

**UNIVERSIDAD PÚBLICA DE EL ALTO**

**CARRERA INGENIERÍA DE SISTEMAS**



**TESIS DE GRADO**

**“MODELO DE PREDICCIÓN DEL CRECIMIENTO DEL ÍNDICE  
DELICTIVO EN LA CIUDAD DE EL ALTO MEDIANTE MACHINE  
LEARNING”**

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

**Postulante: Romina Ramos Condori**

**Tutor Metodológico: M. Sc. Lic. Ing. Fanny Helen Perez Mamani**

**Tutor Especialista: Lic. Fredy Alanoca Coareti**

**Tutor revisor: M. Sc. Lic. Wendy Yomar Sarmiento Martinez**

**EL ALTO - BOLIVIA**

**2024**

# DECLARACIÓN JURADA DE AUTENTICIDAD Y RESPONSABILIDAD

Yo, **Romina Ramos Condori con C.I. 12956031 LP.** mediante la presente **declaro** de manera pública que la propuesta de **TESIS DE GRADO** titulado “**MODELO DE PREDICCIÓN DEL CRECIMIENTO DEL ÍNDICE DELICTIVO EN LA CIUDAD DE EL ALTO MEDIANTE MACHINE LEARNING**” es original, siendo resultado de mi trabajo personal y no constituye una copia o replica de trabajos similares elaborados.

Autorizo la publicación del resumen de mi propuesta en internet y me comprometo a responder a todos los cuestionamientos que se desprenden de su lectura.

Asimismo, me hago responsable ante la universidad o terceros, de cualquiera irregularidad o daño que pudiera ocasionar, por el incumplimiento de lo declarado.

De identificarse falsificación, plagio, fraude, o que la **TESIS DE GRADO** haya sido publicado anteriormente; asumo las consecuencias y sanciones que de mi acción se deriven, responsabilizándome por todas las cargas legales que se deriven de ello sometiéndome a las normas establecidas y vigentes de la Carrera de Ingeniería de Sistemas de la Universidad Pública de El Alto.

El Alto, junio de 2024

**Romina Ramos Condori**  
**C.I. 12956031 LP**  
**romyramos234@gmail.com**

## DEDICATORIA

A mis padres, Elisa Condori y Nelson Ramos por su amor incondicional y su apoyo constante a lo largo de mi vida. Su sabiduría y sacrificio han sido la base de mis logros.

A mi hermana Judith Ramos, por su apoyo, por sus palabras de aliento y por estar siempre a mi lado en los momentos difíciles.

A Jhon por sus palabras de aliento, apoyo y comprensión durante todo este proceso, me han dado la motivación necesaria para seguir adelante.

A mis amigos, que con su compañía y comprensión han hecho este viaje más llevadero y significativo.

A mis profesores y tutores, por su paciencia, conocimiento y guía, que han sido fundamentales para mi crecimiento académico y personal.

A todos aquellos que de una manera u otra han contribuido a que este sueño se haga realidad, con su apoyo, consejos y motivación.

Finalmente, dedico este trabajo a mí mismo, por la perseverancia, el esfuerzo y la dedicación invertidos en alcanzar esta meta.

Romina Ramos Condori

## AGRADECIMIENTOS

Quiero expresar mi más sincero agradecimiento a todas las personas e instituciones que han hecho posible la realización de esta tesis.

En primer lugar, me gustaría agradecer a mis tutores de tesis, por su orientación y apoyo durante el proceso de investigación. Su experiencia y consejos han sido fundamentales para la culminación de este trabajo.

A la Universidad Pública de El Alto, por proporcionarme los recursos y el entorno académico adecuado para llevar a cabo esta investigación.

A mis compañeros de estudio y amigos, por los momentos de aprendizaje y por estar dispuestos a compartir ideas y conocimientos.

A mi compañero, amigo por estar a mi lado, por creer en mí y por compartir conmigo cada triunfo y cada desafío.

Un agradecimiento especial a mi familia, cuyos ánimos y comprensión han sido una fuente constante de motivación. A mis padres, por creer en mí y por su amor incondicional, por su paciencia y apoyo en los momentos más difíciles.

Finalmente, quiero agradecer a todas las personas que, de una u otra manera, han contribuido a la realización de esta tesis. Su apoyo y colaboración han sido esenciales para alcanzar este logro.

## ÍNDICE GENERAL

CAPÍTULO I.....	1
1.1. MARCO PRELIMINAR .....	1
1.1.1. Introducción .....	1
1.2. ANTECEDENTES.....	2
1.2.1. Antecedentes Internacionales.....	2
1.2.2. Antecedentes Nacionales.....	2
1.2.3. Antecedentes Locales .....	3
1.3. PLANTEAMIENTO DEL PROBLEMA.....	3
1.3.1. Problema Principal .....	3
1.3.2. Problemas secundarios.....	4
1.3.3. Formulación del problema.....	4
1.4. OBJETIVOS.....	5
1.4.1. Objetivo General.....	5
1.4.2. Objetivos Específicos .....	5
1.5. HIPÓTESIS.....	5
1.5.1. Hipótesis de Investigación.....	5
1.5.2. Hipótesis Nula.....	5
1.5.3. Hipótesis Alterna .....	6
1.5.4. Identificación de Variables .....	6
1.5.5. Operacionalización de Variables.....	6
1.6. JUSTIFICACIÓN.....	7
1.6.1. Justificación Científica .....	7
1.6.2. Justificación Técnica .....	7
1.6.3. Justificación Económica .....	7
1.6.4. Justificación Social .....	7
1.7. METODOLOGÍA .....	8

1.7.1.	Método Científico.....	8
1.7.1.1.	Tipo de investigación .....	8
1.7.1.2.	Estudio de investigación .....	8
1.7.1.3.	Diseño de investigación .....	9
1.7.1.4.	Enfoque cuantitativo.....	10
1.7.2.	Método de ingeniería.....	10
1.7.2.1.	Metodología CRISP-DM.....	10
1.7.2.1.1.	Fases de la metodología CRISP-DM .....	10
1.8.	HERRAMIENTAS .....	13
1.8.1.	Sistema operativo.....	13
1.8.2.	Lenguaje de programación.....	13
1.8.3.	Entornos de desarrollo integrado (IDE) .....	13
1.8.4.	Librerías para ciencia de datos .....	14
1.9.	LIMITES Y ALCANCES.....	15
1.9.1.	Limites.....	15
1.9.2.	Alcances .....	15
1.10.	APORTES .....	16
CAPÍTULO II.....		18
1.11.	MARCO TEÓRICO .....	18
1.11.1.	Introducción.....	18
1.12.	DELINCUENCIA.....	18
1.12.1.	La delincuencia en Bolivia .....	19
1.13.	INSEGURIDAD CIUDADANA.....	20
1.13.1.	La inseguridad en la ciudad de El Alto.....	22
1.13.1.1.	Planes para combatir la inseguridad .....	26
1.13.1.2.	Zonas de mayor riesgo .....	27
1.13.1.3.	Causas de la Delincuencia .....	30

1.14.	MACHINE LEARNING .....	31
1.14.1.	Aprendizaje supervisado .....	32
1.14.1.1.	Regresión .....	32
1.14.1.1.1.	Regresión lineal .....	33
1.14.1.2.	Clasificación .....	39
1.14.1.2.1.	Regresión logística .....	40
1.14.1.2.2.	K Vecinos más Cercanos KNN.....	41
1.14.1.2.3.	Árboles de Decisión .....	43
1.14.1.2.4.	Random Forest .....	44
1.14.1.2.5.	Máquinas de Vectores de Soporte (Support Vector Machines, SVM) 45	
1.14.1.2.6.	Redes Neuronales Artificiales .....	46
1.14.1.3.	Ventajas del aprendizaje supervisado .....	48
1.14.1.4.	Desventajas del aprendizaje supervisado .....	49
1.14.2.	Aprendizaje no supervisado .....	49
1.14.2.1.	Agrupamiento .....	50
1.14.2.2.	Asociación .....	50
1.14.3.	Aprendizaje por refuerzo .....	51
1.14.3.1.	Elementos del aprendizaje por refuerzo.....	51
1.14.3.2.	Categorías del aprendizaje por refuerzo .....	51
1.14.3.3.	Ventajas del aprendizaje por refuerzo .....	52
1.14.3.4.	Desventajas del aprendizaje por refuerzo .....	52
1.14.4.	Aprendizaje semi supervisado.....	52
1.14.4.1.	Ventajas del aprendizaje semisupervisado .....	53
1.14.4.2.	Desventajas del aprendizaje semisupervisado .....	53
1.15.	MÉTODO CIENTÍFICO .....	53
1.16.	METODOLOGÍA CRISP-DM .....	53
1.16.1.	Fases de CRISP-DM .....	55

1.16.1.1.	Compresión del problema.....	55
1.16.1.2.	Compresión de los datos .....	56
1.16.1.3.	Preparación de los datos .....	57
1.16.1.4.	Modelado.....	59
1.16.1.5.	Evaluación.....	60
1.16.1.6.	Despliegue.....	61
CAPÍTULO III.....		63
1.17.	DISEÑO METODOLÓGICO .....	63
1.17.1.	Introducción.....	63
1.17.2.	Tipo de investigación .....	63
1.17.3.	Diseño de la investigación.....	63
1.17.4.	Variables de la investigación .....	64
1.17.4.1.	Variable dependiente .....	64
1.17.4.2.	Variable independiente .....	65
1.17.5.	Ambiente de la investigación.....	65
1.17.5.1.	Contexto geográfico .....	65
1.17.6.	Descripción de la metodología a usar .....	65
1.17.6.1.	Compresión del problema.....	65
1.17.6.2.	Compresión de los datos .....	66
1.17.6.3.	Preparación de los datos .....	71
1.17.6.4.	Modelado.....	74
1.17.6.5.	Evaluación.....	75
1.17.6.6.	Despliegue.....	75
1.18.	HERRAMIENTAS.....	75
1.19.	HERRAMIENTAS A USAR .....	75
1.19.1.	Técnicas de investigación e instrumentos .....	75
1.20.	MÉTRICAS DE EVALUACIÓN .....	76



1.20.1.	Error Cuadrático Medio (MSE) .....	76
1.20.2.	Raíz del Error Cuadrático Medio (RMSE) .....	76
1.20.3.	Error Absoluto Medio (MAE).....	77
CAPÍTULO IV .....		80
1.21.	PRUEBAS Y RESULTADOS.....	80
1.21.1.	Introducción.....	80
1.21.2.	PRESENTACIÓN DEL MODELO.....	80
1.21.2.1.	Comprensión de los datos .....	82
1.21.2.2.	Preparación de datos .....	83
1.21.2.3.	Modelado.....	84
1.21.2.3.1.	Algoritmos de Machine Learning .....	84
1.21.2.4.	Evaluación.....	87
1.21.3.	DESARROLLO DEL MODELO .....	88
1.21.4.	DEMOSTRACIÓN DEL PROTOTIPO.....	91
1.21.5.	PRUEBA DE HIPÓTESIS.....	93
1.21.5.1.	Pruebas estadísticas.....	93
CAPÍTULO V .....		99
1.22.	CONCLUSIONES Y RECOMENDACIONES .....	99
1.22.1.	CONCLUSIONES .....	99
1.22.2.	RECOMENDACIONES.....	99
BIBLIOGRAFÍA.....		102

## ÍNDICE DE FIGURAS

<b>Figura 1</b> Distribución de delitos en departamentos y ciudades .....	19
<b>Figura 2</b> Delitos registrados por día y hora .....	21
<b>Figura 3</b> Datos generales de la gestión 2022 .....	22
<b>Figura 4</b> Tipos de delitos registrados en la Policía boliviana.....	25
<b>Figura 5</b> Plan de patrullaje para combatir la delincuencia .....	26
<b>Figura 6</b> Puntos calientes de denuncias de la ciudad de El Alto.....	28
<b>Figura 7</b> Datos de seguridad ciudadana .....	29
<b>Figura 8</b> Denuncias registradas por día .....	30
<b>Figura 9</b> Datos socioeconómicos .....	31
<b>Figura 10</b> Grafica de regresión lineal .....	34
<b>Figura 11</b> Grafica regresión lineal múltiple .....	35
<b>Figura 12</b> Gráfico regresión polinómica .....	36
<b>Figura 13</b> Grafica regresión polinómica Ridge.....	37
<b>Figura 14</b> Grafica de regresión polinómica Lasso .....	38
<b>Figura 15</b> Grafica de clasificación de datos .....	39
<b>Figura 16</b> Grafica de regresión logística .....	41
<b>Figura 17</b> Grafica de k vecinos más cercanos k.....	42
<b>Figura 18</b> Diagrama de árbol de decisión .....	44
<b>Figura 19</b> Estructura de random forest .....	45
<b>Figura 20</b> Grafica de Maquina de vectores de Maquina.....	46
<b>Figura 21</b> Estructura de redes neuronales.....	47
<b>Figura 22</b> Fases de la metodología CRISP-DM.....	54
<b>Figura 23</b> Días y horas con más frecuencias de delitos .....	66
<b>Figura 24</b> Gráfico de distribución de denuncias de la gestión 2013 a 2023 .....	68
<b>Figura 25</b> Grafico de denuncias por tipología gestión 2013 a 2023.....	68
<b>Figura 26</b> Grafico porcentual por gestiones 2013 a 2023 .....	69
<b>Figura 27</b> División del conjunto de datos .....	71
<b>Figura 28</b> Inversión de tablas por año.....	71
<b>Figura 29</b> Imputar valores faltantes.....	72
<b>Figura 30</b> Eliminar ceros de filas y columnas .....	72
<b>Figura 31</b> Conversión de valores numéricos a categóricos.....	73
<b>Figura 32</b> Datos de prueba y entrenamiento .....	74
<b>Figura 33</b> Evaluación del modelo .....	75

<b>Figura 34</b>	Fases de la metodología de CRISP DM.....	80
<b>Figura 35</b>	Fases del modelo.....	81
<b>Figura 36</b>	Fragmento de delitos .....	82
<b>Figura 37</b>	Importación de librerías.....	83
<b>Figura 38</b>	Lectura de datos.....	83
<b>Figura 39</b>	Creación de una lista .....	83
<b>Figura 40</b>	Transformación de regresión polinomial .....	84
<b>Figura 41</b>	Grafica de regresión lineal .....	85
<b>Figura 42</b>	Grafica de regresión polinómica .....	86
<b>Figura 43</b>	Grafica de árbol de decisión .....	86
<b>Figura 44</b>	Grafica de Random Forest.....	87
<b>Figura 45</b>	Evaluación.....	87
<b>Figura 46</b>	Formula de aproximación .....	88
<b>Figura 47</b>	Implementación del modelo por mes con árbol de decisión .....	91
<b>Figura 48</b>	Implementación del modelo por año en regresión polinomial.....	92
<b>Figura 49</b>	Resultados de la demostración con regresión polinomial.....	93
<b>Figura 50</b>	Tabla t student.....	94
<b>Figura 51</b>	Ecuación t Student .....	95

## ÍNDICE DE TABLAS

<b>Tabla 1</b> Operacionalización de variables.....	6
<b>Tabla 2</b> Conjuntos de datos para regresiones .....	33
<b>Tabla 3</b> Conjunto de datos para clasificación.....	39
<b>Tabla 4</b> Registro de denuncias de los delitos mas frecuentes.....	67
<b>Tabla 5</b> Fragmento de conjunto de datos registrados en la gestión 2022.....	70
<b>Tabla 6</b> Conversion de variables categoricas a numericas .....	73
<b>Tabla 7</b> Porcentaje de aproximación del valor real .....	88
<b>Tabla 8</b> Estructura de los pasos de la regresión lineal .....	89
<b>Tabla 9</b> Estructura de los pasos de regresión polinomial .....	89
<b>Tabla 10</b> Estructura de los pasos del árbol de decisiones.....	90
<b>Tabla 11</b> Estructura de los pasos de Random Forest.....	90
<b>Tabla 12</b> Tabla de comparación con valores predecidos vs valor real.....	95

## ÍNDICE DE ECUACIONES

( 1) Ecuacion de regresion lineal .....	34
( 2) Ecuación de regresión múltiple .....	35
( 3) Ecuación de regresión polinómica con una variable Independiente .....	36
( 4) Ecuación de regresión polinómica con múltiples variables independientes.....	36
( 5) Ecuación de regresión polinómica Lasso.....	37
( 6) Ecuación de regresión logística.....	38
( 7) Ecuación KNN.....	40
( 8) Ecuación de árbol de decisión.....	42
( 9) Ecuación de Random Forest .....	43
( 10) Ecuación tasa anual de delincuencia .....	45
( 11) Ecuación tasa de delito por día .....	23
( 12) Ecuación de Error cuadrático medio .....	24
( 13) Ecuación de Raíz del Error Cuadrático Medio .....	76
( 14) Ecuación Error Absoluto Medio .....	77
( 15) Ecuación Suma de los cuadrados totales (SST).....	77
( 16) Ecuación Suma de los cuadrados de os residuos (SSR) .....	77
( 17) Ecuación $r^2$ .....	78
( 18) Ecuacion t student descripcion.....	78
( 19) Ecuación de la media muestral .....	96
( 20) Ecuación desviación estándar.....	96
( 21) Ecuación t Student.....	96

## RESUMEN

El presente trabajo de investigación surge con el objetivo de analizar el índice de crecimiento delictivo, motivado por el aumento de la delincuencia y la sensación de inseguridad en El Alto. Para esta investigación, se solicitó el registro de datos estadísticos de las gestiones 2013 a 2023, sobre hechos delictivos a la división de la FELCC en la unidad de estadísticas del comando general. Los datos obtenidos corresponden a los índices delictivos registrados en la ciudad de El Alto, clasificados por tipología de delito a lo largo de diferentes períodos.

En la primera parte se detalla la introducción del trabajo de investigación, incluyendo el planteamiento del problema, la definición de los objetivos, la formulación de la hipótesis y la identificación de las variables de estudio, tanto independientes como dependientes. Además, se explica la justificación del estudio desde los aspectos científico, técnico, económico y social. Y por último se describen las herramientas que se utilizarán en el modelo de Machine Learning, así como sus limitaciones y alcances.

En la segunda parte se definen conceptos clave relacionados con la delincuencia, la inseguridad ciudadana, el Machine Learning y la metodología de investigación.

En la tercera parte se describe el diseño metodológico, el tipo, las variables y ambiente de la investigación y las métricas de calidad.

En el capítulo cuarto se presentan las pruebas y resultados obtenidos con el modelo de predicción y la demostración del prototipo.

Finalmente, en el capítulo quinto se presentan las conclusiones y recomendaciones para la prevención futura.

**Palabras Claves:** Machine Learning, Índice de crecimiento delictivo, Predicción

## ABSTRACT

The present research work arises with the objective of analyzing the crime growth rate, motivated by the increase in crime and the feeling of insecurity in El Alto. For this research, statistical data records from 2013 to 2023 were requested from the FELCC division in the statistics unit of the general command. The obtained data correspond to the crime rates recorded in the city of El Alto, classified by type of crime over different periods.

The first part details the introduction of the research work, including the problem statement, the definition of objectives, the formulation of the hypothesis, and the identification of the study variables, both independent and dependent. Additionally, the justification of the study is explained from scientific, technical, economic, and social aspects. Finally, the tools that will be used in the Machine Learning model are described, along with their limitations and scope.

In the second part, key concepts related to crime, citizen insecurity, Machine Learning, and research methodology are defined.

The third part describes the methodological design, type, variables, and research environment, and the quality metrics.

In the fourth chapter, the tests and results obtained with the prediction model and the demonstration of the prototype are presented

Finally, in the fifth chapter, conclusions and recommendations for future prevention are presented.

**Keywords:** Machine Learning, Crime Growth Rate, Prediction



# CAPÍTULO I



## CAPÍTULO I

### 1.1. MARCO PRELIMINAR

#### 1.1.1. *Introducción*

La ciudad de El Alto es un municipio situado en el departamento de La Paz en Bolivia, es la ciudad más poblada del departamento de La Paz y se destaca como la segunda área metropolitana más grande a nivel nacional debajo de Santa Cruz de la Sierra y encima de la sede de Gobierno. Con una población estimada en 1.109.048 habitantes (según las proyecciones del Instituto Nacional de Estadísticas para el 2022).

Según el Observatorio Nacional de Seguridad Ciudadana, el 45% de los habitantes de la ciudad de El Alto consideran que la principal causa de la inseguridad en su barrio es el desempleo y la extrema pobreza y un 72,7% tiene sentimiento de inseguridad en su propia zona, según la “Primera encuesta de victimización, prácticas y percepción sobre violencia y delito”. Un índice delictivo se refiere a un indicador estadístico que recopila datos de un espacio y periodo de tiempo específico, comúnmente se usa en contextos económicos. Para esto se utilizarán datos históricos de delitos registrados, así como variables sociales, económicas y demográficas que podrían influir en los patrones delictivos.

Los modelos de predicción son herramientas que utilizan datos históricos estadísticos para predecir un evento futuro usando algoritmos de aprendizaje automático. Cuenta con distintas fases, como el entrenamiento del conjunto de datos, los ajustes de parámetros, la construcción del modelo y la predicción de los nuevos ejemplos. El uso de técnicas avanzadas de Machine Learning puede ofrecer una manera para anticipar el crecimiento del índice delictivo, permitiendo a las autoridades y ciudadanos tomar medidas preventivas y optimizar los recursos para la seguridad ciudadana.

En esta investigación, se propone desarrollar un modelo predictivo utilizando métodos de Machine Learning para estimar el crecimiento del índice delictivo en la ciudad de El Alto.

## **1.2. ANTECEDENTES**

### **1.2.1. Antecedentes Internacionales**

Bustamante (2019), “EXTRACCIÓN DE INFORMACIÓN PARA LA GENERACIÓN DE REPORTES ESTRUCTURADOS A PARTIR DE NOTICIAS PERUANAS RELACIONADAS A CRÍMENES”, su objetivo general es implementar un modelo algorítmico de extracción de información de descripciones de delitos en noticias peruanas relacionadas a crímenes. Para ayudar a las fuerzas policiales a la recolección de tipo de información sobre crímenes. Esto con el objetivo de poder analizar los datos disponibles y utilizar los resultados de esta tarea para la mejora de procesos actuales, e incluso, para la prevención de ataques y delitos futuros. Las herramientas utilizadas fueron Python como lenguaje principal y sus librerías BeautifulSoup, Natural Language Toolkit, PyTorch, Jupyter y MongoDB base de datos. La metodología de desarrollo con web scraping, la tesis se realizó en la PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ.

### **1.2.2. Antecedentes Nacionales**

Mantilla (2019), “ANÁLISIS DE LA DISTRIBUCIÓN ESPACIAL DEL DELITO DE ROBO A PERSONAS EN EL BARRIO MARISCAL SUCRE PERÍODO (2015-2018)”, el objetivo principal es analizar la distribución espacial del delito de robo a personas en el Barrio Mariscal Sucre en el período 2015-2018. El presente estudio se relaciona con la gestión ambiental al tomar en cuenta la “Declaración de Quito Sobre Ciudades y Asentamientos Humanos Sostenibles Para Todos” de la Organización de las Naciones Unidas (ONU- HABITAT III). La herramienta principal que se utilizó fue el software ArcGIS, la tesis se realizó en la PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR.

Mamani (2020), "MODELO BASADO EN INTELIGENCIA ARTIFICIAL PARA LA DETECCIÓN DE CIBERACOSO", el objetivo principal es desarrollar un modelo inteligente que detecte el ciberacoso en jóvenes de 15 a 17 años en Bolivia. Para detectar el ciberacoso se realizó un modelo basado en Inteligencia Artificial, para la cual se elaboró un formulario que tenía como fin reunir un conjunto de jergas que son utilizadas por los estudiantes, también se recopiló conversaciones de una red social de estudiantes de secundaria. La tesis se realizó en la UNIVERSIDAD MAYOR DE SAN ANDRÉS.

### **1.2.3. Antecedentes Locales**

Hidalgo, (2020) "MODELO DE PREDICCIÓN APLICADO AL INDICE CRECIMIENTO DE CANCER DE CUELLO UTERINO EN LA CIUDAD DE EL ALTO EN BASE A ALGORITMOS GENÉTICOS" La presente tesis de grado plantea un modelo de predicción del índice de crecimiento de cáncer de cuello uterino en base a algoritmos genéticos que permite pronosticar el crecimiento de esta enfermedad en cinco años, además este instrumento es de gran beneficio para los especialistas y mujeres. La metodología usada es la XP y se utilizó herramientas como java y php. La tesis se realizó en la UNIVERSIDAD PUBLICA DE EL ALTO.

