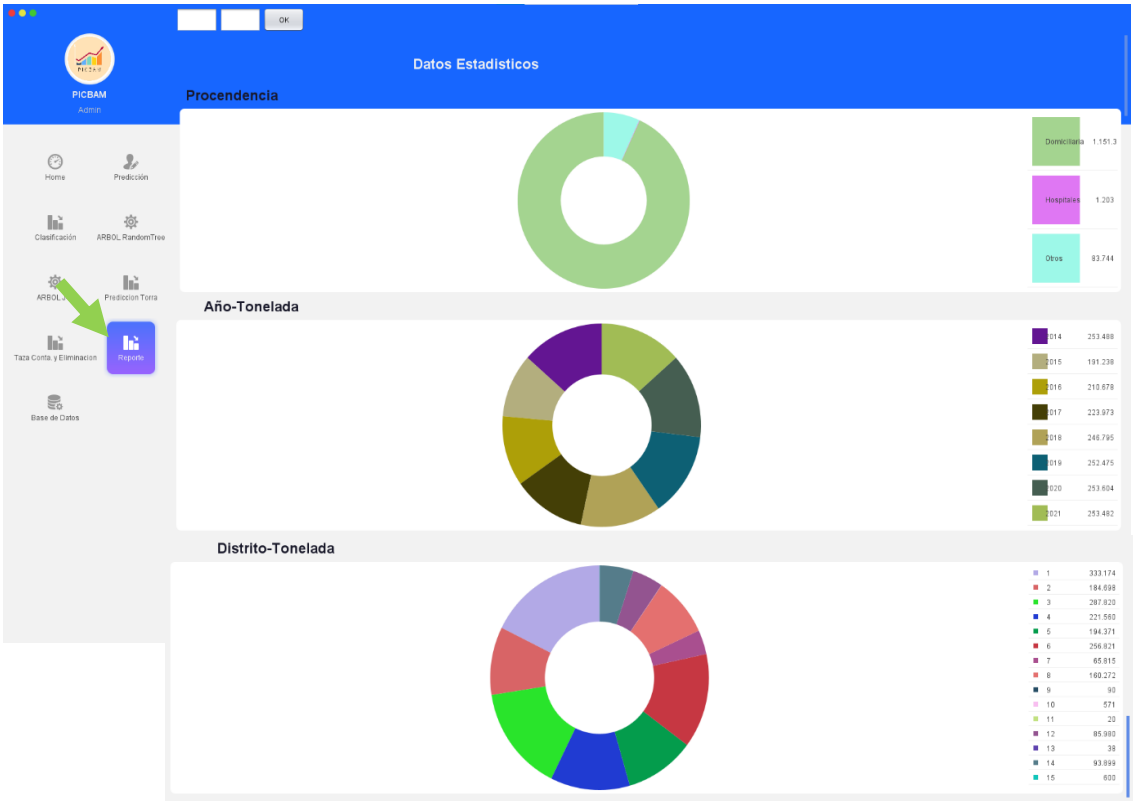
Muestra información de la forma de eliminación de los residuos y el tipo de residuo más generado por la población.



REPORTES

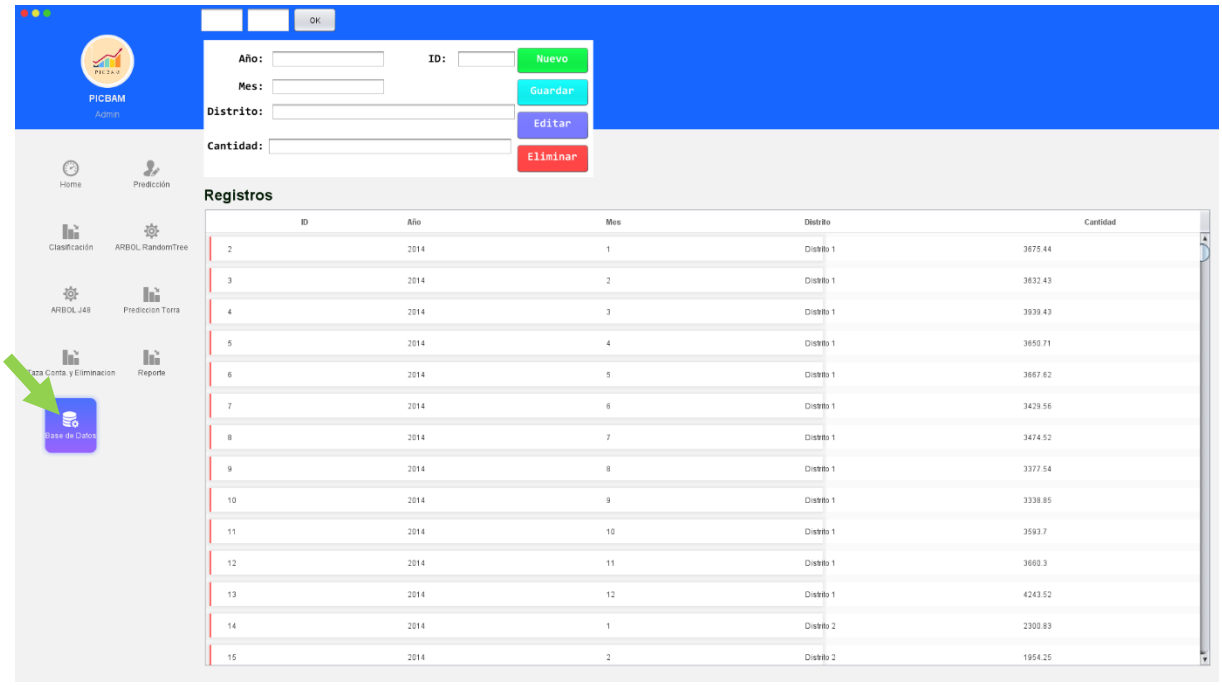
Muestra datos estadísticos desde la base de datos como ser:

- ¿Dónde votamos más basura? Resp. Domicilios
- ¿En qué año se generó más basura? Resp. 2020
- ¿Qué distrito genera más basura? Resp. Distrito 1



BASE DE DATOS

Muestra los datos de la base de datos donde podemos realizar el CRUD (Create, Read, Update, Delete) actualmente tenemos 1016 datos.



Manual Técnico

PICBAM

Modelo de Proyección del índice de contaminación de la basura de la Ciudad de El Alto

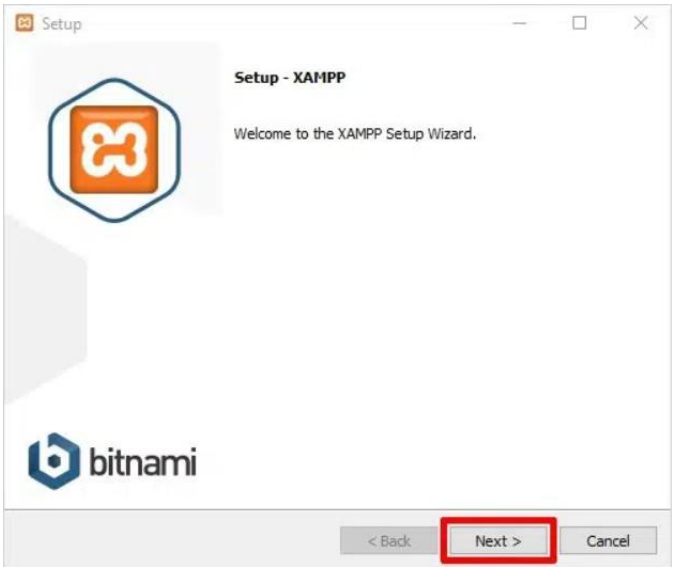
INSTALACION DE XAMPP

Paso 1: Descargar e instalar XAMPP

Para descargar e instalar XAMPP, diríjase a este enlace https://www.apachefriends.org/es/index.html desde donde podrá descargar XAMPP. Verá que XAMPP está listo para ser descargado para diferentes plataformas como Windows, Linux, Mac OS X. Como estamos hablando de cómo instalar XAMPP en Windows 10, vamos a elegir la opción de Windows como se muestra a continuación.

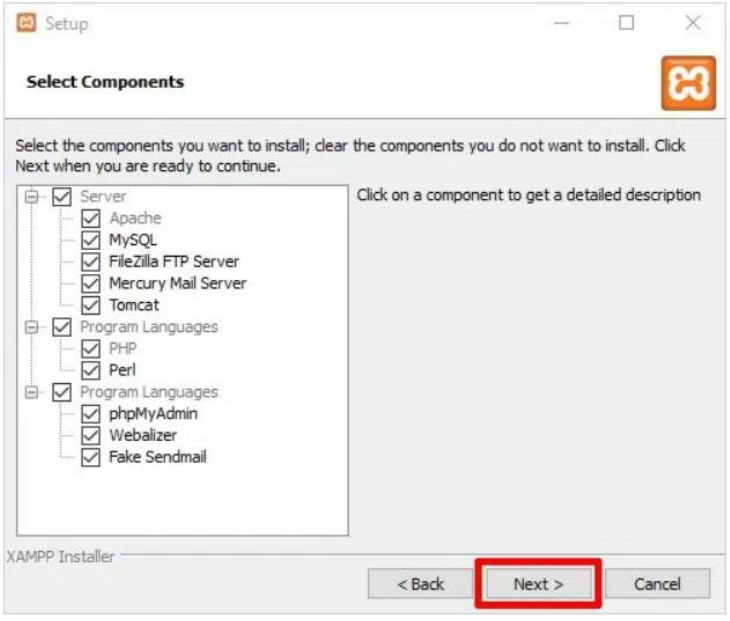
Paso 2: Ejecutar el instalador para instalar XAMPP

Durante el proceso de instalación, es posible que aparezcan algunas ventanas emergentes de advertencia. Justo después haga clic en «Sí» para iniciar el proceso de instalación. Una vez descargado, se abrirá el asistente de instalación de XAMPP. Ahora haga clic en el botón «Siguiente» para continuar.



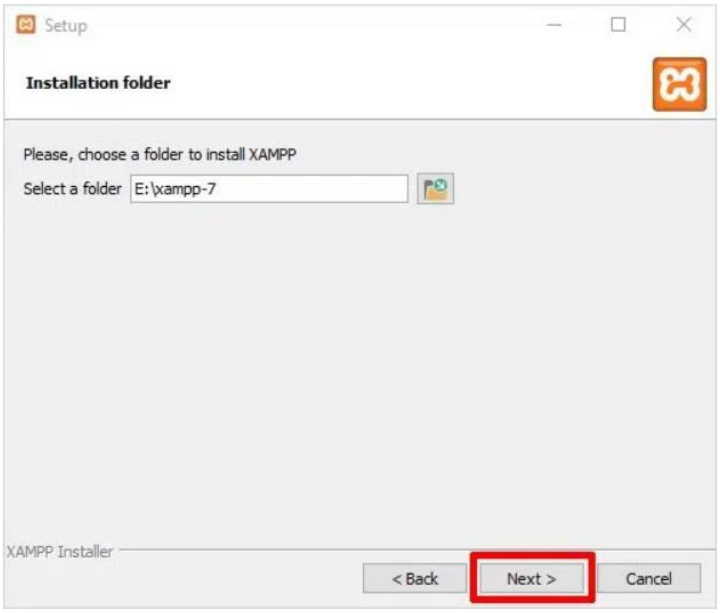
Seleccione los componentes

A continuación, tendrá que marcar los componentes que quiera instalar y puede desmarcar o dejar como está lo que le interese. Verá que hay algunas opciones de color gris claro. Estas son las opciones que se requieren para ejecutar el software y se instalarán automáticamente. Ahora haga clic en el botón «Siguiente» para continuar.



