# Del Click a la Compra ML en Mercado Libre

Lucas Arbues - Agustin Alfie - Abril Noguera



#### **AGENDA**

01

**INTRODUCCIÓN** 

04

DATOS Y PROCESAMIENTO

02

ENTENDIMIENTO DEL NEGOCIO

05

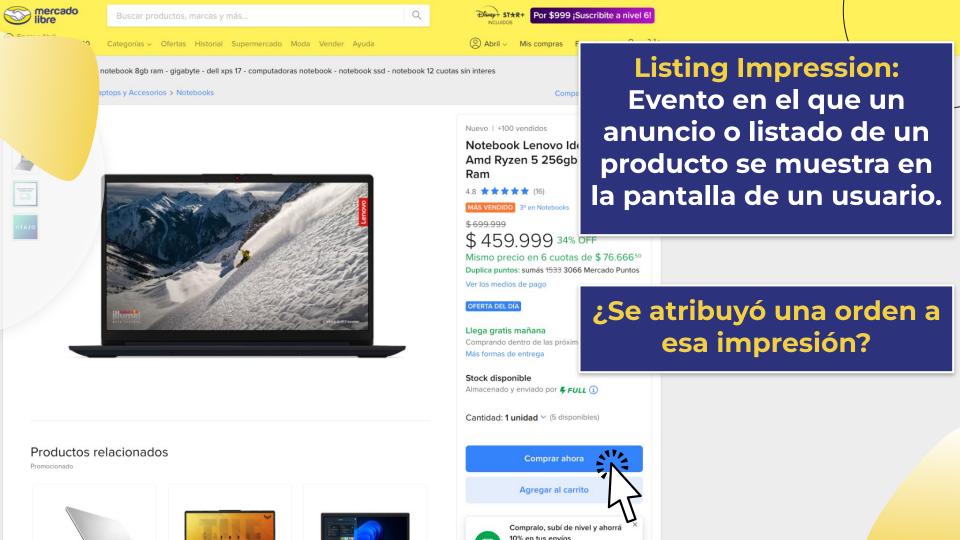
MODELOS PREDICTIVOS

03

DEFINICIÓN DEL PROBLEMA

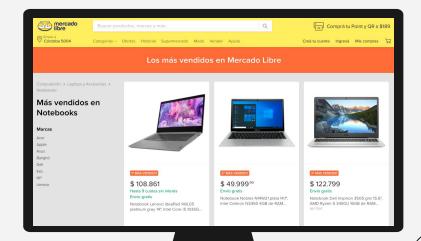
06

**RESULTADOS** 



# Tasa de Conversión

Porcentaje de visitantes que realizan una compra.



# Mejorar El Entendimiento Del Negocio



#### **OBJETIVO**

Reconocer características de usuarios para optimizar las estrategias de marketing y ventas aumentando la tasa de conversión de los usuarios de Mercado Libre.



# CONTEXTO DEL MERCADO

Comprender el comportamiento del consumidor en el comercio electrónico, tendencias del mercado, competencia, y cómo Mercado Libre se posiciona en este escenario.



#### IMPACTO FINANCIERO

Aumento de ventas, optimización del presupuesto de marketing, reducción de costos y mayores ingresos.

# ¿Qué Datos se Extraen del Click?

fulfillment title avg\_asp\_item\_bday accepts\_mercadopago site id available\_quantity avg\_gmv\_item\_domain\_30days health print\_position total\_orders\_domain\_30days avg\_asp\_item\_domain full\_name qty\_items\_dom avg\_asp\_item\_sel offset total\_si\_sel\_30days total\_orders\_sel\_30days user id qty\_items\_sel date category\_id original\_price avg visits item avg\_orders\_item\_bday main\_picture avg\_si\_item\_sel\_30day total\_gmv\_seller total\_visits\_domain total\_items\_seller total\_si\_item\_30days total\_orders\_item\_30days benefit boosted free\_shipping total si domain 30days logistic\_type tags avg\_gmv\_item\_bday deal\_print\_id avg\_qty\_orders\_item\_sel\_30days aty items dom total visits item avg\_si\_item\_sel\_30day is pdp conversion domain id ROW ID sold\_quantity total\_items\_domain avg\_gmv\_item\_sel total\_gmv\_domain\_30days total\_asp\_item\_sel\_30days total\_asp\_item\_domain\_30days listing\_type\_id total\_visits\_seller avg\_qty\_orders\_item\_domain\_30days avg\_gmv\_seller\_bday avg\_si\_item\_bday warranty item id total\_gmv\_item\_30days total\_gmv\_domain\_bday product id print\_server\_timestamp avg\_orders\_seller\_bday

## ¿Qué Datos se Extraen del Click?

#### **Temporales:**

Print\_server\_timestamp date

#### **Booleanas:**

free\_shipping accepts\_mercadopago fulfillment boosted

#### **Unicas:**

ROW\_ID site\_id user\_id deal\_print\_id

#### Categoricas:

benefit logistic\_type main picture listing\_type\_id is\_pdp domain id item id category\_id product id warranty title tags full name