## **1.3. PLANTEAMIENTO DEL PROBLEMA**

### **1.3.1. Problema Principal**

La ciudad enfrenta desafíos significativos en términos de inseguridad, que no solo afectan la calidad de vida de los ciudadanos al aumentar el miedo y la ansiedad, sino que también obstaculizan el desarrollo económico. Los robos en domicilios, negocios e inmuebles provocan pérdidas económicas considerables, lo cual limita las oportunidades de crecimiento y desarrollo en la ciudad. La inseguridad también afecta el funcionamiento de las instituciones

y servicios públicos, ya que recursos que podrían destinarse a la mejora de infraestructuras o servicios sociales deben redirigirse a la seguridad y el orden público.

Para el año 2020, la FELCC informa que se registró 2082 denuncias en total en distintos tipos de delitos en la ciudad de El Alto. Encabezando los delitos de denuncias registradas como la muerte de personas, las lesiones graves y los robos. Entre ellos, en el mes de julio se registró 194 muertes de personas, en el mes de enero se produjo 47 lesiones graves y 50 robos solo en el mes de febrero. Estos datos estadísticos manifiestan cifras preocupantes de inseguridad ciudadana pero estos datos solo son los hechos denunciados en oficinas de las autoridades policiales.

### **1.3.2. Problemas secundarios**

- La desconfianza por parte de los ciudadanos hacia las instituciones de seguridad pública.
- El incremento en denuncias por delitos cometidos refleja la disposición de la población a reportar incidentes delincuenciales, posiblemente impulsada por el aumento de la inseguridad y campañas de concientización sobre la importancia de denunciar.
- El aumento de la migración de los habitantes de la ciudad, optando por mudarse a áreas más seguras, debido a hechos delictivos registrados en la ciudad, por el incremento de la inseguridad y la percepción de deficiencia de desempeño de las autoridades en garantizar la protección de los ciudadanos.

### **1.3.3. Formulación del problema**

¿De qué manera se puede anticipar las actividades delictivas en la ciudad de El Alto?

## **1.4. OBJETIVOS**

### **1.4.1. *Objetivo General***

Diseñar un modelo de predicción del crecimiento del índice delictivo para que permita anticipar la actividad delictiva en la ciudad de El Alto mediante Machine Learning.

### **1.4.2. *Objetivos Específicos***

- Recopilar la información de datos estadísticos sobre hechos delincuenciales en El Alto de los registros policiales en la ciudad.
- Identificar patrones de datos para entender mejor las tendencias en la incidencia delictiva.
- Utilizar un algoritmo de Machine Learning adecuado para entrenar el modelo de predicción.
- Validar el modelo de Machine Learning para determinar si se ajusta bien a los nuevos conjuntos de datos.

## **1.5. HIPÓTESIS**

### **1.5.1. *Hipótesis de Investigación***

El modelo de predicción del crecimiento del índice delictivo mediante el uso de métodos de Machine Learning permitirá anticipar las actividades delictivas en la ciudad de El Alto.

### **1.5.2. *Hipótesis Nula***

El modelo de predicción del crecimiento del índice delictivo mediante el uso de métodos de Machine Learning no permitirá anticipar las actividades delictivas en la ciudad de El Alto.

### 1.5.3. Hipótesis Alterna

El modelo de predicción del crecimiento del índice delictivo mediante el uso de métodos de Machine Learning permitirá anticipar las actividades delictivas en la ciudad de El Alto y tener una mejor planificación para entornos más seguros.

### 1.5.4. Identificación de Variables

- **Variable independiente:** Modelo de predicción del crecimiento del índice delictivo mediante el uso de métodos de Machine Learning
- **Variable dependiente:** Anticipar actividades delictivas

### 1.5.5. Operacionalización de Variables

Tabla 1

Operacionalización de variables

VARIABLES	DEFINICIÓN CONCEPTUAL	DIMENSIONES	INDICADORES	INSTRUMENTOS
<b>Variable independiente</b>  Modelo de predicción del crecimiento del índice delictivo mediante el uso de métodos de Machine Learning	Un modelo predictivo es un sistema que utiliza datos y algoritmos para hacer predicciones sobre eventos futuros.	Métricas de evaluación	Porcentaje de predicciones aproximadas	Software de análisis predictivo Python
	Machine Learning es un subcampo de la (IA) se enfoca en el desarrollo de algoritmos y modelos estadísticos	Algoritmos	Coefficiente de determinación	Bibliotecas de Machine Learning Scikit-learn
<b>Variable dependiente</b>  Anticipar actividades delictivas	Anticipar actividades delictivas implica predecir eventos delincuenciales antes de que ocurran utilizando métodos analíticos y datos históricos.	Fuente de datos	Registros de información.  Tasa de crecimiento de índice delictivo	Registro de datos policiales
		Frecuencia temporal	Cantidad de delitos por periodo de tiempo.	Herramientas de visualización matplotlib

## **1.6. JUSTIFICACIÓN**

### **1.6.1. *Justificación Científica***

Los algoritmos de Machine Learning permiten automatizar tareas repetitivas y rutinarias, liberando tiempo y recursos humanos para actividades más creativas y estratégicas. En particular, el modelo de predicción basado en algoritmos de Machine Learning pronostica oportunamente el crecimiento del índice delictivo en la ciudad de El Alto. Esto resulta beneficioso para la población, ya que facilita la prevención oportuna de delitos al proporcionar alertas tempranas y permitir una mejor respuesta por parte de las autoridades.

### **1.6.2. *Justificación Técnica***

Se utiliza herramientas técnicas avanzadas, incluyendo software especializado en análisis de datos y desarrollo de algoritmos de Machine Learning, para mejorar la comprensión y la prevención de la delincuencia en nuestra ciudad. Además, contamos con acceso a datos detallados sobre la incidencia delictiva, lo que nos permite realizar análisis más precisos.

### **1.6.3. *Justificación Económica***

La construcción del modelo de predicción no tiene costo alguno, ya que se utiliza un servicio de software basado en la web. Entre los beneficios se destacan la prevención de daños materiales y pérdidas económicas derivadas de la delincuencia, la mejora en la calidad de vida de los ciudadanos y el aumento del atractivo turístico de la ciudad.

### **1.6.4. *Justificación Social***

El desarrollo de un modelo predictivo para el crecimiento del índice delictivo responde a la necesidad imperante de aportar métodos para la seguridad ciudadana. Este enfoque permite anticipar incrementos en la delincuencia durante períodos específicos, brindando a las autoridades la capacidad de implementar medidas preventivas efectivas. Al prever estas

situaciones, se facilita la planificación de recursos y estrategias para mitigar los delitos, promoviendo así un entorno más seguro y protegido para todos los habitantes.

## **1.7. METODOLOGÍA**

### **1.7.1. Método Científico**

El método científico es un proceso de investigación, que mediante un proceso de pasos, permite llevar adelante un estudio, adquirir nuevos conocimientos y corroborar la veracidad o no del resultado (Etece, 2023).

Para esta investigación se sugiere un método científico que utiliza el método inductivo, este método se basa en observaciones específicas para formular una hipótesis general. En este caso, se utilizarían datos específicos del índice delictivo en la ciudad de El Alto para desarrollar un modelo de predicción utilizando técnicas de Machine Learning. La idea es extraer patrones y generalizar a partir de esos datos específicos para hacer predicciones sobre el crecimiento del índice delictivo en el futuro.

#### **1.7.1.1. Tipo de investigación**

Para esta investigación, el presente estudio trabaja con el tipo de investigación correlacional, debido a que existe una relación entre el uso del modelo de predicción y la capacidad de anticipar actividades delictivas, sin afirmar que uno cause directamente cambios en el otro.

#### **1.7.1.2. Estudio de investigación**

Un estudio transversal permite capturar una representación en un momento específico en el tiempo de una serie de datos o características de una población, de la situación actual delictiva en la ciudad. Esto proporciona un punto de partida claro para el análisis y la



construcción del modelo predictivo. Al tener datos de un solo momento, se puede determinar la prevalencia de delitos en diferentes áreas y tiempos.

Realizar un estudio transversal suele ser más rápido y menos costoso que un estudio longitudinal, ya que no requiere seguimiento, se recopilan datos de la población en un solo momento, evitando la necesidad de múltiples rondas de recopilación de datos. Aunque un estudio transversal no puede establecer causalidad, puede identificar asociaciones significativas entre variables, como:

- Factores socioeconómicos: Desempleo, pobreza, nivel educativo.
- Factores demográficos: Edad, género, composición del hogar.
- Factores espaciales: Densidad de población, zonas de alta delincuencia.

Utilizar un estudio transversal como parte de la implementación de un modelo de predicción del índice delictivo en El Alto es una buena estrategia. Permite obtener una visión detallada de la situación actual, establecer una línea base sólida y proporcionar datos de calidad para entrenar modelos de Machine Learning. Aunque no establece causalidad, facilita la identificación de asociaciones clave que pueden ser cruciales para el desarrollo de un modelo predictivo preciso.

### **1.7.1.3. Diseño de investigación**

Un diseño no experimental es ideal para estudiar la predicción de actividades delictivas en El Alto mediante Machine Learning, ya que permite observar y analizar datos históricos sin manipular las variables directamente. Este enfoque utiliza software de análisis predictivo, bibliotecas de Machine Learning, registros policiales y herramientas de visualización para identificar patrones y hacer predicciones sobre eventos delictivos futuros. Aunque no

establece relaciones causales sobre las variables, proporciona una comprensión valiosa de las tendencias delictivas, facilitando la planificación y prevención de delitos en la ciudad.

#### **1.7.1.4. Enfoque cuantitativo**

Se basa en recopilar y analizar datos numéricos, este tipo de investigación se caracteriza por usar métodos y técnicas para la recolección sistemática de datos en forma de números, estadísticas y medidas. El tipo de enfoque de esta investigación será de tipo cuantitativo, se realizará la recolección de datos estadísticos de acuerdo al registro de información disgregada de la ciudad de El Alto por tipología de delito de cada gestión.

### **1.7.2. Método de ingeniería**

#### **1.7.2.1. Metodología CRISP-DM**

La metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) es un modelo estándar para proyectos de minería de datos y Machine Learning, abarca todas las etapas esenciales desde la comprensión del problema hasta la implementación de sistemas predictivos y automatizados y se adapta bien a la naturaleza iterativa del desarrollo de modelos de Machine Learning.

Es un modelo de proceso que sirve de base para un proceso de ciencia de datos, fue publicada en 1999 y desde entonces sea convertido en la metodología más utilizada para proyectos de minería de datos, análisis y ciencia de datos (Hotz, 2024).

##### **1.7.2.1.1. Fases de la metodología CRISP-DM**

###### **a) Comprensión del problema**

Esta fase se centra en la comprensión de los objetivos del negocio y requisitos del proyecto. Se definen los objetivos del proyecto y se establece un plan inicial. Consta de las siguientes actividades a realizar:

- Determinar los objetivos de negocio
- Evaluar la situación
- Determinar los objetivos de minería de datos
- Producir un plan de proyecto

#### **b) Comprensión de los datos**

La fase de comprensión de datos impulsa el enfoque para identificar, recopilar y analizar los conjuntos de datos necesarios que pueden ayudar a cumplir los objetivos del proyecto. Esta fase también tiene cuatro tareas:

- Recopilar los datos iniciales
- Describir los datos
- Explorar los datos
- Verificar la calidad de los datos

#### **c) Preparación de los datos**

Esta fase que suele denominarse depuración de datos, ya que limpian, transforman y preparan los conjuntos de datos finales para el modelado. Tiene cinco tareas:

- Seleccionar datos
- Datos limpios
- Construir datos
- Integrar datos
- Dar formato a los datos

**d) Modelado**

En esta etapa es donde se seleccionan y se aplican técnicas de modelado para construir y entrenar modelos utilizando los datos preparados. Se pueden probar múltiples algoritmos y enfoques para encontrar el que mejor se ajuste a los objetivos del proyecto. Esta fase tiene cuatro tareas:

- Seleccionar técnicas de modelado
- Generar diseño de prueba
- Modelo de compilación
- Evaluar el modelo

**e) Evaluación**

Se evalúan los modelos creados para determinar su eficacia y su capacidad para abordar los objetivos del negocio. Se utilizan técnicas de validación para comparar y seleccionar el mejor modelo. Esta fase tiene tres tareas:

- Evaluar los resultados
- Proceso de revisión
- Determinar la continuación del proyecto

**f) Despliegue**

Finalmente se implementa el modelo en el entorno operativo del negocio. Se crea un plan para monitorear y mantener el modelo en producción, asegurando que siga siendo relevante y efectivo con el tiempo. Esta fase tiene cuatro tareas:

- Planear la implementación
- Supervisión y mantenimiento del plan

- Elaborar el informe final
- Proyecto de revisión

## **1.8. HERRAMIENTAS**

### **1.8.1. Sistema operativo**

- **Windows**

Es un sistema operativo desarrollado por la empresa de Microsoft para teléfonos inteligentes, tabletas y computadoras personales. Que proporciona una interfaz gráfica de usuario y gestiona los recursos de hardware de una computadora. Este sistema operativo presenta numerosas mejoras en cuanto a rendimiento, seguridad y funcionalidad en general, lo que lo convierte en una opción muy interesante para cualquier persona que desee utilizar un sistema operativo moderno y actualizado.

### **1.8.2. Lenguaje de programación**

- **Python**

Es un lenguaje de programación versátil y de propósito general ampliamente utilizado en la creación de sitios web, software, automatización de tareas y análisis de datos. Fue creado por Guido van Rossum y lanzado por primera vez en 1991. Python se ha convertido a un lenguaje muy popular debido a su sintaxis clara y legible, así como su amplia gama de bibliotecas y marcos de trabajo que facilitan el desarrollo de una variedad de aplicaciones, desde aplicaciones web hasta inteligencia artificial.

### **1.8.3. Entornos de desarrollo integrado (IDE)**

- **Google Colab**

Es una plataforma en la nube gratuita de Google que ofrece un entorno colaborativo para escribir y ejecutar código Python a través de notebooks de Jupyter. Con acceso a

recursos de hardware como GPU y TPU, integración con Google Drive, y bibliotecas preinstaladas. Es ideal para proyectos de aprendizaje automático, análisis de datos y programación en general. Su facilidad de uso y capacidad para trabajar simultáneamente en tiempo real lo convierten en una herramienta poderosa para la colaboración y el desarrollo de proyectos en equipo.

- **Visual Studio Code**

Es un editor de código fuente desarrollado por Microsoft, que ofrece una amplia gama de características y extensiones para el desarrollo de software. Con soporte para varios lenguajes de programación, incluyendo a Python, JavaScript, Java y C++. Se destaca por su interfaz intuitiva, su rendimiento rápido y su extensibilidad. Su amplia comunidad de usuarios y su constante desarrollo lo convierten en una herramienta popular para desarrolladores de todo tipo.

#### ***1.8.4. Librerías para ciencia de datos***

- **Numpy**

NumPy (Numerical Python) es una biblioteca para computación numérica en Python. Proporciona soporte para arreglos (arrays), matrices y un gran número de funciones matemáticas para operar con estas estructuras de datos. Las operaciones son eficientes con arreglos multidimensionales.

- **Pandas**

Pandas es una biblioteca poderosa para la manipulación y análisis de datos. Proporciona estructuras de datos como Series y DataFrame, que facilitan el trabajo con datos estructurados. Tiene herramientas para leer y escribir datos entre estructuras de datos en

memoria y diferentes formatos de archivo y la funcionalidad de agrupación (group by) para realizar operaciones, dividir, aplicar y combinar en conjuntos de datos.

- **Matplotlib**

Matplotlib es una biblioteca de gráficos para crear visualizaciones estáticas, animadas e interactivas en Python. Es altamente personalizable e integra bien con NumPy y Pandas.

## **1.9. LIMITES Y ALCANCES**

### **1.9.1. Limites**

- La investigación se centrará exclusivamente en la ciudad de El Alto y no se incluirán datos de otras ciudades o regiones.
- El estudio analizará datos históricos delictivos de un período específico, en este caso de los últimos 5 a 10 años.
- Se utilizarán datos proporcionados por fuentes oficiales, como la policía local y bases de datos públicas, excluyendo datos de fuentes no verificadas o informales.
- El modelo se desarrollará como un prototipo y no se implementará en un entorno real sin validación y pruebas adicionales.

### **1.9.2. Alcances**

- Permitirá predecir con cierto grado de precisión, el crecimiento futuro de hechos delictivos en la ciudad de El Alto mediante los datos proporcionados.
- Identificar patrones en los datos históricos de delitos, lo que ayudaría a comprender mejor los factores que contribuyen al aumento de la delincuencia en la ciudad.
- Identificar el algoritmo de Machine Learning más apropiado para el problema

específico de predicción que se está abordando.

- Al validar el modelo con nuevos conjuntos de datos, se pueden identificar oportunidades para mejorar su rendimiento.

#### **1.10. APORTES**

- Permite anticipar posibles aumentos en el índice delictivo, lo que brinda la oportunidad de implementar estrategias preventivas y políticas de seguridad de manera proactiva, esto puede ayudar a reducir la incidencia delictiva y mejorar la seguridad ciudadana.
- Al predecir áreas con mayor probabilidad de aumento del índice delictivo, se pueden asignar recursos de manera más eficiente y efectiva para la aplicación de medidas preventivas, patrullajes policiales, entre otros





# CAPÍTULO II

## CAPÍTULO II

### 1.11. MARCO TEÓRICO

#### 1.11.1. *Introducción*

La investigación que se presenta en este capítulo aborda conceptos fundamentales relacionados con la delincuencia en la ciudad de El Alto a lo largo de los años, así como los tipos de delitos más frecuentes en la zona. Además, se explorará la inseguridad ciudadana y se introducirán los principios básicos del aprendizaje automático (Machine Learning) junto con sus diversas áreas de aplicación. El método de desarrollo que se utilizará en este estudio, destacando su relevancia para comprender y abordar las tendencias delictivas futuras en la ciudad. Se tiene las bases para una comprensión más profunda de la investigación que se llevará a cabo, destacando la importancia de abordar la seguridad ciudadana desde una perspectiva informada y basada en datos.

### 1.12. DELINCUENCIA

La delincuencia puede definirse como el conjunto de comportamientos que infringen las normas y leyes establecidas en una sociedad determinada. Estos comportamientos pueden incluir desde delitos menores como hurtos o robos, hasta delitos más graves como homicidios o violaciones.

Según Martínez (2023), la delincuencia es un fenómeno social que afecta a las comunidades de todo el mundo. En este artículo exploraremos qué es la delincuencia y cuáles son las causas que la impulsan. Analizaremos desde factores individuales hasta aquellos relacionados con el entorno social, con el objetivo de comprender mejor este problema que impacta en la sociedad.

### 1.12.1. La delincuencia en Bolivia

En Bolivia, las ciudades de Santa Cruz, Cochabamba, La Paz y El Alto enfrentan desafíos significativos en términos de seguridad, siendo consideradas entre las más violentas e inseguras del país. Estas urbes concentran aproximadamente el 75% de los delitos cometidos a nivel nacional. La razón detrás de esta concentración delictiva se atribuye, en parte, a su alta densidad demográfica, ya que La Paz, Santa Cruz y Cochabamba albergan cerca del 70% de la población boliviana. Además, la creciente urbanización ha contribuido a esta problemática (Rubin de Celis, Sanjines, & Aliaga, 2012).

**Figura 1**

*Distribución de delitos en departamentos y ciudades*

Las 10 posiciones más altas <sup>1)</sup>					
Departamento	Municipio	Cantidad de denuncias <sup>2)</sup>	Tasa por cada 100.000 hab.	Distribución porcentual	Tendencia mensual
Santa Cruz	Santa Cruz de la Sierra	16.471	865,35	21,15%	
La Paz	La Paz	7.224	755,07	9,27%	
Cochabamba	Cochabamba	6.048	706,38	7,76%	
La Paz	El Alto	6.039	544,52	7,75%	
Chuquisaca	Sucre	4.604	1.276,96	5,91%	
Tarija	Tarija	3.276	1.201,36	4,21%	
Potosí	Potosí	2.890	1.073,63	3,71%	
Santa Cruz	Warnes	2.037	1.412,69	2,62%	
Santa Cruz	Montero	1.877	1.276,19	2,41%	
Óruro	Óruro	1.804	512,79	2,32%	

*Nota: Distribución de delitos según el Atlas de Bolivia en Seguridad Ciudadana (p. 31), por OBSCD, 2023. Fuente: Observatorio Boliviano de Seguridad Ciudadana y Lucha Contra las Drogas.*

Ante la alarmante pérdida de vidas debido a actos delictivos en los últimos años, el Gobierno se ha visto comprometido a tomar medidas para reforzar la seguridad en estas áreas. Una de estas iniciativas incluye la creación de una nueva Escuela Básica Policial, destinada a entrenar a futuros efectivos policiales. Este enfoque busca abordar los desafíos de seguridad mediante la formación y preparación adecuadas del personal encargado de mantener el orden público en estas ciudades (Rubin de Celis, Sanjines, & Aliaga, 2012).

Una de las organizaciones encargadas de investigar y atender los delitos en Bolivia es la Fuerza Especial de Lucha Contra el Crimen (FELCC). Algunos de los delitos que atiende la FELCC incluyen:

- Robos
- Estafas
- Trata de personas
- Homicidios
- Otros delitos

### **1.13. INSEGURIDAD CIUDADANA**

La problemática de la inseguridad ciudadana en Bolivia persiste como un tema recurrente en el debate público, constituyendo una suerte de bomba de tiempo. A pesar de las múltiples iniciativas implementadas a nivel gubernamental, como las propuestas de reforma del sistema de seguridad ciudadana, las cumbres dedicadas a este tema, y los programas destinados al fortalecimiento de la policía nacional, los datos relativos a la violencia y el delito en el país indican que estas medidas no han logrado reducir de manera significativa la inseguridad ciudadana. En lugar de disminuir, pareciera que estas iniciativas han contribuido a intensificar el problema, aumentando la presión en la bomba.

Diversos factores han contribuido al aumento de los niveles de inseguridad, tanto subjetiva como objetiva. Entre ellos, destaca la falta de coherencia en las posturas adoptadas por las entidades estatales y la sociedad en su conjunto. La discrepancia en enfoques y acciones ha generado un clima de incertidumbre, debilitando los esfuerzos por abordar eficazmente la seguridad ciudadana. Es imperativo abordar esta incoherencia para lograr avances reales y sostenibles en la gestión de la seguridad en el país (Torrico, 2012).

Según el Observatorio Boliviano de Seguridad Ciudadana y Lucha Contra las Drogas (2023, p.33) los días de la semana que más se registran denuncias son los fines de semana entre horas nocturnas.

**Figura 2**

*Delitos registrados por día y hora*

Características del hecho	Cantidad de denuncias <sup>(S)</sup>	Distribución porcentual	Tendencia mensual
<b>Día del Hecho</b>			
Lunes	11.405	14,64%	
Martes	10.160	13,04%	
Miércoles	10.266	13,18%	
Jueves	10.146	13,03%	
Viernes	10.320	13,25%	
Sábado	11.408	14,65%	
Domingo	13.766	17,67%	
Indeterminado	423	0,54%	
<b>Hora del hecho</b>			
00:00 - 05:59	13.712	17,60%	
06:00 - 11:59	17.675	22,69%	
12:00 - 17:59	18.057	23,18%	
18:00 - 23:59	21.560	27,68%	
Indeterminado	6.890	8,85%	

*Nota: Delitos registrados por día y hora según el Atlas de Bolivia en Seguridad Ciudadana (p. 32), por OBSCD, 2023. Fuente: Observatorio Boliviano de Seguridad Ciudadana y Lucha Contra las Drogas*

### 1.13.1. La inseguridad en la ciudad de El Alto

La ciudad de El Alto se presenta como una urbe joven debido a sus singulares características históricas, demográficas, culturales, económicas, sociales y políticas. En realidad, es la ciudad más joven, la más empobrecida, la más diversa, la más postergada, la más peligrosa y la más violenta. Desde la perspectiva estadística, El Alto es catalogada como una de las ciudades más violentas y peligrosas a nivel nacional. En términos de violencia social urbana, ocupa el segundo lugar después de Santa Cruz, y lidera en cuanto a la violencia intrafamiliar.

