

PROYECTO

El Desafío de Don Rene

Integrantes:

- Consuelo Rojas N.
- Daniel Carmona G.



TABLE OF CONTENTS

01.

INTRO

Descripción del
problema que aqueja a
Don Rene

02.

EDA

Exploración de los datos
y relaciones

03.

PRE PROCESAMIENTO

Codificación y
conversión de datos

04.

MODELOS

Modelos usados y sus
parámetros

05.

RESULTADOS

Experimentales y de la
competencia

06.

CONCLUSIÓN

Sobre el modelo,
resultados y proyecto



01. Don Rene

Don René quiere ver si se puede predecir la clasificación de un juego y sus ventas si se tienen algunos datos previos, como lo son su precio

01. DATOS

OBJECTS

Características con strings o listas con strings. Ejemplo: name, developer, publisher, platforms, etc.

O

I

INT

Características numéricas. Como: required_age, achievements, average_playtime, price

RATING

Valoración por el público del juego. Cuenta con 5 clases: negativos, mixtos, mayormente positivos, positivos y muy positivos.

R

E

ESTIMATED_SELLS

Estimación de las ventas que produciría un juego en específico.

02.EDA

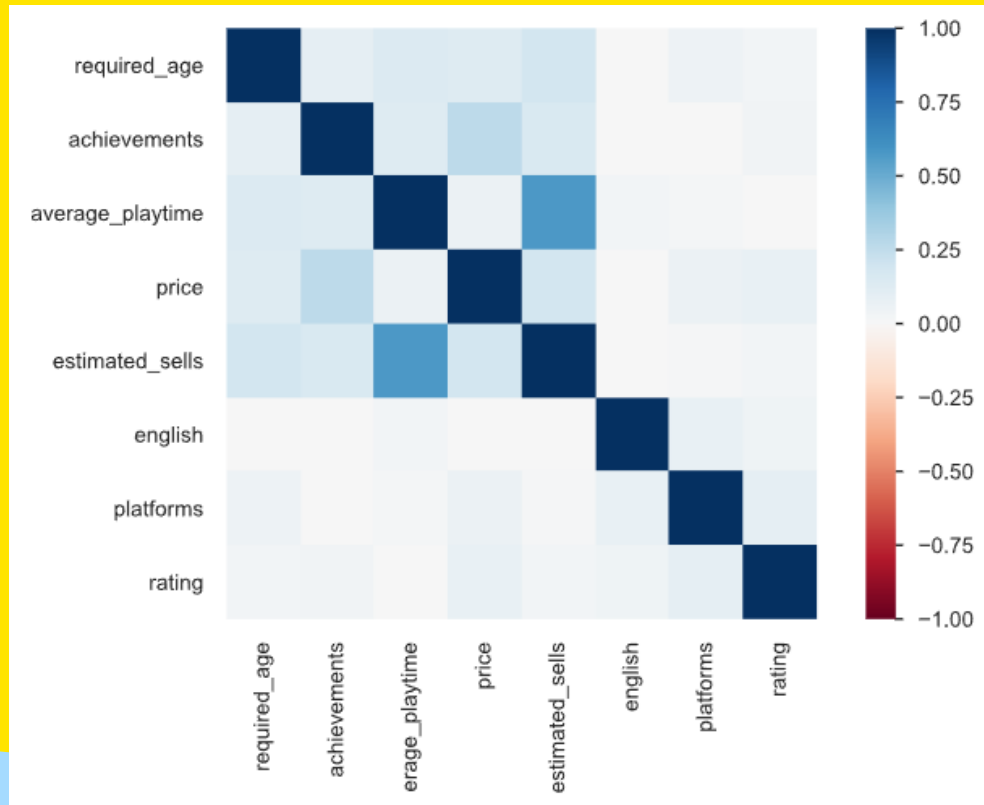
Análisis exploratorio de los datos



02. COMPOSICIÓN POR PLATAFORMA

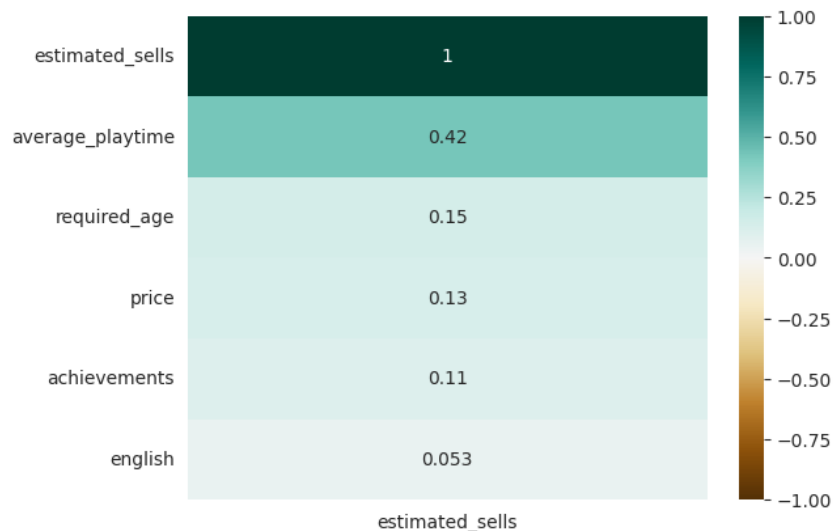


02. MATRIZ DE CORRELACIÓN

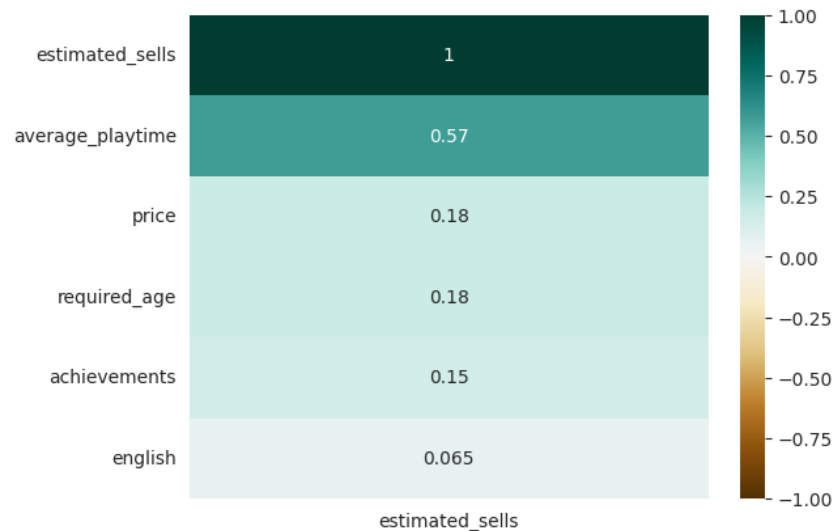


02. CORRELACIÓN

Features Correlating with estimated sells by Kendall method



Features Correlating with estimated sells by Spearman method



02. DISTRIBUCIÓN DE VENTAS Y RELACIÓN





03. PRE PROCESAMIENTO

WHAT SETS US APART?



NLP

CountVectorizer() y
StemmerTokenizer(),
Uni, bi y tri-gramas



CODIFICACIÓN

Diccionario con entidades
codificadas



NORMALIZACIÓN

Aplicación de
MixMaXScaler(),
StandarScaler()

04. MODELOS



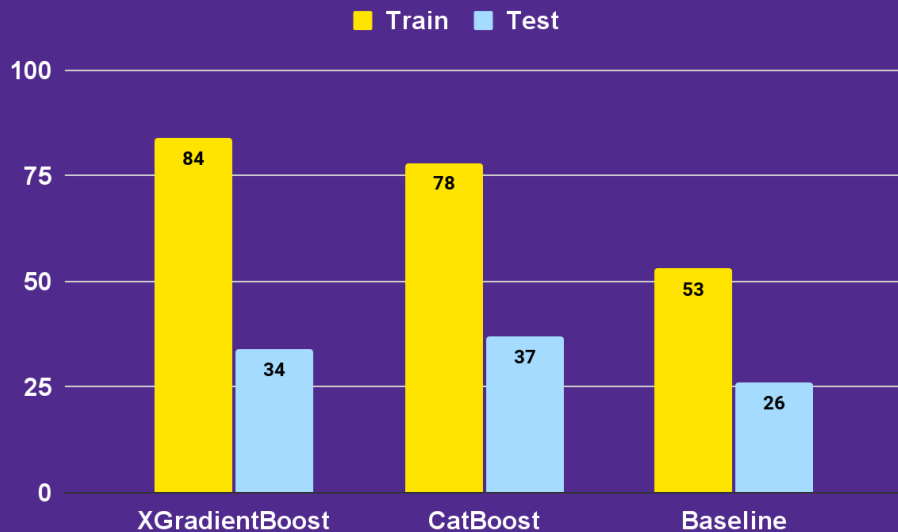
04. MODELOS

	CATBOOST	XGBOOST
TREE SYMMETRY	Symmetric	Asymmetric, level-wise tree growth
SPLITTING METHOD	Greedy method	Pre-sorted and histogram-based
TYPE OF BOOSTING	Ordered	-
Categorical Support	Support (perform one-hot encoding)	Do not support



