

APPLICATION RECORD UN DATASET BANCARIO



APPLICATION RECORD

ABSTRACT

Hoy en día, y desde siempre, las entidades bancarias se financian por medio de los créditos que otorgan a personas particulares o empresas. En principio como créditos que se solicitaban en el banco, y con el paso del tiempo aparecieron las tarjetas de crédito, las cuales se han mantenido en el tiempo y son útiles tanto para los bancos, como para los clientes. Se trata de una modalidad de prestamos donde los gastos tienen un financiamiento fijo (en cuotas) o variable según el gasto y el pago mensual.

Una problemática importante es decidir a quien se le otorga este beneficio, en base a sus ingresos y sus gastos. Hay que predecir de alguna manera como le ira al cliente porque la idea principal es que siempre la use y siempre pague aunque sea el mínimo para poder sumarle intereses.

APPLICATION RECORD DATASET

El dataset escogido se obtiene mediante la página web de kaggle

Posee 438,557 registros divididos en 18 columnas, de las cuales dos son de tipo flotante, ocho son de tipo entero y ocho son de tipo cadena de texto

Vamos luego a enriquecer el dataset actual con 8 columnas de tipo flotante de otro dataset, obtenido desde una API de kaggle

APPLICATION RECORD UN DATASET BANCARIO

Las principales interrogantes de este proyecto son las siguientes:

- ▶ Clientes por tipo de ingreso
- ▶ Ingreso anual por tipo de educación y género
- ▶ Ingreso anual por tamaño de grupo familiar
- ▶ Ingreso anual por tiempo trabajado en días

