

# Kubo technical assessment

---

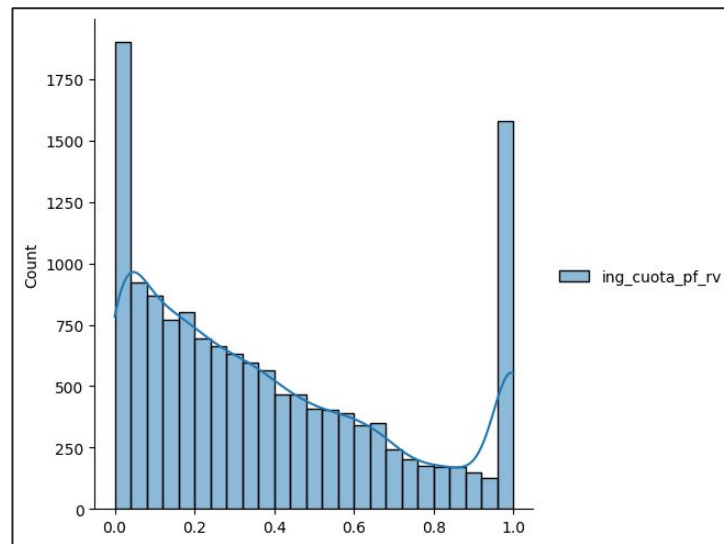
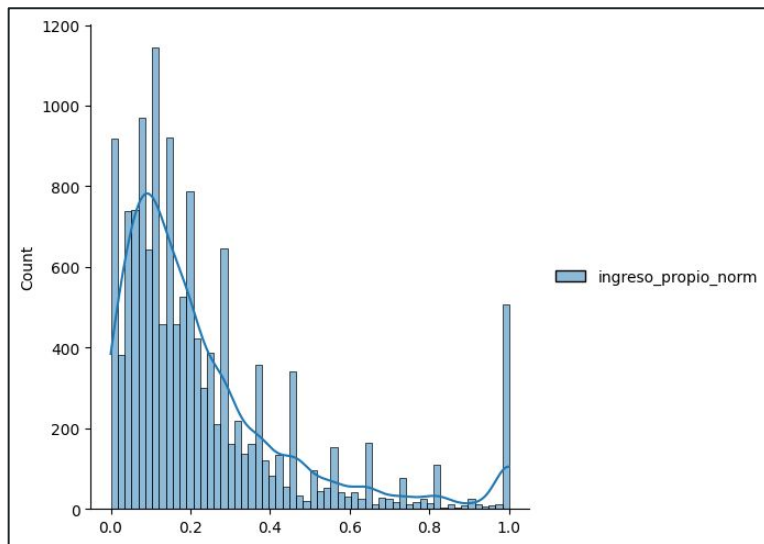
Presents: Daniel Pérez

# Sobre el EDA y limpieza/selección de variables

---

# Valores nulos en variables numéricas

- Un total de 9 filas se vieron afectadas por valores nulos en dos columnas. Se usó la mediana para rellenar los valores nulos en sus respectivas columnas
- Se removieron 3 variables: "hora\_reg\_norm", "dia\_reg", "destino"
- Los datos ya se encontraban normalizados.

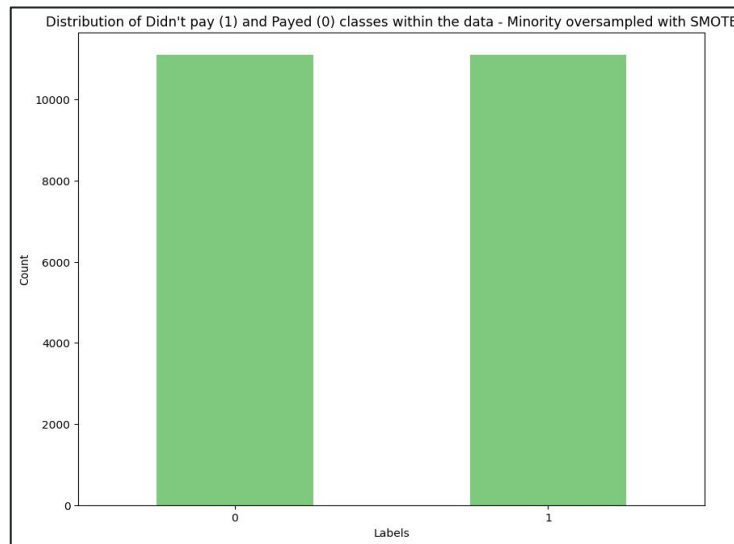
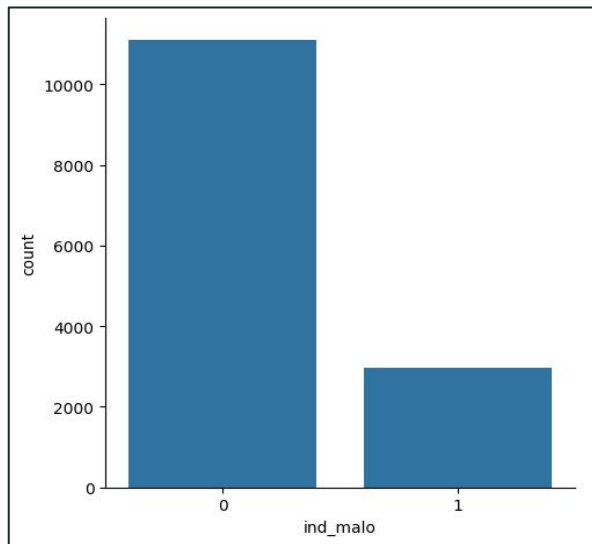