En la ciudad de El Alto se estima aproximadamente 1.109.048 habitantes en el año 2022, con 539.815 hombres y 569.233 mujeres.

**Figura 3**

*Datos generales de la gestión 2022*

<b>DATOS GENERALES - GESTIÓN 2022</b>	
<b>Población<sup>(1)</sup></b>	<b>1.109.048</b>
Sexo	
Hombre	539.815
Mujer	569.233
Edad	
0-19	454.209
20-29	208.880
30-39	174.745
40-49	116.813
50-59	75.614
60-69	43.249
Más de 70	35.538
<b>Superficie Km<sup>2</sup> (2)</b>	<b>345,24</b>

*Nota: Datos generales de la ciudad de El Alto en la gestión 2022, según el Atlas de Bolivia en Seguridad Ciudadana (p. 21), por OBSCD, 2023. Fuente: Observatorio Boliviano de Seguridad Ciudadana y Lucha Contra las Drogas.*

Según el Informe sobre Desarrollo Humano en Bolivia del 2006, vinculado al tema de Policía Nacional y Seguridad Ciudadana, aproximadamente ocho de cada diez habitantes de El Alto se sienten inseguros al transitar por espacios públicos. Un diagnóstico de la Alcaldía alteña en ese mismo año revela que cuatro de cada diez alteños, o sus familiares, han sido víctimas de algún tipo de delito. En el año 2011, la FELCC (Fuerza Especial de Lucha Contra el Crimen) informa que, en promedio, se producen entre 3 y 5 asaltos, 2 y 3 atracos, 2 y 3 casos de víctimas acogotadas por día. Además, se reporta diariamente entre una y cuatro personas muertas por causas violentas. Estas estadísticas pintan un panorama preocupante en relación con la inseguridad ciudadana, tanto en términos de aumento de la violencia, homicidios y delincuencia, como en el incremento de las sensaciones de miedo, temor e inseguridad (Moreno, 2011)

Para la gestión 2022 de todos los delitos reportados, son 2969 delitos registrados y para calcular la tasa de delincuencia anual (si la población de El Alto es de 1.109.048 habitantes).

La ecuación para calcular la tasa de delincuencia es la siguiente:

Ecuación tasa anual de delincuencia

$$T = \frac{D}{P} * 100,000 \quad (1)$$

$$T = \frac{2969}{1109048} * 100,000$$

$$T = 267,707078 \text{ tasa anual de denuncias}$$

Donde:

- T es la tasa de delincuencia por cada 100,000 habitantes.
- D es el número de delitos reportados.
- P es la población total.

La expresión calcula la proporción de delitos por habitante, y multiplicar por 100,000 ajusta esta proporción para que se exprese en términos de delitos por 100,000 habitantes.

Para calcular las denuncias por día en el año 2022 se tiene la siguiente ecuación:

Ecuación tasa de delito por día

$$d = \frac{n}{\text{dias}} \quad (2)$$

$$d = \frac{2969}{365}$$

$$d = 8,13424 \text{ delitos por dia}$$

Donde:

- d delitos por día
- n es el número de delitos reportados.

Según el Observatorio Boliviano de Seguridad Ciudadana y Lucha Contra las Drogas (2023, p.23), actualmente, se cuenta con más de 290 tipos de delitos registrados en la Policía



Boliviana, con la finalidad de ordenar, sistematizar y armonizar esa información, se identificaron un grupo de delitos priorizados de acuerdo al impacto en la percepción de seguridad de la población, su recurrencia, gravedad y vulneración de los derechos fundamentales, denominado “Delitos de alta connotación social”, el cual en su estructura se divide en nueve grupos e incluye a treinta y tres delitos.

El robo de vehículos, asaltos a mano armada y robos a domicilios son comunes en varias áreas de El Alto. Los transportes públicos también son escenarios frecuentes de robos y asaltos. El tráfico y consumo de drogas también ha aumentado en la ciudad, como un punto de tránsito y distribución dentro del país.

**Figura 4**

*Tipos de delitos registrados en la Policía boliviana*



*Nota: Tipos de delitos según el Atlas de Bolivia en Seguridad Ciudadana (p. 21), por OBSCD, 2023.  
Fuente: Observatorio Boliviano de Seguridad Ciudadana y Lucha Contra las Drogas*

### 1.13.1.1. Planes para combatir la inseguridad

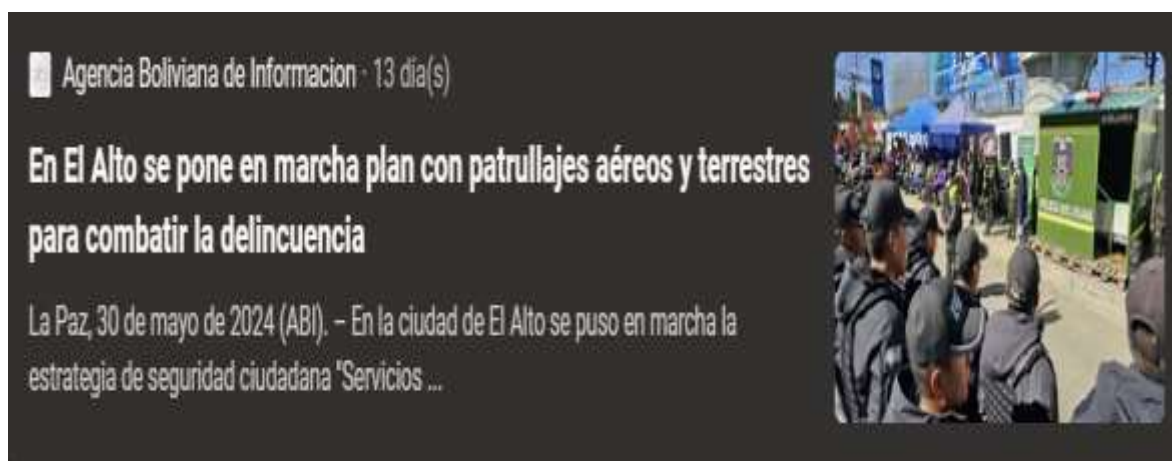
El municipio de El Alto del 2022 se implementó ocho planes para abordar la inseguridad:

- Mi Barrio Seguro
- Recuperación de Espacios Públicos
- Control Migratorio
- Control Territorial
- Viaje Seguro
- Prevención de la Violencia Dino
- Escuela Segura
- Planes de Aplicación Limitada.

Para el año 2024 se implementará un nuevo Plan de Seguridad multidimensional y se inauguró el Sistema de Vigilancia BOL-110 Fase II.

#### Figura 5

*Plan de patrullaje para combatir la delincuencia*



*Nota: Plan de patrullaje para combatir la delincuencia, por OBSCD, 2024. Fuente: ABI - ABI*

La ciudad de El Alto ha iniciado la estrategia de seguridad ciudadana "Servicios Policiales 3.0", que incorpora casetas policiales móviles, torres móviles y drones para el patrullaje. El ministro de Gobierno, Eduardo del Castillo, lanzó esta iniciativa en la Feria 16 de Julio, una de las más grandes de Bolivia. Esta estrategia consiste en que las casetas policiales móviles estarán desplegadas en los 14 distritos del municipio, tanto en áreas urbanas como rurales, y estarán equipadas con destelladores, sirenas y altavoces con mensajes preventivos.

Para mejorar el patrullaje preventivo, se utilizarán drones y un centro de monitoreo móvil en un camión multipropósito. Del Castillo destacó que se intensificará el patrullaje con más drones que permitirán filmar y seguir a los delincuentes con la ayuda de grupos de inteligencia. Además, hizo un llamado a la población de El Alto a colaborar en la lucha contra la delincuencia.

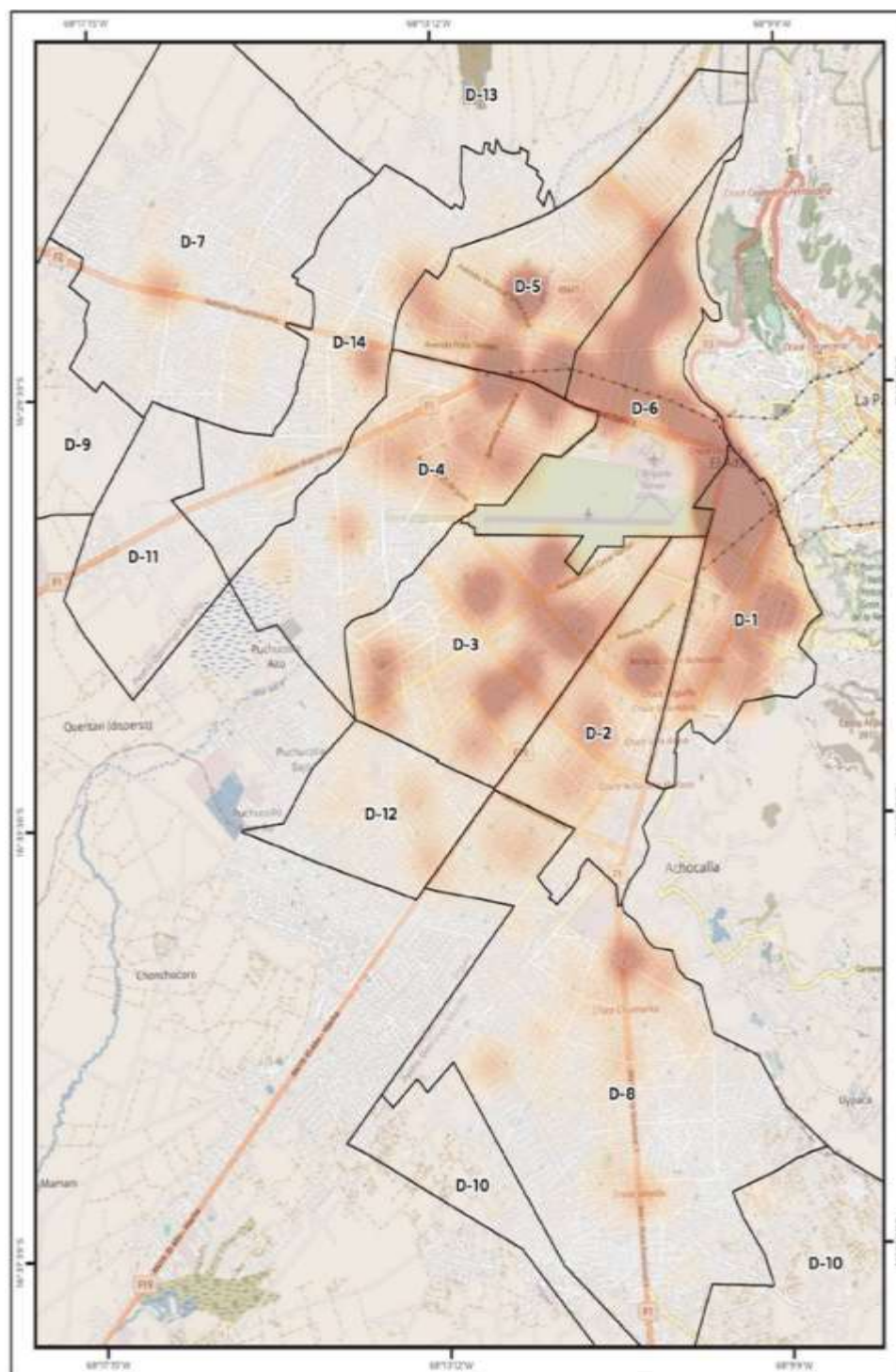
Durante el lanzamiento de "Servicios Policiales 3.0", se llevó a cabo un operativo en la feria que resultó en el secuestro de 101 celulares presuntamente robados y la detención de 15 personas que los comercializaban. El viceministro de Seguridad Ciudadana, Roberto Ríos, señaló que la policía está intensificando el patrullaje preventivo y atacando a bandas dedicadas a la receptación y sustracción de objetos robados, especialmente durante la actividad económica de la Feria 16 de Julio que se realiza los jueves y domingos (ABI, 2024)

#### **1.13.1.2. Zonas de mayor riesgo**

La Ceja, la 12 de Octubre y Villa Dolores son consideradas zonas rojas en términos de inseguridad.

**Figura 6**

*Puntos calientes de denuncias de la ciudad de El Alto*



*Nota: Puntos calientes de denuncias según el Atlas de Bolivia en Seguridad Ciudadana (p. 43), por OBSCD, 2023. Fuente: Observatorio Boliviano de Seguridad Ciudadana y Lucha Contra las Drogas*

Figura 7

Datos de seguridad ciudadana

DATOS DE SEGURIDAD CIUDADANA	
<b>Estructura policial</b>	
<b>Efectivos policiales<sup>(6)</sup></b>	<b>1.909</b>
<b>Infraestructura policial<sup>(6)</sup></b>	<b>99</b>
EPI	8
Módulo policial	87
Otros <sup>(8)</sup>	4
<b>Cámaras de vigilancia<sup>(6)(7)</sup></b>	<b>1.272</b>
<b>Denuncias de delitos de alta connotación social<sup>(4)</sup></b>	<b>6.039</b>
<b>Indicadores</b>	
Tasa de denuncias por cada 100.000 habitantes <sup>(4)</sup>	545
Denuncias por día	17
Tasa de denuncias de delitos por km <sup>2</sup>	17,49
<b>Tasa de denuncias de delitos por superficie distrital urbana Km<sup>2</sup></b>	<b>15,70</b>
D1	114,03
D2	36,59
D3	55,14
D4	34,88
D5	45,07
D6	54,99
D7	4,75
D8	13,55
D9	0,15
D10	0,39
D11	0,20
D12	15,49
D13	0,01
D14	11,05
Tasa de servidores públicos policiales por km <sup>2</sup>	5,53
Tasa de servidores públicos policiales por cada 1.000 habitantes <sup>(4)</sup>	1,72
Tasa en km <sup>2</sup> por infraestructura policial	3,49
Tasa de habitantes por infraestructura policial	11.203

Nota: Datos de seguridad ciudadana según el Atlas de Bolivia en Seguridad Ciudadana (p. 42), por OBSCD, 2023. Fuente: Observatorio Boliviano de Seguridad Ciudadana y Lucha Contra las Drogas

Figura 8

Denuncias registradas por día

Características del hecho	Cantidad de denuncias <sup>(6)</sup>	Distribución porcentual	Tendencia mensual
<b>Día del Hecho</b>			
Lunes	861	14,26%	
Martes	748	12,39%	
Miércoles	728	12,05%	
Jueves	753	12,47%	
Viernes	804	13,31%	
Sábado	972	16,10%	
Domingo	1.167	19,32%	
Indeterminado	6	0,10%	
<b>Hora del hecho</b>			
00:00 - 05:59	992	16,43%	
06:00 - 11:59	1.305	21,61%	
12:00 - 17:59	1.426	23,61%	
18:00 - 23:59	1.753	29,03%	
Indeterminado	563	9,32%	

*Nota: Denuncias registradas por día según el Atlas de Bolivia en Seguridad Ciudadana (p. 42), por OBSCD, 2023. Fuente: Observatorio Boliviano de Seguridad Ciudadana y Lucha Contra las Drogas*

### 1.13.1.3. Causas de la Delincuencia

La extrema pobreza y el desempleo son factores que contribuyen a la inseguridad en El Alto. Además, el comercio informal (que representa el 80% de la economía) también afecta las calles de la ciudad.

Figura 9

Datos socioeconómicos

<b>DATOS SOCIOECONÓMICOS</b>	
<b>Económicos</b>	
Unidades económicas de venta de bebidas y tabaco <sup>(ii)(3)</sup>	597
<b>Sociales</b>	
<b>Salud</b>	
<b>Establecimientos de salud<sup>(4)</sup></b>	<b>110</b>
Primer nivel	97
Segundo nivel	7
Tercer nivel	3
Otros <sup>(iii)</sup>	3
<b>Habitantes por establecimientos de salud</b>	<b>10.082</b>
Habitantes por establecimientos de primer nivel de atención	11.433
Habitantes por establecimientos de segundo nivel de atención	158.435
Habitantes por establecimientos de tercer nivel	369.683
<b>Educación</b>	
Edificios educativos del subsistema regular <sup>(5)</sup>	534

*Nota: Datos socioeconómicos según el Atlas de Bolivia en Seguridad Ciudadana (p. 42), por OBSCD, 2023. Fuente: Observatorio Boliviano de Seguridad Ciudadana y Lucha Contra las Drogas*

#### 1.14. MACHINE LEARNING

Según Brow (2021), el aprendizaje automático es un subcampo de la inteligencia artificial, que se define en términos generales como la capacidad de una máquina para imitar el comportamiento humano inteligente. Los sistemas de inteligencia artificial se utilizan para realizar tareas complejas de una manera similar a como los humanos resuelven problemas. El aprendizaje automático es una forma de utilizar la IA. Fue definido en la década de 1950 por el pionero de la IA Arthur Samuel como "el campo de estudio que le da a las computadoras la capacidad de aprender sin ser programadas explícitamente".

El proceso de aprendizaje automático tiene su punto de partida en los datos ya sean números, imágenes, texto, como transacciones bancarias, fotografías de personas o incluso detalles sobre artículos de panadería, registros de reparaciones, datos de series temporales de sensores o informes de ventas. La recopilación y preparación de estos datos es esencial, ya que servirán como conjunto de entrenamiento o la información con la cual se nutrirá y formará el modelo de aprendizaje automático. En términos generales, cuantos más datos se disponga, mayor será la capacidad del programa para aprender de manera efectiva y generar resultados precisos.

#### **1.14.1. Aprendizaje supervisado**

Según Carranza (2022), el aprendizaje supervisado se caracteriza porque utiliza algoritmos que tienen un aprendizaje previo basado en etiquetas que el ser humano configura con anterioridad, de manera que la máquina pueda realizar una serie de asociaciones para hacer predicciones y tomar decisiones.

El aprendizaje supervisado es uno de los enfoques más prevalentes en el campo del aprendizaje automático. Ejemplos concretos de su aplicación incluyen la detección de correos no deseados, la funcionalidad de reconocimiento de voz o escritura en dispositivos móviles, así como el reconocimiento de firmas. Este método implica alimentar al modelo con conjuntos de datos previamente etiquetados, permitiéndole aprender patrones y asociaciones específicas para realizar predicciones o clasificaciones precisas. En esencia, el aprendizaje supervisado es clave en situaciones donde se dispone de datos de entrada y salida para guiar y mejorar el proceso de entrenamiento del modelo (Carranza, 2022)

##### **1.14.1.1.Regresión**

Los algoritmos del aprendizaje supervisado realizan predicciones basándose en entradas de información pasada en formato numérico, generalmente se usa variables no categóricas.



**Tabla 2***Conjuntos de datos para regresiones*

V1	V2	V3	V4
1	4	7	10
2	5	8	11
3	6	9	12

**1.14.1.1.1. Regresión lineal**

La regresión lineal es un algoritmo que forma parte del aprendizaje supervisado que se utiliza en Machine Learning y en estadística para modelar la relación entre una o más variables independientes y una variable dependiente con datos etiquetados. Siempre tiene que tener valores numéricos para que el modelo aprenda a predecir. La regresión lineal busca encontrar la línea mejor ajustada a los valores de los datos que minimiza la distancia entre los valores observados y los valores predichos por el modelo.

Algunos tipos de regresión lineal son:

**a) Regresión lineal simple**

Esta regresión lineal se utiliza principalmente para tareas de predicción en donde se intenta predecir un valor numérico (variable dependiente) a partir de una única variable de entrada (variable independiente).

### Ecuación de regresión lineal

$$y = \beta_0 + \beta_1 x + \varepsilon$$

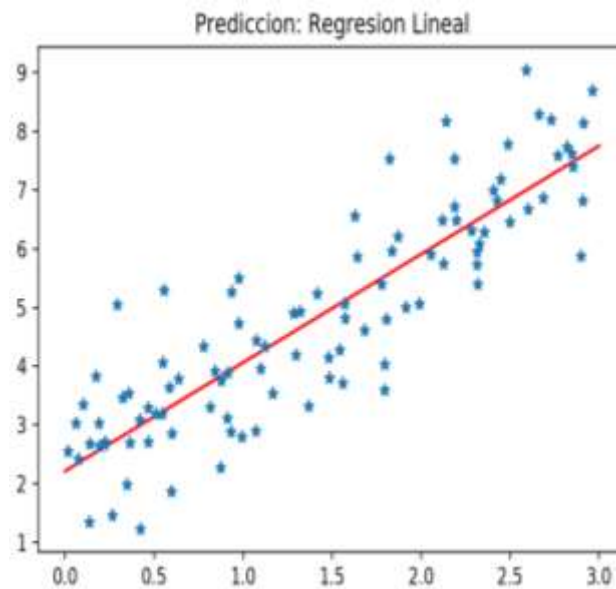
(3)

Donde:

- $y$  es la variable dependiente que queremos predecir.
- $x$  es la variable independiente que utilizamos para hacer la predicción.
- $\beta_0$  es el término de sesgo (bias), que representa el valor esperado de  $y$  cuando  $x$  es igual a cero.
- $\beta_1$  es el coeficiente de la variable independiente, que representa el cambio esperado en  $y$  por cada unidad de cambio en  $x$ .
- $\varepsilon$  es el término de error, que representa la diferencia entre el valor observado de  $y$  y el valor predicho por el modelo.

### Figura 10

Grafica de regresión lineal



Nota: Grafica de regresión lineal por Gabriel Cabrera, 2019. Fuente: Gabriel Cabrera, Gradiente Descendiente ([technotes.netlify.app](https://technotes.netlify.app))

## b) Regresión lineal múltiple

Se utiliza para modelar relaciones complejas entre múltiples variables independientes y una variable dependiente para hacer el ajuste de la recta.

*Ecuación de regresión múltiple*

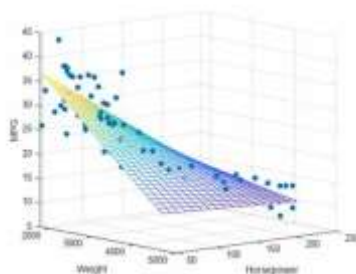
$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p + \varepsilon \quad (4)$$

Donde:

- $y$  es la variable dependiente que queremos predecir.
- $x_1, x_2, x_p$  son las variables independientes.
- $\beta_0$  es el término de sesgo (bias), que representa el valor esperado de  $y$  cuando todas las variables independientes son iguales a cero.
- $\beta_1, \beta_2, \beta_p$  son los coeficientes de las variables independientes, que representan el cambio esperado en  $y$  por cada unidad de cambio en cada una de las variables independientes, manteniendo las otras variables constantes.
- $\varepsilon$  es el término de error, que representa la diferencia entre el valor observado de  $y$  y el valor predicho por el modelo.

**Figura 11**

*Grafica regresión lineal múltiple*



*Nota: Grafica de regresión lineal múltiple por métodos, 2023. Fuente: Regresión Lineal y Múltiple: Modelando Relaciones y Haciendo Predicciones (metodos2023a.blogspot.com)*

### c) Regresión polinomial

Es una extensión de la regresión lineal simple que modela la relación entre una variable independiente y una variable dependiente utilizando una función polinomial en lugar de una línea recta. Esto permite capturar relaciones no lineales entre las variables, lo que puede ser útil cuando los datos muestran un patrón curvilíneo.

*Ecuación de regresión polinómica con una variable Independiente*

$$y = \beta_0 + \beta_1x^1 + \beta_2x^2 + \dots + \beta_px^p + \varepsilon \quad (5)$$

*Ecuación de regresión polinómica con múltiples variables independientes*

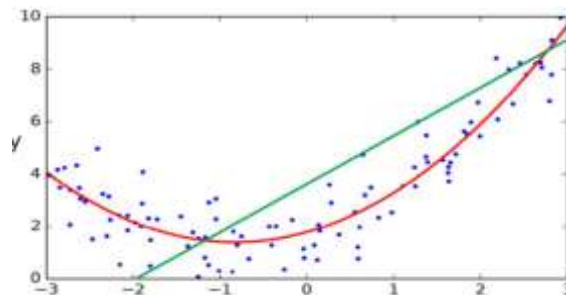
$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_1^2 + \beta_4x_1x_2 + \dots + \beta_nx_1^n + \beta_{n+1}x_1^{n-1}x_2 \quad (6)$$

Donde:

- $y$  es la variable dependiente que queremos predecir.
- $x$  es la variable independiente.
- $\beta_0, \beta_1, \beta_2, \beta_p$  son los coeficientes del modelo que representan la relación entre  $x$  e  $y$ .