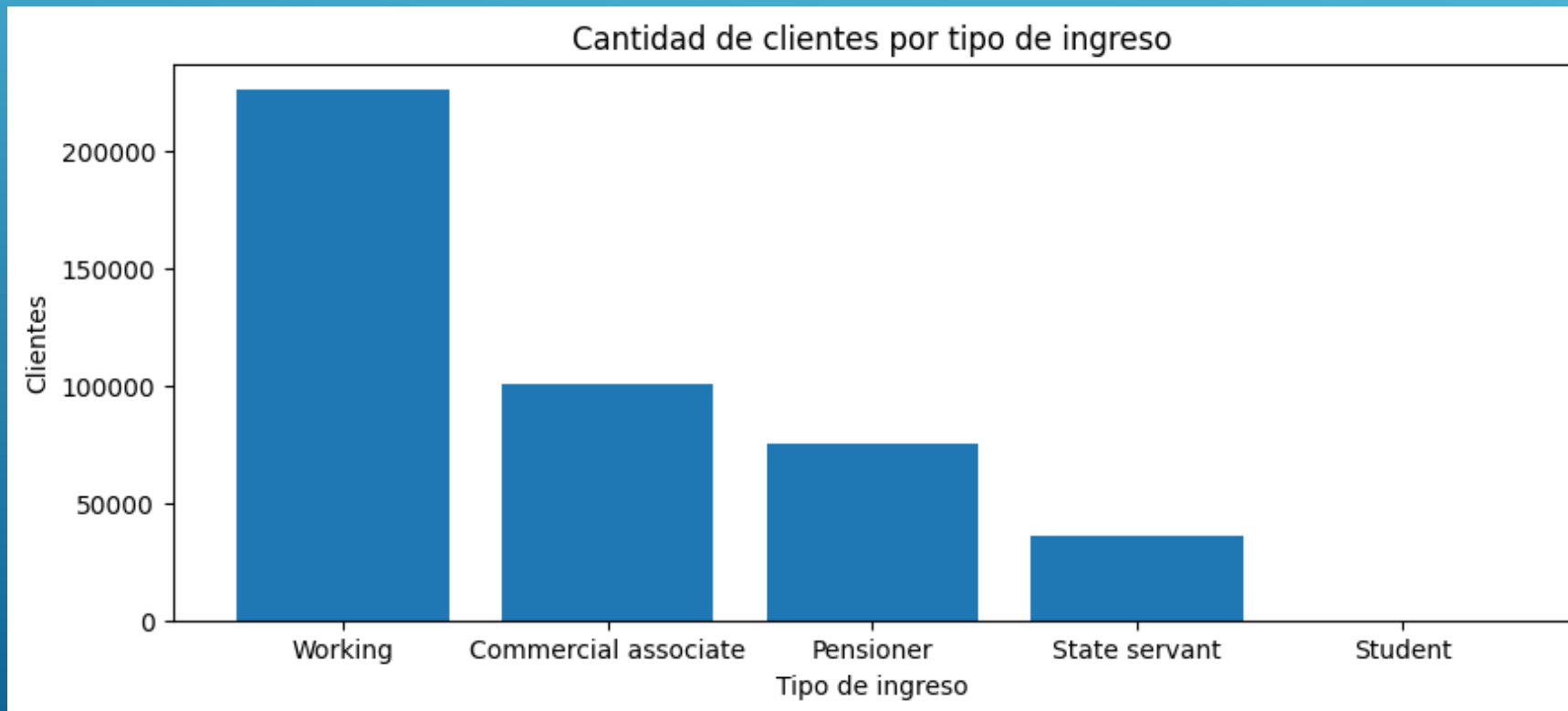
ANÁLISIS EXPLORATORIO DE DATOS



ANÁLISIS EXPLORATORIO DE DATOS

► Clientes por tipo de ingreso

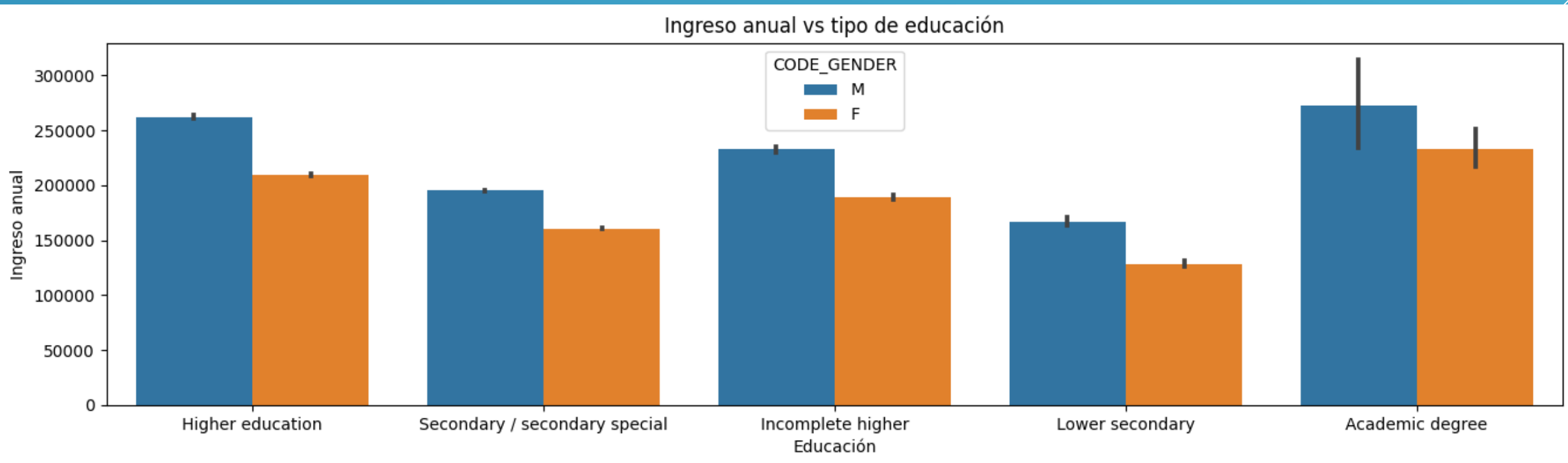
La mayor cantidad de clientes tienen tipo de ingreso Working