Seleccione la carpeta de instalación

Ahora elegirá la carpeta donde puede instalar XAMPP. Elija la ubicación predeterminada o seleccione cualquier ubicación de su elección y pulse «Siguiente» para seguir adelante.



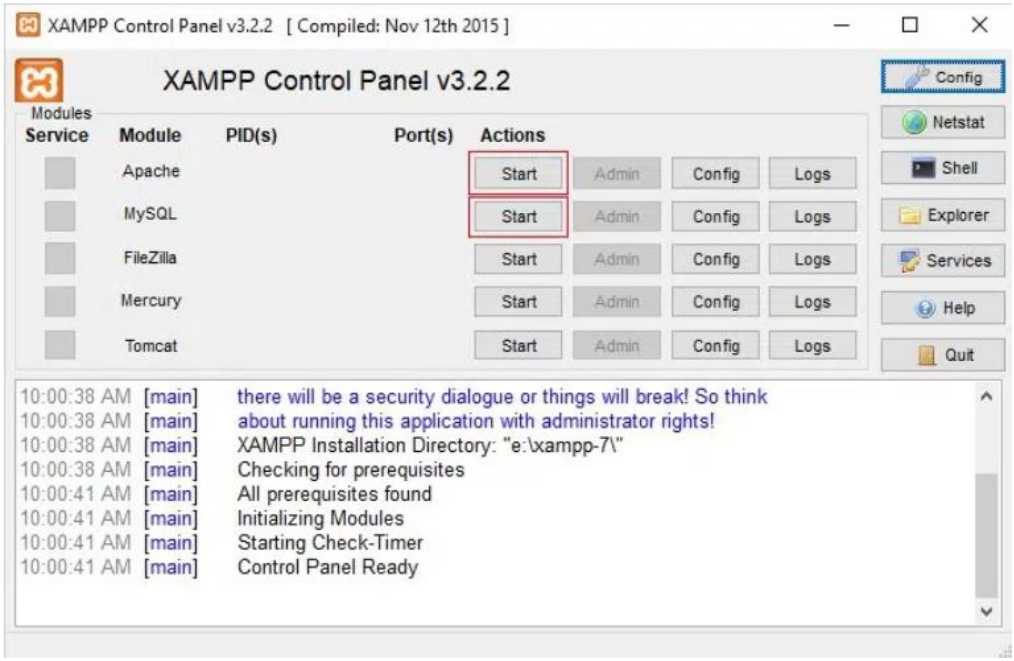
Bienvenida del asistente de XAMPP

Ahora tenga paciencia y espere a que se complete la instalación.

Instalación de XAMPP completada

Paso 3: XAMPP está ahora instalado en Windows, ejecútelo

Si todo el proceso de instalación de XAMPP fue correcto, entonces el panel de control se abrirá sin problemas. Ahora haga clic en el botón ‘Start’ correspondiente a Apache y MySQL.



Guía para la instalación y configuración de NetBeans12.1

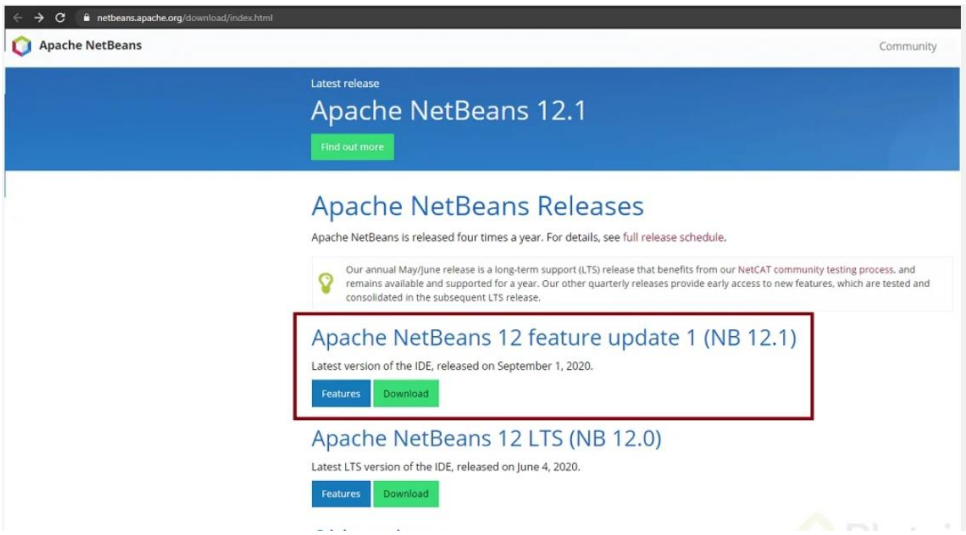
El objetivo principal de esta guía es ofrecer una relación de los pasos necesarios para instalar y configurar NetBeans 12.1 bajo Sistema operativo Windows 10, creando un proyecto básico en el que se configure el driver de mysql.

Pre requisitos:

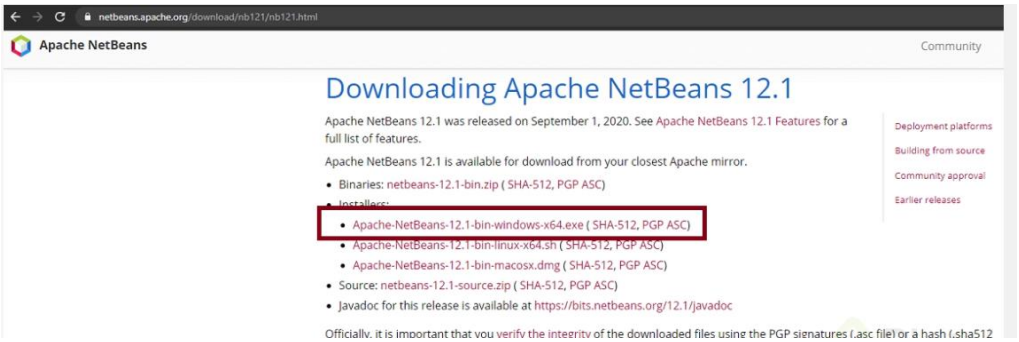
- JDK 8 o superior instalado

Pasos:

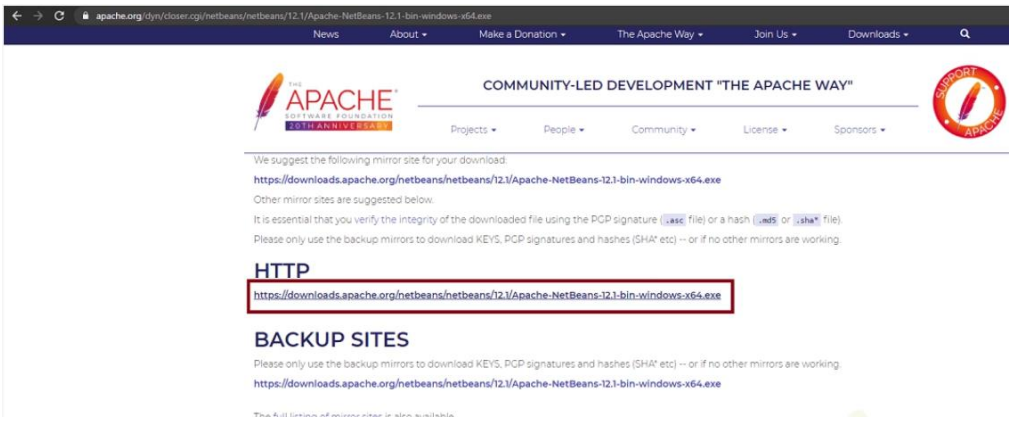
- Acceder al enlace de descarga: https://netbeans.apache.org/download/index.html



- Descargar la última versión acorde al Sistema Operativo, actualmente es la 12.1.