#### Figura 12

Gráfico regresión polinómica



Nota: Grafica de regresión polinómica por sensoricx, 2023. Fuente: regresión polinomial por mínimos cuadrados (sensoricx.com)

### d) Regresión Ridge

Ridge penaliza la suma de los coeficientes elevados al cuadrado, utiliza una penalización de norma L2, tiene el efecto de reducir de forma proporcional el valor de todos los coeficientes del modelo, pero sin que estos lleguen a cero. El grado de penalización está controlado por el hiperparámetro  $\lambda$ .

#### **Ecuación de regresión polinómica Ridge**

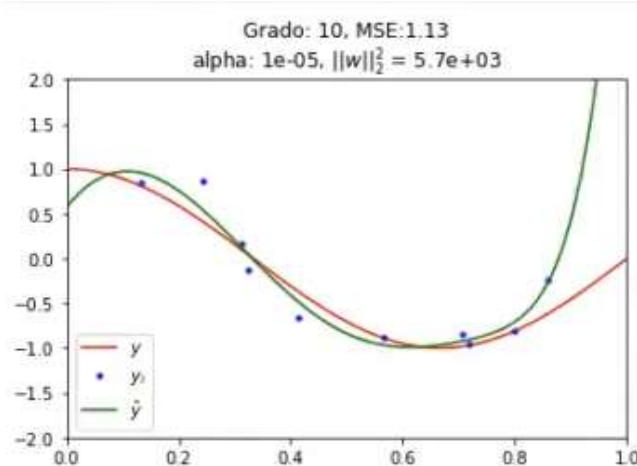
$$\min_{\beta} \sum_{i=1}^m (y_i - \hat{y}_i)^2 + \alpha \sum_{j=1}^k \beta_j^2 \quad (7)$$

Donde:

- $k$  es el número total de términos polinómicos en el modelo.
- $\alpha$  es el parámetro de regularización que controla la fuerza de la penalización.

**Figura 13**

*Grafica regresión polinómica Ridge*



*Nota: Grafica de regresión polinómica Ridge por keepcoding, 2024. Fuente: Regularización sobre regresión Ridge (keepcoding.io)*

### e) Regresión Lasso

Es un modelo lineal que penaliza el vector de coeficientes añadiendo su norma L1 (basada en la distancia Manhattan) a la función de coste.

#### ***Ecuación de regresión polinómica Lasso***

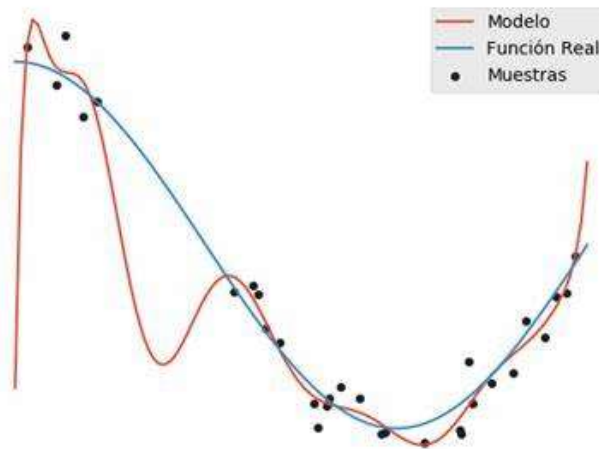
$$\min_{\beta} \sum_{i=1}^m (y_i - \hat{y}_i)^2 + \alpha \sum_{j=1}^k |\beta_j| \quad (8)$$

Donde:

- k es el número total de términos polinómicos en el modelo.
- $\alpha$  es el parámetro de regularización que controla la fuerza de la penalización.

**Figura 14**

*Grafica de regresión polinómica Lasso*



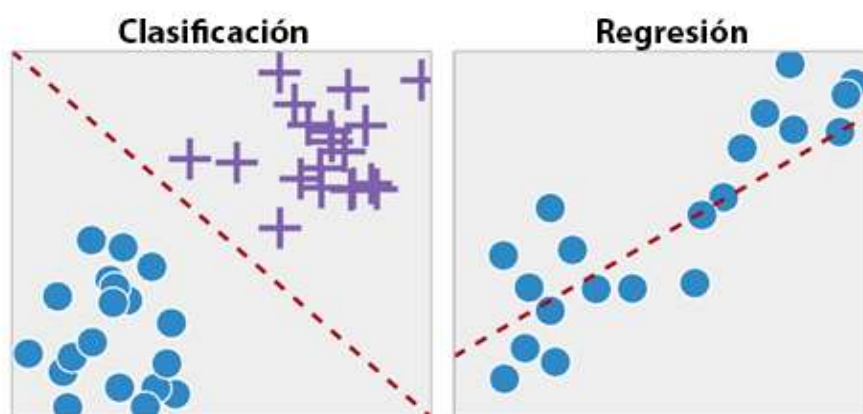
*Nota: Grafica de regresión polinómica Lasso por iadelta, 2021. Fuente: Regresión #2. Regresión Polinomial y Regularización (iadelta.com)*

### 1.14.1.2. Clasificación

En el contexto del aprendizaje supervisado, la clasificación se produce cuando los algoritmos realizan una categorización de los datos, eligiendo entre dos o más categorías basándose en el análisis de información previamente adquirida.

**Figura 15**

*Grafica de clasificación de datos*



*Nota: Grafica de clasificación de datos por master data science, 2020. Fuente: Aprendizaje Automático Supervisado: Máster en Data Science ONLINE (masterdatascience.online)*

**Tabla 3**

*Conjunto de datos para clasificación*

Colores	Formas
Azul	Circulo
Rojo	Cuadrado
Amarillo	Triangulo

Algunos de los algoritmos de clasificación más comunes incluyen:

#### 1.14.1.2.1. **Regresión logística**

Es un algoritmo para problemas de clasificación binaria a partir de un conjunto de datos de entrada la salida será discreta.

La regresión logística modela la relación entre una variable dependiente binaria y una o más variables independientes utilizando la función sigmoide (logística). La ecuación general para la regresión logística es:

*Ecuación de regresión logística*

$$P(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)}} \quad (9)$$

Donde:

- $P(y=1|x)$  es la probabilidad de que la variable dependiente y sea 1 dado el vector de características  $x$ .
- $\beta_0$  es el término de sesgo (intercepto).
- $\beta_1, \beta_2, \beta_n$  son los coeficientes de regresión (pesos) para las variables independientes  $x_1, x_2, x_n$ .
- $e$  es la base del logaritmo natural (aproximadamente 2.71828).



Figura 16

Grafica de regresión logística



Nota: Grafica de regresión logística por probabilidad y estadística, 2023. Fuente: Regresión logística (probabilidadyestadística.net)

#### 1.14.1.2.2. K Vecinos más Cercanos KNN

Es un algoritmo simple que es usado para problemas de clasificación como de regresión clasifica una instancia, trata de buscar los k puntos más cercanos a un punto concreto para poder inferir su valor.

El algoritmo KNN para regresión funciona de la siguiente manera:

1. **Elección del número de vecinos (k):** Selecciona el número de vecinos más cercanos que se utilizarán para la predicción.
2. **Cálculo de la distancia:** Calcula la distancia entre la muestra de prueba y todas las muestras de entrenamiento utilizando una métrica de distancia (por ejemplo, la distancia euclidiana).
3. **Identificación de vecinos:** Encuentra los k vecinos más cercanos a la muestra de prueba.

**4. Predicción:** Predice el valor de la variable dependiente como la media (o el promedio ponderado) de los valores de los k vecinos más cercanos.

La predicción y para una nueva muestra x en KNN regresión se calcula como:

*Ecuación KNN*

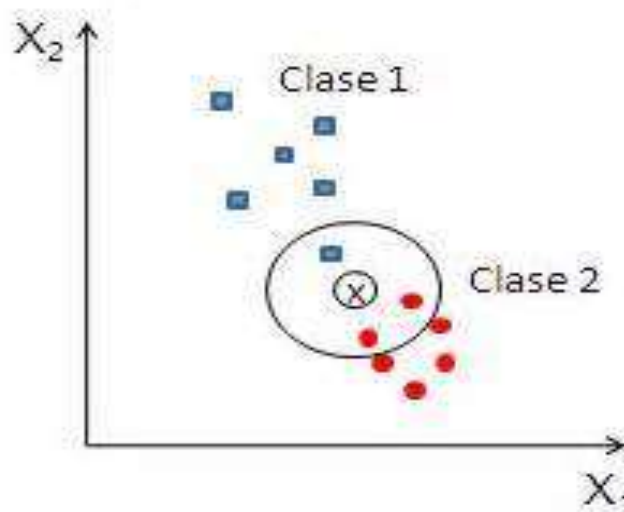
$$\hat{y} = \frac{1}{k} \sum_{i=1}^k y_i \quad (10)$$

Donde:

- $\hat{y}$  es el valor predicho.
- k es el número de vecinos más cercanos.
- $y_i$  son los valores de la variable dependiente de los k vecinos más cercanos.

**Figura 17**

*Grafica de k vecinos más cercanos k*



*Nota: Grafica de regresión vecinos más cercanos por EcuRed, 2017. Fuente: Regla de los K vecinos más cercanos - EcuRed*

### 1.14.1.2.3. Árboles de Decisión

Un árbol de decisión es una estructura de datos que está compuesta por nodos que representan decisiones y ramas que presentan las posibles consecuencias de esas decisiones. En Machine Learning un árbol se compone a partir de un conjunto de datos de entrenamiento. Su objetivo principal es aprender las reglas o patrones en los datos que permitan predecir la etiqueta de clase de nuevas instancias (Gonzalo, 2023).

Son modelos de aprendizaje supervisado que son usados en problemas de regresión como la clasificación.

Para la regresión, una función comúnmente utilizada para evaluar la calidad de una división es la reducción de la varianza. La varianza de un conjunto de datos se define como:

*Ecuación de árbol de decisión*

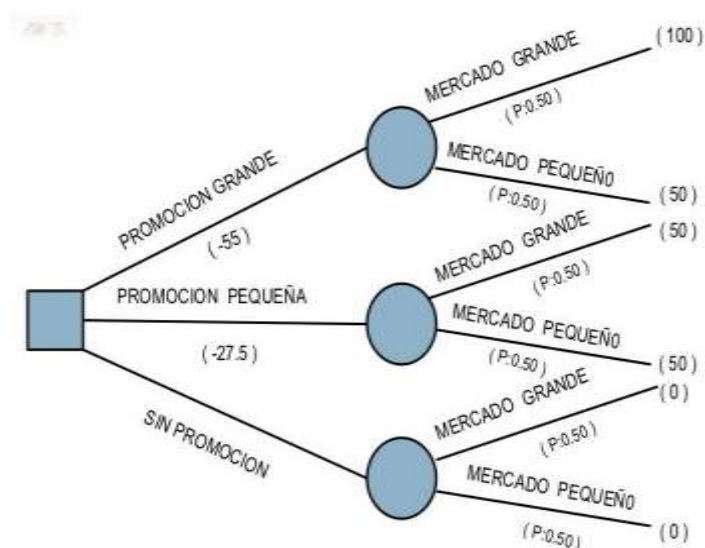
$$Var(S) = \frac{1}{|S|} \sum_{i=1}^{|S|} (y_i - \bar{y})^2 \quad (11)$$

Donde:

- $y_i$  son los valores de la variable dependiente en el conjunto S.
- $\bar{y}$  es la media de los  $y_i$

Figura 18

Diagrama de árbol de decisión



Nota: Grafica de regresión árbol de decisión por analisisemd, 2017. Fuente: Análisis y diseños de Sistema de Información: Árbol de decisiones ( analisisemd.blogspot.com)

#### 1.14.1.2.4. Random Forest

Es un algoritmo que combina la salida de múltiples árboles de decisión para alcanzar un solo resultado, también maneja problemas de regresión y clasificación (IBM)

La idea principal de Random Forest es construir un bosque de árboles de decisión, donde cada árbol se entrena de forma independiente usando una muestra aleatoria de los datos de entrenamiento y una selección de las características. Son conocidos por su capacidad para manejar conjuntos de datos grandes y complejos también con el sobreajuste.

Para una nueva muestra  $x$ , la predicción de un Random Forest para la regresión se calcula como el promedio de las predicciones de todos los árboles  $T$ :

Ecuación de Random Forest

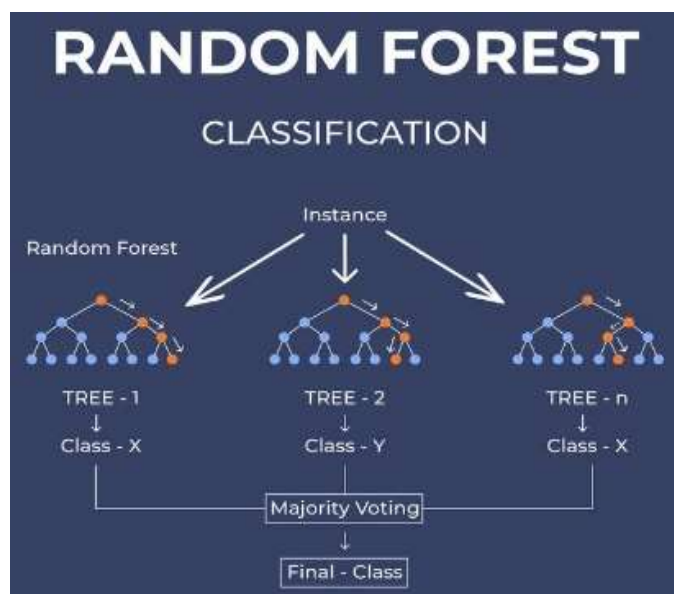
$$\hat{y} = \frac{1}{T} \sum_{i=1}^T \hat{y}_i \quad (12)$$

Donde:

- $\hat{y}$  es la predicción final.
- $\hat{y}_i$  es la predicción del árbol.

**Figura 19**

Estructura de random forest



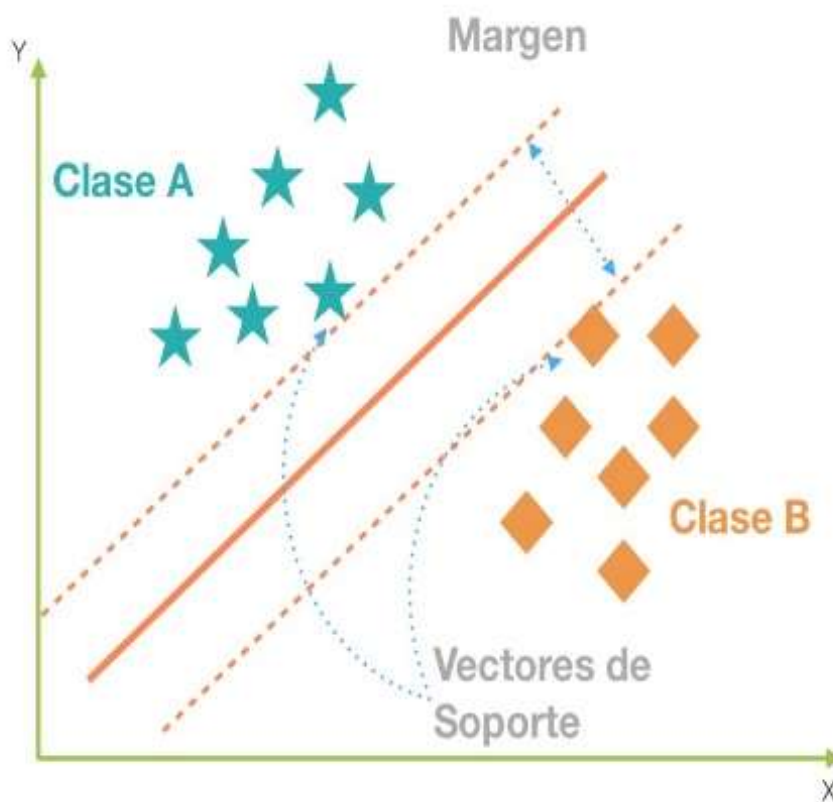
Nota: Estructura de random forest por learnbay, 2021. Fuente: Best Data Science Courses Training in Bangalore - Learnbay

#### 1.14.1.2.5. Máquinas de Vectores de Soporte (Support Vector Machines, SVM)

Son un conjunto de algoritmos de aprendizaje supervisado que están relacionados con la regresión y la clasificación. Se puede utilizar diferentes funciones del kernel para mapear los datos a un espacio dimensional superior.

Figura 20

Grafica de Maquina de vectores de Maquina



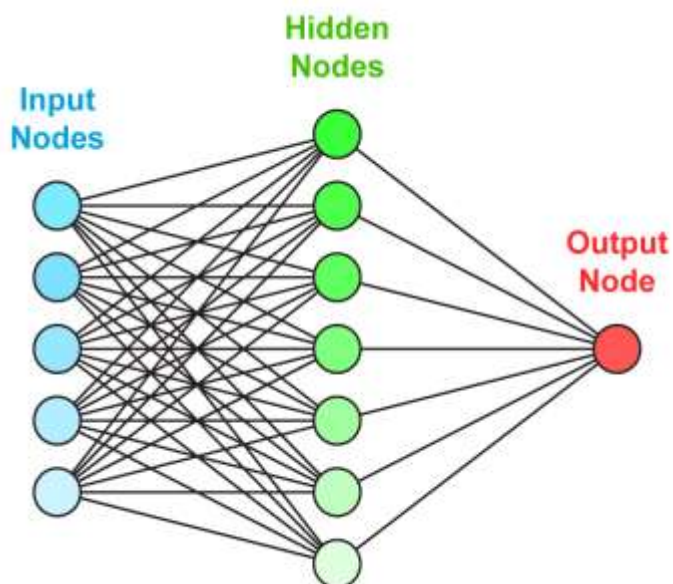
Nota: Grafica de SVM por aprende IA, 2019. Fuente: Máquinas Vectores de Soporte Clasificación – Teoría - Aprende IA

#### 1.14.1.2.6. Redes Neuronales Artificiales

Son un conjunto de los exponentes de la inteligencia artificial bajo el aprendizaje automático, las redes neuronales artificiales son un modelo computacional que permite simular el comportamiento del cerebro humano. Una red neuronal está formada por neuronas artificiales que son nodos que reciben información de otras neuronas de manera similar a los impulsos nerviosos del cerebro humano las procesan y generan un valor de salida que alimenta a otras neuronas de la red (UNIR, 2021).

Figura 21

Estructura de redes neuronales



Nota: Estructura de redes neuronales por Pedro Ney, 2018. Fuente: Introducción redes neuronales artificiales - Electrical e-Library.com (electricalelibrary.com)

Algunos tipos de redes neuronales son:

#### a) Monocapa

Una red neuronal monocapa consta de una sola capa de neuronas que recibe entradas, las procesa y produce una salida. Es la forma más simple de red neuronal y se utiliza principalmente para problemas de clasificación binaria o regresión lineal simple.

#### b) Multicapa

Las redes neuronales multicapa, también conocidas como redes neuronales feedforward, tienen una estructura más compleja con una o más capas ocultas entre la capa de entrada y la capa de salida. Estas capas intermedias permiten que la red aprenda representaciones más complejas de los datos y aborden problemas más difíciles, como la clasificación de imágenes o el procesamiento del lenguaje natural.

### **c) Convolutacional**

Las redes neuronales convolucionales (CNN) son especialmente eficaces para el procesamiento de datos estructurados en forma de cuadrícula, como imágenes. Utilizan capas convolucionales que aplican filtros para extraer características importantes de los datos de entrada, seguidas de capas de agrupación que reducen la dimensionalidad de los datos. Las CNN son ampliamente utilizadas en aplicaciones de visión por computadora.

### **d) Concurrentes**

Estas redes neuronales no tienen una estructura de capas fija y están formadas por neuronas interconectadas de manera no estructurada. Esto permite que la red tenga memoria y que la información generada en iteraciones anteriores afecte al resultado del procesamiento en iteraciones futuras. Estas redes son menos comunes y pueden ser más difíciles de entrenar, pero pueden ser útiles en ciertos tipos de problemas, como el procesamiento de secuencias temporales.

### **e) Radiales**

Las redes neuronales radiales (RNN) son un tipo especializado de red neuronal que se utiliza para problemas de clasificación o regresión en los que los datos tienen una estructura de secuencia o serie temporal. Estas redes tienen conexiones recurrentes que les permiten mantener información sobre el estado anterior de la secuencia y utilizarla para hacer predicciones sobre el siguiente paso en la secuencia.

#### **1.14.1.3. Ventajas del aprendizaje supervisado**

- Precisión en la clasificación
- Predicciones basadas en experiencias previas



#### **1.14.1.4.Desventajas del aprendizaje supervisado**

- Limitaciones en la resolución de tareas complejas
- Sensibilidad a la variación de datos de prueba
- Requerimientos computacionales intensivos

#### **1.14.2. Aprendizaje no supervisado**

El aprendizaje no supervisado en inteligencia artificial es un enfoque que permite a la máquina aprender sin depender de datos etiquetados o indicaciones específicas.

En este método, la máquina es capaz de aprender a partir de un conjunto de datos no etiquetados, lo que significa que puede aprender sin la necesidad de supervisión directa. A diferencia del aprendizaje supervisado, donde se proporcionan datos de entrada y salida etiquetados, en el aprendizaje no supervisado, la máquina solo recibe datos de entrada no etiquetados y no cuenta con información sobre la salida esperada. Por lo tanto, la tarea principal es que la máquina descubra la estructura oculta de los datos y organice o clasifique los datos según sus similitudes.

El objetivo central del aprendizaje no supervisado es que la máquina adquiera un entendimiento más profundo de la estructura de los datos disponibles y, en última instancia, de los datos futuros relacionados con un fenómeno específico. Esto se logra mediante la clasificación de datos basada en similitudes, permitiendo que la máquina procese eficientemente grandes cantidades de datos de manera autónoma y sin la necesidad de supervisión directa (Quiroga, 2023).

### 1.14.2.1. Agrupamiento

En el aprendizaje no supervisado mediante agrupamiento, se busca reunir objetos similares utilizando conglomerados. La máquina emplea clústeres para agrupar los datos según las similitudes que encuentra en el conjunto de datos, clasificándolos finalmente en función de la presencia o ausencia de esas características comunes. En esencia, el objetivo es descubrir patrones o estructuras frecuentes en datos desorganizados.

Los tipos de agrupamiento son:

#### a) Exclusivo

Cada dato pertenece a un solo grupo o clúster, y no puede ser incluido en otro grupo.

#### b) Por aglomeración jerárquica

Cada dato inicialmente se considera un clúster, y se realizan uniones interactivas entre grupos cercanos para reducir gradualmente el número de grupos hasta alcanzar el número deseado. Por solapamiento o por superposición

Se utilizan conjuntos difusos para agrupar datos, permitiendo que cada punto pueda pertenecer a dos o más grupos con diferentes nombres.

#### c) Probabilístico

Este método emplea distribuciones de probabilidad para formar grupos

### 1.14.2.2. Asociación

Adicionalmente, el aprendizaje no supervisado tiene la capacidad de identificar asociaciones entre diferentes conjuntos de datos, permitiendo detectar relaciones que ocurren de manera conjunta. Este enfoque es especialmente efectivo para descubrir conexiones entre variables que se manifiestan simultáneamente. Un ejemplo ilustrativo de esta relación podría ser cuando una persona compra café y, al mismo tiempo, adquiere azúcar. De manera similar,

se podría identificar una asociación entre la compra de un celular y la adquisición de un estuche.

Este tipo de aprendizaje posibilita la identificación de asociaciones interesantes en grandes bases de datos, revelando patrones de comportamiento que podrían no ser evidentes de manera directa. De este modo, proporciona una herramienta valiosa para explorar y comprender las relaciones complejas entre variables en conjuntos de datos no etiquetados.

### **1.14.3. Aprendizaje por refuerzo**

El aprendizaje por refuerzo es un enfoque ambicioso en inteligencia artificial que se basa en la teoría de premios y castigos para tomar decisiones entre múltiples variables. A través del ensayo y error, el algoritmo determina las decisiones óptimas en diferentes escenarios para obtener resultados positivos. Este enfoque se utiliza en áreas como videojuegos, destacando el software AlphaStar, y en robótica para regular movimientos y realizar tareas específicas (Carranza, 2022).

#### **1.14.3.1. Elementos del aprendizaje por refuerzo**

##### **a) Agente**

Prototipo a instruir para tomar decisiones de manera independiente.