ANÁLISIS EXPLORATORIO DE DATOS

- Ingreso anual por tipo de educación y género

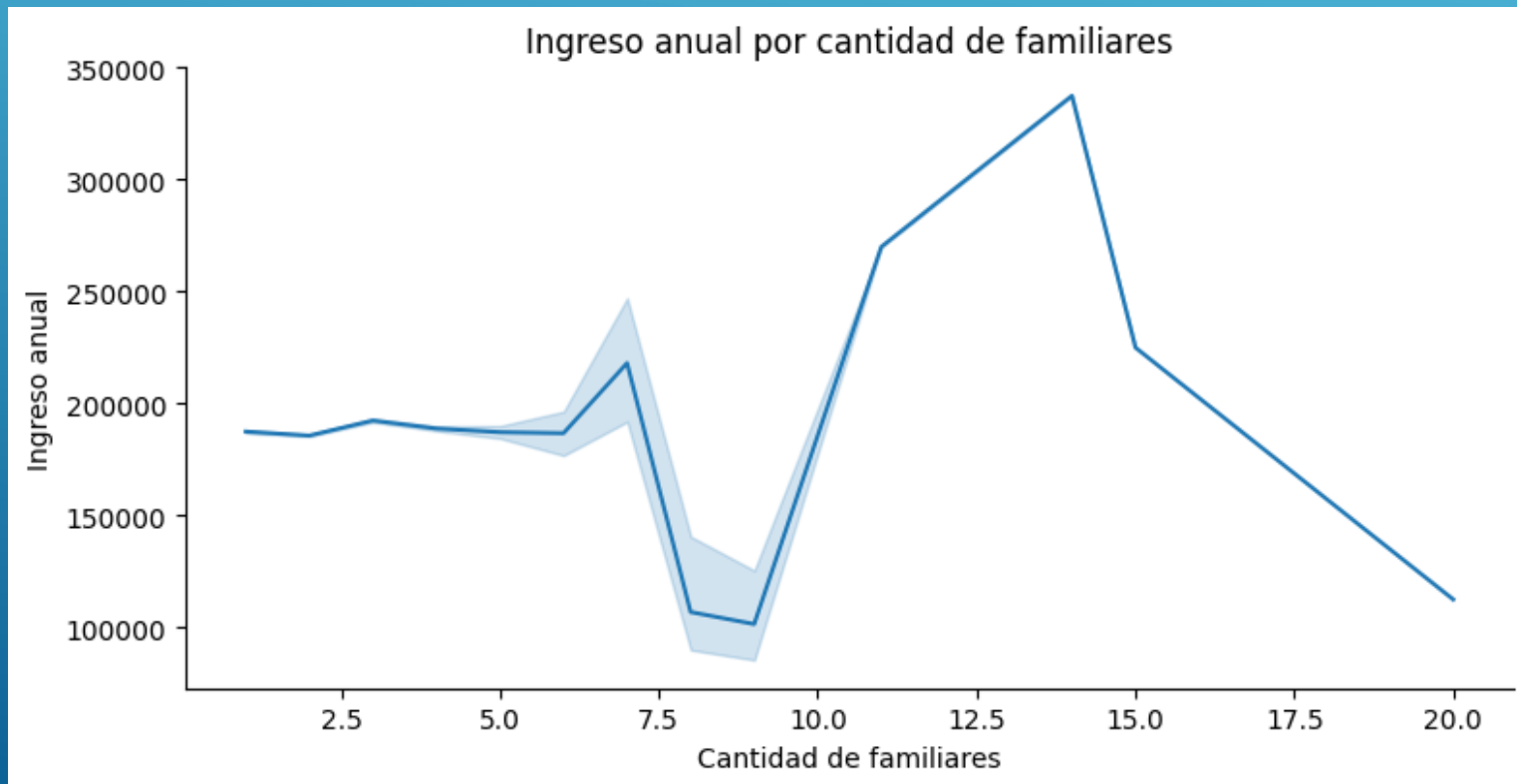
Del gráfico anterior analizamos el ingreso anual por tipo de educación, donde se observa que los ingresos mas altos son para los de grado académico y educación superior. Además se observa que los de sexo masculino tienen mas alto ingreso que los de sexo femenino



ANÁLISIS EXPLORATORIO DE DATOS

► Ingreso anual por tamaño de grupo familiar

Podemos decir que el mayor ingreso está en familias con mas de 12 miembros, y que el menor ingreso es de familias con 6 miembros aproximadamente



ANÁLISIS EXPLORATORIO DE DATOS

- Ingreso anual por tiempo trabajado en días
Para los solteros con grado académico de estudios

El mayor ingreso anual se observa entre los 1500 y 2700 días trabajados, luego hay una tendencia lineal en los ingresos a partir de los 3100 días de trabajo



INSIGHTS Y RECOMENDACIONES



INSIGHTS Y RECOMENDACIONES

Los principales insights y recomendaciones de este análisis exploratorio son los siguientes:

- ▶ Ofrecer promociones crediticias desde aplicaciones telefónicas a clientes trabajadores
- ▶ Incentivar el uso de créditos para consumo familiar para los clientes con un tamaño de grupo familiar de entre 11 y 15 personas
- ▶ Verificar los créditos aprobados para los clientes que solo tienen secundario
- ▶ Verificar los límites de crédito de los clientes solteros con grado académico que tengan entre 2000 y 3000 días trabajando
- ▶ Hacer seguimiento a los gastos de los clientes solteros con grado académico, que tienen más de 4000 días trabajando

MODELO SELECCIONADO

Luego de utilizar varios modelos, el modelo de predicción seleccionado es el XGBoost Classifier. El cual luego de agregar hiperparámetros para un mejor desempeño, arrojó las siguientes métricas:

	precision	recall	f1-score	support
0	0.76	0.85	0.80	29200
1	0.68	0.55	0.61	17104
accuracy			0.74	46304
macro avg	0.72	0.70	0.70	46304
weighted avg	0.73	0.74	0.73	46304

Estas métricas representan un alto desempeño en cuanto a precisión, exactitud, y a la capacidad del modelo predictivo de encontrar los valores relevantes dentro del dataset, por esto los valores no presentan tanta varianza, además de ser un modelo mejorado (Boost) para estos casos

MODELO SELECCIONADO

Durante todo el proyecto se utilizó Feature Selection y PCA para seleccionar los campos más significativos, lo cual complementa el modelo seleccionado

Finalmente y haciendo un resumen de todas las evaluaciones realizadas, se puede concluir que el modelo escogido representa un muy buen ajuste sobre los datos de entrenamiento (incluyendo el agregado de hiperparámetros), y que aunque se agreguen más datos significativos, las principales métricas se mantienen.

Con esto validamos que nuestro modelo servirá perfectamente para predecir la aprobación de una tarjeta de crédito para futuros clientes de este banco