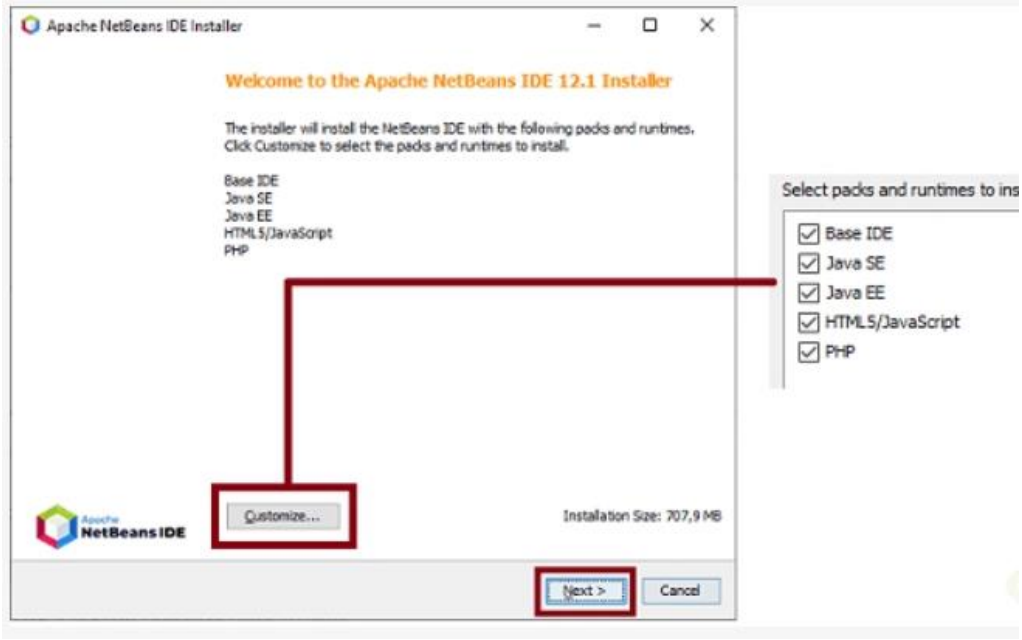


- Seleccionar el proveedor de descarga

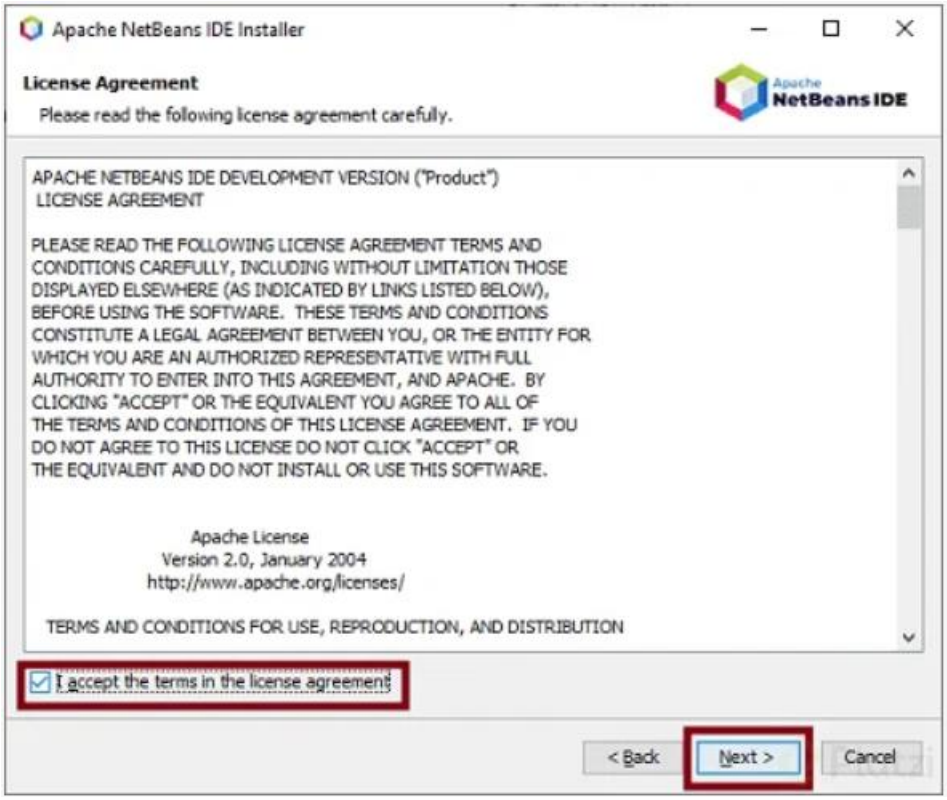


- **Ejecutar el instalador**: La instalación por defecto incluye Las características básicas de Java SE, Java Enterprise Edition(Java Web), Html/Javascript y

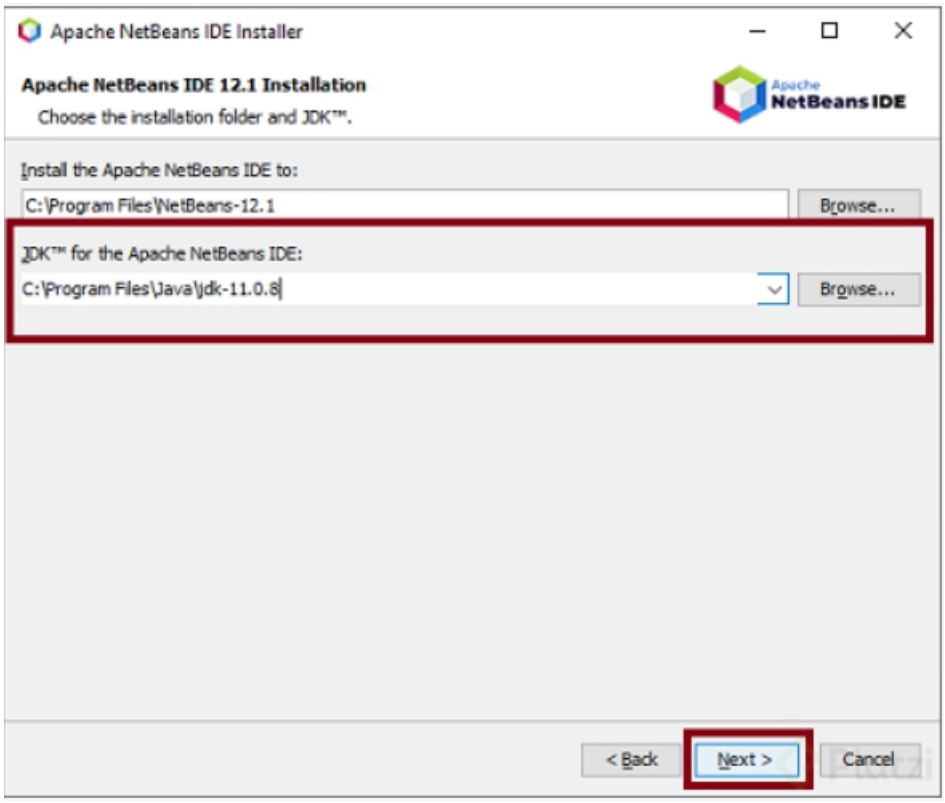
Php. Haciendo clic en el botón Customize es posible marcar o desmarcar las opciones por defecto.



- Aceptar términos de licencia



- **Seleccionar ubicación** y versión del JDK: Es posible seleccionar la ubicación para la instalación de Netbeans y la versión del JDK (En caso de tener instalada más de una).



GOBIERNO AUTÓNOMO MUNICIPAL DE EL ALTO

INFORME

CITE: GAMEA/SMASGAR/DGIR/USC/INF/204/2020

A: Ing. Sergio Nava Orellana

SECRETARIO MUNICIPAL DE AGUA, SANEAMIENTO, GESTIÓN AMBIENTAL Y RIESGOS

GOBIERNO AUTÓNOMO MUNICIPAL DE EL ALTO

Via: Ing. Belinda Zenteno

JEFE UNIDAD DE SEGUIMIENTO Y CONTROL

GOBIERNO AUTÓNOMO MUNICIPAL DE EL ALTO

De: Ing. David Quispe Flores

TÉCNICO EN CONTROL Y SEGUIMIENTO

GOBIERNO AUTÓNOMO MUNICIPAL DE EL ALTO

Ref: **SOLICITUD DE DATOS DE LA CONTAMINACIÓN DE LA BASURA EN LA CIUDAD DE EL ALTO**

Fecha: El alto 30 de marzo del 2022

I. ANTECEDENTES

En fecha 14 de marzo la Universitaria Maribel Huchani Silvestre con CI 9965652 LP mediante una nota a la Secretaria de Agua, Saneamiento Gestión Ambiental y Riesgos (SMASGAR) solicita: **SOLICITUD DE DATOS DE LA CONTAMINACIÓN DE LA BASURA EN LA CIUDAD DE EL ALTO**

II. DESARROLLO

En atención a la Nota con Cite: **SMASGAR/DGIR/2831/2022** cumplimos en informar lo siguiente:

1. ¿Cuanta basura se genera en la ciudad de el alto por distritos y en general por gestiones?

Generación de basura por distritos

Gestión 2014

Periodo	Distrito 1	Distrito 2	Distrito 3	Distrito 4	Distrito 5	Distrito 6	Distrito 7	Distrito 8	Distrito 9	Distrito 10	Distrito 11	Distrito 12	Distrito 13	Distrito 14	Varios distritos*	Total
ene-14	3.675,44	2.300,83	3.716,16	2.818,67	2.322,44	2.730,82	933,47	1.813,76	-	-	-	1.222,01	-	1.273,02	-	22.806,62
feb-14	3.632,43	1.954,25	3.223,55	2.439,13	2.242,16	2.536,13	863,29	1.919,27	-	9,74	-	1.004,97	-	1.162,66	-	20.987,58
mar-14	3.939,43	2.111,23	3.394,43	2.768,14	2.362,96	2.715,05	857,27	1.950,04	-	-	-	1.033,40	-	1.205,09	-	22.337,04
abr-14	3.650,71	2.041,49	3.195,85	2.619,23	2.124,53	2.519,77	853,77	1.783,89	6,04	5,99	-	1.005,53	6,54	1.135,80	-	20.949,14
may-14	3.667,62	2.018,84	3.094,04	2.680,16	2.204,97	2.539,57	956,08	1.751,09	-	-	-	983,97	-	1.152,08	-	21.048,42
jun-14	3.429,56	2.059,20	3.098,21	2.647,14	2.201,39	2.504,08	895,52	1.581,83	-	-	-	1.102,15	-	1.129,26	-	20.648,34
jul-14	3.474,52	2.137,85	3.134,79	2.534,98	2.208,99	2.543,33	968,14	1.689,49	5,50	8,07	-	1.100,14	-	1.165,27	-	20.971,07
ago-14	3.377,54	1.956,68	2.897,00	2.440,57	2.070,85	2.475,22	891,33	1.748,73	-	19,03	-	1.016,69	-	1.084,34	-	19.977,98
sep-14	3.338,85	1.916,74	2.888,65	2.393,22	2.040,95	2.477,54	912,15	1.763,28	6,22	-	-	1.011,55	-	1.080,96	-	19.830,11
oct-14	3.593,70	1.969,73	2.973,89	2.649,79	2.090,62	2.481,17	905,08	1.808,48	2,78	-	-	1.104,14	-	1.101,35	-	20.680,73
nov-14	3.660,30	2.025,60	3.080,23	2.600,98	2.250,12	2.587,08	891,71	1.858,25	-	-	-	1.056,24	3,62	1.160,23	-	21.174,36
dic-14	4.243,52	2.138,65	3.369,10	3.052,72	2.697,95	2.953,38	929,41	2.147,27	-	-	-	1.159,44	-	1.397,78	-	24.089,22
																255.500,61

