



# Modelo de Aprendizaje Automático para la Probabilidad de Impago

Walter Quispe Vargas, PhD

DAME-UNSAAC

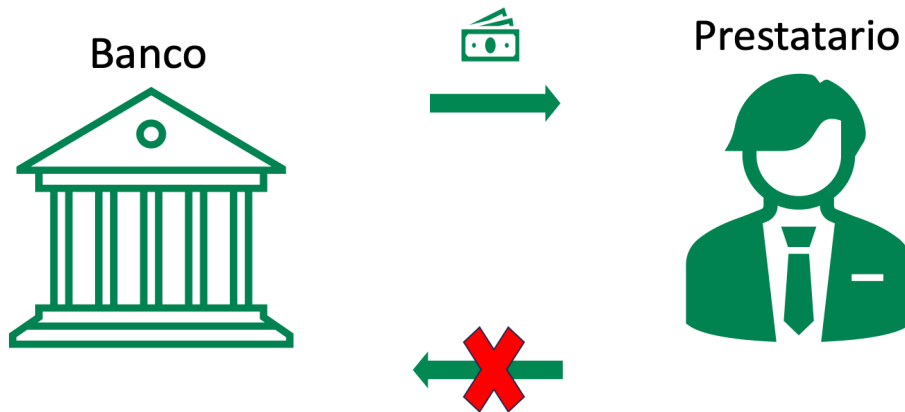
2023-09-13



# Contenido

- Riesgo de Crédito y Probabilidad de Impago (*Default*)
- Desarrollo de Modelos en Aprendizaje Automatizado (*Machine Learning Modeling*)
  - Pre-procesamiento de Datos
  - Selección de Variables y Modelo
  - *Extreme Gradient Boosting (XGBoost)*
  - Explicabilidad del Modelo.

# Riesgo de Crédito



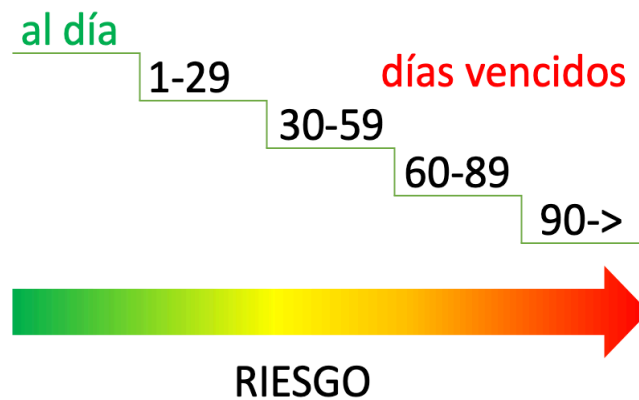
- La posibilidad de una **pérdida**, resultante de que el prestatario no pague un préstamo o no cumpla con sus obligaciones contractuales.

# Pérdida Crediticia Esperada (ECL)

- $ECL = PD \times LGD \times EAD$
- *PD: Probabilidad de Impago (Probability of Default)*
  - Modelos de Clasificación
- *LGD: Pérdida dado el Impago (Loss Given Default)*
- *EAD: Exposición al Impago (Exposure At Default)*
  - Modelos de Regresión (Tobit, Beta)

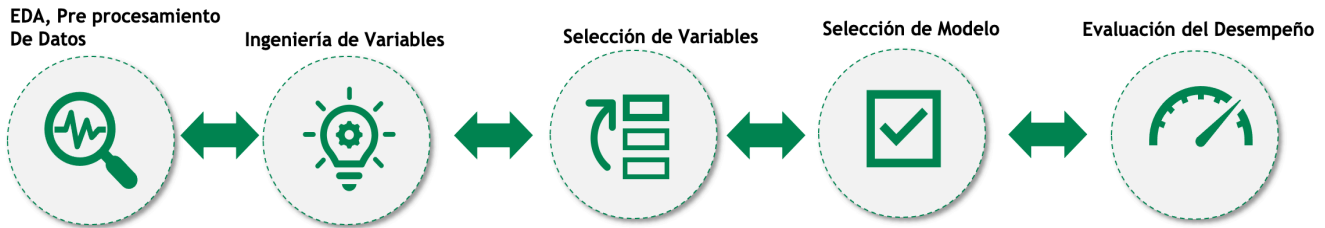
# Probabilidad de Impago (PD)

