

# Video Games Sales

Eduardo Rodríguez Gil, ITC A01274913, Tecnológico de Monterrey

Este documento muestra la implementación y explicación del dataset de Video game sales.

## I. INTRODUCCIÓN

Los Videojuegos son una enorme industria que ha ido creciendo a lo largo de los años. Donde año tras año su popularidad crece gracias a la gran variedad de juegos que se llegan a vender en el mercado. Todo esto es posible por la variedad de tipos de juegos que existen como los shooter, battle royal, open world, etc. Por esta razón veremos las ventas del mundo y de Norteamérica en millones, de algunas copias de los videojuegos más famosos que actualmente existen.



Figura 1. Video games.

## II. DATA SET

El dataset lo obtuve de Kaggle, donde contiene un database de más de 15,000 videojuegos diferentes, donde están separados en diferentes categorías como Global\_Sales, Na\_Sales, etc. El dataset se redujo a 600 videojuegos aproximadamente para poder hacer uso de los datos más fácil.

	Global_Sales	NA_Sales
0	82.74	41.49
1	40.24	29.08
2	35.82	15.85
3	33.00	15.75
4	31.37	11.27

Figura 2. Data set de vgsales.

## III. MODELO EN TRAIN, TEST Y VALIDATION

Para la realización del modelo separamos los datos en dos grupos. El primer grupo es el de Train model, donde se agregan varios datos para poder realizar las predicciones del modelo y la gráfica de regresión lineal (Figura 3).

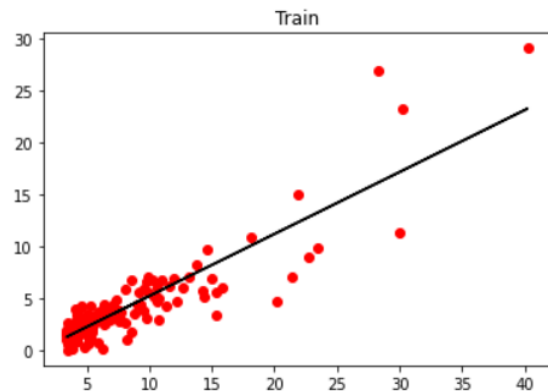


Figura 3. Train model.

Nuestro segundo modelo es el test model, donde de igual manera se agregan datos aleatorios para de igual forma poder realizar las predicciones del modelo y la gráfica de regresión lineal (Figura 4).

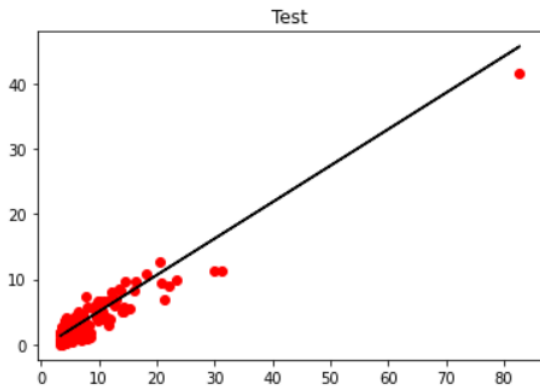


Figura 4. Test model.

Por último, nuestros últimos datos fueron los de validation que igual que los anteriores modelos tomábamos en muestra datos de manera aleatoria para la realización de la predicción y su respectiva gráfica (Figura 5).

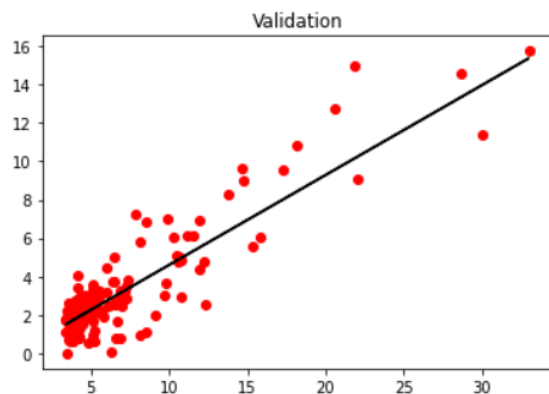


Figura 5. Validation model.

En la siguiente figura (Figura 6) podemos ver los datos que tomamos para cada una de las pruebas tanto de train, test y validation juntas en una misma gráfica.

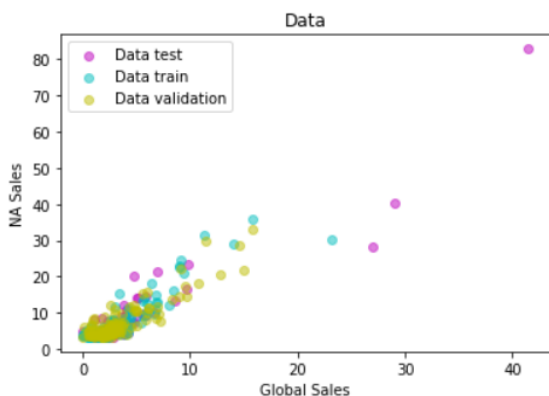


Figura 6. Data.

