**Escuela de Estudios de Postgrado**

**Facultad de Ingeniería**

**Universidad San Carlos de Guatemala**

**Maestría en Ingeniería para la Industria con**

**Especialización en Ciencias de la Computación.**

**Minería de Datos**

**Ing. Ms. Kevin Lajpop**

**PROYECTO 2**

**Análisis de Faltas Judiciales por Multas de Tránsito en Guatemala (2015)**

<https://github.com/obduliogonzalez/proyecto2_mineria_datos>

**Néstor Obdulio González López**

**Carné: 999013150**

Introducción

El presente proyecto tiene como objetivo el análisis de un conjunto de datos sobre faltas judiciales relacionadas con multas de tránsito en Guatemala durante el año 2015. Este análisis se realiza utilizando técnicas de minería de datos para identificar patrones, realizar predicciones y proponer estrategias de mejora en la gestión y prevención de infracciones de tránsito. La información extraída de este estudio podría ayudar a los responsables de la toma de decisiones a implementar políticas más efectivas en el área de seguridad vial.

Objetivo

El objetivo principal de este proyecto es aplicar técnicas de minería de datos para analizar el conjunto de datos de faltas judiciales por multas de tránsito, con el fin de realizar predicciones y proponer medidas que mejoren la gestión y control de infracciones de tránsito. Para ello, se emplearán modelos de aprendizaje automático, como los algoritmos de árboles de decisión y bosques aleatorios, para proporcionar una visión más clara y proactiva de la situación.

Descripción del Conjunto de Datos

El conjunto de datos utilizado en este análisis se refiere a las faltas judiciales por multas de tránsito registradas en Guatemala durante el año 2015. Los datos se tomaron de la base de datos oficial proporcionada por el Instituto Nacional de Estadística de Guatemala (INE), que está disponible en el siguiente enlace:

https://www.ine.gob.gt/bases-de-datos/faltas-judiciales/

El conjunto de datos incluye variables como:

Tipo de infracción: Clasificación de la falta (por ejemplo, exceso de velocidad, uso del celular, falta de documentación, etc.).

Fecha y hora de la infracción.

Ubicación geográfica: Departamento y municipio donde ocurrió la infracción.

Detalles del infractor: Edad, género, tipo de vehículo, entre otros.

Resultado judicial: Multa impuesta, pago realizado, proceso judicial, etc.

Antes de proceder al análisis, se llevó a cabo un proceso de limpieza de datos para eliminar registros incompletos o inconsistentes y estandarizar las variables necesarias.

Predicciones con Árboles de Decisión

4.1. Método y Algoritmo Utilizado

El algoritmo de Árboles de Decisión fue utilizado para realizar predicciones sobre la probabilidad de ciertos eventos relacionados con las infracciones de tránsito. Este modelo es fácil de interpretar y permite visualizar claramente cómo las diferentes variables afectan la toma de decisiones.

Situaciones Analizadas

Se identificaron cuatro situaciones distintas en las que se aplicó el modelo de árboles de decisión para predecir el resultado de las faltas judiciales por multas de tránsito:

Situación 1: Predicción de la probabilidad de que un infractor pague una multa al momento de ser sancionado, para un conductor masculino de 35 años que comete una infracción por exceso de velocidad en la ciudad de Guatemala.

Situación 2: Predicción de la probabilidad de que una multa por falta de documentación se convierta en un proceso judicial, para una mujer de 28 años en el departamento de Sacatepéquez.

Situación 3: Predicción de la probabilidad de que una infracción por uso del celular al conducir sea reportada en una zona rural, para un conductor de 40 años.

Situación 4: Predicción de la probabilidad de que un infractor reincida en la misma infracción durante el mismo año, para un hombre de 50 años que ha cometido más de tres infracciones previas.

Resultados y Gráficas

Los árboles de decisión generados fueron visualizados y analizados para interpretar la relevancia de cada variable en las predicciones. Las gráficas presentadas mostraron la estructura de las decisiones y las tasas de acierto para cada situación analizada.

Interpretación de los Resultados

Los resultados indicaron que ciertas variables, como la edad y el tipo de infracción, tienen un impacto significativo en la probabilidad de que un infractor pague la multa o enfrente un proceso judicial. Por ejemplo, se observó que los conductores jóvenes tienen mayor probabilidad de incumplir y de reincidir en infracciones, mientras que la probabilidad de pago es mayor en infractores de mediana edad.

Predicciones con Bosques Aleatorios

5.1. Método y Algoritmo Utilizado

El algoritmo de Bosques Aleatorios (Random Forest) fue utilizado para realizar un análisis más robusto y minimizar el riesgo de sobreajuste, proporcionando una comparación con los resultados obtenidos de los árboles de decisión individuales.

Situaciones Analizadas

Se profundizó en dos situaciones de predicción:

Situación 1: Predicción de la probabilidad de que una infracción por exceso de velocidad en una carretera de alta circulación sea pagada.

Situación 2: Predicción de la probabilidad de que una infracción por uso del celular en zonas urbanas se convierta en un proceso judicial.

Resultados y Gráficas

Se generaron gráficas y tablas que muestran la importancia de las variables y la precisión del modelo de bosques aleatorios. Comparado con los árboles de decisión, los bosques aleatorios proporcionaron una mayor precisión en las predicciones al integrar múltiples árboles de decisión y tomar el promedio de sus resultados.

Interpretación de los Resultados

Los resultados del análisis con bosques aleatorios mostraron que las variables más influyentes en las predicciones fueron el tipo de infracción y la ubicación. La combinación de múltiples árboles de decisión permitió una mejor identificación de patrones complejos y un mayor nivel de precisión en las predicciones.

Propuestas de Mejora y Conclusión

A partir del análisis realizado, se propone la implementación de medidas de control y prevención basadas en los hallazgos obtenidos. Las principales recomendaciones incluyen:

Campañas de concienciación para los grupos de edad con mayor incidencia de infracciones.

Mejoras en la supervisión en áreas de alto riesgo para reducir la frecuencia de infracciones graves.

Optimización de los procedimientos judiciales para agilizar el proceso de pago y minimizar los casos que se convierten en procesos judiciales.

La aplicación de técnicas de minería de datos y modelos de aprendizaje automático ha permitido un análisis profundo de las faltas judiciales por multas de tránsito en Guatemala, brindando un enfoque basado en datos para mejorar la seguridad vial y la gestión de infracciones en el país.