Gestión 2015

Periodo	Distrito 1	Distrito 2	Distrito 3	Distrito 4	Distrito 5	Distrito 6	Distrito 7	Distrito 8	Distrito 9	Distrito 10	Distrito 11	Distrito 12	Distrito 13	Distrito 14	Varios distritos*	Total
ene-15	3.597,06	1.828,11	2.857,41	2.186,48	1.978,61	2.987,48	460,70	1.485,17	-	-	-	580,40	-	828,15	-	18.789,55
feb-15	3.168,07	1.542,00	2.480,62	1.871,09	1.626,12	2.548,91	382,02	1.188,90	-	-	-	422,29	-	621,73	-	15.851,74
mar-15	3.183,07	1.556,52	2.466,57	1.789,85	1.570,41	2.563,53	426,76	1.307,60	-	-	-	468,39	-	693,47	-	16.026,15
abr-15	3.094,19	1.481,70	2.393,81	1.772,26	1.448,58	2.471,48	420,54	1.250,47	-	-	-	509,37	-	597,73	-	15.440,10
may-15	3.178,08	1.503,45	2.465,89	1.837,35	1.627,76	2.605,80	425,68	1.131,01	-	-	-	486,26	-	624,96	-	15.886,23
jun-15	3.050,11	1.502,23	2.296,68	1.759,26	1.520,06	2.201,99	306,74	1.159,52	-	-	-	502,99	-	550,76	-	14.850,31
jul-15	3.157,54	1.410,31	2.360,35	1.869,81	1.608,72	2.682,06	403,09	1.196,38	-	-	-	589,68	-	621,56	-	15.899,47
ago-15	2.947,80	1.404,42	2.376,44	1.849,18	1.536,03	2.478,10	377,31	1.158,48	-	-	-	528,67	-	614,64	-	15.271,06
sep-15	2.778,74	1.424,12	2.308,17	1.762,17	1.520,10	2.537,18	380,13	1.154,31	-	-	-	517,90	-	619,52	-	15.002,32
oct-15	2.887,19	1.507,60	2.483,04	1.755,55	1.486,23	2.594,76	394,67	1.305,96	-	-	-	536,27	5,15	637,10	-	15.593,50
nov-15	3.069,69	1.452,86	2.396,70	1.856,36	1.627,17	2.710,73	421,69	1.325,26	-	-	-	581,32	-	710,85	-	16.152,59
dic-15	3.128,11	1.537,81	3.018,20	1.809,73	1.640,14	2.520,29	400,03	1.255,82	-	-	-	478,71	-	686,49	-	16.475,32
																191.238,30

Gestión 2016

Periodo	Distrito 1	Distrito 2	Distrito 3	Distrito 4	Distrito 5	Distrito 6	Distrito 7	Distrito 8	Distrito 9	Distrito 10	Distrito 11	Distrito 12	Distrito 13	Distrito 14	Varios distritos*	Total
ene-16	4.047,79	2.262,20	3.307,79	2.759,35	2.508,96	3.098,29	550,67	2.006,55	-	-	-	1.074,59	-	873,52	-	22.489,69
feb-16	3.629,30	1.861,83	2.728,29	2.156,04	1.957,80	2.778,13	527,40	1.494,24	-	-	-	644,33	-	777,60	-	18.554,94
mar-16	3.379,88	1.983,88	3.089,78	2.095,56	1.829,86	2.775,84	443,99	1.609,70	-	-	-	622,01	-	777,50	-	18.607,98
abr-16	2.834,89	1.588,07	2.800,04	2.390,77	1.730,12	2.371,64	466,66	1.560,74	-	-	-	532,89	-	698,86	-	16.974,64
may-16	3.074,12	1.755,29	2.718,79	1.895,58	1.599,37	2.558,38	421,37	1.296,25	-	-	-	571,05	-	650,87	-	16.541,05
jun-16	2.934,40	1.642,26	2.536,94	1.817,92	1.568,53	2.458,74	418,71	1.347,89	-	-	-	528,03	-	708,49	-	15.961,90
jul-16	2.992,29	1.685,63	2.519,80	1.838,65	1.588,16	2.524,24	409,92	1.443,32	-	-	5,22	536,92	-	702,94	-	16.247,08
ago-16	2.873,36	1.730,51	2.554,40	1.797,28	1.519,27	2.386,63	387,81	1.343,50	-	-	-	558,77	-	644,54	600,22	15.796,05
sep-16	2.931,59	1.699,44	2.519,01	1.824,31	1.552,21	2.365,04	429,13	1.316,95	-	-	-	584,09	-	653,48	-	15.875,24
oct-16	3.111,44	1.758,44	2.635,12	1.940,37	1.662,59	2.518,24	503,78	1.405,61	-	-	-	663,42	-	692,64	-	16.891,62
nov-16	3.097,42	1.689,80	2.572,73	1.961,43	1.708,91	2.425,49	489,88	1.431,21	-	-	-	633,04	-	682,23	-	16.692,11
dic-16	3.706,75	1.949,28	3.050,93	2.151,73	1.970,73	2.830,47	525,12	1.690,14	-	-	-	766,58	-	804,51	-	19.446,21
																210.078,47

Gestión 2017

Periodo	Distrito 1	Distrito 2	Distrito 3	Distrito 4	Distrito 5	Distrito 6	Distrito 7	Distrito 8	Distrito 9	Distrito 10	Distrito 11	Distrito 12	Distrito 13	Distrito 14	Varios distritos*	Total
ene-17	3.764,51	1.945,65	2.940,52	2.299,68	2.092,23	2.889,22	644,03	1.765,47	-	61,48	-	800,21	-	941,30	-	20.144,27
feb-17	3.181,73	1.568,08	2.370,33	1.853,60	1.746,75	2.526,08	461,59	1.390,46	-	8,82	-	685,70	-	775,02	-	16.568,13
mar-17	3.811,09	1.964,79	3.138,07	2.304,83	2.136,18	2.970,60	649,27	1.903,60	-	-	-	839,42	-	960,67	-	20.678,50
abr-17	3.421,29	1.805,93	2.753,97	1.974,18	1.793,25	2.573,94	484,44	1.664,20	-	-	-	675,56	-	825,78	-	17.972,52
may-17	3.470,36	1.830,19	2.742,87	2.261,33	1.984,31	2.613,09	486,76	1.647,50	-	-	-	784,87	-	842,79	-	18.664,04
jun-17	2.828,88	2.407,56	2.514,43	1.819,62	1.717,26	2.176,04	423,71	1.448,05	-	-	-	719,22	-	788,38	-	16.843,13
jul-17	3.812,44	2.158,23	3.349,94	2.337,74	2.137,29	3.219,86	594,70	1.783,83	-	-	-	923,15	-	1.008,90	-	21.326,05
ago-17	2.944,39	1.745,53	2.645,38	1.750,02	1.601,99	2.480,46	421,11	1.501,71	-	-	-	667,32	-	660,77	-	16.418,66
sep-17	2.414,96	3.080,36	2.484,15	1.916,69	1.609,41	1.952,53	472,10	1.377,19	-	-	-	793,24	-	712,05	-	16.812,66
oct-17	3.462,78	1.823,54	2.874,34	2.055,58	1.890,72	2.604,40	501,86	1.628,85	-	-	-	790,69	-	805,26	-	18.437,99
nov-17	3.387,65	1.892,81	2.999,79	2.097,13	1.961,76	2.786,48	532,22	1.626,60	-	-	-	774,99	-	793,19	-	18.852,61
dic-17	3.966,97	2.224,93	3.269,54	2.331,14	2.200,41	3.012,05	608,47	1.795,59	-	-	-	901,47	-	944,22	-	21.254,78
																223.973,30

Gestión 2018

Periodo	Distrito 1	Distrito 2	Distrito 3	Distrito 4	Distrito 5	Distrito 6	Distrito 7	Distrito 8	Distrito 9	Distrito 10	Distrito 11	Distrito 12	Distrito 13	Distrito 14	Varios distritos*	Total
ene-18	4.012,27	2.035,98	3.414,32	2.613,00	2.410,67	3.068,77	700,84	2.189,62	-	-	-	960,80	-	1.100,99	-	22.507,24
feb-18	3.714,28	1.955,85	3.043,30	2.279,65	2.107,37	2.751,81	625,17	1.801,12	-	-	-	814,68	-	1.000,83	-	20.094,02
mar-18	3.937,54	2.005,89	3.405,51	2.378,28	2.282,05	2.958,62	757,00	2.231,32	-	-	-	942,52	-	1.142,86	-	22.041,56
abr-18	3.692,33	1.964,56	2.980,92	2.138,93	2.140,48	2.854,39	688,77	1.776,03	-	-	-	910,98	-	934,53	-	20.081,90
may-18	3.757,34	2.072,95	3.101,94	2.164,75	2.064,87	2.955,15	661,81	1.803,20	-	-	-	1.098,17	-	900,46	-	20.580,62
jun-18	3.248,89	1.877,99	2.934,32	2.206,49	1.973,20	2.769,84	597,21	1.679,67	-	-	-	997,29	-	1.137,08	-	19.421,97
jul-18	3.128,98	1.577,74	2.530,20	1.968,34	1.733,72	2.499,05	553,88	1.420,67	-	-	-	839,98	-	844,26	-	17.096,82
ago-18	3.376,09	3.775,18	3.048,18	2.257,89	2.127,38	2.853,74	691,64	1.705,56	-	-	-	1.027,97	-	1.059,92	-	21.923,55
sep-18	3.207,00	1.732,71	2.998,31	2.051,33	1.852,52	2.689,68	538,56	1.562,86	5,25	27,88	-	852,59	-	881,30	-	18.399,99
oct-18	3.466,90	1.898,55	3.079,52	2.312,21	2.126,43	3.063,50	621,25	1.732,33	-	13,26	-	929,25	2,51	1.176,99	-	20.422,70
nov-18	3.573,37	1.942,15	3.069,31	2.294,80	2.124,54	2.899,15	750,00	1.833,32	-	20,09	7,39	974,76	2,18	1.739,92	-	21.230,98
dic-18	4.179,12	2.059,31	3.503,88	2.570,92	2.443,31	3.263,49	773,85	1.885,39	-	20,58	-	1.020,88	-	1.273,06	-	22.993,79
																246.795,13

Gestión 2019

Periodo	Distrito 1	Distrito 2	Distrito 3	Distrito 4	Distrito 5	Distrito 6	Distrito 7	Distrito 8	Distrito 9	Distrito 10	Distrito 11	Distrito 12	Distrito 13	Distrito 14	Varios distritos*	Total
ene-19	4.468,34	2.327,35	3.828,23	2.676,04	2.526,86	3.487,84	1.062,02	2.263,62	-	-	-	1.414,51	-	1.365,42	-	25.420,23
feb-19	3.751,19	1.899,15	3.240,35	2.320,43	2.158,72	2.933,91	789,95	1.906,08	-	23,19	-	1.119,91	-	1.158,63	-	21.301,51
mar-19	4.338,31	2.235,13	3.577,97	2.673,92	2.618,03	3.287,57	876,91	2.101,90	30,30	35,57	-	1.185,64	-	1.184,79	-	24.146,03
abr-19	3.836,69	1.925,99	3.250,27	2.350,37	2.099,47	2.948,74	820,77	1.827,43	3,11	19,47	-	1.048,55	-	1.036,11	-	21.166,97
may-19	3.916,80	1.957,71	3.027,30	2.301,79	2.226,73	2.932,81	810,95	1.828,60	-	29,94	-	994,54	-	1.089,88	-	21.117,05
jun-19	3.494,51	1.829,88	3.045,36	2.119,45	1.903,84	2.686,73	697,15	1.573,23	-	33,08	-	1.000,47	4,07	972,91	-	19.360,68
jul-19	3.672,15	1.933,11	3.107,09	2.429,59	2.086,76	2.862,75	743,81	1.416,60	-	19,49	-	1.143,43	-	1.002,46	-	20.417,24
ago-19	3.376,04	1.905,81	3.120,21	2.196,00	1.953,59	2.661,00	693,59	1.735,28	-	31,79	-	961,08	-	909,71	-	19.544,10
sep-19	3.455,00	1.883,63	2.799,30	2.142,31	1.955,77	2.690,28	783,67	1.680,24	-	27,83	-	949,78	-	968,05	-	19.335,86
oct-19	3.564,52	2.103,61	3.194,73	2.273,64	2.030,68	2.756,45	855,26	1.613,04	-	44,96	-	1.054,34	-	1.005,74	-	20.496,97
nov-19	2.930,20	1.660,89	2.558,60	1.930,77	1.701,32	2.210,76	632,74	1.241,23	-	6,44	-	814,55	-	820,32	-	16.507,82
dic-19	4.407,56	2.194,00	3.495,38	2.603,61	2.437,56	3.190,13	936,88	2.032,73	-	24,31	-	1.195,17	-	1.149,26	-	23.666,59
																252.481,05