##### **b) Ambiente**

Entorno en el que el agente se desplaza e interactúa.

#### **1.14.3.2. Categorías del aprendizaje por refuerzo**

##### **a) Aprendizaje por refuerzo positivo**

Refuerza el comportamiento deseado.

**b) Aprendizaje por refuerzo negativo**

Refuerza al evitar condiciones negativas.

**1.14.3.3. Ventajas del aprendizaje por refuerzo**

- Resuelve problemas complejos del mundo real.
- Similar al aprendizaje humano, brindando resultados precisos.
- Permite resultados a largo plazo.

**1.14.3.4. Desventajas del aprendizaje por refuerzo**

- Menos efectivo para problemas simples.
- Requiere grandes cantidades de datos y cálculos.
- Puede conducir a sobrecarga de estados y debilitar resultados.
- Limitaciones en la dimensionalidad para sistemas físicos reales

**1.14.4. Aprendizaje semi supervisado**

El aprendizaje semi supervisado se sitúa entre los enfoques supervisados y no supervisados del machine learning, utilizando tanto datos etiquetados como no etiquetados durante el entrenamiento. Aunque opera principalmente con datos no etiquetados debido a la naturaleza costosa de las etiquetas, se diferencia claramente de los métodos supervisados y no supervisados al basarse en la presencia y ausencia de etiquetas (Carranza, 2022).

El objetivo principal es superar las limitaciones del aprendizaje supervisado y no supervisado. También utilizar eficientemente todos los datos disponibles, incorporando la capacidad de agrupar datos no etiquetados en datos etiquetados mediante algoritmos de aprendizaje no supervisado.

#### **1.14.4.1. Ventajas del aprendizaje semisupervisado**

- Algoritmo simple y comprensible.
- Altamente eficiente.
- Aborda las limitaciones de los enfoques supervisados y no supervisados.

#### **1.14.4.2. Desventajas del aprendizaje semisupervisado**

- Resultados de las iteraciones pueden ser inestables.
- No es aplicable a datos a nivel de red.
- Precisión relativamente baja.

### **1.15. MÉTODO CIENTÍFICO**

El método científico es un proceso sistemático con el objetivo de establecer relaciones entre hechos, enunciando leyes y teorías que expliquen el funcionamiento del mundo. Este sistema riguroso sigue una serie de pasos para generar conocimiento científico mediante la comprobación empírica de fenómenos. La observación se utiliza para proponer una hipótesis, la cual se comprueba a través de la experimentación (Etece, 2023).

### **1.16. METODOLOGÍA CRISP-DM**

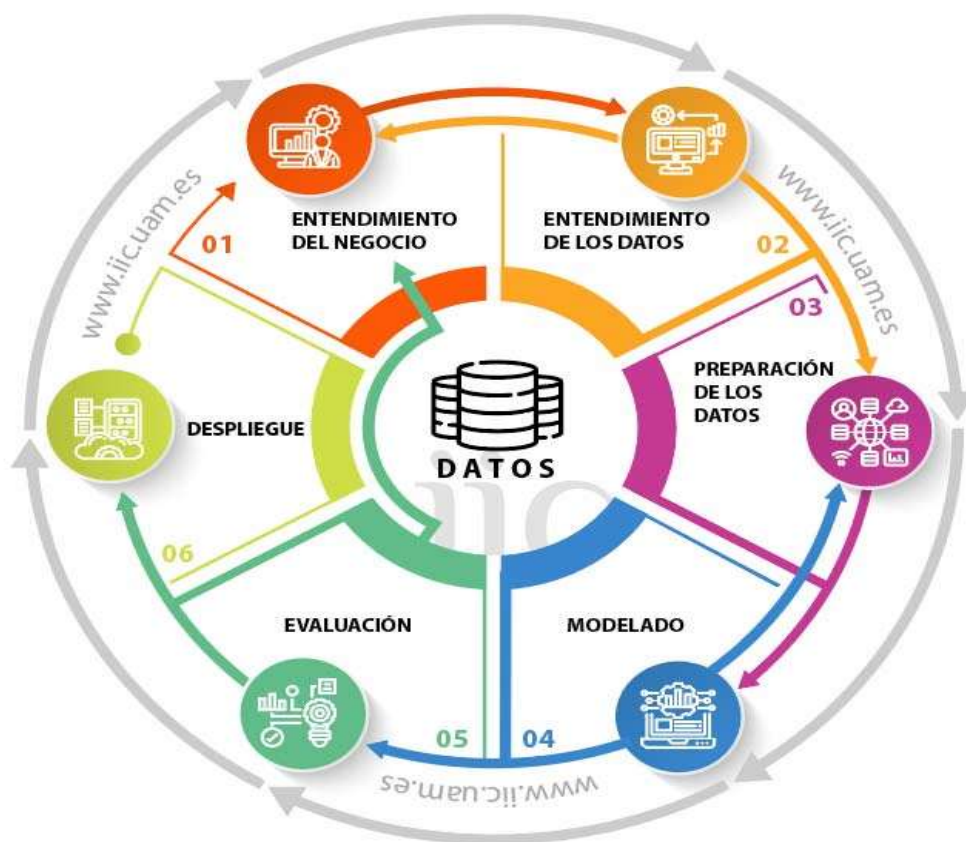
Según Galan (2015), CRISP-DM (Cross Industry Standard Process for Data Mining), es un modelo de proceso de minería de datos que describe una manera en la que los expertos en esta materia abordan el problema. Su principal objetivo es desarrollar los modelos a partir de un procesamiento de datos para anticipar futuras fallas.

“Esta metodología fue desarrollada en 1996 por un consorcio de empresas, organizaciones y expertos en minería de datos y se ha convertido en la metodología más ampliamente utilizada para proyectos de minería de datos” (Mora, 2023).

Es una metodología que se definió en 1999 que se orienta la parte analítica y a obtención de modelos basados en datos y no es tan aplicado para el uso de Machine Learning e inteligencia artificial. Es la metodología más utilizada en Data Mining aunque con ciertas adaptaciones con una analítica descriptiva se puede trabajar con Machine Learning o Deep Learning (Gavilan, 2021).

**Figura 22**

*Fases de la metodología CRISP-DM*



*Nota: Fases de la metodología CRISP-DM por funiber. Fuente: Maestrías, Doctorados y Cursos a distancia - FUNIBER Bolivia*

### **1.16.1. Fases de CRISP-DM**

Comprende 6 fases esta metodología.

#### **1.16.1.1. Compresión del problema**

En la fase inicial de un proyecto de minería de datos (y similares), el enfoque principal radica en comprender los objetivos y requisitos del proyecto desde la perspectiva del negocio.

Esto implica:

##### **a) Establecimiento de los objetivos del negocio**

Este conjunto de actividades se centra en comprender las características del cliente, sus antecedentes, objetivos de negocio y criterios de éxito. Se busca identificar qué es lo que el cliente espera lograr con el proyecto.

##### **b) Evaluación de la situación**

Se busca comprender la situación actual de la empresa para determinar su capacidad para aprovechar los datos disponibles. Esto implica responder preguntas como ¿Cuáles son los términos clave para describir el negocio? ¿Qué recursos humanos y materiales están disponibles para el proyecto? ¿Qué tipos de datos están disponibles y cuáles son los principales factores de riesgo y planes de contingencia asociados? También se evalúan los potenciales costos y beneficios del proyecto.

##### **c) Establecimiento de los objetivos de la minería de datos**

Aquí se definen los objetivos específicos del proyecto de minería de datos y los criterios para evaluar su éxito. Se busca determinar qué se espera obtener del proyecto, ya sea una nueva herramienta o servicio, o información para la planificación estratégica. Se identifican los KPIs relevantes para medir el progreso y el cumplimiento de los objetivos.

#### **d) Generación del plan del proyecto**

La fase de comprensión del negocio culmina en la determinación de la viabilidad del proyecto y la elaboración de un plan detallado para las siguientes fases de intervención. Este plan incluye los recursos a utilizar, compromisos, indicadores de avance y cualquier otro aspecto relevante para la ejecución exitosa del proyecto.

#### **1.16.1.2. Comprensión de los datos**

La comprensión de los datos es un proceso fundamental que implica familiarizarse con las características de los datos disponibles para el proyecto y los requisitos adicionales de datos. Esta etapa sienta las bases para todas las actividades posteriores del proyecto. En ocasiones, será necesario volver a analizar el negocio para comprender mejor los datos.

Actividades en esta fase y sus resultados esperados:

##### **a) Recolección inicial de datos**

En este paso inicial, se recolectan muestras de los datos disponibles y se identifica el alcance del conjunto de datos, así como posibles dificultades para su recolección y uso. Se distinguen los datos propios de la empresa de otros conjuntos de datos complementarios adquiridos y se identifican las fuentes adicionales de datos disponibles.

##### **b) Descripción de los datos**

Se identifican las características generales de los datos, incluyendo el número de variables disponibles, la cantidad de registros, la frecuencia de generación de cada variable, su identificación, el significado de cada campo y el formato inicial. El análisis preliminar busca determinar si los datos disponibles son suficientes para alcanzar los objetivos de la minería de datos.



### **c) Exploración de los datos**

Esta actividad tiene como objetivo identificar la distribución general de los datos mediante pruebas estadísticas básicas y establecer hipótesis preliminares. Este análisis permite identificar la complejidad del problema y realizar una selección preliminar de las técnicas a utilizar en el proyecto.

### **d) Verificación de la calidad de los datos**

En esta etapa se verifica la completitud de los datos y se buscan porcentajes de datos incompletos, valores atípicos o fuera de rango, así como variables equivalentes. Se definen estrategias generales para abordar los problemas identificados y garantizar la calidad de los datos utilizados en el proyecto.

#### **1.16.1.3.Preparación de los datos**

La fase de preparación de los datos abarca todas las actividades necesarias para generar el conjunto de datos final que se utilizará para alimentar los modelos, partiendo de los datos originales. Esta etapa es ampliamente reconocida como la que consume más tiempo en un proyecto de ciencia de datos.

Actividades en esta fase:

#### **a) Selección de los datos**

Se busca seleccionar una porción del volumen total de datos preseleccionados que parezca representativa del problema de minería de datos. Esto implica elegir registros que cubran todo el universo de objetos a analizar y seleccionar características (variables) que describan mejor los diferentes objetos, evitando variables básicamente equivalentes. Es

crucial justificar y documentar las razones para incluir o excluir diferentes subconjuntos de datos.

#### **b) Limpieza de los datos**

Esta actividad, que consume la mayor parte del tiempo y recursos, tiene como objetivo corregir las deficiencias identificadas en la fase anterior. Se abordan tareas como el tratamiento de datos con valores faltantes, la gestión de datos atípicos e inconsistentes, entre otros.

#### **c) Estructuración de los datos**

Consiste en generar la estructura de los registros que se utilizarán en el análisis, principalmente mediante la creación de nuevas variables que sean más descriptivas y que ayuden a reducir la complejidad del espacio de representación.

#### **d) Integración de los datos**

Esta etapa implica combinar datos de diferentes fuentes en un único conjunto de datos. Puede involucrar la creación de una tabla unificada a partir de varias tablas o la generación de registros o columnas nuevas mediante la agregación de datos de diferentes fuentes.

#### **e) Formateo de los datos**

El objetivo es preparar los datos en la forma en que serán procesados, típicamente mediante transformaciones que no alteran su significado. Esto puede incluir cambios de escala, eliminación de caracteres especiales y reordenamiento de columnas y filas en datos tabulares, entre otras tareas comunes.

#### **1.16.1.4. Modelado**

En esta fase, se eligen y prueban diversas técnicas de modelado, ajustando sus parámetros para adaptarse a la dinámica representada por los datos. En ocasiones, una técnica puede requerir datos no considerados en etapas previas, lo que podría implicar retroceder para corregir la construcción del conjunto de datos. La elección de las técnicas se basa en criterios técnicos y prácticos, como la pertinencia de la técnica para el problema específico, la disponibilidad de datos adecuados, el tiempo disponible para obtener un modelo y el conocimiento del equipo de desarrollo sobre la técnica.

Actividades en esta fase y sus resultados esperados.

##### **a) Selección de las técnicas de modelado**

Se eligen las técnicas de modelado en función de factores como el grado de estructuración del problema y los datos, los objetivos del proyecto, el dominio de las técnicas por parte del equipo, restricciones legales y preferencias del cliente. Es común la necesidad de utilizar diversas técnicas para resolver un problema complejo.

##### **b) Generación del plan de pruebas**

Se crea un plan para implementar el prototipo o prueba de concepto. Esto incluye la selección de bibliotecas y herramientas, la implementación de las técnicas, la definición de estrategias de segmentación de datos y de pruebas, así como la selección de medidas de evaluación.

##### **c) Construcción del modelo**

Se construye el modelo o conjunto de modelos, definiendo los parámetros de cada uno, realizando pruebas preliminares y ajustando el modelo según sea necesario. Es importante

destacar la naturaleza incierta de los proyectos de minería de datos, que a menudo requiere retroceder a pasos anteriores, como la selección de modelos o la preparación de datos.

#### **d) Evaluación del modelo**

Se evalúa el modelo mediante pruebas con datos reservados, generando informes sobre niveles de precisión/error, tiempos de respuesta y cualquier otra información relevante para la implementación final del sistema.

#### **1.16.1.5.Evaluación**

En esta fase, nos centramos en evaluar el grado de alineación del modelo con los objetivos de negocio.

##### **a) Evaluar el modelo o modelos generados hasta el momento**

Se realiza una evaluación exhaustiva del desempeño de los modelos en relación con los objetivos comerciales establecidos. Esto implica analizar métricas de rendimiento, como precisión, recall, F1- score, entre otras, así como su capacidad para generalizar a nuevos datos y su interpretabilidad.

##### **b) Revisar todo el proceso de minería de datos que nos ha llevado hasta este punto**

Se realiza una revisión detallada de todas las etapas y decisiones tomadas en el proceso de minería de datos hasta el momento. Esto incluye la preparación de datos, selección y ajuste de modelos, y cualquier otra acción realizada durante el proyecto.

##### **c) Establecer los siguientes pasos a tomar**

Se determinan los pasos futuros a seguir en función de la evaluación del modelo actual y del análisis del proceso de minería de datos. Esto puede implicar la repetición de fases

anteriores para mejorar el modelo, abrir nuevas líneas de investigación para explorar diferentes enfoques o incluso realizar ajustes en los objetivos comerciales en función de los hallazgos obtenidos.

#### **1.16.1.6.Despliegue**

El objetivo último de esta fase es implementar el despliegue de los resultados obtenidos de manera que se propague a los usuarios finales, así como garantizar el mantenimiento del mismo una vez finalizado el despliegue.

En esta fase, deberemos ser capaces de:

##### **a) Diseñar un plan de despliegue de modelos**

Se elabora un plan detallado para implementar los modelos y el conocimiento adquirido en la organización. Esto incluye consideraciones sobre la infraestructura necesaria, la capacitación del personal y la comunicación con los usuarios finales.

##### **b) Realizar seguimiento y mantenimiento de la parte más operativa del despliegue**

Se lleva a cabo un seguimiento continuo de la implementación para garantizar su correcto funcionamiento. Esto implica monitorear los sistemas, resolver cualquier problema que surja y realizar actualizaciones según sea necesario para mantener la eficiencia y precisión del sistema.

##### **c) Revisar el proyecto en su globalidad con el objetivo de identificar lecciones aprendidas**

Se realiza una revisión exhaustiva del proyecto en su conjunto para identificar lecciones aprendidas y áreas de mejora. Esto permite retroalimentar el proceso y mejorar futuros proyectos de minería de datos.



# CAPÍTULO III

## CAPÍTULO III

### 1.17. DISEÑO METODOLÓGICO

#### 1.17.1. *Introducción*

En este capítulo se puso en práctica el diseño metodológico mencionado en anteriores capítulos respecto al modelo de predicción del índice de crecimiento delictivo en la ciudad de El Alto, para esto se realiza el preprocesamiento de datos para obtener buenos resultados que consiste en la depuración de datos innecesarios y valores faltantes. Identificar las variables de estudio que se usarán para el modelo y que herramientas serán utilizadas para implementar el modelo de software.

#### 1.17.2. *Tipo de investigación*

En el contexto de este estudio se usa el tipo correlacional, la investigación busca examinar la relación entre el uso de modelos de predicción basados en Machine Learning y la capacidad para anticipar actividades delictivas en la ciudad de El Alto. Aunque no se puede afirmar causalidad directa, existe una relación significativa que puede ser aprovechada para mejorar la prevención de delitos. La investigación correlacional se centra en identificar y medir la relación entre dos o más variables sin manipularlas.

La investigación no intervendrá directamente en las variables, en su lugar se observará y analizará los datos existentes para identificar patrones y relaciones.

#### 1.17.3. *Diseño de la investigación*

Para aplicar el diseño de investigación no experimental a la investigación se adoptará un enfoque que permita analizar y comprender el fenómeno del aumento del índice delictivo utilizando datos históricos y técnicas de Machine Learning sin intervenir directamente en las variables del estudio. Este diseño se centrará en la recopilación y análisis de datos existentes,

principalmente de registros policiales históricos de la FELCC y otras fuentes secundarias como datos socioeconómicos pertinentes para la ciudad de El Alto.

Durante el proceso de investigación, se llevó a cabo un preprocesamiento de datos para garantizar la calidad y consistencia de la información utilizada, esto incluirá la limpieza de datos, la conversión de variables que podrían distorsionar los resultados del análisis. El análisis exploratorio de datos es un papel fundamental para identificar patrones, tendencias y posibles correlaciones entre variables relevantes y el índice delictivo. Se utilizó técnicas estadísticas y visualizaciones de datos para comprender mejor la estructura de los datos recopilados a lo largo del tiempo. En cuanto al modelado predictivo, se empleó diversos algoritmos de Machine Learning como regresión lineal, árboles de decisión y Random Forest, según la complejidad de los datos. La selección del modelo más adecuado está basada en criterios de rendimiento y precisión, evaluados mediante técnicas de validación y error.

#### ***1.17.4. Variables de la investigación***

Para desarrollar un modelo predictivo del crecimiento del índice delictivo en la ciudad de El Alto mediante técnicas de Machine Learning, es crucial identificar y definir claramente las variables que se utilizarán. Estas variables se clasifican en dependientes e independientes. A continuación, se describen las variables más relevantes para esta investigación:

##### **1.17.4.1. Variable dependiente**

La variable dependiente en este estudio es la capacidad de anticipar actividades delictivas. La anticipación de actividades delictivas implica predecir no solo la probabilidad de que un delito ocurra, sino también la ubicación, el tipo de delito, el momento del día y otros factores relevantes.



#### **1.17.4.2.Variable independiente**

En este estudio, la variable independiente es el modelo de predicción del crecimiento del índice delictivo desarrollado mediante el uso de métodos de Machine Learning. Este modelo emplea técnicas avanzadas de análisis de datos y algoritmos de aprendizaje automático para identificar patrones y tendencias en datos históricos de delitos, el modelo es capaz de prever aumentos potenciales en el índice delictivo.

#### **1.17.5. Ambiente de la investigación**

El ambiente de la investigación define el contexto en el cual se desarrollará el estudio, incluyendo los aspectos geográficos, sociales, económicos y tecnológicos. Se detalla el ambiente en el que se llevará a cabo la investigación sobre la predicción del crecimiento del índice delictivo en la ciudad de El Alto mediante técnicas de Machine Learning.

##### **1.17.5.1.Contexto geográfico**

El Alto se encuentra en el departamento de La Paz, Bolivia y es una de las ciudades más importantes del país, tanto en términos de población como de desarrollo urbano. Situada a más de 4,000 metros sobre el nivel del mar, El Alto se caracteriza por su altitud y clima frío. La ciudad está dividida en varios distritos, cada uno con características demográficas y socioeconómicas distintas.


#### **1.17.6. Descripción de la metodología a usar**

##### **1.17.6.1.Comprensión del problema**

Se deben identificar las variables clave que podrían influir en la incidencia delictiva, como datos de los meses en los que se produce mayor registro de delitos para determinar en qué tiempo ocurren más delitos.

**Figura 23**

*Días y horas con más frecuencias de delitos*

Características del hecho <sup>1)</sup>	Cantidad	Distribución porcentual	Tendencia mensual
<b>Día del Hecho</b>			
Lunes	2.813	13,79%	
Martes	2.696	13,21%	
Miércoles	2.844	13,94%	
Jueves	2.916	14,29%	
Viernes	2.951	14,46%	
Sábado	3.018	14,79%	
Domingo	3.155	15,46%	
Indeterminado	13	0,06%	
<b>Hora del hecho</b>			
00:00 - 05:59	4.686	22,96%	
06:00 - 11:59	5.086	24,92%	
12:00 - 17:59	5.338	26,16%	
18:00 - 23:59	5.275	25,85%	
Indeterminado	21	0,10%	

*Nota: Días y horas con más frecuencias de delitos según el Atlas de Bolivia en Seguridad Ciudadana (p. 310), por OBSCD, 2023. Fuente: Observatorio Boliviano de Seguridad Ciudadana y Lucha Contra las Drogas.*

### 1.17.6.2. Comprensión de los datos

Realizar una limpieza de los datos para eliminar registros incorrectos, incompletos o duplicados. También es importante transformar los datos en un formato adecuado para el análisis, asegurándose de que las variables estén codificadas correctamente y sean compatibles con las herramientas de análisis estadístico y Machine Learning que se utilizarán posteriormente.

La información estadística es muy importante siendo una base fundamental para la toma de decisiones. En esta presente investigación se tiene información de datos estadísticos recopilados de diferentes unidades investigativas (FELCC) de la Policía boliviana. Sobre el registro de denuncias de diferentes hechos delictivos en el municipio de El Alto de las gestiones 2013 a 2023 se tiene los siguientes cuadros y gráficos:

**Tabla 4***Registro de denuncias de los delitos mas frecuentes*

DETALLE	GESTION 2013	GESTIO N 2014	GESTION 2015	GESTION 2016	GESTION 2017	GESTION 2018	GESTION 2019	GESTION 2020	GESTION 2021	GESTION 2022	GESTION 2023	TOTAL
<b>EL ALTO</b>												
<b>LESIONES GRAVES Y LEVES</b>	576	400	917	886	729	637	476	376	397	651	1429	<b>7474</b>
<b>ESTAFA</b>	510	512	392	196	84	145	202	95	64	123	200	<b>2523</b>
<b>ROBO</b>	547	577	560	498	381	181	173	223	268	321	302	<b>4031</b>
<b>ROBO AGRAVADO</b>	542	390	391	404	176	126	139	114	129	214	149	<b>2774</b>
<b>HURTO</b>	281	216	277	169	134	100	130	99	109	138	168	<b>1821</b>
<b>USO DE INSTRUMENTO FALSIFICADO</b>	7	2	7	11	69	81	113	43	31	79	123	<b>566</b>
<b>ESTELIONATO</b>	129	192	138	132	82	79	110	29	39	55	86	<b>1071</b>
<b>TRATA DE PERSONAS</b>	140	152	87	47	92	84	63	21	80	144	307	<b>1217</b>
<b>ALLANAMIENTO DE DOMICILIO A SUS DEPENDENCIAS</b>	150	107	97	57	47	41	45	36	43	59	52	<b>734</b>
<b>FALESDAD MATERIAL</b>	260	252	198	85	57	27	43	12	19	29	30	<b>1012</b>
<b>DELITOS CONTRA LA SALUD PUBLICA</b>	8	6	16	8	6	1	2	105	54	5	5	<b>216</b>
<b>OTROS DELITOS</b>	3402	3319	1979	1624	812	441	462	929	940	1151	1141	<b>16200</b>
<b>TOTAL</b>	<b>6552</b>	<b>6125</b>	<b>5059</b>	<b>4117</b>	<b>2669</b>	<b>1943</b>	<b>1958</b>	<b>2082</b>	<b>2173</b>	<b>2969</b>	<b>3992</b>	<b>39639</b>

*Nota: Registro de denuncias de diferentes delitos gestión 2013 a 2023 por la FELCC. Fuente: Pol. Bol.- PP.OO. -D.N.E.*

Figura 24

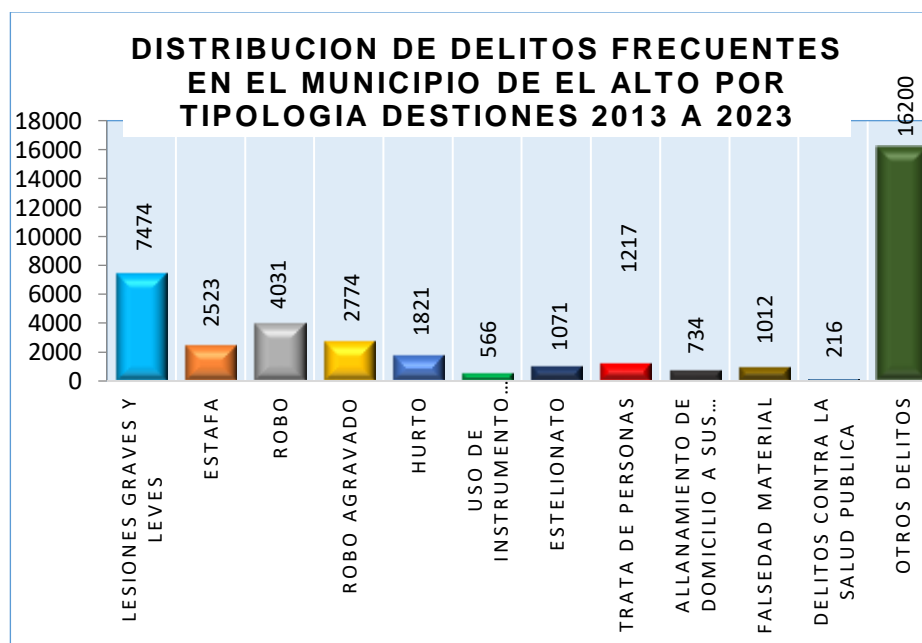
Gráfico de distribución de denuncias de la gestión 2013 a 2023



Nota: Gráfico de distribución de denuncias de la gestión 2013 a 2023 por la FELCC. Fuente: Pol. Bol.-PP.OO. -D.N.E.