- De acuerdo con IFRS 9 y CECL:
  - “El impago ocurre cuando la institución bancaria considera que es poco probable que el deudor pague la totalidad de sus obligaciones crediticias al grupo bancario”. (Bellini, 2019)



# Machine Learning Modeling

## PROCESO DE DESARROLLO DEL MODELO



- **CONTEXTO**
- Observar el comportamiento de pagos de los clientes en un periodo de 12 meses, si durante este periodo, el atraso máximo a nivel clientes supera el umbral de 60 días, diremos que el cliente ha hecho **default**.

# Pre-procesamiento de Datos

[Diccionario](#)[DataFrame](#)[Skim\(\)](#)[Imbalance](#)[EDA](#)

VARIABLE	DESCRIPCIÓN
FLG_CLI_DEF60	Default
DIF_BU_MA_100_12M	DIFERENCIA ENTRE MESES BUEN Y MAL COMPORTAMIE CON MATERIALIDAD>100  Diferer entre meses de buen comportamiento(calificaci y mal comportamiento(calificaci en los últimos 12 meses (un materialidad=100)
DIF_BU_MA_100_24M	DIFERENCIA ENTRE MESES BUEN Y MAL COMPORTAMIE CON MATERIALIDAD>100  Diferer entre meses de buen comportamiento(calificaci y mal comportamiento(calificaci en los últimos 24 meses (un



materialidad=100)

DIF\_BU\_MA\_12M

DIFERENCIA ENTRE MESES  
BUEN Y MAL COMPORTAMIE  
CON  
MATERIALIDAD>0||Diferenci  
entre meses de buen  
comportamiento(calificació  
y mal  
comportamiento(calificació  
en los últimos 12 meses (un  
materialidad=0)

DIF\_BU\_MA\_500\_12M

DIFERENCIA ENTRE MESES  
BUEN Y MAL COMPORTAMIE  
CON  
MATERIALIDAD>500||Diferen  
entre meses de buen  
comportamiento(calificació  
y mal  
comportamiento(calificació  
en los últimos 12 meses (un  
materialidad=500)

DIF\_BU\_MA\_ALI\_12M

DIFERENCIA ENTRE MESES  
BUEN Y MAL COMPORTAMIE  
ALINEADO CON  
MATERIALIDAD>0||Diferenci

entre meses de buen comportamiento alineado(calificación alineada=0) y mal comportamiento(calificación alienada>0) en los últimos 12 meses (umbral materialidad)

FLG\_DVNCD\_24M

FLAG DEUDA VENCIDA||Indica existencia de deuda vencida en los últimos 24 meses

INC\_MAX\_DTOTSH\_ACTU\_24M

INCREMENTO DE DEUDA TC S/HIP. (CONSIDERANDO MÁXIMO)||Máxima deuda total s/hip. en los últimos 24 meses respecto a la deuda total s/hip. actual

INC\_MAX\_ENT\_ACTU\_24M

INCREMENTO DE ENTIDADES ACREEDORAS (CONSIDERANDO MÁXIMO)||Máximo nro. de entidades en los últimos 24 meses respecto al nro. de entidades actual

INC\_SUM\_ACTSH\_ACTU\_24M

INCREMENTO DE DEUDA ACTIVA S/HIP. (CONSIDERANDO SUMA)||Total deuda activa s/hip. en los últimos 24 meses respecto a la deuda activa s/hip. actual

en los últimos 24 meses  
respecto a la deuda activa s  
actual

MAX_ATR_24M	ATRASO MÁXIMO  Nro. máxi de dias de atraso en los últi 24 meses
MAX_ATR_I_12M	MAX_ATR  Máximo atrasos e últimos 12 meses
MAX_ATR_I_18M	MAX_ATR  Máximo atrasos e últimos 18 meses
MAX_ATR_I_U6K_24M	MAX_ATR  Máximo atrasos e últimos 6 meses de comportamiento en los últi 24 meses
MAX_DEUNOREV_12M	MÁXIMA DEUDA NO REVOLVENTE  Máxima deud asociada al concepto 'Prést no revolvente' en los último meses
MAX_DEU_JUD_24M	MÁXIMA DEUDA JUDICIAL  Máxima deuda judicial en los últimos 24 m
MAX_ENT_12M	MÁXIMA CANTIDAD DE ENTIDADES ACREEDORAS  M

