

**UNIVERSIDAD TECNOLÓGICA DE PANAMÁ**

FACULTAD DE INGENIERÍA DE SISTEMAS COMPUTACIONALES

FACULTAD DE INGENIERÍA INDUSTRIAL

**PROYECTO FINAL**

MODELOS PREDICTIVOS

**ESTUDIANTES:**

JAVIER GONZALEZ 7-710-1757

**FACILITADOR:**

JUAN CASTILLO

**AÑO:**

2024

INTRODUCCION

En las últimas décadas, el delito de robo ha emergido como una preocupación significativa en muchos países de América Latina, y Panamá no es la excepción. A pesar de los esfuerzos de las autoridades para combatir esta problemática, el robo sigue siendo uno de los delitos más comunes y perturbadores en el país, afectando tanto a individuos como a comunidades enteras. La incidencia de estos delitos no solo impacta la seguridad y el bienestar de los ciudadanos, sino que también tiene repercusiones en la economía y en la percepción de seguridad general.

Para abordar este problema de manera efectiva, es esencial contar con herramientas analíticas que permitan entender mejor los patrones y las tendencias asociadas con el delito de robo. En este contexto, el uso de modelos predictivos puede ofrecer una perspectiva valiosa sobre cómo se desarrollan estos delitos.

En este proyecto final, utilizaremos un dataset estadístico sobre robos en Panamá de la PGN para explorar diferentes modelos predictivos con el objetivo de identificar el modelo más eficaz para anticipar y comprender estos eventos delictivos.

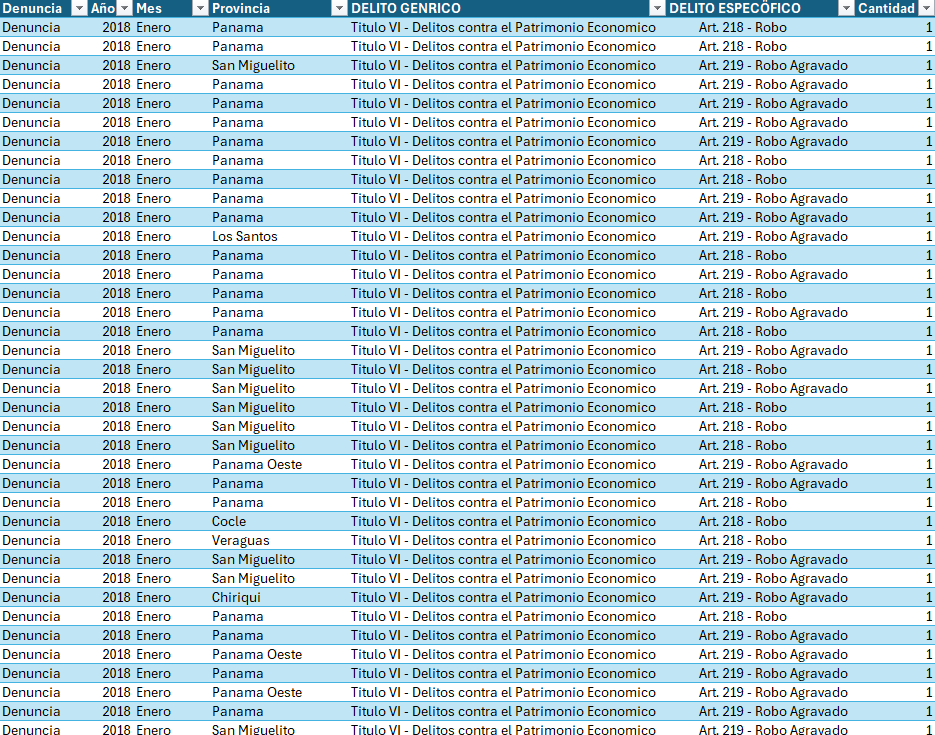
Los modelos predictivos que se evaluarán incluyen técnicas tanto tradicionales como avanzadas. A través de este análisis, buscaremos determinar cuál de estos modelos ofrece la mejor precisión en sus predicciones y también proporcionar los pronósticos realizados.