Gestión 2020

Periodo	Distrito 1	Distrito 2	Distrito 3	Distrito 4	Distrito 5	Distrito 6	Distrito 7	Distrito 8	Distrito 9	Distrito 10	Distrito 11	Distrito 12	Distrito 13	Distrito 14	Varios distritos*	Total
ene-20	4.519,97	2.271,93	3.579,89	2.670,61	2.705,59	3.263,09	1.063,00	2.153,02	-	15,75	-	1.305,98	-	1.313,33	-	24.862,16
feb-20	4.203,69	2.176,53	3.361,74	2.365,22	2.282,64	2.929,76	949,05	1.987,04	-	-	-	1.132,74	-	1.153,20	-	22.541,61
mar-20	3.907,56	2.240,51	3.463,11	2.660,84	2.490,98	3.066,59	927,91	2.058,39	-	-	-	1.176,08	-	1.204,85	-	23.196,82
abr-20	2.773,36	2.099,75	3.164,61	2.402,97	2.203,53	2.490,36	725,66	1.966,37	-	7,11	-	1.130,89	-	1.087,71	-	20.052,32
may-20	2.707,34	2.088,66	3.115,73	2.378,59	2.081,58	2.342,81	701,34	1.765,75	-	-	-	1.107,57	-	1.054,37	-	19.343,74
jun-20	3.135,36	2.567,90	3.099,88	2.399,87	2.054,77	2.410,23	702,83	1.640,86	-	-	-	989,95	-	1.057,38	-	20.059,03
jul-20	2.650,19	1.780,20	2.826,94	2.217,83	1.908,49	2.156,07	602,75	1.326,59	8,15	-	-	945,93	-	974,43	-	17.397,57
ago-20	2.950,37	2.264,61	2.970,47	2.105,91	1.907,50	2.261,56	627,73	1.585,09	-	-	-	875,84	-	975,35	-	18.524,43
sep-20	3.572,30	2.233,58	3.353,31	2.891,60	2.223,93	2.594,17	929,96	1.870,02	-	8,28	-	1.020,35	-	1.163,51	-	21.861,01
oct-20	3.510,64	2.397,13	3.447,43	3.298,47	2.098,13	2.763,79	984,52	1.571,71	-	-	-	1.037,81	-	1.128,24	-	22.237,87
nov-20	3.138,21	2.101,99	3.287,56	3.020,12	2.123,25	2.523,42	747,32	1.523,58	-	8,57	-	1.047,38	4,71	1.031,47	-	20.557,58
dic-20	3.912,57	2.289,41	3.643,18	2.855,15	2.311,07	2.818,62	821,39	1.954,43	-	-	7,93	1.084,10	-	1.272,16	-	22.970,01
																253.604,15

Gestión 2021

Periodo	Distrito 1	Distrito 2	Distrito 3	Distrito 4	Distrito 5	Distrito 6	Distrito 7	Distrito 8	Distrito 9	Distrito 10	Distrito 11	Distrito 12	Distrito 13	Distrito 14	Varios distritos*	Total
ene-21	3.675,44	2.300,83	3.716,16	2.818,67	2.322,44	2.730,82	933,47	1.813,76	-	-	-	1.222,01	-	1.273,02	-	22.806,62
feb-21	3.632,43	1.954,25	3.223,55	2.439,13	2.242,16	2.536,13	863,29	1.919,27	-	9,74	-	1.004,97	-	1.162,66	-	20.987,58
mar-21	3.939,43	2.111,23	3.394,43	2.768,14	2.362,96	2.715,05	857,27	1.950,04	-	-	-	1.033,40	-	1.205,09	-	22.337,04
abr-21	3.650,71	2.041,49	3.195,85	2.619,23	2.124,53	2.519,77	853,77	1.783,89	6,04	5,99	-	1.005,53	6,54	1.135,80	-	20.949,14
may-21	3.667,62	2.018,84	3.094,04	2.680,16	2.204,97	2.539,57	956,08	1.751,09	-	-	-	983,97	-	1.152,08	-	21.048,42
jun-21	3.429,56	2.059,20	3.098,21	2.647,14	2.201,39	2.504,08	895,52	1.581,83	-	-	-	1.102,15	-	1.129,26	-	20.648,34
jul-21	3.474,52	2.137,85	3.134,79	2.534,98	2.208,99	2.543,33	968,14	1.689,49	5,50	8,07	-	1.100,14	-	1.165,27	-	20.971,07
ago-21	3.377,54	1.956,68	2.897,00	2.440,57	2.070,85	2.475,22	891,33	1.748,73	-	19,03	-	1.016,69	-	1.084,34	-	19.977,98
sep-21	3.338,85	1.916,74	2.888,65	2.393,22	2.040,95	2.477,54	912,15	1.763,28	6,22	-	-	1.011,55	-	1.080,96	-	19.830,11
oct-21	3.593,70	1.969,73	2.973,89	2.649,79	2.090,62	2.481,17	905,08	1.808,48	2,78	-	-	1.104,14	-	1.101,35	-	20.680,73
nov-21	3.660,30	2.025,60	3.080,23	2.600,98	2.250,12	2.587,08	891,71	1.858,25	-	-	-	1.056,24	3,62	1.160,23	-	21.174,36
dic-21	4.243,52	2.138,65	3.369,10	3.052,72	2.697,95	2.953,38	929,41	2.147,27	-	-	-	1.159,44	-	1.397,78	-	24.089,22
																255.500,61

III. CONCLUSIONES

Se realizo las respuestas a las preguntas de manera puntual.

Es en cuanto informo para fines consiguientes.

Cuadro Nº 6.05.03
BOLIVIA: RESIDUOS SÓLIDOS RECOLECTADOS, POR CIUDADES, SEGÚN TIPO DE PROCEDENCIA, 2010 - 2020
(En toneladas)