	maximo de entidades acreedoras en los últimos 12 meses
NMES_ATR15_I_U6K_24M	N_MESES_ATR  Número de meses con atrasos mayores iguales a 15 en los últimos 6 meses de comportamiento en los últimos 24 meses
NMES_UATR15_I_U6K_24M	N_MESES_UATR  Número de meses desde el último atraso mayor o igual a 15 días, en los últimos 6 meses de comportamiento en una ventana de 24 meses (distancia)
NMES_UATR1_I_24M	N_MESES_UATR  Número de meses desde el último atraso mayor o igual a 1 días de atraso observado en los 24 últimos meses (distancia)
NMES_UATR3_I_24M	N_MESES_UATR  Número de meses desde el último atraso mayor o igual a 3 días de atraso observado en los 24 últimos meses (distancia)
N_BU1000_01_18M	MESES BUEN

COMPORTAMIENTO (0 ó 1) C  
 MATERIALIDAD>1000||Nro. l  
 Meses con calificación 'NOF  
 o 'CPP' y materialidad>1000  
 los últimos 18 meses

N\_BU1000\_ALI\_18M

MESES BUEN  
 COMPORTAMIENTO ALINEA  
 CON MATERIALIDAD > 1000|  
 de meses con buen  
 comportamiento alineado e  
 últimos 18 meses (umbral d  
 materialidad = 1000 soles)

N\_BU100\_12M

MESES BUEN  
 COMPORTAMIENTO CON  
 MATERIALIDAD > 100||Nro. c  
 meses con buen  
 comportamiento en los últi  
 12 meses (umbral de  
 materialidad = 100 soles)

N\_BU\_01\_12M

MESES BUEN  
 COMPORTAMIENTO (0 ó 1) C  
 MATERIALIDAD>0||Nro. De M  
 con calificación 'NORMAL' o  
 'CPP' y materialidad>0 en l  
 últimos 12 meses

N\_BU\_12M

MESES BUEN  
COMPORTAMIENTO CON  
MATERIALIDAD > 0||Nro. de  
meses con buen  
comportamiento en los últi  
12 meses (umbral de  
materialidad = 0 soles)

N\_DEU\_ACTSH\_18M

NRO. DE MESES CON DEUDA  
ACTIVA SIN HIPOTECARIA||N  
de meses con deuda activa  
deuda hipotecaria en los úl  
18 meses

N\_NEG\_24M

NRO. DE MESES CON DEUDA  
NEGOCIO||Nro. de meses co  
deuda negocio(micro,peque  
mediana) en los últimos 24  
meses

N\_NOR\_12M

NRO. DE MESES CON  
CALIFICACIÓN "NORMAL"||N  
de meses que el cliente fue  
calificado como 'NORMAL'. I  
últimos 12 meses

N\_NOR\_24M

NRO. DE MESES CON  
CALIFICACIÓN "NORMAL"||N  
de meses que el cliente fue

calificado como 'NORMAL'. I  
últimos 24 meses

N\_ULT\_DIFNOR\_12M

NRO. DE MESES DESDE QUE  
ÚLTIMA VEZ QUE SE REPOR  
CALIFICACIÓN DIFERENTE  
"NORMAL"||Nro. De meses c  
última vez que se reportó  
calificación diferente a 'Nor  
en los últimos 12 meses

N\_ULT\_DIFNOR\_24M

NRO. DE MESES DESDE QUE  
ÚLTIMA VEZ QUE SE REPOR  
CALIFICACIÓN DIFERENTE  
"NORMAL"||Nro. De meses c  
última vez que se reportó  
calificación diferente a 'Nor  
en los últimos 24 meses

N\_ULT\_NOR\_ALI\_24M

NRO. DE MESES DESDE QUE  
ÚLTIMA VEZ QUE SE REPOR  
CALIFICACIÓN ALINEADA  
"NORMAL"||Nro. De Meses d  
última vez que se reportó  
calificación alineada 'Norm  
los últimos 24 meses