CONTENIDO

Dataset: Informe Estadístico de Delitos de Robo en Panamá

El dataset seleccionado para el proyecto final cuenta con 7 columnas detalladas de la siguiente manera:

* Denuncia: todos los registros son denuncias que fueron obtenidas o realizadas.
* Año: el periodo del dataset va desde 2018 hasta 2023
* Mes: cuenta con los doce meses del calendario.
* Provincia: muestra los registros de robo en todas las provincias de Panamá
* Delito Genérico: todas las filas corresponden a “Delitos contra el Patrimonio Económico”.
* Delito Especifico: los registros muestran dos tipos de delitos que son Robo y Robo Agravado
* Cantidad: es una columna que se agrego en la que cuenta los registros de robo para cada fila.



Antes de empezar con el análisis, primero se eliminaron las columnas que no iban a ser utilizadas para continuar, las cuales son:

* Denuncia
* Provincia
* Delito Genérico
* Delito Especifico

Las razones por las cuales no se tomaron en consideración es por que algunas no aportan un valor al análisis a realizar y otras como provincia y delito especifico, fueron eliminadas ya que el pronostico que se buscaba realizar tomaba como referencia todos los valores en general para no hacer ninguna segregación con respecto a los robos en cada provincia.

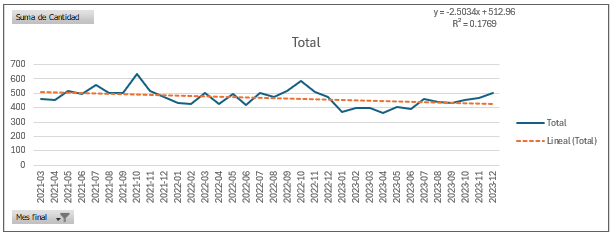
Grafica de tendencia

Como contamos con información que va desde 2018 hasta 2023, al graficar estos valores, nos damos cuenta de que existe una tendencia decreciente o un sesgo para los periodos de Febrero-2020 hasta Febrero-2021, una de las principales razones por las que puede existir es por la pandemia y que quizás para este periodo el delito que se daba era el ciberdelito.

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Se toma la decisión de obviar estos valores y empezar el análisis con los valores desde Marzo-2021.



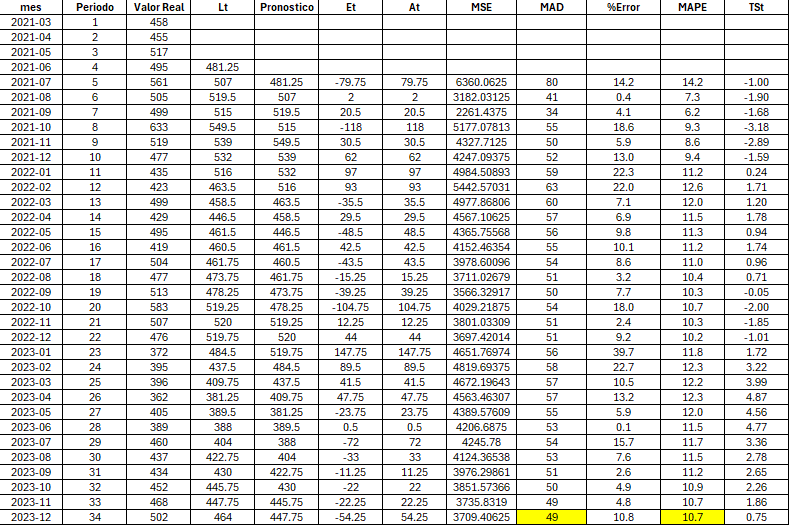
Al tomar los valores y graficarlos nuevamente, nos damos cuenta de que no existe sesgo y que podemos empezar a aplicar los modelos predictivos.

Modelos en Excel

Mediante la herramienta de Excel, se utilizaron tres modelos predictivos para el dataset:

* Modelo Promedio Movil
* Modelo Suavizamiento Exponencial
* Modelo Holt

Para el promedio Movil, los resultados obtenidos fueron los siguientes:



Para el Modelo de Suavizamiento Exponencial:

Tabla

Descripción generada automáticamente

Modelo Holt

Tabla

Descripción generada automáticamente

Modelo en Python

Al investigar sobre que otros modelos podían implementarse para un dataset con información de robos realizados en distintos periodos, el modelo ARIMA (AutoRegressive Integrated Moving Average) puede ser útil en ciertos contextos para analizar datos de series temporales.

ARIMA es un modelo utilizado para análisis y pronóstico de series temporales que combina componentes autoregresivos, de media móvil e integración (diferenciación). Es eficaz para predecir valores futuros basándose en patrones históricos en datos de series temporales.