TIPO DE PROCEDENCIA	TOTAL	SUCRE	LA PAZ	COCHABAMBA	ORURO	POTOSÍ	TARIJA	SANTA CRUZ	TRINIDAD	COBIJA	EL ALTO
2010	**995.519**	**n.d.**	**177.817**	**131.866**	**44.277**	**58.670**	**47.709**	**359.826**	**24.264**	**7.794**	**143.296**
Domiciliarios	757.988		144.290	113.628	38.744	32.150	34.473	233.866	18.531	3.128	139.178
Áreas Públicas	70.469		10.504	7.788	2.951	11.680	8.821	23.737	4.194	794	n.d.
Mercados	73.263		10.643	9.556	2.435	9.795	3.378	33.645	1.451	2.360	n.d.
Establecimientos de salud	9.642		3.269	894	147	3.365	309	852	88	594	124
Otros[1]	84.158		9.112	n.d.	n.d.	1.680	728	67.726	n.d.	918	3.993
2011	**1.009.839**	**n.d.**	**177.629**	**136.428**	**44.473**	**50.459**	**51.764**	**363.808**	**22.113**	**9.452**	**153.712**
Domiciliarios	782.164		144.839	117.917	39.447	33.480	37.057	238.312	16.933	5.055	149.124
Áreas Públicas	57.273		10.121	8.030	2.668	4.075	10.160	17.480	3.868	870	n.d.
Mercados	68.428		10.699	9.556	2.201	4.840	3.624	33.966	1.225	2.316	n.d.
Establecimientos de salud	11.267		3.364	925	157	4.104	228	1.820	86	458	125
Otros[1]	90.708		8.605	n.d.	n.d.	3.960	696	72.230	n.d.	753	4.464
2012	**1.045.675**	**n.d.**	**181.267**	**140.233**	**47.996**	**53.914**	**53.794**	**376.507**	**26.424**	**3.756**	**161.785**
Domiciliarios	875.353		147.890	121.725	41.288	33.558	41.187	307.180	24.363	1.585	156.577
Áreas Públicas	39.287		10.254	8.030	3.504	3.782	8.036	5.081	77	523	n.d.
Mercados	73.516		10.991	9.556	2.898	5.713	3.631	39.319	472	936	n.d.
Establecimientos de salud	12.104		3.420	922	306	4.354	248	988	1.512	222	133
Otros[1]	45.414		8.712	n.d.	n.d.	6.506	692	23.939	n.d.	490	5.075
2013	**1.113.048**	**n.d.**	**186.378**	**166.849**	**49.389**	**49.918**	**49.668**	**400.928**	**24.290**	**n.d.**	**185.627**
Domiciliarios	942.884		151.486	148.344	37.767	31.770	38.132	334.265	22.630	n.d.	178.490
Áreas Públicas	32.866		10.300	8.030	4.384	3.331	6.629	11	182	n.d.	0
Mercados	77.953		11.207	9.556	5.459	5.173	3.494	42.790	274	n.d.	0
Establecimientos de salud	11.959		3.541	919	1.780	3.111	211	1.039	1.204	n.d.	156
Otros[1]	47.385		9.845	0	0	6.533	1.203	22.823	0	n.d.	6.981
2014	**1.179.894**	**n.d.**	**187.650**	**178.034**	**55.855**	**65.076**	**47.001**	**430.103**	**24.322**	**n.d.**	**191.853**
Domiciliarios	983.238		153.035	152.326	33.180	42.353	36.502	358.137	22.794	n.d.	184.910
Áreas Públicas	32.937		7.184	10.631	5.551	5.080	4.315	0	176	n.d.	0
Mercados	96.181		8.694	14.398	13.743	6.549	3.798	48.783	217	n.d.	0
Establecimientos de salud	16.893		7.019	679	3.380	3.200	136	1.177	1.135	n.d.	167
Otros[1]	50.644		11.718	0	0	7.894	2.250	22.006	0	n.d.	6.776
2015	**1.319.375**	**56.575**	**206.308**	**177.517**	**53.710**	**62.949**	**53.459**	**488.737**	**19.805**	**n.d.**	**200.315**
Domiciliarios	1.090.406	43.435	168.297	150.968	33.209	43.390	42.630	398.360	17.133	n.d.	192.984
Áreas Públicas	33.165	0	11.898	10.795	3.174	3.792	3.457	0	49	n.d.	0
Mercados	128.476	11.680	12.451	15.196	17.034	4.252	4.348	63.345	169	n.d.	0
Establecimientos de salud	12.477	365	3.439	559	293	3.725	193	1.253	2.453	n.d.	198
Otros[1]	54.851	1.095	10.222	0	0	7.790	2.830	25.780	0	n.d.	7.134
2016	**1.426.988**	**60.987**	**212.554**	**171.337**	**57.044**	**47.335**	**56.648**	**558.229**	**28.069**	**17.950**	**216.836**
Domiciliarios	1.185.712	47.379	175.604	145.927	39.513	31.260	45.329	452.349	26.114	14.359	207.877
Áreas Públicas	28.854	n.d.	11.237	10.349	809	3.160	1.843	n.d.	1.366	90	0
Mercados	139.799	11.481	12.619	14.454	13.151	3.280	4.039	78.532	449	1.795	0
Establecimientos de salud	11.311	1.019	3.321	607	427	3.634	254	1.334	140	359	215
Otros[1]	61.312	1.108	9.773	n.d.	3.144	6.001	5.182	26.013	n.d.	1.346	8.743
2017	**1.521.884**	**57.199**	**224.433**	**192.008**	**60.512**	**47.717**	**59.060**	**599.853**	**19.553**	**16.511**	**245.038**
Domiciliarios	1.246.981	43.187	184.860	163.312	42.910	30.330	48.607	483.302	17.766	11.558	221.149
Áreas Públicas	29.909	n.d.	12.216	11.655	213	3.600	1.986	n.d.	207	33	n.d.
Mercados	161.815	11.859	13.588	16.435	13.768	6.667	3.877	91.255	1.065	3.302	n.d.
Establecimientos de salud	10.026	1.017	2.745	607	357	2.405	365	1.417	516	380	218
Otros[1]	73.153	1.137	11.025	n.d.	3.264	4.715	4.225	23.879	n.d.	1.238	23.670
2018	**1.616.728**	**60.652**	**236.370**	**212.196**	**61.492**	**45.810**	**62.881**	**636.928**	**26.100**	**17.042**	**257.257**
Domiciliarios	1.363.874	48.430	198.119	179.573	44.961	27.350	53.537	530.198	23.724	11.980	246.001
Áreas Públicas	27.331	n.d.	9.175	13.733	185	3.400	796	n.d.	8	34	n.d.
Mercados	158.757	10.074	15.426	18.239	12.899	8.170	3.799	84.767	2.032	3.351	n.d.
Establecimientos de salud	9.802	825	2.775	651	479	2.360	343	1.507	205	394	264
Otros[1]	56.965	1.323	10.875	n.d.	2.969	4.530	4.406	20.456	130	1.283	10.992
2019	**1.600.938**	**22.461**	**230.674**	**199.123**	**64.781**	**43.460**	**64.317**	**672.341**	**30.642**	**18.359**	**254.780**
Domiciliarios	1.333.568	18.005	192.377	169.289	49.012	26.010	55.975	562.460	27.515	12.852	220.072
Áreas Públicas	23.511	0	6.388	12.087	530	4.470	0	0	0	37	0
Mercados	156.389	3.654	16.470	17.046	12.999	7.510	4.208	88.017	2.813	3.672	0
Establecimientos de salud	8.210	168	2.483	702	390	1.725	308	1.542	214	422	255
Otros[1]	79.260	633	12.957	0	1.849	3.745	3.826	20.322	100	1.375	34.453
2020[p]	**1.620.982**	**39.505**	**234.939**	**192.140**	**62.923**	**45.300**	**67.800**	**659.547**	**36.978**	**19.497**	**262.353**
Domiciliarios	1.405.022	31.667	195.810	163.409	48.003	31.650	60.868	570.159	33.593	13.648	256.215
Áreas Públicas	23.257	0	7.308	11.668	307	2.530	0	1.392	13	39	0
Mercados	135.405	6.427	16.847	16.447	12.206	5.085	3.359	67.949	3.184	3.899	0
Establecimientos de salud	9.826	296	2.528	617	471	3.160	311	1.556	187	448	251
Otros[1]	47.473	1.114	12.446	0	1.936	2.875	3.261	18.491	0	1.462	5.886

Fuente: Instituto Nacional de Estadística
Gobiernos Autónomos Municipales
(p): Preliminar
n.d.: Información no disponible
[1] Agrupa residuos generados en industria y mataderos.
Nota: A partir del 2015, la información considera la ciudad de Sucre por tener desagregación por tipo de procedencia.

BOLIVIA: RECOLECCIÓN DE RESIDUOS SÓLIDOS POR CIUDADES, SEGÚN AÑO Y MES, 2005 - 2022
(En toneladas)

PERIODO	TOTAL	Sucre	La Paz	Cochabamba	Oruro	Potosí	Tarija	Santa Cruz	Trinidad	Cobija	El Alto
2005	**785.653**	**34.168**	**157.526**	**115.260**	**34.769**	**19.008**	**26.967**	**310.389**	**17.639**	**757**	**69.169**
Enero	68.840		15.288	10.841	3.197	1.730	2.348	27.807	1.352	64	6.214
Febrero	60.565		13.247	9.286	3.191	1.423	2.175	23.976	1.353	62	5.852
Marzo	65.162		14.327	10.422	3.024	1.707	2.375	25.784	1.384	64	6.075
Abril	61.868		13.328	8.923	3.121	1.560	2.289	25.007	1.459	65	6.117
Mayo	60.407		12.810	9.260	2.929	1.563	2.173	25.000	1.394	61	5.217
Junio	60.523		11.130	10.297	2.635	1.537	2.119	26.174	1.524	63	5.044
Julio	60.296		12.716	9.022	2.226	1.550	2.143	24.820	1.551	64	6.205
Agosto	61.878		12.302	9.943	2.631	1.572	2.234	25.353	1.631	62	6.150
Septiembre	59.015		12.105	9.263	2.784	1.556	1.973	23.694	1.331	63	6.245
Octubre	61.183		12.508	9.807	2.722	1.505	2.171	25.428	1.518	65	5.459
Noviembre	63.682		13.018	9.233	3.016	1.538	2.396	26.690	1.574	62	6.156
Diciembre	68.065		14.748	8.963	3.294	1.768	2.573	30.655	1.568	62	4.435
2006	**850.731**	**35.534**	**169.666**	**114.467**	**37.845**	**20.555**	**28.886**	**315.881**	**22.413**	**686**	**104.798**
Enero	71.756		15.396	9.285	3.535	1.713	2.591	29.165	1.975	59	8.038
Febrero	62.156		12.989	7.572	3.673	1.582	2.298	23.533	1.765	60	8.683
Marzo	71.418		14.675	9.623	3.440	1.826	2.729	27.784	1.668	59	9.614
Abril	65.606		13.001	8.595	3.370	1.341	2.331	25.923	1.680	61	9.304
Mayo	68.562		13.331	10.423	3.193	1.670	2.467	25.957	1.853	54	9.614
Junio	66.964		12.772	9.176	2.945	1.306	2.233	27.053	1.923	57	9.500
Julio	67.620		12.680	9.902	2.601	1.545	2.404	27.132	1.999	58	9.300
Agosto	66.220		13.019	9.729	2.709	1.963	2.346	25.083	2.016	55	9.300
Septiembre	62.386		12.755	9.848	2.801	1.630	2.062	23.975	1.878	58	7.379
Octubre	67.804		15.322	9.258	2.786	1.737	2.504	26.832	1.888	54	7.423
Noviembre	69.145		15.815	10.324	3.251	1.960	2.306	26.209	1.877	59	7.345
Diciembre	75.559		17.910	10.732	3.540	2.281	2.617	27.235	1.891	54	9.300

2007	**887.814**	**38.801**	**168.205**	**117.473**	**38.794**	**33.488**	**30.143**	**329.337**	**20.803**	**938**	**109.830**
Enero	73.970		16.054	10.264	3.805	1.857	2.847	27.780	1.983	80	9.300
Febrero	65.389		14.016	8.995	3.603	2.745	2.386	23.491	1.681	72	8.400
Marzo	73.470		14.469	10.538	3.650	1.752	2.643	29.037	2.010	71	9.300
Abril	69.261		13.951	10.155	3.409	2.274	2.420	25.930	2.047	74	9.000
Mayo	71.693		14.533	10.481	3.224	2.513	2.491	27.897	1.478	76	9.000
Junio	70.465		13.515	9.885	3.225	2.639	2.769	27.621	1.594	80	9.137
Julio	67.904		13.026	9.947	3.078	2.656	2.297	27.030	1.666	78	8.126
Agosto	67.451		12.945	9.664	3.021	3.288	2.438	25.667	1.720	77	8.631
Septiembre	67.656		12.919	9.422	2.756	3.001	2.352	26.576	1.616	73	8.939
Octubre	72.382		13.802	9.793	2.995	3.109	2.505	29.419	1.686	77	8.996
Noviembre	70.722		13.619	8.415	2.696	3.023	2.320	28.696	1.673	87	10.193
Diciembre	78.650		15.355	9.914	3.332	4.631	2.675	30.191	1.650	93	10.808
2008	**913.963**	**40.354**	**164.849**	**122.013**	**38.631**	**37.405**	**36.630**	**328.232**	**18.817**	**1.018**	**126.013**
Enero	77.260		15.563	10.448	3.854	3.302	2.880	29.205		88	11.921
Febrero	68.008		13.625	9.802	3.437	2.720	2.640	24.016	1.568	86	10.116
Marzo	71.361		13.939	10.896	3.325	1.714	2.946	26.259	1.821	93	10.369
Abril	69.792		13.538	10.723	2.765	2.249	2.902	25.328	1.846	87	10.355
Mayo	71.435		13.825	10.927	2.914	2.518	2.982	26.507	1.805	85	9.873
Junio	64.874		13.429	10.272	3.132	2.783	3.065	20.941	1.746	87	9.419
Julio	76.220		13.415	10.244	2.977	2.677	3.155	31.627	1.736	86	10.304
Agosto	74.048		13.049	9.917	3.026	3.445	3.156	29.826	1.545	82	10.003
Septiembre	70.373		13.294	8.766	3.079	3.175	3.118	26.153	1.698	75	11.016
Octubre	74.412		13.403	10.040	2.897	3.170	3.215	29.279	1.771	84	10.553
Noviembre	72.871		12.601	9.417	3.333	3.967	3.185	28.456	1.593	82	10.238
Diciembre	82.955		15.168	10.561	3.893	5.686	3.386	30.637	1.690	85	11.848
2009	**995.945**	**41.316**	**168.285**	**125.182**	**42.810**	**37.287**	**40.464**	**381.681**	**20.381**	**n.d.**	**138.539**
Enero	79.146		15.161	10.778	3.759	4.347	3.190	28.150	1.655	n.d.	12.107
Febrero	70.383		13.451	9.763	3.746	3.174	2.935	25.308	1.763	n.d.	10.244
Marzo	84.181		14.758	10.697	3.911	2.212	3.233	35.246	1.715	n.d.	12.410
Abril	79.178		13.809	10.444	3.607	2.429	3.286	32.638	1.688	n.d.	11.277
Mayo	75.264		13.551	10.618	3.347	2.684	3.400	28.250	1.684	n.d.	11.730
Junio	73.357		13.417	10.602	3.482	2.647	3.309	27.097	1.616	n.d.	11.187
Julio	80.082		13.496	10.513	3.526	2.663	3.280	33.156	1.745	n.d.	11.703
Agosto	79.371		12.971	10.752	3.280	3.821	3.409	32.604	1.554	n.d.	10.980
Septiembre	78.839		12.985	10.548	3.202	3.174	3.461	32.352	1.703	n.d.	11.415
Octubre	82.689		14.040	10.202	3.331	3.237	3.627	35.005	1.679	n.d.	11.570
Noviembre	79.801		13.996	9.400	3.434	2.341	3.619	34.305	1.607	n.d.	11.100
Diciembre	92.338		16.652	10.865	4.186	4.559	3.717	37.571	1.973	n.d.	12.816