PROM\_DEUREV\_24M

PROMEDIO DEUDA  
REVOLVENTE||Promedio de

deuda revolving en los últimos 24 meses

PROM\_DEUVNCD\_12M

PROMEDIO DEUDA VENCIDA||Promedio de deuda vencida en los últimos 12 meses

RMAX\_DACT\_DACTSH\_24M

MÁXIMA DEUDA ACTIVA RESPECTO A LA MÁXIMA DE ACTIVA S/HIP.||Ratio de la máxima deuda activa respecto a la máxima deuda activa sin deuda hipotecaria en los últimos 24 meses

RMAX\_DREV\_DDIR\_24M

MÁXIMA DEUDA REVOLVENTE RESPECTO A LA MÁXIMA DE DIRECTA||Ratio de la máxima deuda revolving sobre la máxima deuda directa en los últimos 24 meses

RMAX\_DVNCD\_DDIR\_12M

MÁXIMA DEUDA VENCIDA RESPECTO A LA MÁXIMA DE DIRECTA||Ratio de la máxima deuda vencida sobre la máxima deuda directa en los últimos 12 meses

RTOT\_DACT\_DTOT\_24M

TOTAL DEUDA ACTIVA RESPECTO A LA MÁXIMA DE ACTIVA S/HIP.



	AL TOTAL DE DEUDA TOTAL  Ratio del total de la deuda activa sobre el total de deuda total en los últimos 24 meses
RTOT_DVNCD_DDIR_12M	TOTAL DEUDA VENCIDA RESPECTO AL TOTAL DE DEUDA DIRECTA  Ratio del total de deuda vencida sobre el total de la deuda directa en los últimos 12 meses
R_DREV_DDIR_24M	DEUDA REVOLVENTE PROMEDIO RESPECTO A LA DEUDA DIRECTA PROMEDIO  Ratio del promedio de la deuda revolvente sobre el promedio de la deuda directa en los últimos 24 meses
R_DVNCD_DDIR_12M	DEUDA VENCIDA PROMEDIO RESPECTO A LA DEUDA DIRECTA PROMEDIO  Ratio del promedio de la deuda vencida sobre el promedio de la deuda directa en los últimos 12 meses
VAR_MAX_CAL_ACTU_24M	VARIACIÓN DE CALIFICACIÓN (CONSIDERANDO)

MÁXIMO)||Máxima calificación en los últimos 24 meses (se consideran nulos) respecto a calificación actual

VAR\_PROM\_DEUDIR\_ACTU\_12M VARIACIÓN DE DEUDA DIRECTA (CONSIDERANDO PROMEDIO)||Deuda directa promedio en los últimos 12 meses respecto a la deuda directa actual

VAR\_PROM\_ENT\_ACTU\_24M VARIACIÓN DE ENTIDADES ACREEDORAS (CONSIDERANDO PROMEDIO)||Promedio de número de entidades en los últimos 24 meses respecto al nro. de entidades actual

# Selección de Variables

[Métodos](#)[Python](#)[Resultados](#)

- Reducción Secuencial:
  - Descartar Variables Altamente Correlacionadas (*collinear*)
  - Descartar Variables con Cero Importancia (*catboost*)
  - Descartar Variables con Cero Información Mutua (*entropy*)
- Importancia de las Variables:
  - Recursive Feature Elimination usando Logistic regression
  - Logistic Regression L1 Penalty (Ridge)
  - Logistic Regression L2 Penalty (Lasso)
  - Random Forest
  - Boosting Machines:
    - AdaBoost
    - LighGBM
    - XGBoost
    - CatBoost
- Algoritmo: “**Boruta**” usando XGBoost