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Imagen que contiene Diagrama

Descripción generada automáticamente Texto

Descripción generada automáticamente con confianza media

Imagen que contiene Diagrama

Descripción generada automáticamenteCon este modelo, se pronosticó para los proximos 6 periodos y se obtuvieron los valores presentados en las imágenes anteriores.

Adicional, tambiens se realizó el calculo de la media movil y la desviacion estandar movil:

Gráfico, Gráfico de líneas

Descripción generada automáticamente

También se hizo una decomposicion de serie temporal con el modelo especificado para graficar la tendencia, la estacionariedad y los residuos en una gráfica de cuatro subplots para mostrar la serie temporal original, representar la tendencia de la serie, mostrar la estacionariedad con la media y representar los residuos de la serie con su media.

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Se aplicó la prueba de Dickey-Fuller Aumentada y se volvieron a calcular la media móvil y la desviación estándar:

Gráfico, Gráfico de líneas

Descripción generada automáticamente Texto

Descripción generada automáticamente

Como resultados de la prueba de Dickey-Fuller Aumentada, el estadistico nos da -1.7693, este valor es mayor a todos los valores criticos que se obtuvieron para los distintos niveles de significancia. Esto indica que no podemos rechazar la hipotesis nula de que la serie tiene una raíz unitaria, el p-valor es 0.3958, que es mayor a los niveles de significancia y refuerza la conclusion de que no puede rechazarse la hipotesis nula.

Modelos en Weka

Weka es un software de código abierto para minería de datos y aprendizaje automático desarrollado en la Universidad de Waikato en Nueva Zelanda. Su nombre proviene de una especie de ave nativa de Nueva Zelanda, el weka, conocida por su curiosidad.

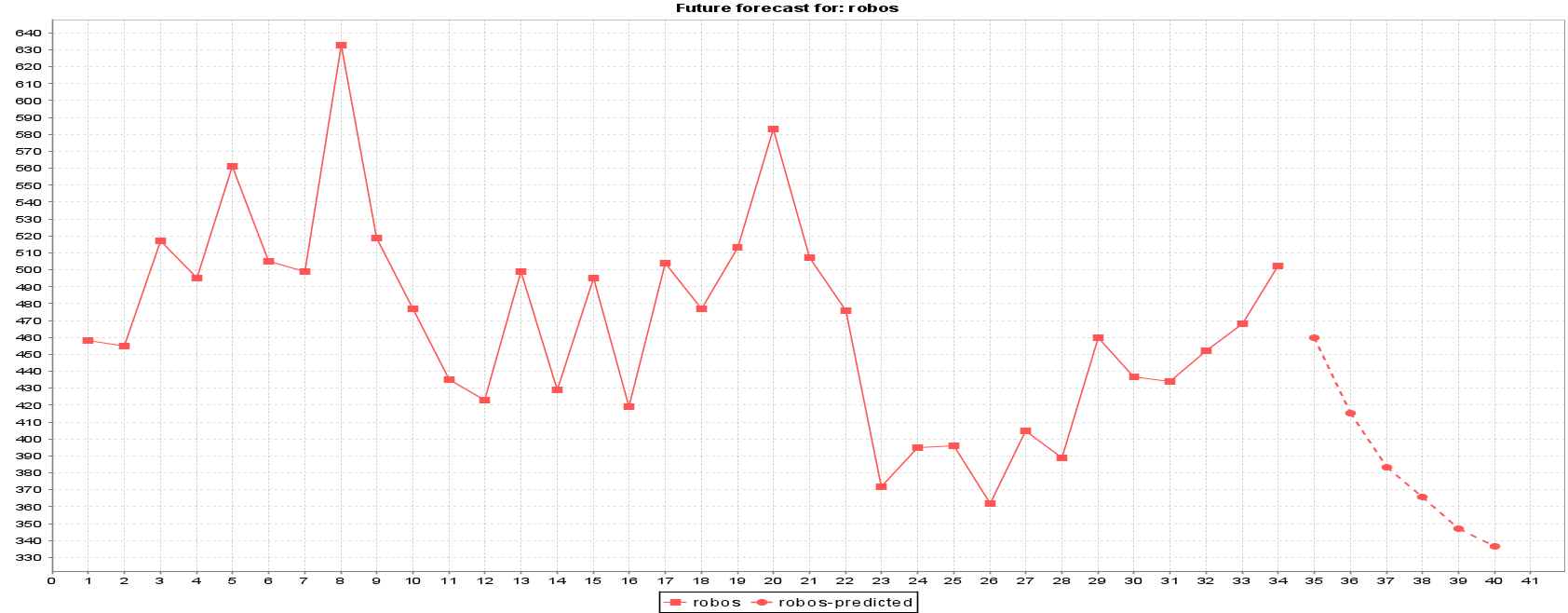
Weka es ampliamente utilizado para explorar conjuntos de datos y descubrir patrones o tendencias mediante técnicas de minería de datos.