2010	**1.040.484**	**44.965**	**177.817**	**131.866**	**44.277**	**58.670**	**47.709**	**359.826**	**24.264**	**7.794**	**143.296**
Enero	92.110		16.729	10.029	4.152	5.323	3.714	36.276	2.020	475	13.393
Febrero	85.599		15.488	10.313	3.912	5.474	3.447	32.424	1.863	452	12.225
Marzo	96.372		16.479	12.056	4.081	5.037	3.624	38.986	2.062	455	13.593
Abril	78.230		14.867	10.921	3.724	5.423	3.386	27.590	1.942	500	9.878
Mayo	83.680		14.775	11.866	3.555	5.087	3.593	28.673	1.994	545	13.592
Junio	82.464		14.386	11.480	3.763	5.257	3.719	29.076	2.041	658	12.084
Julio	79.619		14.106	12.042	3.613	5.485	3.857	26.602	2.060	683	11.172
Agosto	76.163		13.929	11.365	3.374	2.528	3.994	27.128	1.996	716	11.132
Septiembre	76.326		13.356	11.015	3.405	4.004	4.024	26.740	2.057	750	10.977
Octubre	80.515		13.774	8.772	3.394	5.029	4.403	31.073	2.106	816	11.147
Noviembre	77.537		14.078	10.875	3.431	4.159	4.747	26.202	1.975	846	11.224
Diciembre	86.903		15.851	11.132	3.873	5.865	5.201	29.057	2.148	898	12.878
2011	**1.058.681**	**48.842**	**177.629**	**136.428**	**44.473**	**50.459**	**51.764**	**363.808**	**22.113**	**9.452**	**153.712**
Enero	87.248		17.026	11.468	4.065	4.200	5.160	28.669	2.861	851	12.948
Febrero	80.854		16.008	9.445	3.638	3.920	4.573	29.436	1.584	780	11.471
Marzo	86.014		15.857	10.965	4.263	3.900	4.661	30.636	1.879	712	13.140
Abril	82.510		14.539	11.125	3.776	3.990	4.263	29.733	1.922	785	12.378
Mayo	80.585		14.371	11.649	3.511	3.960	4.092	27.824	1.931	745	12.502
Junio	78.591		13.703	11.467	3.540	4.290	3.763	27.419	1.871	790	11.748
Julio	81.821		13.716	11.757	3.415	3.895	3.659	30.101	1.791	774	12.714
Agosto	84.598		13.981	11.828	3.502	4.425	3.509	31.596	1.815	760	13.182
Septiembre	79.547		13.333	11.375	3.397	4.170	3.462	29.327	1.139	755	12.589
Octubre	83.000		13.885	11.670	3.293	4.040	3.945	30.979	1.614	775	12.799
Noviembre	88.357		14.310	11.575	3.690	4.199	5.264	33.300	1.715	785	13.520
Diciembre	96.714		16.900	12.104	4.384	5.470	5.413	34.790	1.991	940	14.721
2012	**1.099.716**	**54.041**	**181.267**	**140.233**	**47.996**	**53.914**	**53.794**	**376.507**	**26.424**	**3.756**	**161.785**
Enero	91.519		16.546	11.596	4.583	4.860	4.793	31.767	2.637	851	13.886
Febrero	86.097		15.558	10.465	4.453	6.300	4.295	28.421	1.856	965	13.783
Marzo	93.156		16.102	11.726	4.538	5.700	4.454	34.044	1.938	687	13.968
Abril	84.431		14.820	11.069	4.036	4.310	4.269	30.083	1.773	621	13.450
Mayo	87.835		14.926	11.872	4.089	4.680	4.204	31.766	2.637	632	13.028
Junio	82.346		14.325	11.307	3.848	4.610	4.062	28.811	2.637	n.d.	12.746
Julio	83.413		14.426	11.475	3.922	3.613	4.301	30.090	2.637	n.d.	12.948
Agosto	83.989		13.891	12.024	3.764	3.935	4.207	30.851	1.988	n.d.	13.330
Septiembre	81.880		13.552	11.545	2.691	3.716	4.418	31.068	1.996	n.d.	12.894
Octubre	87.203		14.823	12.637	3.543	4.128	4.731	32.663	1.879	n.d.	12.799
Noviembre	88.420		15.196	12.457	4.122	4.038	4.875	31.691	2.315	n.d.	13.726
Diciembre	95.386		17.102	12.060	4.406	4.024	5.185	35.252	2.131	n.d.	15.227

2013	**1.167.095**	**54.047**	**186.378**	**166.849**	**49.389**	**49.918**	**49.668**	**400.928**	**24.290**	**n.d.**	**185.627**
Enero	98.767		17.458	15.039	4.568	3.895	4.795	34.362	2.050	n.d.	16.599
Febrero	89.131		15.378	11.289	4.497	3.690	4.295	34.104	1.169	n.d.	14.710
Marzo	90.393		15.996	13.088	4.707	3.450	4.454	31.603	2.100	n.d.	14.996
Abril	91.839		15.289	13.632	4.047	3.761	4.272	33.630	2.193	n.d.	15.015
Mayo	91.773		15.261	13.422	4.068	3.909	4.206	33.960	1.919	n.d.	15.028
Junio	87.871		14.563	13.800	3.846	3.725	4.171	31.685	2.127	n.d.	13.954
Julio	94.883		15.533	14.394	3.904	3.863	4.439	35.065	2.127	n.d.	15.558
Agosto	90.457		14.641	14.143	3.805	4.891	4.494	31.655	2.159	n.d.	14.670
Septiembre	86.032		14.050	13.798	2.859	4.152	3.255	30.471	2.098	n.d.	15.349
Octubre	93.212		15.173	14.677	3.710	3.951	3.602	34.058	2.121	n.d.	15.920
Noviembre	93.824		15.350	14.940	4.456	4.091	3.511	32.736	2.087	n.d.	16.652
Diciembre	104.866		17.687	14.627	4.922	6.540	4.173	37.599	2.142	n.d.	17.176
2014	**1.234.103**	**54.209**	**187.650**	**178.034**	**55.855**	**65.076**	**47.001**	**430.103**	**24.322**	**n.d.**	**191.853**
Enero	107.516		16.893	15.886	5.061	5.605	4.289	38.192	2.153	n.d.	19.437
Febrero	97.429		15.246	14.871	5.110	6.170	3.811	33.887	1.863	n.d.	16.473
Marzo	101.366		15.798	15.172	6.203	5.584	3.694	35.754	2.126	n.d.	17.036
Abril	97.841		15.327	13.914	4.532	5.480	3.545	37.005	2.102	n.d.	15.935
Mayo	95.710		15.369	14.159	2.747	5.372	3.731	35.594	1.839	n.d.	16.899
Junio	94.736		14.720	14.242	4.206	5.226	3.566	35.286	2.038	n.d.	15.453
Julio	95.603		15.857	14.326	4.242	4.995	3.829	35.285	2.038	n.d.	15.031
Agosto	94.733		14.938	15.473	4.152	4.932	3.621	34.904	2.067	n.d.	14.645
Septiembre	94.098		14.311	14.892	4.464	5.172	3.927	33.832	2.010	n.d.	15.491
Octubre	91.279		15.483	14.053	4.702	5.085	4.213	34.446	2.032	n.d.	11.264
Noviembre	95.697		15.665	13.619	4.677	4.980	4.049	34.618	1.999	n.d.	16.090
Diciembre	113.886		18.044	17.427	5.758	6.475	4.726	41.300	2.056	n.d.	18.099
2015	**1.319.375**	**56.575**	**206.308**	**177.517**	**53.710**	**62.949**	**53.459**	**488.737**	**19.805**	**n.d.**	**200.315**
Enero	119.340	4.805	18.568	16.102	5.636	5.895	4.757	39.650	3.668	n.d.	20.259
Febrero	103.835	4.340	16.273	14.426	5.025	5.450	4.327	35.025	1.616	n.d.	17.354
Marzo	110.470	4.805	17.743	15.816	4.663	5.160	4.515	38.658	2.426	n.d.	16.685
Abril	109.710	4.650	16.839	15.744	3.897	5.150	4.260	40.957	2.079	n.d.	16.133
Mayo	108.034	4.805	16.946	15.226	4.339	4.970	4.193	39.416	1.611	n.d.	16.528
Junio	108.643	4.650	16.728	14.384	4.919	5.290	4.415	40.729	2.259	n.d.	15.269
Julio	105.586	4.805	16.387	13.745	4.316	2.470	4.424	40.886	2.038	n.d.	16.515
Agosto	107.286	4.805	16.235	13.488	3.988	6.370	4.195	40.254	2.099	n.d.	15.853
Septiembre	105.219	4.650	16.284	13.338	4.021	5.290	4.200	39.820	2.010	n.d.	15.606
Octubre	109.148	4.805	16.985	14.410	4.383	5.830	4.291	42.123	n.d.	n.d.	16.321
Noviembre	108.923	4.650	17.347	14.530	3.896	5.980	4.447	41.297	n.d.	n.d.	16.777
Diciembre	123.181	4.805	19.972	16.309	4.627	5.094	5.435	49.923	n.d.	n.d.	17.016