# Selección de Modelo

Challenge

Comparación

Linear/Logistic  
Regression

Decision  
Trees

Random  
Forest

XGBoost

Neural  
Networks



Explainability

Prediction Power

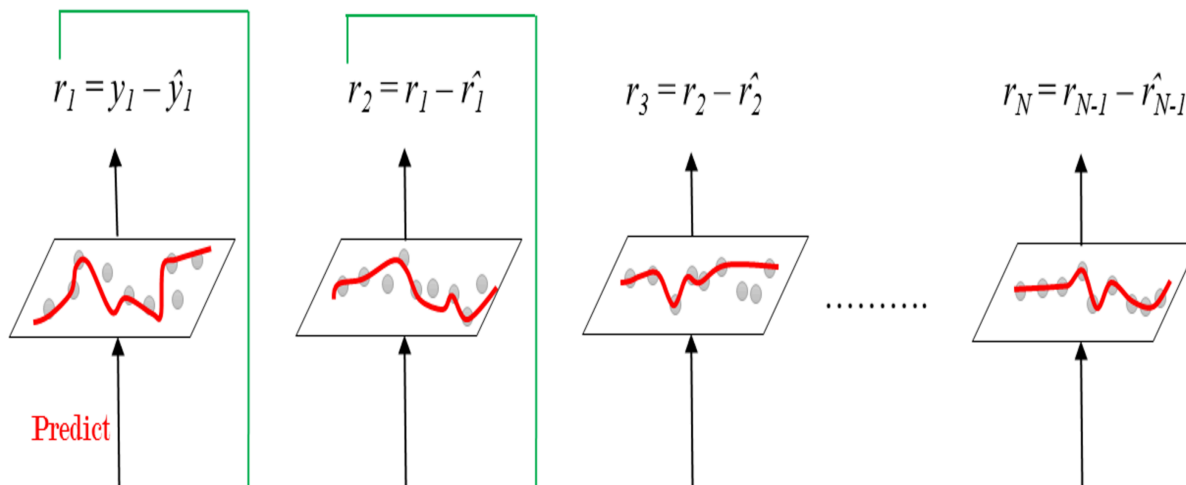


# Extreme Gradient Boosting (XGBoost)

## Fundamentos

## R

- **XGBoost**: Algoritmo de *Machine Learning* para resolver problemas de Clasificación y Regresión.

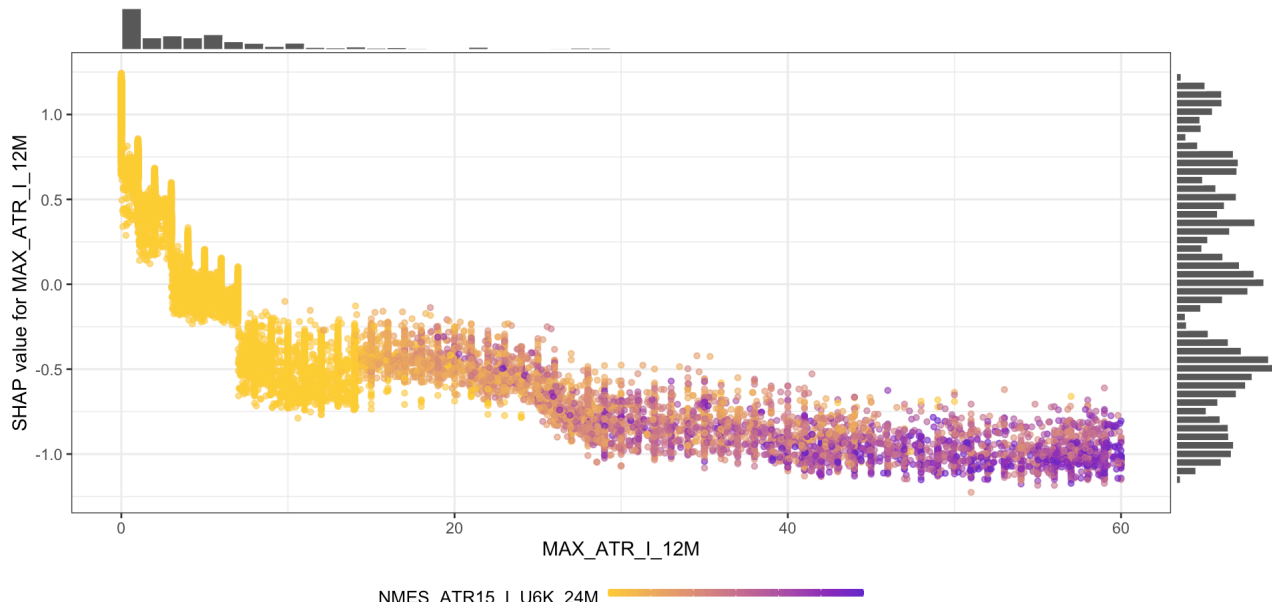


# Explicabilidad del Modelo

SHAP

En General

Parcial



# Preguntas