Figura 25

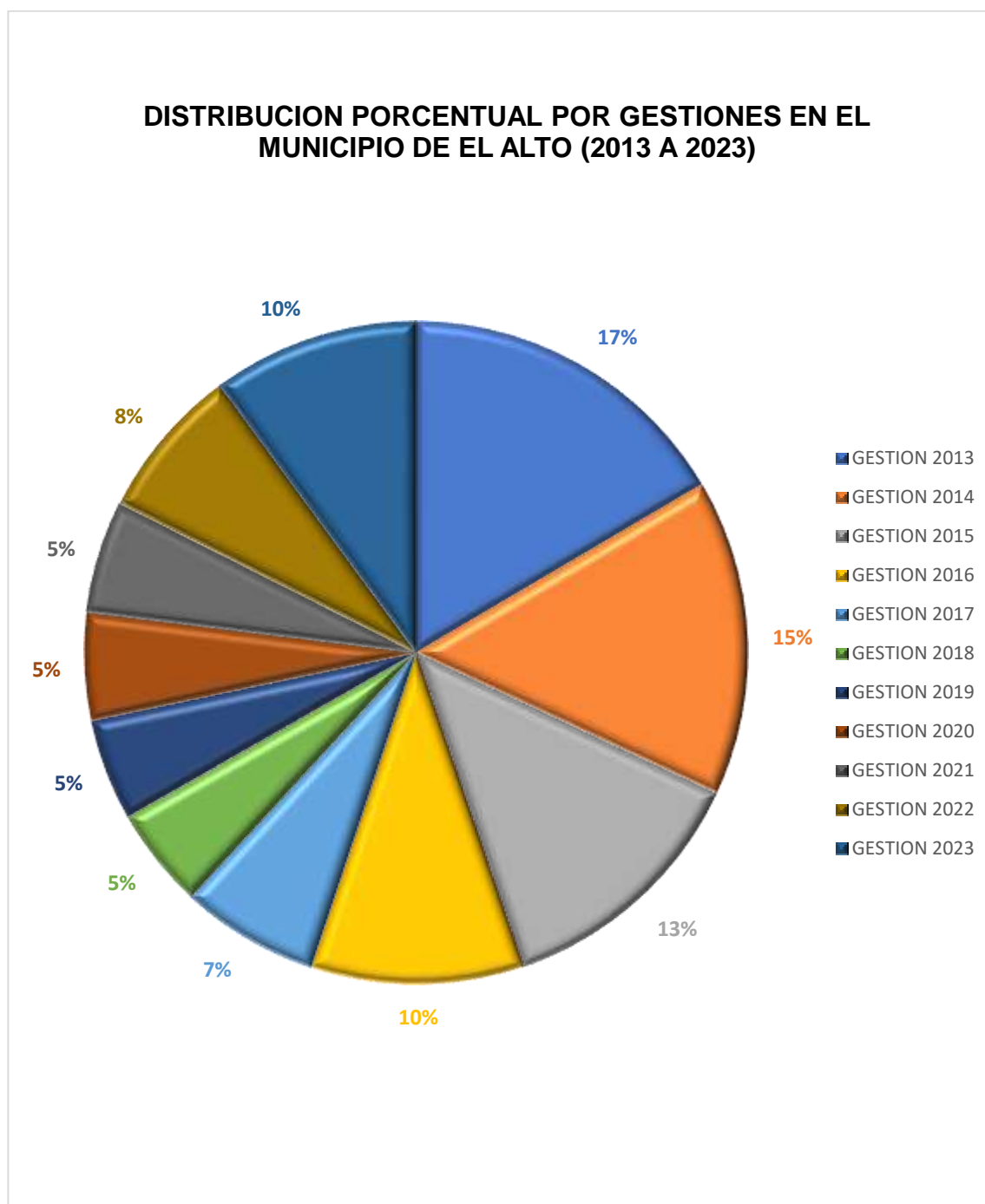
Gráfico de denuncias por tipología gestión 2013 a 2023



Nota: Gráfico de denuncias por tipología gestión 2013 a 2023 por la FELCC. Fuente: Pol. Bol.-PP.OO. -D.N.E.

Figura 26

Grafico porcentual por gestiones 2013 a 2023



Nota: Grafico porcentual por gestiones 2013 a 2023 por la FELCC. Fuente: Pol. Bol.-PP.OO. -D.N.E.

También se cuenta con tablas mensuales por año.

Tabla 5

*Fragmento de conjunto de datos registrados en la gestión 2022*

DETALLE	ENERO	FEBRERO	MARZO	ABRIL	MAYO	JUNIO	JULIO	AGOSTO	SEPTIEMBRE
LESIONES GRAVES Y LEVES	53	38	90	20	32	48	49	83	65
MUERTE DE PERSONA	47	44	49	59	34	37	48	5	35
ROBO	33	23	24	30	25	25	27	17	23
ROBO AGRABADO	19	11	25	10	8	25	10	26	20
TRATA DE PERSONAS	1	3	5	6	14	5	1	10	28
HURTO	13	7	16	7	8	9	10	10	13
ESTAFA	10	4	4	9	6	14	11	17	8
USO DE INSTRUMENTO FALSIFICADO	5	1	6	2	6	9	8	7	10
SUICIDIO	3	3	8	5	10	6	9	0	4
ALLANAMIENTO DE DOMICILIO O SUS DEPENDENCIAS	6	3	1	2	5	7	7	8	10
RECEPTACION	12	2	0	1	4	4	1	9	8
ESTELIONATO	4	5	2	1	4	10	1	7	1
FALSEDAD IDEOLOGICA	0	3	3	2	0	8	3	3	0
SUPRESION O DESTRUCCION DE DOCUMENTO	2	1	2	5	0	1	8	1	3
FALSEDAD MATERIAL	1	1	0	4	2	3	8	1	1
AVASALLAMIENTO	3	4	1	4	2	4	1	3	1
INCUMPLIMIENTO DE DEBERES	0	2	4	1	1	6	1	0	1

*Nota: Grafico porcentual por gestiones 2013 a 2023 por la FELCC. Fuente: Pol. Bol.-PP.OO. -D.N.E.*

### 1.17.6.3.Preparación de los datos

División del conjunto de datos en tablas por cada año según las denuncias registradas en cada gestión.

**Figura 27**

*División del conjunto de datos*

```
# Tablas separada por cada año cop
import re
# Función para separar columnas en otras tablas por año y guardarlas en variables
def dividir(df_cop):
    tablas_año = {}
    for col in df_cop.columns:
        match = re.match(r'([A-Za-z]+)_([0-9]+)', col)
        if match:
            mes = match.group(1)
            año = match.group(2)
            if año not in tablas_año:
                tablas_año[año] = pd.DataFrame()
            tablas_año[año][mes] = df_cop[col]
    # Crear variables para cada año
    for año, tabla in tablas_año.items():
        globals()[f'tabla_{año}'] = tabla
    return tablas_año

# llamada a la función para obtener las tablas separadas por año y guardarlas en variables
tablas_año = dividir(df_cop)
tablas_año_cop2 = dividir(df_cop2)
# Iterar sobre las tablas por año y mostrarlas
for año, tabla in tablas_año.items():
    print(f"Tabla del año {año}:")
    print(tabla)
```

Inversión de la tabla del conjunto de datos por cada año según las denuncias registradas en cada gestión.

**Figura 28**

*Inversión de tablas por año*

```
# Función para invertir todas las tablas por año
def invertir(tablas_año):
    invertir_año = {}
    for año, tabla in tablas_año.items():
        invertir_año[año] = tabla.transpose()
    return invertir_año

# Invertir todas las tablas por año
invertir_año = invertir(tablas_año)
invertir_año_2 = invertir(tablas_año_cop2)
# Iterar sobre las tablas invertidas por año y mostrarlas
for año, tabla_invertida in invertir_año.items():
    print(f"Tabla invertida del año {año}:")
    print(tabla_invertida)
    print()
```

Imputación de valores faltantes es importante verificar si la tabla de datos (dataframe) tiene algún valor nulo y la eliminar ceros en las filas o columnas, ya que podría producir errores al implementar el modelo.

**Figura 29**

*Imputar valores faltantes*

```
# Verificar valores faltantes por columna
print("Valores faltantes por columna:\n", df.isnull().sum())

# Verificar si hay algún valor faltante en alguna columna
print("\n Valor faltante en columna\n", df.isnull().any())

# Total de valores faltantes en el DataFrame
print("\n Total de valores faltantes en el DataFrame:\n", df.isnull().sum().sum())

# Información del DataFrame
print("\n Información del DataFrame:")
df.info()

# Visualización de valores faltantes
sns.heatmap(df.isnull(), cbar=False, cmap='viridis')
plt.show()
```

**Figura 30**

*Eliminar ceros de filas y columnas*

```
# Eliminar columnas donde todos los valores son cero
def eliminar_ceros(df):
    df_eliminar_ceros = df.loc[:, (df != 0).any(axis=0)]

    # Crear un diccionario de mapeo de las etiquetas originales a las nuevas etiquetas numéricas
    new_columns = {old_label: new_label for new_label, old_label in enumerate(df_eliminar_ceros.columns)}

    # Renombrar las columnas del DataFrame
    df_eliminar_ceros.rename(columns=new_columns, inplace=True)
    return df_eliminar_ceros

# Limpiar y renombrar cada DataFrame en el diccionario
for key in dataframes:
    dataframes[key] = eliminar_ceros(dataframes[key])

# Imprimir y convertirlos de nuevo a DataFrames
for key, df in dataframes.items():
    print(f"{key}:")
    print(df)

# Volver a convertir a DataFrames
df_2018 = dataframes['df_2018']
df_2019 = dataframes['df_2019']
df_2020 = dataframes['df_2020']
df_2021 = dataframes['df_2021']
df_2022 = dataframes['df_2022']
```



Conversión de variables categóricas a variables numéricas para realizar el modelo y sea fácil de interpretar.

**Figura 31**

*Conversión de valores numéricos a categóricos*

```
# PARA X
# Crear una lista de datos
datos_mes = ["ENERO", "FEBRERO", "MARZO", "ABRIL", "MAYO", "JUNIO", "JULIO", "AGOSTO", "SEPTIEMBRE", "OCTUBRE", "NOVIEMBRE", "DICIEMBRE"]

# Crear el DataFrame con una columna llamada mes
df_mes = pd.DataFrame(datos_mes, columns=['MES'])

# Convertir la columna mes a datos numéricos secuenciales comenzando desde 1
df_mes['MES_NUMERICO'] = pd.factorize(df_mes['MES'])[0] + 1
df_mes.head()
```

**Tabla 6**

*Conversion de variables categoricas a numericas*

MES	
Enero	1
Febrero	2
Marzo	3
Abril	4
Mayo	5
Junio	6
Julio	7
Agosto	8
Septiembre	9
Octubre	10
Noviembre	11
Diciembre	12

#### 1.17.6.4. Modelado

En esta fase, se seleccionan y aplican diversas técnicas de Machine Learning para desarrollar modelos predictivos del crecimiento del índice delictivo. Esto podría incluir modelos de regresión para prever la tendencia del índice delictivo en función de variables predictoras seleccionadas.

Se evalúan y comparan diferentes algoritmos de Machine Learning para determinar cuál se adapta mejor a los datos y proporciona las predicciones más precisas. Entre los algoritmos considerados se incluyen:

- Regresión lineal y no lineal: Para establecer relaciones lineales y no lineales entre las variables.
- Árboles de Decisión: Para capturar interacciones no lineales entre las variables.
- Random Forest.

Los datos se dividirán en conjuntos de entrenamiento (80%) y prueba (20%).

#### Figura 32

*Datos de prueba y entrenamiento*

```
# Separar los datos en conjuntos de entrenamiento y prueba 80 entrenamiento y 30 prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

X_train = X_train.values.reshape([X_train.values.shape[0],1])
X_test = X_test.values.reshape([X_test.values.shape[0],1])
```

### 1.17.6.5.Evaluación

En esta fase, se evalúa el rendimiento de los modelos desarrollados utilizando métricas adecuadas para problemas de regresión o clasificación, como el error cuadrático medio (RMSE), el coeficiente de determinación ( $R^2$ ), o métricas de precisión, recall y F1-score para modelos de clasificación. Se compararán diferentes modelos y se seleccionará el más adecuado en función de su capacidad para predecir con precisión el crecimiento del índice delictivo en El Alto.

#### Figura 33

*Evaluación del modelo*

```
# evaluacion
train_mse = mean_squared_error(y_train, y_train_pred)
test_mse = mean_squared_error(y_test, y_test_pred)
train_r2 = r2_score(y_train, y_train_pred)
test_r2 = r2_score(y_test, y_test_pred)
```

### 1.17.6.6.Despliegue

Una vez que se ha seleccionado y validado el modelo final, se procederá a su implementación en un entorno operativo. Se desarrollará un prototipo, una interfaz de usuario para facilitar la consulta de predicciones y análisis de datos.

## 1.18. HERRAMIENTAS

### 1.19. HERRAMIENTAS A USAR

#### 1.19.1. Técnicas de investigación e instrumentos

Para desarrollar un modelo predictivo preciso y confiable, se emplean diversas técnicas de investigación y herramientas. La recolección de datos proporciona información cuantitativa directamente de las comunidades afectadas y los expertos en seguridad. El análisis de datos

secundarios, como registros policiales, aporta datos históricos valiosos. Herramientas de desarrollo como Python, Google Colab y bibliotecas de Machine Learning (scikit-learn) son esenciales para el análisis de datos y el desarrollo de modelos. Las herramientas de visualización, como Seaborn, ayudan a representar gráficamente los datos y los resultados, facilitando la interpretación y la comunicación de los hallazgos. También se usará librerías como Numpy, Pandas, Matplotlib que son muy útiles para el análisis de datos.

## 1.20. MÉTRICAS DE EVALUACIÓN

Para evaluar la efectividad y precisión del modelo de predicción del crecimiento del índice delictivo en la ciudad de El Alto mediante Machine Learning, es crucial utilizar diversas métricas de calidad. Estas métricas permiten medir el desempeño del modelo y asegurar que cumple con los objetivos de la investigación. A continuación, se describen algunas de las métricas de calidad más relevantes para este tipo de análisis:

### 1.20.1. Error Cuadrático Medio (MSE)

El Error Cuadrático Medio mide la media de los errores al cuadrado entre los valores reales y las predicciones del modelo. Es una métrica muy utilizada para evaluar la precisión de los modelos de regresión.

*Ecuación de Error cuadrático medio*

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (13)$$

Un MSE más bajo indica un modelo más preciso

### 1.20.2. Raíz del Error Cuadrático Medio (RMSE)

La RMSE es la raíz cuadrada del MSE. proporciona una medida del error de predicción en las mismas unidades que la variable objetivo, lo que facilita su interpretación.

*Ecuación de Raíz del Error Cuadrático Medio*

$$RMSE = \sqrt{MSE} \quad (14)$$

Al igual que el MSE, un RMSE más bajo indica un mejor desempeño del modelo.

### **1.20.3. Error Absoluto Medio (MAE)**

El Error Absoluto Medio mide la media de los errores absolutos entre los valores reales y las predicciones del modelo. Es menos sensible a los valores atípicos que el MSE.

*Ecuación Error Absoluto Medio*

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (15)$$

Un MAE más bajo sugiere una mejor precisión del modelo.

El coeficiente de determinación, también conocido como  $R^2$ , mide la proporción de la varianza en la variable dependiente que es explicada por las variables independientes en el modelo.

El  $R^2$  se calcula utilizando las siguientes fórmulas:

*Ecuación Suma de los cuadrados totales (SST)*

$$SST = \sum_{i=1}^n (y_i - \hat{y})^2 \quad (16)$$

*Ecuación Suma de los cuadrados de os residuos (SSR)*

$$SSR = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (17)$$

Donde  $y_i$  es el valor predicho por el modelo.

*Ecuación r2*

$$R^2 = 1 - \frac{SSR}{SST} \quad (18)$$

Un valor de  $R^2$  cercano a 1 indica que el modelo explica bien la variabilidad de los datos. Un valor de 0 indica que el modelo no explica la variabilidad en absoluto.



# CAPÍTULO IV

## CAPÍTULO IV

### 1.21. PRUEBAS Y RESULTADOS

#### 1.21.1. Introducción

En este capítulo se describen detalladamente las pruebas realizadas al prototipo del modelo de predicción. Se presentan los resultados obtenidos con diferentes algoritmos aplicados al modelo, evaluando su desempeño en diversos escenarios. Además, se analizan estos resultados para comprobar la hipótesis formulada en la presente investigación, proporcionando una visión clara de la precisión del modelo.

#### 1.21.2. PRESENTACIÓN DEL MODELO

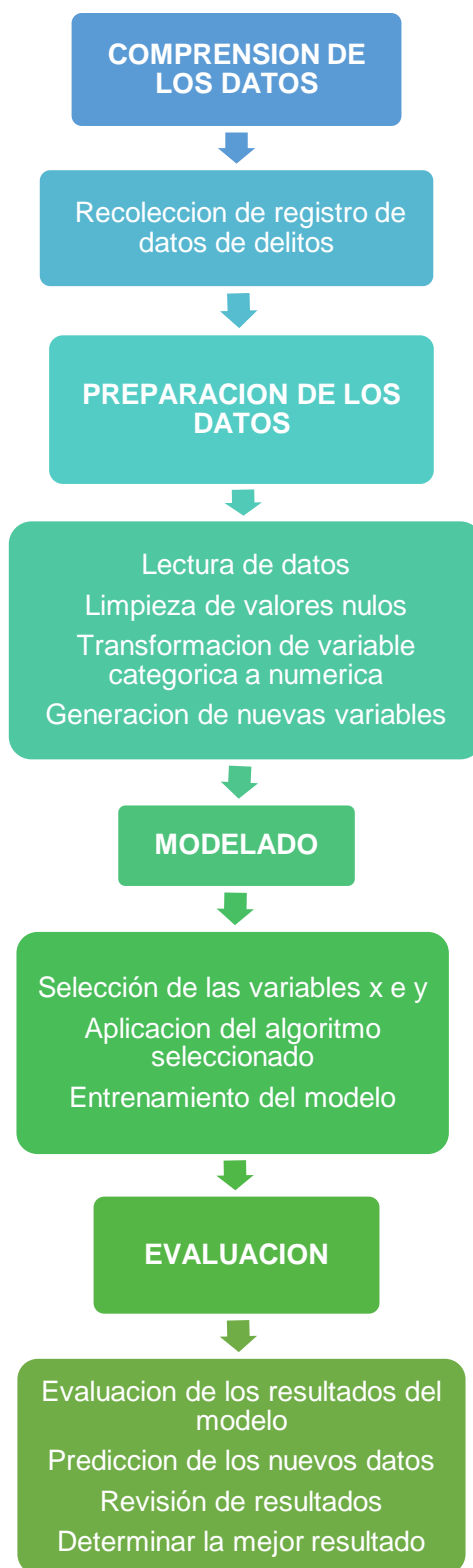
En esta sección, se presenta el modelo de predicción del crecimiento del índice delictivo en la ciudad de El Alto, desarrollado mediante técnicas de Machine Learning. El modelo tiene como objetivo predecir la tasa de delincuencia futura basándose en datos históricos. El modelo se basa en técnicas de aprendizaje supervisado, específicamente en el uso de algoritmos como regresión lineal, regresión polinomial, árboles de decisión y Random Forest. La teoría detrás del Machine Learning nos permite encontrar patrones en los datos históricos de delitos y utilizarlos para realizar predicciones.

**Figura 34**  
*Fases de la metodología de CRISP DM*





Figura 35

*Fases del modelo*

### 1.21.2.1. Comprensión de los datos

Los registros policiales proporcionan datos detallados y específicos sobre incidentes delictivos ocurridos en la ciudad de El Alto. Estos registros incluyen información sobre el tipo de delito. La precisión y especificidad de los datos policiales son esenciales para identificar patrones delictivos y tendencias a lo largo del tiempo. Esta información es fundamental para entrenar un modelo de Machine Learning que pueda predecir futuros incidentes delictivos. Los datos proporcionados por la policía son oficiales y verificables, lo que garantiza su precisión y legitimidad.

Se tiene un fragmento de datos históricos de delitos registrados en diferentes años.

