

Bitácora 1: Análisis exploratorio para proyecto de investigación.

Tema de interés

Fraudes en transacciones de créditos La identificación del fraude en las transacciones de tarjetas de crédito es un tema de relevancia, ya que puede ayudar a detectar potencial daño del uso no autorizado de recursos financieros tanto de las compañías crediticias como de los usuarios dueños de las tarjetas de crédito.

Criminalidad por regiones en Costa Rica El análisis de la criminalidad por regiones es de interés, ya que se pueden identificar distintos patrones utilizados por los delincuentes y cuáles áreas se ven más afectadas, de esta forma se podrían implementar nuevas políticas de seguridad que protejan los derechos humanos, asimismo que otorguen más tranquilidad y protección a los ciudadanos en su vida cotidiana.

Recursos encontrados

“Credit card fraud detection using naive Bayesian and C4.5 decision tree classifiers”:

En este ‘paper’, se explora la implementación de distintos modelos de ‘machine learning’ en la detección de fraudes de tarjetas de crédito. Uno de los modelos que explica es el Naive-Bayes, el cual utiliza probabilidades condicionales y el teorema de Bayes, desde la suposición de que las variables son independientes para dar una probabilidad de que la transacción sea fraudulenta.

“A supervised ‘machine learning’ algorithm for detecting and predicting fraud in credit card transactions”: Este artículo tiene como propósito evaluar tres modelos de ‘machine learning’ en la detección de fraudes en las transacciones de tarjetas de crédito. Los tres modelos que se prueban son: regresión logística, bosques aleatorios y árboles de decisión.

“KNN Model-Based Approach in Classification”: Este artículo profundiza en lo que consiste el algoritmo de k-Vecinos Más Cercanos (KNN); asimismo, propone un modelo que ayuda a mejorar la eficiencia y precisión de este algoritmo con tal de obtener una mejor clasificación de los datos.

Tema potencial

El tema de la investigación será implementar modelos de ‘machine learning’ para detectar fraude en tarjetas de crédito. Se plantea seguir la tesis *“Credit Card Fraud detection using ‘machine learning’ ”* y replicar los modelos utilizados (Naive-Bayes, KNN y regresión logística) para llegar a los mismos porcentajes de predicción que se llega en el artículo con la base de datos dada, e incluso aplicar otros modelos para hacer más comparación más detallada. Algunos aspectos positivos sobre escoger este tema son:

- El fraude en tarjetas de crédito representa una amenaza diaria para todos los agentes del sistema financiero, por lo que su detección es de interés común para la sociedad; por lo que este estudio genera un precedente para las empresas y los agentes financieros para aplicar los modelos de ‘machine learning’ a gran escala.
- Al recrear este estudio se exploran diversos modelos de ‘machine learning’ lo que genera un conocimiento más amplio de diversos algoritmos y ofrece la oportunidad de compararlos entre ellos y destacar las ventajas y deficiencias de cada modelo.
- El proyecto presenta una oportunidad para explorar las diversas librerías y herramientas que ofrece Python, de modo que permite profundizar más en las formas en las cuales se puede utilizar este lenguaje de programación.