2016	**1.426.988**	**60.987**	**212.554**	**171.337**	**57.044**	**47.335**	**56.648**	**558.229**	**28.069**	**17.950**	**216.836**
Enero	127.843	5.341	19.532	16.108	5.280	4.530	4.753	45.424	2.094	1.612	23.168
Febrero	116.273	4.827	18.122	15.140	5.048	3.200	4.881	42.137	2.190	1.560	19.169
Marzo	123.940	5.084	18.675	15.740	5.407	3.060	4.756	48.022	2.404	1.488	19.304
Abril	119.646	5.076	17.643	14.019	4.411	3.865	5.250	47.947	2.372	1.410	17.652
Mayo	116.886	5.066	17.840	14.588	4.438	4.039	4.269	45.772	2.173	1.488	17.213
Junio	112.831	5.076	17.055	13.819	4.254	3.905	4.277	44.275	2.133	1.380	16.657
Julio	116.978	5.060	17.004	14.760	4.311	4.155	4.492	46.329	2.462	1.426	16.978
Agosto	116.750	5.060	16.822	13.993	4.111	4.130	4.662	47.096	2.310	1.446	17.120
Septiembre	111.074	5.050	16.180	13.182	4.254	4.030	4.603	43.167	2.530	1.422	16.657
Octubre	116.319	5.076	17.137	13.678	4.864	3.850	4.566	45.584	2.458	1.465	17.643
Noviembre	118.958	5.063	16.609	14.250	4.987	3.930	4.795	48.036	2.263	1.627	17.398
Diciembre	129.490	5.208	19.935	12.060	5.680	4.641	5.343	54.440	2.681	1.626	17.876
2017	**1.521.884**	**57.199**	**224.433**	**192.008**	**60.512**	**47.717**	**59.060**	**599.853**	**19.553**	**16.511**	**245.038**
Enero	131.380	5.281	19.766	16.335	5.572	4.265	5.427	49.929	2.558	1.416	20.832
Febrero	117.159	4.776	16.896	13.953	4.836	4.375	4.743	46.965	1.989	1.400	17.225
Marzo	136.539	4.494	19.471	17.158	5.316	3.954	5.302	55.526	2.383	1.447	21.489
Abril	123.718	4.581	18.075	16.131	5.300	4.050	4.644	49.033	1.892	1.348	18.664
Mayo	134.967	4.701	19.600	17.218	5.279	3.850	4.929	54.127	2.017	1.352	21.895
Junio	131.776	4.401	18.970	16.332	5.163	4.030	4.616	51.372	1.854	1.356	23.680
Julio	124.000	4.736	17.946	15.330	4.836	3.638	4.657	50.722	1.501	1.332	19.302
Agosto	116.268	4.628	17.711	15.154	4.767	4.240	4.656	44.943	1.759	1.341	17.068
Septiembre	119.149	4.707	17.872	14.733	4.499	3.635	4.623	41.821	1.813	1.340	24.107
Octubre	124.899	4.700	18.515	16.022	4.456	3.530	4.838	50.417	1.788	1.368	19.265
Noviembre	125.945	4.936	18.388	15.864	4.848	3.980	4.886	52.082	n.d.	1.387	19.575
Diciembre	136.084	5.259	21.224	17.777	5.640	4.170	5.739	52.917	n.d.	1.423	21.936
2018	**1.616.728**	**60.652**	**236.370**	**212.196**	**61.492**	**45.810**	**62.881**	**636.928**	**26.100**	**17.042**	**257.257**
Enero	139.517	5.684	20.699	17.349	5.430	4.380	5.389	53.319	2.639	1.447	23.182
Febrero	126.094	5.821	19.088	17.049	5.285	3.800	5.178	45.448	2.068	1.547	20.810
Marzo	136.007	5.701	20.941	18.505	5.625	3.610	4.853	50.106	2.353	1.554	22.758
Abril	130.935	5.689	19.100	16.709	5.061	3.570	4.915	51.508	2.104	1.456	20.822
Mayo	133.634	5.780	19.376	16.602	5.326	3.490	5.761	52.554	2.096	1.384	21.265
Junio	126.030	4.647	18.619	16.752	5.201	3.425	5.367	48.447	2.016	1.355	20.200
Julio	129.250	5.030	18.945	17.602	4.273	3.345	4.939	51.412	2.026	1.364	20.313
Agosto	133.842	4.506	18.757	17.850	4.860	4.300	4.774	52.081	2.061	1.364	23.290
Septiembre	131.885	4.542	18.320	18.620	4.591	3.860	4.853	54.595	2.083	1.364	19.058
Octubre	138.256	3.497	20.106	17.600	4.686	3.750	5.399	58.412	2.252	1.368	21.185
Noviembre	140.460	4.400	19.980	17.280	5.165	3.880	5.254	58.958	2.088	1.404	22.052
Diciembre	150.818	5.355	22.438	20.278	5.990	4.400	6.198	60.087	2.314	1.435	22.322

2019	**1.600.938**	**22.461**	**230.674**	**199.123**	**64.781**	**43.460**	**64.317**	**672.341**	**30.642**	**18.359**	**254.780**
Enero	141.365	n.d.	14.593	20.042	6.407	4.380	5.690	62.780	2.666	1.596	23.211
Febrero	139.435	n.d.	23.900	17.321	5.728	4.290	5.361	57.009	2.266	1.617	21.943
Marzo	141.689	3.492	20.929	18.418	6.114	3.880	5.681	56.211	2.592	1.539	22.835
Abril	132.080	n.d.	20.100	17.387	5.503	3.665	5.546	54.331	2.282	1.513	21.754
Mayo	140.150	5.244	20.100	16.927	5.218	3.665	5.457	57.846	2.417	1.508	21.767
Junio	127.069	n.d.	18.683	16.421	5.109	3.490	5.031	54.735	2.153	1.436	20.011
Julio	129.955	n.d.	19.137	17.665	5.257	3.370	5.048	54.403	2.589	1.445	21.041
Agosto	127.840	n.d.	18.079	16.506	4.948	3.730	4.955	55.477	2.581	1.450	20.114
Septiembre	128.147	3.630	17.119	16.319	4.880	3.390	4.918	54.026	2.512	1.464	19.889
Octubre	125.635	3.145	19.714	16.307	4.684	2.410	5.064	48.759	2.880	1.548	21.125
Noviembre	120.395	3.975	15.996	12.902	4.716	3.340	5.384	52.865	2.885	1.572	16.760
Diciembre	147.178	2.975	22.325	12.909	6.217	3.850	6.184	63.898	2.820	1.670	24.330
2020	**1.621.303**	**39.505**	**234.939**	**192.140**	**62.923**	**45.300**	**68.121**	**659.547**	**36.978**	**19.497**	**262.353**
Enero	153.645	3.180	22.699	18.909	6.315	4.395	6.505	61.543	2.932	1.670	25.498
Febrero	140.110	2.509	21.779	18.408	6.021	4.120	5.835	53.864	2.856	1.611	23.106
Marzo	142.177	2.850	22.264	18.442	5.999	3.640	6.277	54.224	3.078	1.645	23.757
Abril	118.214	3.103	17.957	14.464	4.838	3.600	4.997	44.495	3.112	1.609	20.038
Mayo	119.900	3.022	17.910	12.722	4.078	3.775	4.972	48.789	3.125	1.607	19.902
Junio	125.706	3.531	18.827	13.896	4.916	3.500	5.111	50.599	3.140	1.578	20.608
Julio	131.959	4.340	17.983	13.144	4.266	3.550	4.997	58.643	3.123	1.583	20.330
Agosto	127.404	3.700	16.872	12.804	4.068	3.540	5.243	57.700	2.940	1.593	18.945
Septiembre	139.278	4.370	18.354	16.278	5.206	3.650	5.311	58.826	3.080	1.640	22.564
Octubre	141.696	3.270	19.128	17.393	5.635	3.670	6.157	58.729	3.175	1.647	22.894
Noviembre	134.506	4.007	19.271	17.399	5.374	3.710	5.835	52.927	3.190	1.651	21.142
Diciembre	146.709	1.624	21.896	18.282	6.208	4.150	6.881	59.208	3.228	1.663	23.569
2021[p]	**1.604.240**	**60.035**	**241.199**	**228.311**	**72.680**	**44.130**	**75.554**	**565.348**	**31.626**	**22.648**	**262.709**
Enero	149.835	4.899	21.787	19.853	8.182	4.010	6.628	56.385	3.230	1.670	23.190
Febrero	136.446	5.583	19.084	18.114	5.517	3.790	5.944	51.955	3.228	1.678	21.553
Marzo	142.738	5.946	21.582	18.295	6.148	3.550	6.764	52.965	2.815	1.673	23.000
Abril	133.716	5.417	18.227	19.487	5.960	3.490	6.197	49.311	2.434	1.676	21.518
Mayo	134.536	4.535	21.582	18.839	5.418	3.680	6.028	47.546	2.627	2.734	21.548
Junio	125.411	4.707	19.327	17.605	5.467	3.600	6.008	42.147	2.493	2.744	21.312
Julio	126.950	4.713	19.627	18.118	5.702	3.670	6.231	42.909	2.528	1.735	21.717
Agosto	123.815	4.546	20.162	18.068	5.425	3.700	5.807	41.745	2.072	1.735	20.555
Septiembre	124.589	4.724	18.357	18.514	5.557	3.540	6.315	42.943	2.438	1.737	20.463
Octubre	127.602	4.903	19.358	19.211	5.777	3.680	5.721	43.525	2.308	1.766	21.354
Noviembre	129.664	4.912	19.607	19.740	6.257	3.630	6.765	42.577	2.614	1.745	21.817
Diciembre	148.939	5.149	22.500	22.466	7.270	3.790	7.146	51.341	2.839	1.755	24.683

2022[p]	**1.049.807**	**44.176**	**154.092**	**158.250**	**52.691**	**30.350**	**51.025**	**352.802**	**24.001**	**8.292**	**174.127**
Enero	143.812	5.851	21.760	21.365	8.067	3.920	7.249	47.468	2.876	896	24.358
Febrero	128.004	5.380	19.621	19.752	6.953	3.830	6.454	41.172	2.533	888	21.420
Marzo	138.202	5.939	21.515	21.437	6.419	3.760	5.971	45.154	2.772	954	24.283
Abril	132.657	5.637	19.736	20.135	6.641	3.680	6.160	44.750	3.067	1.279	21.573
Mayo	129.034	5.439	19.606	18.869	6.236	3.700	6.444	43.549	3.146	965	21.082
Junio	125.333	5.884	18.377	18.891	6.039	3.650	6.197	42.226	3.193	970	19.906
Julio	124.158	5.729	14.080	19.288	6.340	3.730	6.284	44.009	3.142	1.057	20.499
Agosto	128.607	4.317	19.398	18.512	5.996	4.080	6.267	44.473	3.273	1.283	21.007

Fuente: Gobiernos Autónomos Municipales

Instituto Nacional de Estadística

(p): Preliminar

n.d.: no disponible, por falta de pesaje o no emisión del registro.

Nota: Sólo se cuentan con datos anuales, previos al año 2015 para la ciudad de Sucre.