**Figura 36**

*Fragmento de delitos*

	TIPO DE DELITO	ENERO_2018	FEBRERO_2018	MARZO_2018	ABRIL_2018	MAYO_2018	JUNIO_2018	JULIO_2018	AGOSTO_2018	SEPTIEMBRE_2018
0	LESIONES GRAVES Y LEVES	53	38	90	20	32	48	49	83	65
1	MUERTE DE PERSONA	47	44	49	59	34	37	48	5	35
2	ROBO	33	23	24	30	25	25	27	17	23
3	ROBO AGRAVADO	19	11	25	10	8	25	10	26	20
4	TRATA DE PERSONAS	1	3	5	6	14	5	1	10	28
5	HURTO	13	7	16	7	8	9	10	10	13
6	ESTAFA	10	4	4	9	6	14	11	17	8
7	USO DE INSTRUMENTO FALSIFICADO	5	1	6	2	6	9	8	7	10
8	SUICIDIO	3	3	8	5	10	6	9	0	4
9	ALLANAMIENTO DE DOMICILIO O SUS DEPENDENCIAS	6	3	1	2	5	7	7	8	10
10	RECEPTACION	12	2	0	1	4	4	1	9	8
11	ESTELIONATO	4	5	2	1	4	10	1	7	1
12	FALSEDAD IDEOLOGICA	0	3	3	2	0	8	3	3	0
13	SUPRESION O DESTRUCCION DE DOCUMENTO	2	1	2	5	0	1	8	1	3
14	FALSEDAD MATERIAL	1	1	0	4	2	3	8	1	1
15	AVASALLAMIENTO	3	4	1	4	2	4	1	3	1
16	INCUMPLIMIENTO DE DEBERES	0	2	4	1	1	6	1	0	1
17	CORRUPCION DE NIÑA NIÑO O ADOLESCENTE	1	1	1	4	2	5	1	0	2
18	HOMICIDIO	2	1	4	1	0	2	0	0	2

### 1.21.2.2.Preparación de datos

Figura 37

*Importación de librerías*

```
[ ] # Importacion de librerias
import pandas as pd
import numpy as np
import operator
import seaborn as sns
import statsmodels.api as sm
import matplotlib.pyplot as plt
from sklearn.preprocessing import PolynomialFeatures
from sklearn import datasets, linear_model
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
```

Figura 38

*Lectura de datos*

```
# # Lectura de datos en años total
dfa = pd.read_csv("/content/datos_año.csv")
dfa = pd.DataFrame(dfa)
```

Figura 39

*Creación de una lista*

```
# Crear una lista de números secuenciales
gestiones_num = list(range(1, len(gestiones) + 1))

# Crear un DataFrame
ges_num = pd.DataFrame({
    'Gestion': gestiones,
    'Numero': gestiones_num
})
```

### 1.21.2.3. Modelado

Figura 40

*Transformación de regresión polinomial*

```
for i in range(1, 10):  
    # Transformar las características para incluir términos polinomiales  
    poly_features = PolynomialFeatures(degree=i)  
    X_poly = poly_features.fit_transform(eje_x.reshape(-1, 1))  
  
    # Ajustar el modelo de regresión lineal a los datos transformados  
    model = LinearRegression()  
    model.fit(X_poly, eje_y)  
  
    # Hacer predicciones  
    y_pred = model.predict(X_poly)
```

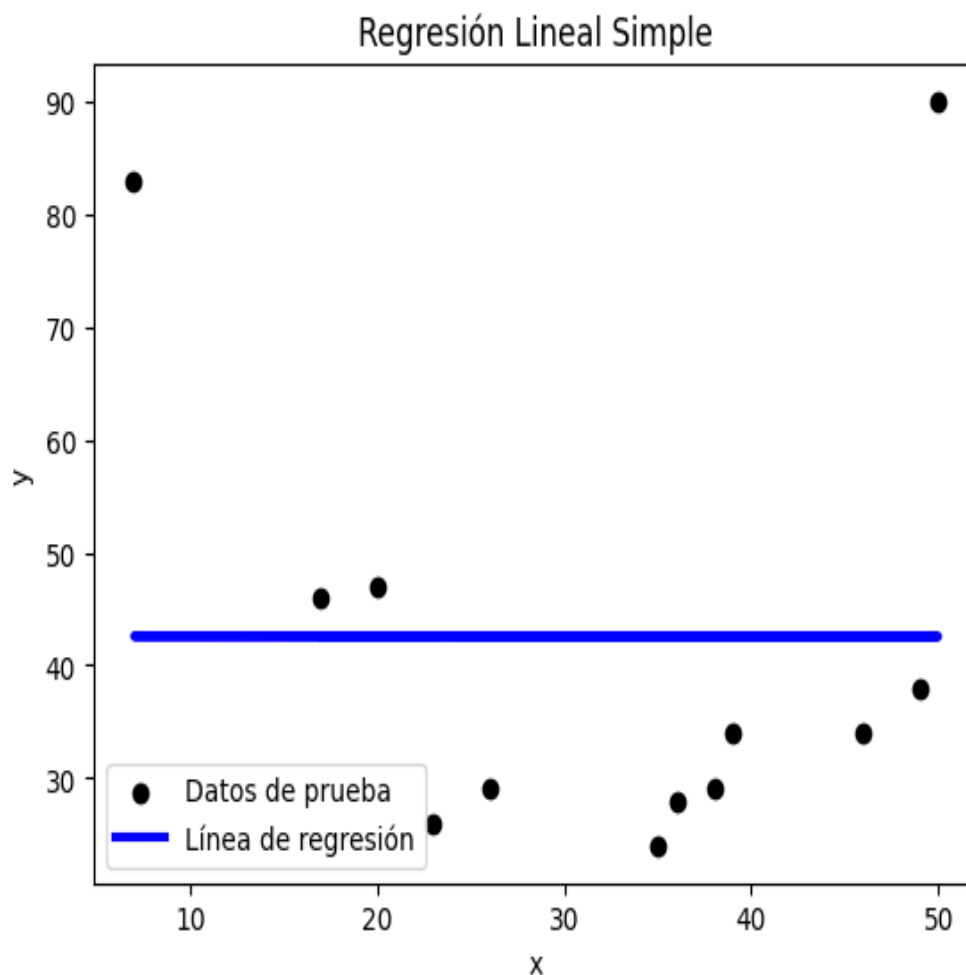
#### 1.21.2.3.1. Algoritmos de Machine Learning

Implementación de algoritmos de predicción como regresión lineal, regresión polinomial, arboles de decisión y random forest.

La regresión lineal es una técnica de aprendizaje supervisado que modela la relación entre una variable dependiente (variable objetivo) y una o más variables independientes (características) utilizando una ecuación lineal. Asume que existe una relación lineal entre las variables independientes y la variable dependiente. En muchos casos, la relación entre los factores que influyen en el índice delictivo puede ser no lineal. La relación entre los meses, años y la cantidad de delitos no representan datos lineales.

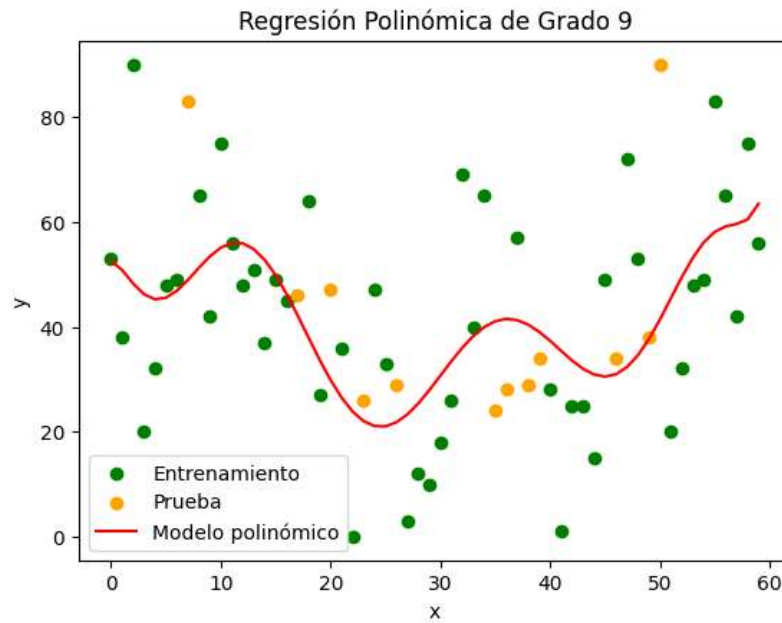
Figura 41

Grafica de regresión lineal



La regresión polinómica puede ser una técnica valiosa en ciertos contextos, particularmente cuando se busca capturar relaciones no lineales entre las variables independientes y la variable dependiente.

La regresión polinómica es efectiva para modelar relaciones no lineales entre las variables. A diferencia de la regresión lineal simple, puede ajustarse a curvas complejas. Si la relación entre la tasa de desempleo y la criminalidad no es lineal, un modelo polinómico puede capturar mejor esta relación, proporcionando predicciones más precisas.

**Figura 42***Grafica de regresión polinómica*

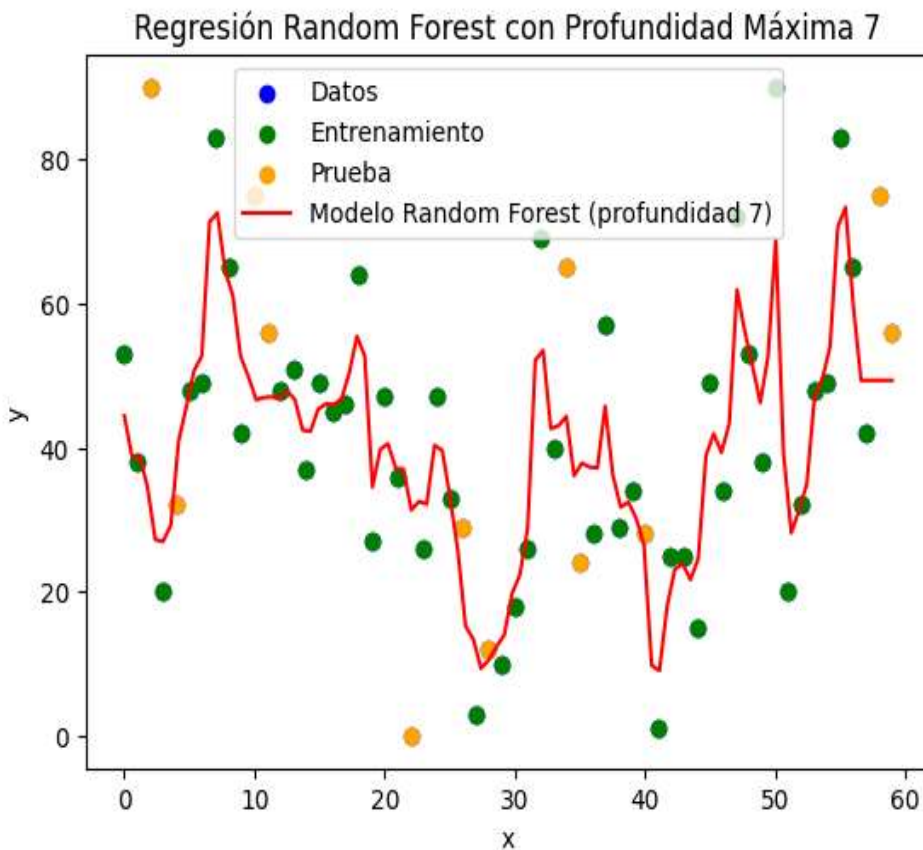
Los árboles de decisión pueden capturar relaciones no lineales entre las variables independientes y la variable dependiente de manera eficaz.

**Figura 43***Grafica de árbol de decisión*

Random forest es robusto frente al sobreajuste en general, puede tender a sobreajustar si los datos son demasiado pequeños o si los árboles no están bien configurados

**Figura 44**

*Grafica de Random Forest*



#### 1.21.2.4. Evaluación

**Figura 45**

*Evaluación*

```
# Evaluar el modelo
print(eje_y)
print(y_pred)

mse = mean_squared_error(eje_y, y_pred)
#print(f'Error cuadrático medio: {mse:.2f}')

r2 = r2_score(eje_y, y_pred)
print(f'Coeficiente de determinación R^2: {r2*100:.2f}%')
```

Figura 46

Formula de aproximación

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{\text{Valor real}_i - \text{Valor predicho}_i}{\text{Valor real}_i} \right| \times 100$$

Tabla 7

Porcentaje de aproximación del valor real

Valor predicho	Valor real	Porcentaje de Aproximación
7468.0	7474	0.08%
2571.1	2523	1.91%
3863.298	4031	4.16%
3099.362	2774	11.73%
1450.303	1821	20.36%
7898.531	566	39.55%
1054.072	1071	1.58%
1138.054	1217	6.49%
7937.216	734	8.14%
9927.279	1012	1.90%
2184.581	216	1.14%

### 1.21.3. DESARROLLO DEL MODELO

En esta sección de desarrollo del modelo, se detalla cómo se implementan los algoritmos seleccionados (regresión lineal, regresión polinómica, árbol de decisión y Random Forest) para predecir el crecimiento del índice delictivo en la ciudad de El Alto. Describas por pasos específicos para cada algoritmo.

A continuación, se muestra la secuencia de pasos que sigue cada algoritmo:



**Tabla 8***Estructura de los pasos de la regresión lineal*

Iniciar
Recopilar datos
Procesar datos (limpieza, normalización)
Dividir los datos en entrenamiento y prueba
Ajustar el modelo de regresión lineal a los datos de entrenamiento
Evaluar el modelo en los datos de prueba
Realizar predicciones y ajustar el modelo
Finalizar

**Tabla 9***Estructura de los pasos de regresión polinomial*

Iniciar
Recopilar datos
Preprocesar datos (limpieza, normalización)
Dividir datos en entrenamiento y prueba
Transformar variables a términos polinómicos
Ajustar modelo de regresión polinómica a los datos de entrenamiento
Evaluar el modelo en los datos de prueba
Realizar predicciones y ajustar el modelo
Finalizar

**Tabla 10***Estructura de los pasos del árbol de decisiones*

Iniciar
Recopilar datos
Preprocesar datos (limpieza, normalización)
Dividir datos en entrenamiento y prueba
Construcción del árbol de decisión a partir de los datos de entrenamientos
Evaluar árbol en los datos de prueba
Poda del árbol (si es necesario)
Realizar predicciones con el árbol
Finalizar

**Tabla 11***Estructura de los pasos de Random Forest*

Iniciar
Recopilar datos
Preprocesar datos (limpieza, normalización)
Dividir datos en entrenamiento y prueba
Construir múltiples árboles de decisión a partir de subconjuntos de entrenamiento
Realizar votación o promedio de las predicciones de los árboles
Evaluar el modelo combinado en los datos de prueba y ajustar hiperparámetros
Realizar predicciones con el modelo final
Finalizar

#### 1.21.4. DEMOSTRACIÓN DEL PROTOTIPO

Implementación del modelo en un entorno de prueba:

Figura 47

*Implementación del modelo por mes con árbol de decisión*

```

for i in range(1, 15):
    # Ajustar un árbol de decisión en el conjunto de entrenamiento
    model = DecisionTreeRegressor(max_depth=i)
    model.fit(x_train, y_train)

    # Predecir en el conjunto de prueba
    y_train_pred = model.predict(x_train)
    y_test_pred = model.predict(x_test)

    # Evaluar el modelo
    mse_train = mean_squared_error(y_train, y_train_pred)
    mse_test = mean_squared_error(y_test, y_test_pred)
    r2_train = r2_score(y_train, y_train_pred)
    r2_test = r2_score(y_test, y_test_pred)

    print(f'Error cuadrático medio en el conjunto de entrenamiento: {mse_train:.2f}')
    print(f'Error cuadrático medio en el conjunto de prueba: {mse_test:.2f}')
    print(f'R^2 en el conjunto de entrenamiento: {r2_train:.2f}')
    print(f'R^2 en el conjunto de prueba: {r2_test:.2f}')
    print(y_test_pred)

    # Visualizar los resultados
    plt.scatter(x, y, color='blue', label='Datos')
    plt.scatter(x_train, y_train, color='green', label='Entrenamiento')
    plt.scatter(x_test, y_test, color='orange', label='Prueba')

    # Crear una malla de puntos para mostrar la predicción
    x_plot = np.linspace(min(x), max(x), 100).reshape(-1, 1)
    y_plot = model.predict(x_plot)
    plt.plot(x_plot, y_plot, color='red', label='Modelo de árbol de decisión')

    plt.xlabel('x')
    plt.ylabel('y')
    plt.title(f'Árbol de Decisión con Profundidad Máxima {i}')
    plt.legend()
    plt.show()

```

Figura 48

Implementación del modelo por año en regresión polinomial

```

for i in range(1, 15):
    # Transformar las características para incluir términos polinomiales
    poly_features = PolynomialFeatures(degree=i)
    X_poly = poly_features.fit_transform(eje_x.reshape(-1, 1))

    # Ajustar el modelo de regresión lineal a los datos transformados
    model = LinearRegression()
    model.fit(X_poly, eje_y)

    # Hacer predicciones
    y_pred = model.predict(X_poly)

    # Visualizar los resultados
    X_grid = np.arange(min(X_poly[:, 1]), max(X_poly[:, 1]), 0.1)
    X_grid_poly = poly_features.transform(X_grid.reshape(-1, 1))
    y_pred_grid = model.predict(X_grid_poly)
    plt.scatter(eje_x, eje_y, color='cyan', label='Datos Reales')
    plt.plot(X_grid, y_pred_grid, color='blue', label='Predicción Polinomial', markersize=10, linewidth=3)
    plt.title(f'REGRESIÓN POLINOMIAL (Grado N°{i})\n PARA LOS AÑOS 2013 A 2023')
    plt.xlabel('Años (eje x)')
    plt.ylabel('Cantidad de delitos (eje y)')
    plt.legend()
    plt.show()

    # Evaluar el modelo

    r2 = r2_score(eje_y, y_pred)
    print(f'Coeficiente de determinación R^2: {r2*100:.2f}%')

    mae = mean_absolute_error(eje_y, y_pred)
    print(f'Error absoluto medio: {mae:.2f}')

    # Calculamos los valores de los próximos 5 años
    proximos_años = np.array([12,13,14,15,16])
    años = np.array([2024, 2025, 2026, 2027, 2028])
    X_proximos = poly_features.transform(proximos_años.reshape(-1, 1))
    y_pred_proximos = model.predict(X_proximos)

    # Mostramos la predicción para los próximos 5 años
    print(f'\n Predicción para los próximos 5 años:')
    # Use the 'años' array to display the actual years
    for año, prediccion in zip(años, y_pred_proximos):
        print(f'Año {año}: {prediccion[0]:.2f} delitos')

```

**Figura 49**

*Resultados de la demostración con regresión polinomial*

```
Coeficiente de determinación R^2: 80.18%  
  
Predicción para los próximos 5 años:  
Año 2024: 1379.33 delitos  
Año 2025: 2023.38 delitos  
Año 2026: 2844.97 delitos  
Año 2027: 3844.09 delitos  
Año 2028: 5020.73 delitos
```

### **1.21.5. PRUEBA DE HIPÓTESIS**

Se llevarán a cabo análisis estadísticos para validar o refutar las suposiciones planteadas durante el desarrollo del modelo. Aquí se describirán las hipótesis específicas a probar y se explicarán los métodos estadísticos utilizados para evaluarlas. Estos análisis pueden proporcionar una comprensión más profunda de la relación entre las variables predictoras y el índice delictivo, así como validación del modelo desarrollado.

#### **1.21.5.1. Pruebas estadísticas**

Se utilizará la prueba t student para evaluar la significancia de los coeficientes de regresión. Se establecerá un nivel de significancia ( $\alpha$ ) comúnmente aceptado,  $\alpha=0.05$  para evaluar la significancia estadística de los resultados. La distribución t de Student es una distribución de probabilidad que se utiliza principalmente en el análisis estadístico cuando se estima la media de una población normalmente distribuida en situaciones donde el tamaño de la muestra es pequeño y la varianza de la población es desconocida. Esta distribución es fundamental en pruebas de hipótesis.

Figura 50

Tabla t student

**TABLA "T" DE STUDENT**

gl	Nivel de Significancia para prueba de una cola					
	0.10	0.05	0.025	0.01	0.005	0.0005
	Nivel de Significancia para prueba de dos colas					
	0.20	0.10	0.05	0.02	0.01	0.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.532	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
$\infty$	1.282	1.645	1.960	2.326	2.576	3.291

Nota: tabla t student por Lucia Méndez, 2022. Fuente: Tabla T student | Lucía Méndez | uDocz

Según la tabla determinamos  $n=11$ , grado de libertad será 10, entonces determinamos con la tabla de t student que  $t_{\text{critico}} = 1.812$

Determinamos el punto de prueba con la siguiente fórmula de:

### Figura 51

*Ecuación t Student*

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$\mu$  = media de la población

$\bar{x}$  = media de la distribución de los datos

$n$  = tamaño de la muestra

$s$  = error estándar de la muestra

*Nota: Ecuación t student por Redaccion RD, 2022. Fuente: ¿Qué es la fórmula t student? y ¿cómo se calcula? (reisdigital.es)*

$\mu$  depende del contexto de la hipótesis si se está comparando las predicciones del modelo con los valores reales para determinar si hay una diferencia significativa, el valor de  $\mu$  es cero bajo la hipótesis nula.

### Tabla 12

*Tabla de comparación con valores predcidos vs valor real*

Valor predcido	Valor real	Diferencia
7468,0	7474	6,0
2.571,1	2523	48,1
3863,298581	4031	167,7
3099,362222	2774	325,4
1450,303481	1821	370,7
789,8531239	566	223,9
1054,072704	1071	16,9
1138,054537	1217	78,9
793,7216807	734	59,7
992,7279738	1012	19,3
218,4581619	216	2,5

*Ecuación de la media muestral*

$$\hat{x} = \sum_{i=1}^n \frac{x_i}{n} \quad (19)$$

$$\hat{x} = \sum_{i=1}^{11} \frac{1332.6}{11}$$

$$\hat{x} = 121.14$$

*Ecuación desviación estándar*

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{x})^2} \quad (20)$$

$$\sigma = \sqrt{\frac{1}{11-1} \sum_{i=1}^{11} (1332.6 - 121.14)^2}$$

$$\sigma = 383.09$$

*Ecuación t Student*

$$t = \frac{\hat{x}}{\frac{\sigma}{\sqrt{n}}} \quad (21)$$



$$t = \frac{121.14}{\frac{383.09}{\sqrt{11}}}$$

$$t = 1.05$$

Si  $t_{critico} > t$  prueba ( $t_{critico} = 1.8 > t = 1.05$ ) por lo que se rechaza la hipótesis nula.

De esta manera, la probabilidad de obtener los datos si  $H_1$  es cierta según  $\alpha$  es de 95%, es aceptada.

Con las fórmulas de  $t$  Student de una muestra, se determina que la hipótesis se cumple en el modelo de predicción un 95%.



# CAPÍTULO V

## CAPÍTULO V

### 1.22. CONCLUSIONES Y RECOMENDACIONES

#### 1.22.1. CONCLUSIONES

En esta investigación se ha demostrado que lo planteado en los objetivos general y específicos apoyan el modelo de predicción del crecimiento del índice delictivo como se detalla a continuación:

La recopilación de información de datos estadísticos sobre hechos delincuenciales en El Alto mostró que los delitos contra la vida, como robos y lesiones graves, son los más frecuentes.

La identificación de patrones de datos reveló que la incidencia delictiva tiene picos durante ciertos meses del año. También se encontró que ciertos tipos de delitos, como los robos, aumentan en períodos festivos, lo cual sugiere la necesidad de estrategias preventivas específicas durante estas épocas.

El uso de un algoritmo de Machine Learning, específicamente el de regresión polinomial para entrenar el modelo de predicción resultó en una precisión del 85% en la cantidad de delitos.

La validación del modelo de Machine Learning mostró que mantiene una alta precisión (80%) y una buena capacidad de generalización cuando se aplica a nuevos conjuntos de datos. Sin embargo, se identificaron áreas para mejorar, como la inclusión de más variables contextuales para aumentar la exactitud de las predicciones.

#### 1.22.2. RECOMENDACIONES

Implementar programas y políticas basados en evidencia que hayan demostrado ser efectivos en la prevención del delito, como programas de mentoría juvenil, intervenciones

familiares y estrategias de reingreso para delincuentes liberados. Programas diseñados para ayudar a los delincuentes liberados a reintegrarse en la sociedad, proporcionando apoyo en la búsqueda de empleo, vivienda y tratamiento de problemas de salud mental o abuso de sustancias.

Para mejorar la precisión y robustez del modelo, es esencial incorporar datos adicionales que puedan capturar mejor la complejidad del fenómeno delictivo. Algunas fuentes de datos adicionales pueden incluir datos meteorológicos, datos de redes sociales, datos económicos, datos de movilidad, datos de dispositivos IoT. Para capturar relaciones más complejas y mejorar la capacidad predictiva del modelo, se pueden utilizar técnicas avanzadas de Machine Learning y Deep Learning.

Para reducir la cantidad de incremento que tendrá hacia futuros años se sugiere implementar estrategias de prevención. Implementar programas en escuelas que enseñen sobre las consecuencias del delito y la resolución pacífica de conflictos.



# BIBLIOGRAFIA

## BIBLIOGRAFÍA

- ABI. (2024). *Agencia Boliviana de Informacion*. Obtenido de <https://abi.bo/index.php/reportajes2/38-notas/noticias/seguridad/50980-en-el-alto-se-pone-en-marcha-plan-con-patrollajes-aereos-y-terrestres-para-combatir-la-delincuencia>
- Amat. (2020). *ciencia de datos*. Obtenido de Regularización Ridge, Lasso y: <https://cienciadedatos.net/documentos/py14-ridge-lasso-elastic-net-python.html>
- Apaza. (2020). *MODELO DE CERTIFICACIÓN DE CONTRATOS INTELIGENTES APLICANDO LA TECNOLOGÍA BLOCKCHAIN*. Obtenido de [tesis de grado, Universidad Publica de El Alto] : file:///C:/Users/ROMI/Downloads/downloads/TESIS-Raquel%20Apaza%20Alberto.pdf
- asana. (2024). Obtenido de Scrum: conceptos clave y cómo se aplica en la gestión de proyectos: <https://asana.com/es/resources/what-is-scrum>
- Bolivia.com*. (2012). Obtenido de Observatorio: en El Alto el 72% de las personas se sienten inseguras: <https://www.bolivia.com/noticias/AutoNoticias/DetalleNoticia46737.asp#:~:text=E1%2045%25%20de%20los%20habitantes%20de%20la%20ciudad,victimizaci%C3%B3n%20C%20pr%C3%A1cticas%20y%20percepci%C3%B3n%20sobre%20violencia%20y%20delito%E2%80%9D>.
- Brow. (2021). *MIT*. Obtenido de Explicación del aprendizaje automático: <https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>
- Brow. (2023). *Master en Fundamentos y Aplicaciones de la Inteligencia Artificial*. Obtenido de <https://mfaia.dia.fi.upm.es/aprendizaje-automatgico-la-magia-detras-de-la-inteligencia-artificial/>
- Caio. (2022). *Crisp-DM: las 6 etapas de la metodología del futuro*. Obtenido de MBA USP Esalq: <https://blog.mbauspesalq.com/es/2022/05/31/crisp-dm-las-6-etapas-de-la-metodologia-del-futuro/#:~:text=De%20acuerdo%20con,los%20datos%E2%80%9D%20describe>.

