**Predicción de fraude en transacciones presenciales**

El proyecto se enfocó en la detección de fraudes en transacciones financieras utilizando un conjunto de datos que contenía información de diversas características de cada transacción. Después de realizar el preprocesamiento de los datos, se observó que había muchas columnas con valores nulos y otras que no se relacionaban con el archivo train.csv. Asimismo, se encontraron columnas duplicadas con diferente nombre, lo que hubiera duplicado la cantidad de columnas en el archivo de train.csv.

Además, se identificó la columna "ID" como un identificador y se eliminó para evitar problemas en el entrenamiento del modelo. También se eliminó la columna "Dist\_HOY" debido a que, según el análisis exploratorio de datos (EDA), los valores de esta columna tenían altos índices de ser aleatorios y no aportarían nada al modelo.

En cuanto al preprocesamiento, se reemplazaron los valores nulos en las columnas de distancia por un valor de 1, y en las columnas donde el número de valores nulos era muy bajo, se eliminaron las filas correspondientes. En el archivo test.csv se encontraron columnas duplicadas, donde una tenía los nulos rellenos por 1 y la otra tenía todos los nulos.

Durante la evaluación de los modelos, se encontró que el promedio de varios modelos era de 73% y el de Random Forest obtuvo una precisión promedio del 94%. No obstante, se seleccionó este modelo debido a su alta precisión del 96.04%. Además, el uso de modelos de aprendizaje automático, como Random Forest, puede ser una herramienta valiosa en la detección de fraudes en transacciones financieras, lo que puede ser de gran importancia para la toma de decisiones en los negocios.

En conclusión.

Basado en los resultados del modelo de Random Forest, que alcanzó una precisión del 96,04%, se puede concluir que es un modelo muy efectivo en la detección de transacciones fraudulentas. Además, la especificidad del modelo (97,9%) indica que es muy bueno en la identificación de transacciones legítimas, lo que significa que es poco probable que se clasifiquen erróneamente transacciones legítimas como fraudulentas. Sin embargo, la tasa de falsos positivos del modelo (2.1%) es relativamente alta, lo que significa que algunas transacciones legítimas pueden ser clasificadas como fraudulentas, lo que podría generar inconvenientes para los clientes y la empresa. Por lo tanto, es importante continuar mejorando con más datos de calidad y ajustando el modelo para reducir aún más la tasa de falsos positivos y mejorar la precisión general del modelo.