## Propuesta

### 1.El nombre de la investigación

Detección predictiva de fraude en transacciones con tarjetas de crédito

### 2. Descripción de la base de datos a utilizar

La base de datos seleccionada es el Credit Card Transactions Dataset de Kaggle, recopilado y compartido por Priyam Choksi. Este dataset contiene aproximadamente 1,295,600 registros de transacciones con tarjetas de crédito, categorizadas como “fraudulentas” o “no fraudulentas”. Cada transacción incluye variables como:

* trans\_date\_trans\_time: Fecha de la transacción
* cc\_num: Número de la tarjeta de crédito
* merchant: Comercio o tienda donde ocurrió la transacción
* category: tipo de la transacción (entretenimiento, supermercado, etc)
* amt: monto de la transacción
* first: primero nombre del dueño de la tarjeta
* last: apellido del dueño de la tarjeta
* gender: génedor del dueño de la tarjeta
* street: dirección del dueño de la tarjeta
* city: ciudad del dueño de la tarjeta
* state: estado del dueño de la tarjeta
* zip: código postal del dueño de la tarjeta
* lat: coordenadas geográficas de la transacción
* long: coordenadas geográficas de la transacción
* city\_pop: Población de la ciudad donde ocurrió la transacción.
* job: ocupación del dueño de la tarjeta
* dob: fecha de nacimiento del dueño de la tarjeta
* trans\_num: número de transacciones único
* unix\_time: fecha de la transacción en formato Unix
* merch\_lat: coordenadas geográficas del comercio
* merch\_long: coordenadas geográficas del comercio
* is\_fraud: indicador de si la transacción es fraudulenta
* merch\_zipcode: código postal del comercio

El dataset incluye datos sintéticos y anonimización de ciertos campos sensibles para preservar la privacidad, lo cual permite su uso en investigación sin comprometer datos personales reales.

### 3. Motivo de la selección de los datos

El dataset fue elegido por las siguientes razones:

* Debido a que representa un problema real en el ambiente financiero, en este caso poder detectar o identificar de alguna forma el fraude
* Es un tema interesante debido a que ya anteriormente he visto este tipo de elementos en el trabajo y me gustaría profundizar un poco más para así obtener nuevos conocimientos tanto para el análisis de comportamiento fraudulento como el desarrollo de un modelo de esta índole

### 4. Breve introducción del contexto

El fraude financiero, en especial de transacciones electrónicas, representa una gran amenaza para bancos, comercios y usuarios, esto debido a que este tipo de ataques representa una gran pérdida para los bancos. Según Q-vision para el 2024 los bancos Panameños reportaron perdidas de 14.2 Millones de dólares por fraude sintético, gracias a la complejidad de estos elementos es necesario la implementación de sistemas inteligentes capaces de detectar patrones de comportamiento sospechosos en tiempo real, sin afectar las operaciones legítimas de los usuarios.

Tradicionalmente, los métodos de detección de fraude se basaban fundamentalmente en reglas estáticas. Sin embargo, estos elementos resultan insuficientes ante la evolución tan rápida del comportamiento fraudulento. Por ello, el uso de modelos predictivos supervisados basados en machine learning se ha convertido en una alternativa eficaz para identificar fraudes de una mejor forma y a su vez más precisa, incluso en conjuntos de datos altamente desbalanceados.

Esta investigación busca explorar, entrenar y evaluar modelos de clasificación que puedan detectar transacciones fraudulentas de manera eficiente y replicable, aun cuando las variables estén anonimizadas.

### 5. Línea de tiempo de investigación