# Variables categóricas

- Se removieron 2 variables: “medio\_contacto\_grupo”, “tipo\_mail\_rango”.
- Posteriormente se les aplicó one-hot encoding para transformarlos a datos numéricos.

Variable	<del>Medio de contacto</del>	<del>Tipo de email</del>	Sexo	Nivel estudios	Estado	Ocupación
count	14067	14067	14067	14067	14067	14067
unique	6	5	2	7	32	2
top	Otro_medio	gma	Hombre	Licenciatura	CDMX	Empleado
freq	4805	6202	7941	8042	3804	10962

# La variable objetivo tiene clases desbalanceadas



- Se realizó un resamdeo mediante la técnica SMOTE previo al entrenamiento para mejorar la capacidad predictiva de los modelos.

# Modelos de ML usados y su rendimiento

---

# Comparación de modelos usados

	#2	#1	#3	#4
Modelo	Gradient Boosting	<b>XG Boost</b>	Random Forest	Logistic Regression
AUC-ROC	0.9088	<b>0.9146</b>	0.8405	0.8226
AUC-PR	0.9291	<b>0.9379</b>	0.8551	0.8340
Gini	0.8177	<b>0.8293</b>	0.6810	0.6452
K-S	0.7034	<b>0.7198</b>	0.5331	0.4979

- AUC: metric used to measure how well a classification model can distinguish between different classes
- Gini: coefficient that ranges from 0 to 1, with 0 representing perfect equality (no discrimination) and 1 representing perfect inequality (perfect discrimination).
- K-S: a measure of the degree of separation between the positive and negative distributions. A higher K-S statistic indicates better separation between classes, which means a better performing model.

# Machine learning y su potencial en el mundo financiero

---



# Estadística y algoritmos como valor agregado



# La importancia de otorgar un crédito

El otorgamiento de crédito es la actividad más importante del sistema bancario.



Las instituciones financieras reciben de individuos y empresas el pago de su préstamo junto con un interés.



A este fenómeno se le llama: **riesgo crediticio.**



El riesgo de prestar dinero consiste en que el deudor falle en pagar su préstamo.

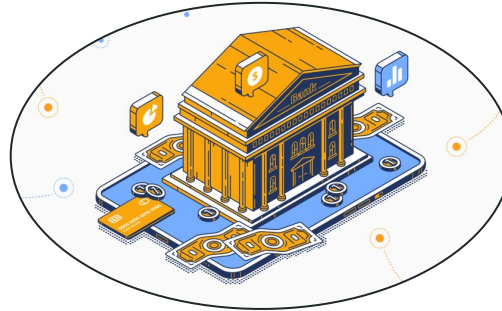
# IA para el manejo de riesgos

Dichos modelos determinan la probabilidad de incumpliente de un cliente.

**La gestión del riesgo puede llevarse a cabo mediante modelos de puntuación crediticia**

Controlarlo permite que una institución mantenga sustentabilidad y estabilidad financiera.

“El uso de machine learning tiene el potencial de mejorar el rendimiento y eficiencia en modelos de puntuación crediticia”<sup>(1)</sup>



Varios casos de éxito han sido los referentes a detección de fraude bancario y acciones contra el lavado de dinero

(1) Montevechi et.al., 2024. Advancing credit risk modelling with Machine Learning: A comprehensive review of the state-of-the-art. Engineering Applications of Artificial Intelligence. <https://doi.org/10.1016/j.engappai.2024.109082>