#### **Target:** conversion

#### **Numericas:**

total\_si\_item\_30days total\_si\_sel\_30days total\_visits\_domain total\_visits\_item total\_visits\_seller health product\_id offset original\_price price print\_position qty\_items\_dom qty\_items\_sel sold\_quantity available\_quantity

avg\_gmv\_item\_domain\_30days avg gmv item sel avg\_gmv\_seller\_bday avg\_qty\_orders\_item\_domain\_30days avg\_qty\_orders\_item\_sel\_30days avg\_si\_item\_sel\_30days total\_asp\_item\_domain\_30days total\_asp\_item\_sel\_30days total\_gmv\_item\_30days total items domain total items seller total orders domain 30days total\_orders\_item\_30days total\_orders\_sel\_30days total si domain 30days

#### **Base de Datos**

#### **Dimensiones**

199.972 registros de impresiones 57 variables

#### **Train & Test Split**

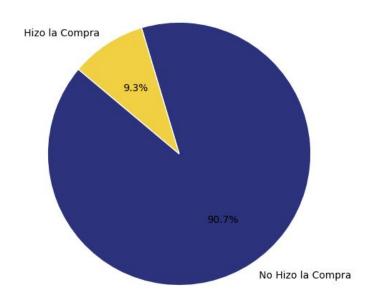
80%

Stratified Split

#### **Cross Validation**

5 folds

#### Proporción de Conversiones:



### **Técnicas Utilizadas**

#### **Análisis de Nulos**

- Imputación con mediana.
- Is\_pdp: comportamiento distinto en caso de ser nulo.

# **Encoding Var. Categoricas**

Se analizó qué técnica de Encoding fue es la más efectiva en cada caso. OHE, LE, OE, FE y TE

#### **Procesamiento**

- Entendimiento y procesamiento de variables de texto.
- Escalamiento de variables numéricas.
  - Analisis de Outliers

#### **Balanceos**

La base está balanceada con un

ratio 10:1. Se realizaron distintas técnicas de balanceo OS, US, LT, NM y SM

# **Encoding Var. Temporales**

Análisis de hora, día de semana y dia cómo variables cíclicas.

#### **Optimizaciones**

Se realizaron distintas técnicas de optimización RS, GS, BS, OP



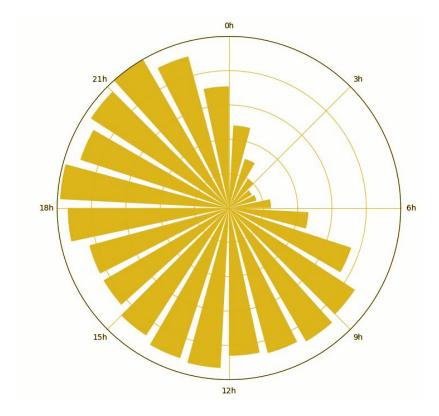
# **Encoding Variables Temporales**



#### Seno y Coseno

- Horas
- Día de la Semana
- Día

2020-03-01 a 2020-04-25



# **Encoding Variables Categoricas**



#### **Target Encoding**

- Category Id
- Domain Id
- Item Id
- Name
- Conversion



#### **One-Hot Encoding**

- Listing Type Id
- Logistic Type
- Platform
- Cat Warranty
- Is Pdp

# Comparación de modelos optimizados

#### **Modelos**

Random Forest XGBoost LightGBM Decision Tree

#### Balanceo

Links de Tomek Near Miss Over Sampling Under Sampling Smote

# Optimización (Auc-Pr)

Random Search Bayes Search Optuna

#### **Mejores Resultados:**

• Modelo: Decision Tree

Balanceo: Smote

• Optimizacion: Random Search

Modelo: Decision Tree

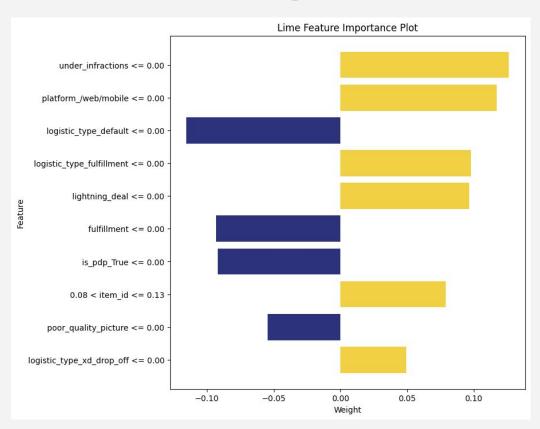
Balanceo: Under Sampling

• Optimizacion: Random Search

ACC Global = 0.67 Recall Clase 1= 0.89

ACC Global = 0.78 Recall Clase 1= 0.83

# Feature importance





## Recomendación Final



Plataforma web



**Infracciones** 



Logística

1

2

3

# **GRACIAS!**

**Preguntas?** 



**CREDITS:** This presentation template was created by <u>Slidesgo</u>, and includes icons by <u>Flaticon</u>, and infographics & images by <u>Freepik</u>

Please keep this slide for attribution