#### IV. BIAS O SESGOS

El bias con el que cuenta el modelo es bajo. Esto se debe a que los errores que llega a obtener el modelo tanto en train y en test son bajos. Diciéndonos que el porcentaje de predicción es muy alto como se puede observar en la (Figura 7), de cinco predicciones que se realizaron con datos aleatorios nos damos cuenta que nuestro modelo es bastante bueno, siendo así 72% en train como la predicción más baja y 76% en test de las cinco veces que se realizó.

Train	Test	Validation
0.8116830 377422517	0.8605875 893203188	0.8947638 261727263
0.8934909 726564324	0.8317349 892082548	0.8182102 977098757
0.8528620 079379314	0.8796299 43029435	0.9029446 896274408
0.8980415 083723439	0.8613453 147837599	0.8461249 584869412
0.8240328 90316858	0.8829078 234255965	0.8457692 418361913

Figura 7. Tabla de predicciones

#### V. VARIANZA

Para nuestro grado de varianza como podemos observar en la (Figura 7) para cada cambio de datos vemos que nunca son grandes los cambios. Siempre llega a cambiar la precisión un mínimo en train y test. Por lo mencionado anteriormente podemos deducir que nuestro grado de varianza es bajo.

#### VI. NIVEL DE AJUSTE

Volviendo a la (Figura 7) nos damos cuenta que en algunas ocasiones por un mínimo porcentaje de error el train model

es más bajo que el test model. Pero ya que los valores de nuestros datos mayormente se encuentran más óptimos podemos decir que el nivel de ajuste de nuestro modelo es óptimo.

## VII. MEJORA DEL MODELO

Para probar una posible mejora con el modelo decidí realizar el método de Hyperparameter. Donde importamos de sklearn (svm) para poder generar un hiperplano óptimo para categorizar nuevos ejemplos.

Dado eso le dimos los posibles valor de nuestro hiper parámetro y le dimos un accuracy a nuestro modelo. Para así realizar el proceso de hyperparameter y obtener nuestro score del modelo.

El método es sencillo de aplicar en el modelo ya que el modelo ya está hecho solo es unirle la extensión del proceso de hiper parámetro.

Lamentablemente el proceso de hiper parámetro no dio los resultados esperados, ya que en mi punto de vista el modelo ya daba un score bastante alto y ya era bastante bueno, ya que lo podíamos llegar a considerar un modelo óptimo. No obstante existen otros métodos de mejora que tal vez sean más eficientes para el modelo y que le den una mejor calidad al modelo.

En la (Figura 9) podemos observar los resultados de 5 pruebas que se hicieron con el método de hiper parámetro y podemos darnos cuenta que aunque no sean tan mal los resultados, el modelo se encuentra menos óptimo llegando a ser overfitting.

Train	Test
0.8237546695458102	0.7540027399105368
0.8747290509586776	0.8056970406721785

0.8006117675596898	0.7648451526190634
0.9001075105014846	0.838366690457036
0.8585578589545844	0.7727610654485644

Figura 9. Tabla de predicciones mejorada.

## VIII. CONCLUSIONES

Es notorio que el modelo realiza buenas predicciones y cuenta con errores demasiado bajos, al grado de ser un modelo óptimo. Esto sucede gracias a la reducción de datos que se hizo en un inicio al dataset y a la mejora que se le fue realizando al modelo en el transcurso aumentando el tamaño del conjunto de entrenamiento. Aunque en algunos puntos todavía se podría hacer una limpieza más a fondo sobre el dataset y esperar aún mejores resultados.

## IX. REFERENCIAS

[1] *Overfitting in Machine Learning* - Javatpoint. (2021). [www.javatpoint.com](https://www.javatpoint.com).

Recuperado 9 de septiembre de 2022, de <https://www.javatpoint.com/overfitting-in-machine-learning#:~:text=%20What%20is%20Overfitting%3F%20%20%20Overfitting%20%26,test%2Funseen%20dataset%20and%20can%E2%80%99t%20generalize%20well.%20More%20>

[2] Canadas, R. (2022, 29 mayo). *Qué es el bias en estadística y machine learning*. abdatum. Recuperado 9 de septiembre de

2022, de  
<https://abdatum.com/tecnologia/que-es-bi>  
[as](#)

[3] Valdés, B. (2022). *Machine Learning*.  
Recuperado 9 de septiembre de 2022, de  
[https://docs.google.com/document/d/1WO](https://docs.google.com/document/d/1WOv6P6BzoFV0x5bN2PW_1MwP4JLdw_To/edit)  
[v6P6BzoFV0x5bN2PW\\_1MwP4JLdw\\_To/](#)  
[edit](#)

[4] Smith, G. (2016, 26 octubre). *Video Game Sales*. Kaggle. Recuperado 9 de  
septiembre de 2022, de  
[https://www.kaggle.com/datasets/gregorut/](https://www.kaggle.com/datasets/gregorut/videogamesales)  
[videogamesales](#)