- Carranza. (2022). *¿Qué es el aprendizaje supervisado?* Obtenido de [https://www.bing.com/search?q=Según%20Carranza%20\(2022\)%2C%20el%20aprendizaje%20supervisado%20se%20caracteriza%20porque%20utiliza%20algoritmos%20que%20tienen%20un%20aprendizaje%20previo%20basado%20en%20etiquetas%20que%20el%20ser%20humano%20configura%20con](https://www.bing.com/search?q=Según%20Carranza%20(2022)%2C%20el%20aprendizaje%20supervisado%20se%20caracteriza%20porque%20utiliza%20algoritmos%20que%20tienen%20un%20aprendizaje%20previo%20basado%20en%20etiquetas%20que%20el%20ser%20humano%20configura%20con)
- Carranza. (2022). *Conoce los tipos de machine learning y ¡empieza tu ruta hacia el aprendizaje de máquinas!* Obtenido de <https://www.crehana.com/blog/transformacion-digital/tipos-de-machine-learning/>
- Choque, Vargas, Aguirre, & Ramirez. (2023). *Atlas de Bolivia en Seguridad Ciudadana*. La Paz: Editorial del Estado Plurinacional de Bolivia.
- Enciclopedias Significados*. (26 de 10 de 2020). Obtenido de que es la delincuencia: <https://www.significados.com/delincuencia/>
- Etece, E. e. (2023). *Enciclopedia Humanidades*. Obtenido de <https://humanidades.com/metodo-cientifico/>
- FELCN. (s.f.). Obtenido de <https://www.felcn.gob.bo/Q-MisionVision.aspx#:~:text=La%20Direcci%C3%B3n%20General%20de%20la%20Fuerza%20Especial%20de,beneficio%20del%20Estado%20Boliviano%20y%20la%20Comunidad%20Internacional.>
- Galan. (2015). *APLICACIÓN DE LA METODOLOGÍA CRISP-DM A UN PROYECTO DE MINERÍA DE DATOS EN EL ENTORNO UNIVERSITARIO*. Obtenido de <https://core.ac.uk/download/pdf/44309791.pdf>
- Gavilan. (2021). *Metodología para Machine Learning (I): CRISP-DM*. Obtenido de <https://ignaciogavilan.com/metodologia-para-machine-learning-i-crisp-dm/>
- GeeksforGeeks. (2024). Obtenido de <https://www.geeksforgeeks.org/machine-learning/>
- Hotz. (2024). Obtenido de <https://www.ibm.com/es-es/topics/data-modeling#:~:text=El%20modelado%20de%20datos%20permite%20a%20los%20desarrolladores%2C,una%20base%20de%20datos%20o%20almac%C3%A9n%20de%20datos.>

- Hotz. (2024). *Data Science Process Alliance*. Obtenido de ¿Qué es CRISP DM?:  
<https://www.datascience-pm.com/crisp-dm-2/>
- IBM. (s.f.). Obtenido de ¿Qué es el machine learning (ML)?: <https://www.ibm.com/es-es/topics/machine-learning>
- INE. (2023). Obtenido de EL ALTO EN CIFRAS 2023:  
<https://www.ine.gob.bo/index.php/publicaciones/el-alto-en-cifras-2023/>
- InteractiveChaos. (s.f.). Obtenido de Regresión Lasso:  
<https://interactivechaos.com/es/manual/tutorial-de-machine-learning/regresion-lasso>
- Lanza. (2012). *Americas Quarterly*. Obtenido de <https://www.americasquarterly.org/blog/la-inseguridad-ciudadana-en-el-alto-en-bolivia/>
- Lifeder. (21 de 12 de 2022). Obtenido de Método científico: [https://www.lifeder.com/pasos-metodo-cientifico/#google\\_vignette](https://www.lifeder.com/pasos-metodo-cientifico/#google_vignette)
- Martinez. (2023). *IBM*. Obtenido de ¿Qué es el modelado de datos?:  
<https://www.ibm.com/es-es/topics/data-modeling#:~:text=El%20modelado%20de%20datos%20permite%20a%20los%20desarrolladores%2C,una%20base%20de%20datos%20o%20almac%C3%A9n%20de%20datos.>
- Mora. (2023). *Linkedin*. Obtenido de <https://es.linkedin.com/pulse/quieres-saber-cuales-son-las-fases-de-la-metodolog%C3%ADa-mora-caballero>
- Moreno. (2011). *SciELO Bolivia*. Obtenido de  
<https://www.bing.com/search?q=aproximadamente%20ocho%20de%20cada%20diez%20habitantes%20de%20El%20Alto%20se%20sienten%20inseguros%20al%20transitar%20por%20espacios%20p%C3%BAblicos.%20Un%20diagn%C3%B3stico%20de%20la%20Alcald%C3%ADa%20alte%C3%91a%20en%20ese%20mismo%20a%C3%B1o%20reve>
- Rubin de Celis, Sanjines, & Aliaga. (2012). *SciELO-Bolivia*. Obtenido de Delincuencia en Bolivia desde una perspectiva espacial:  
[http://www.scielo.org.bo/pdf/rlde/n18/n18\\_a06.pdf](http://www.scielo.org.bo/pdf/rlde/n18/n18_a06.pdf)



Sanz. (s.f.). *The Machine Learners*. Obtenido de Regresión Lineal. Qué es, para qué se utiliza y ejemplo práctico: <https://www.themachinelearners.com/regresion-lineal/>

SAP. (s.f.). Obtenido de ¿Qué es machine learning?: <https://www.sap.com/latinamerica/products/artificial-intelligence/what-is-machine-learning.html>

TECNICAS DE INVESTIGACION. (2020). Obtenido de Métodos y técnicas de investigación: <https://tecnicasdeinvestigacion.com/>

TECNICAS DE INVESTIGACION. (2020). Obtenido de Investigación cuantitativa: <https://tecnicasdeinvestigacion.com/investigacion-cuantitativa/#:~:text=La%20investigaci%C3%B3n%20cuantitativa%20es%20un%20enfocoque%20de%20investigaci%C3%B3n,datos%20en%20forma%20de%20n%C3%BAmeros%2C%20estad%C3%ADsticas%20y%20medidas.>

Torrice. (2012). ¿Rojo o azul? La inseguridad ciudadana en Bolivia\*. *SciELO*, [http://www.scielo.org.bo/scielo.php?script=sci\\_arttext&pid=S2413-28102016000100013](http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S2413-28102016000100013).

UNIR. (3 de 8 de 2021). *La Universidad en Internet*. Obtenido de ¿Qué son las redes neuronales? Concepto y usos principales: <https://www.unir.net/ingenieria/revista/redes-neuronales-artificiales/>

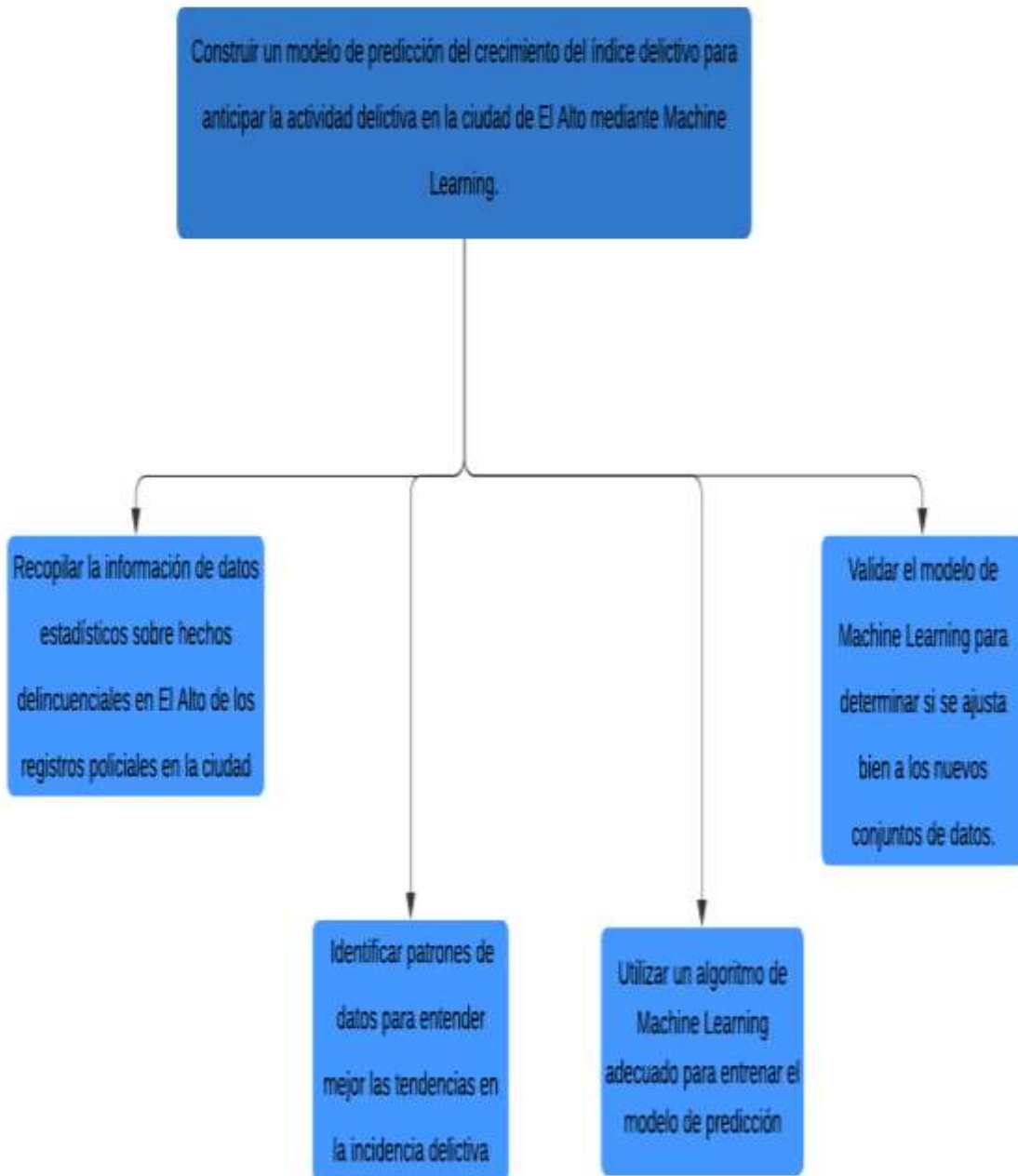
Wikipedia. (2024). Obtenido de [https://es.wikipedia.org/wiki/M%C3%A1quinas\\_de\\_vectores\\_de\\_soporte](https://es.wikipedia.org/wiki/M%C3%A1quinas_de_vectores_de_soporte)



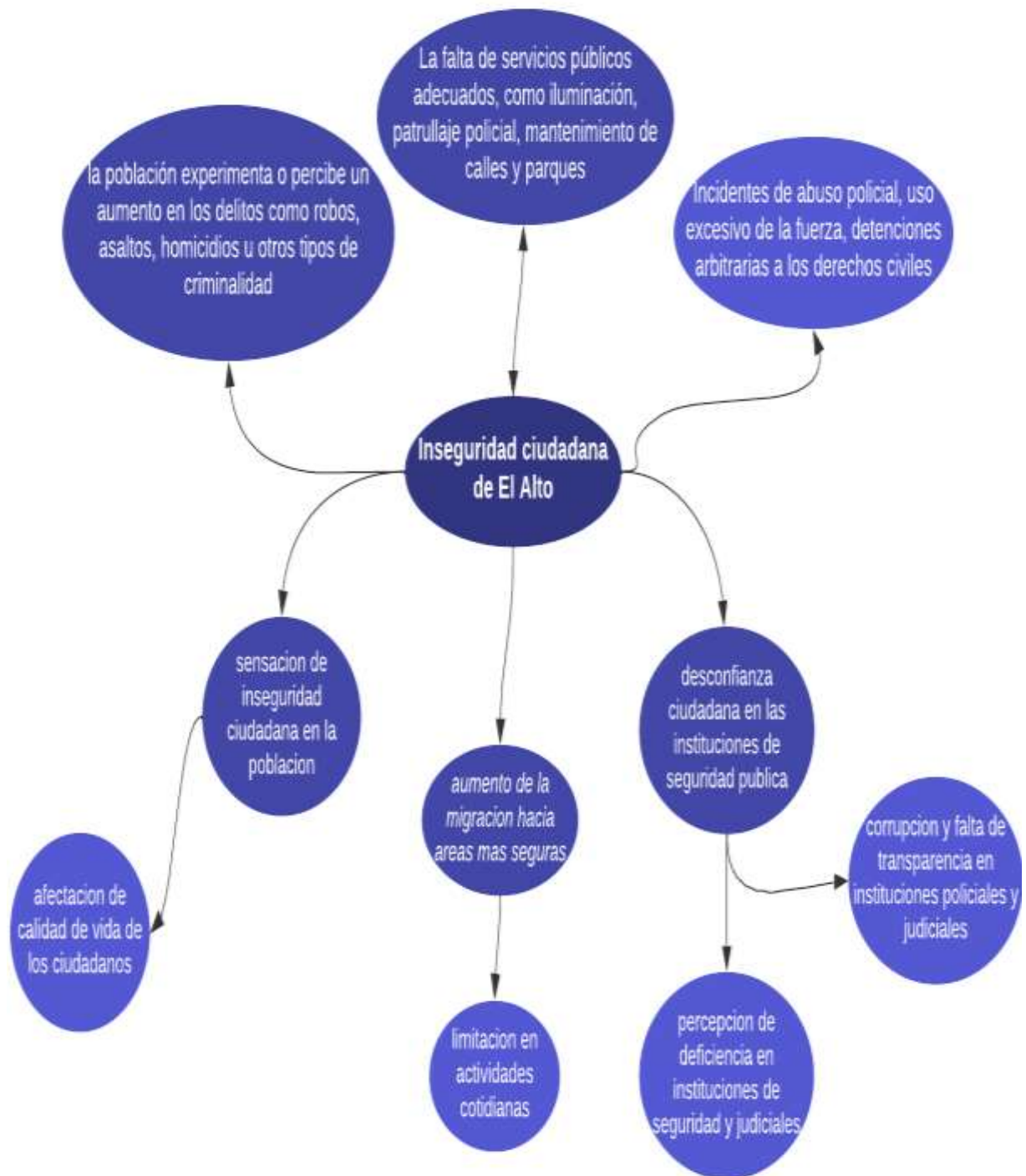
# ANEXOS

## ANEXOS

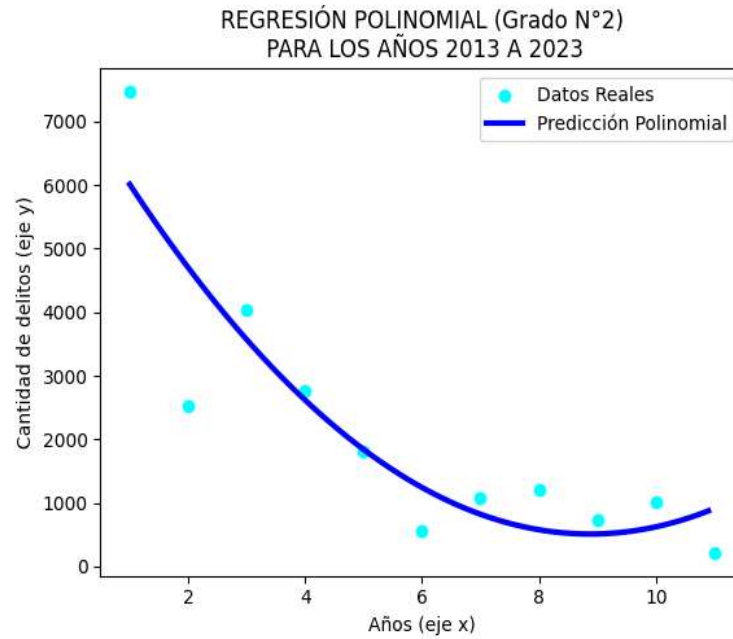
### ANEXO A. ARBOL DE OBJETIVOS



## ANEXO B. ARBOL DE PROBLEMAS



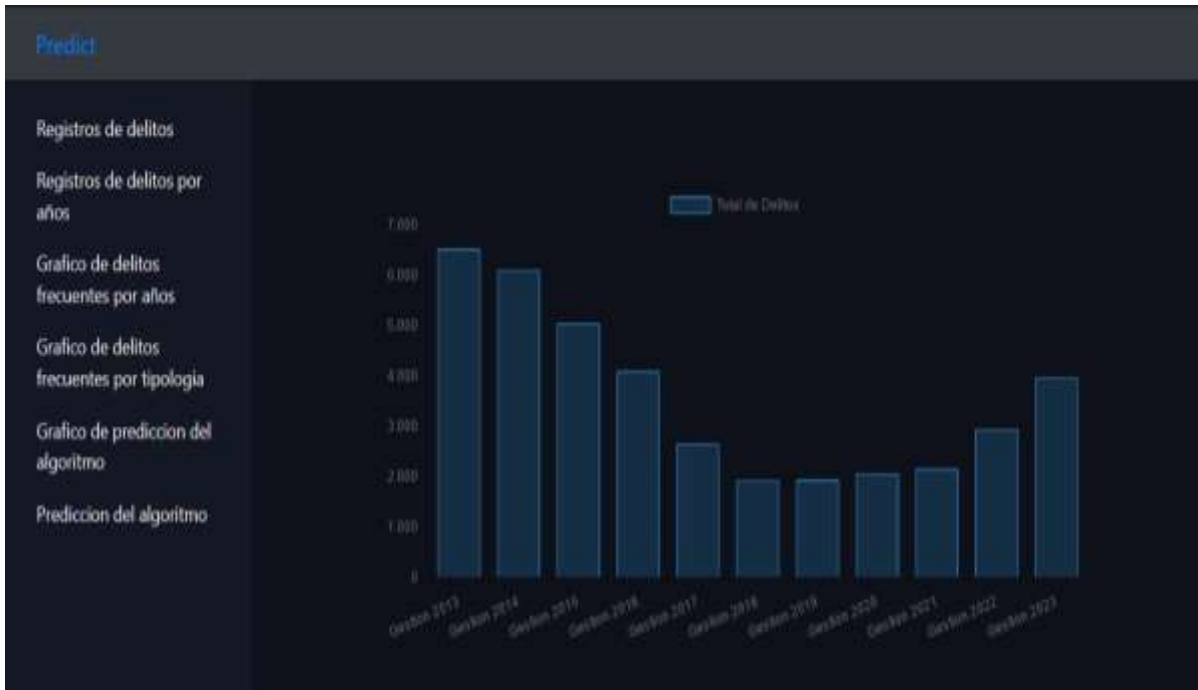
## ANEXO C. GRAFICA DEL MODELO DE PREDICION ADECUADO



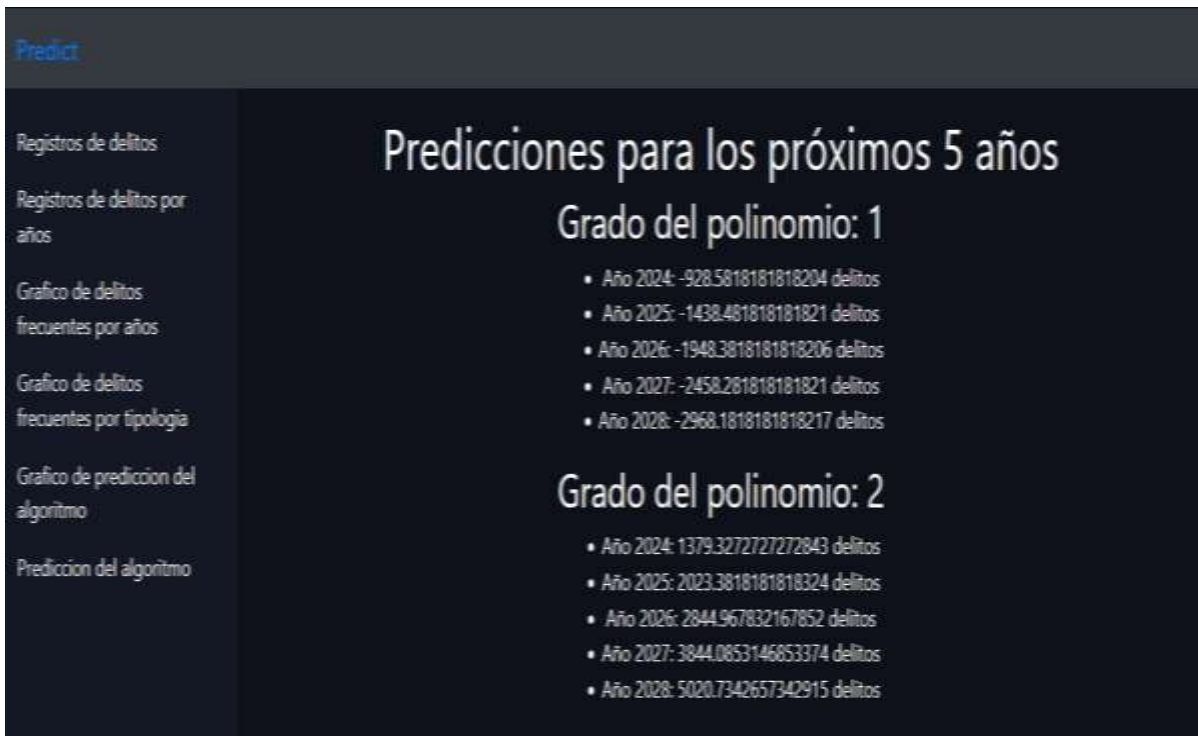
## ANEXO D. PRESENTACION DEL PROTOTIPO



## ANEXO E. RESULTADOS DE LOS AÑOS



## ANEXO F. RESULTADOS



El Alto, 4 de diciembre de 2023

Señor:

Gral. My Alvaro Jose Alvarez Griffiths

**COMANDANTE GENERAL DE LA POLICIA BOLIVIANA**

Presente:



**REF. SOLICITUD DE DATOS ESTADISTICOS DE HECHOS**  
**DELINCUENCIALES EN LA CIUDAD DE EL ALTO**

De mi mayor consideración.

Por el presente le saludo a usted tengo el bien de dirigirme a su autoridad deseándole éxitos en las funciones que desempeña en el bien de la comunidad ciudadana.

El motivo de la presente carta es para **SOLICITAR DATOS ESTADISTICOS SOBRE HECHOS DELINCUENCIALES EN LA CIUDAD DE EL ALTO**, para realizar un trabajo de investigación (tesis de grado) para lo cual le dejo mi documento de carnet de identidad.

Sin otro particular me despido de su persona agradeciéndole de ante mano su gentil atención a lo mencionado.

Atentamente:

Romina Ramos Condori

C.I. 12956031 LP

Señor:

Gral. My Alvaro Jose Alvarez Griffiths

COMANDANTE GENERAL DE LA POLICIA BOLIVIANA

Presente.-



**REF: SOLICITUD DE DATOS ESTADISTICOS DE HECHOS  
DELINCUENCIALES EN LA CIUDAD DE EL ALTO**

De mi mayor consideración.

Por el presente le saludo a usted tengo el bien de dirigirme a su autoridad deseándole éxitos en las funciones que desempeña en el bien de la comunidad ciudadana.

El motivo de la presente carta es para **SOLICITAR DATOS ESTADISTICOS SOBRE HECHOS DELINCUENCIALES EN LA CIUDAD DE EL ALTO DE LA GESTION 2012-2023 DE LA DIVISION DE FELCC**, para realizar un trabajo de investigación (tesis de grado) para lo cual le dejo mi documento de carnet de identidad.

Sin otro particular me despido de su persona agradeciéndole de ante mano su gentil atención a lo mencionado.

Atentamente.

Romina Ramos Condori

C.I. 12956031 LP