Fue utilizado para predecir con tres modelos utilizando su función de forecast:

* Multilayer Perceptron
* SMOreg
* Random Forest

Los resultados para cada uno fueron los siguientes:

Multilayer Perceptron



Texto

Descripción generada automáticamente Tabla

Descripción generada automáticamente

Modelo SMOreg

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Texto

Descripción generada automáticamente Tabla

Descripción generada automáticamente

Random Forest

Gráfico, Gráfico de líneas

Descripción generada automáticamente

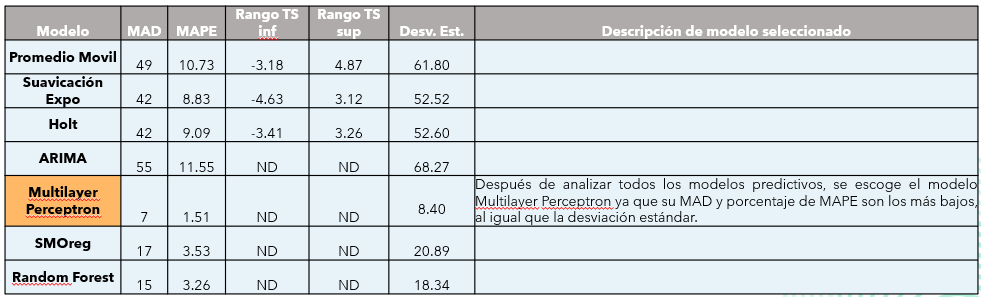
Texto

Descripción generada automáticamente Interfaz de usuario gráfica, Tabla

Descripción generada automáticamente

Resultados de los Modelos

Al terminar de correr todos los modelos se toman las variables de MAD, MAPE, Rango TS Inferios, Range TS Superior y las desviacion estandar para poder tomar una decision de cual modelo es el que mejor se ajusta al dataset escogido para pronosticar:



CONCLUSIONES

En el análisis realizado con el dataset de robos en Panamá, se ha optado por utilizar el modelo de Multilayer Perceptron (MLP) para la tarea de predicción. Este modelo, implementado en Weka, ha demostrado ser una herramienta efectiva para modelar y anticipar el número de robos en función de los datos históricos disponibles.

Los resultados del modelo MLP indican un desempeño sólido en términos de precisión predictiva. En particular, el Error Absoluto Medio (MAD) obtenido fue de 7, lo que sugiere que, en promedio, las predicciones del modelo se desvían en 7 unidades del valor real. Este nivel de error es relativamente bajo, lo cual es indicativo de una buena capacidad del modelo para capturar las tendencias y patrones en los datos.

Adicionalmente, el Error Porcentual Absoluto Medio (MAPE) fue de 1.51%, una cifra que refleja una alta precisión en las predicciones, con un margen de error porcentual bastante reducido. Este resultado es prometedor, ya que un MAPE bajo generalmente indica que el modelo realiza pronósticos bastante precisos y confiables.

La desviación estándar de 8.40 también refuerza la estabilidad del modelo, mostrando que la variabilidad en los errores de predicción no es excesiva y se mantiene en un rango manejable.

Para evaluar la capacidad del modelo MLP en la predicción a futuro, se realizó un pronóstico para los próximos 6 períodos. Los resultados mostraron una media de los valores pronosticados de 384. Esta estimación proporciona una visión útil para la planificación y la estrategia de prevención de robos, ya que permite anticipar el nivel de actividad delictiva en el corto plazo.

En resumen, el modelo Multilayer Perceptron ha demostrado ser una herramienta eficaz para la predicción del número de robos en Panamá, con resultados de precisión aceptables y errores relativamente bajos. Estos hallazgos sugieren que el MLP es una opción sólida para la modelización de datos temporales en este contexto, y que los resultados obtenidos pueden servir como base para la formulación de políticas y estrategias de prevención del delito.