05. RESULTADOS

05. RESULTADOS EXPERIMENTOS



Best f1-weighted: Cat Boost

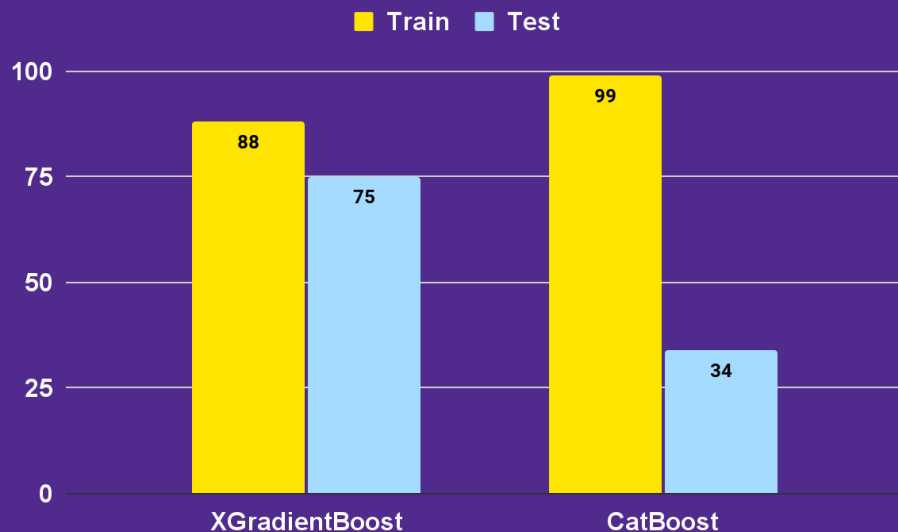


TRAIN



TEST

05. RESULTADOS EXPERIMENTOS



Best r2-Score:
GradientBoost*

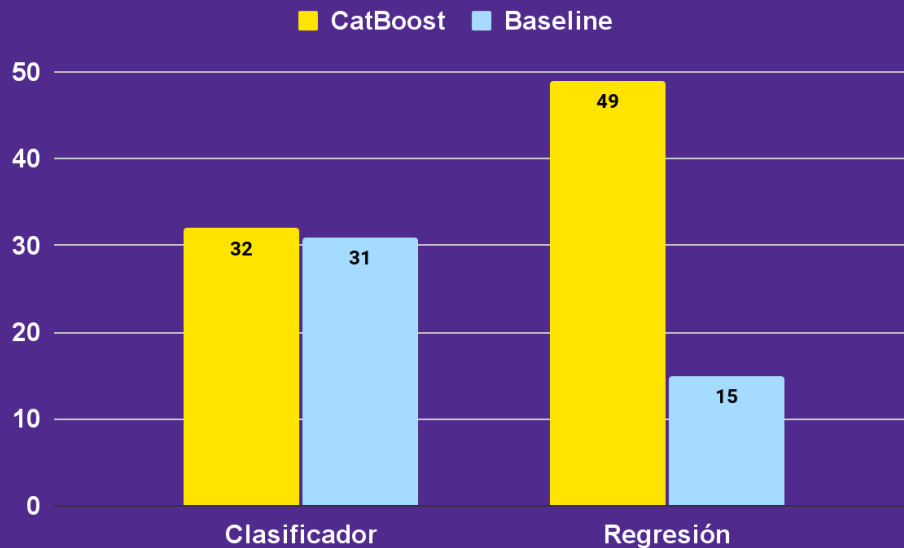


TRAIN



TEST

05. RESULTADOS COMPETENCIA



Best f1-
weighted: Cat
Boost



CATBOOST

BASELINE

06. CONCLUSION





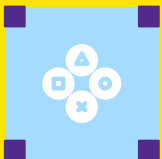
NO CONCLUYENTE

Con estos resultados, no se puede concluir la posibilidad de predecir la clasificación de ratings y estimación de ventas



MODELOS Y RESULTADOS

Los resultados de los modelos seleccionados hacen sentido, dado que CatBoost es una implementación de Gradient Boosting, que mejora aspectos de overfitting.



COMPLEJIDAD

El proyecto era una tarea compleja. La articulación de muchos atributos de distintos tipos sin que pierdan relevancia es complejo.

PROYECTO

El Desafío de Don Rene

Integrantes:

- Consuelo Rojas N.
- Daniel Carmona G.

