

# 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 Y TEST

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).

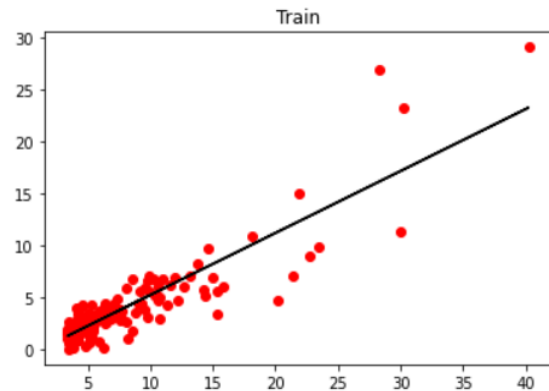


Figura 3. Train model.

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).

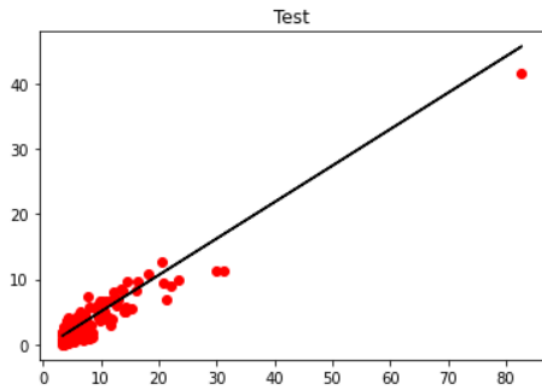


Figura 4. Test model.

#### 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 5), de diez 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 diez veces que se realizó.

Train precision	Test precision
0.811683037742 2517	0.860587589320 3188
0.893490972656 4324	0.781734989208 2548
0.782862007937 9314	0.879629943029 435
0.898041508372 3439	0.761345314783 7599
0.724032890316 858	0.882907823425 5965
0.871997739854 5051	0.787807960363 56
0.869664381631 0312	0.768091714463 6605
0.773474226635	0.877164778409

0756	1166
0.802088647727 0343	0.827514367237 246
0.795681137414 9417	0.870026650035 1539

Figura 5. Tabla de predicciones

#### V. VARIANZA

Para nuestro grado de varianza como podemos observar en la (Figura 5) 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 5) 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 poder lograr una buen desempeño del modelo fue necesario probar con diferentes tamaños del conjunto de entrenamiento, ya que en un inicio el porcentaje de predicción era menor a 50%. Empecé a ampliar el tamaño del conjunto de entrenamiento y me comencé a percatar que la predicciones que tenían eran cada vez mejores. Con esto podemos decir que entre más ampliemos el tamaño del conjunto de entrenamiento mejor va a ser nuestra predicción y por ende un mejor modelo.

## 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](http://www.javatpoint.com). Recuperado 9 de septiembre de 2022, de <https://www.javatpoint.com/overfitting-in-machine-learning#:~:text=%20What%20is%20Overfitting%3F%20%201%20Overfitting%20%26,test%20Unseen%20dataset%20and%20can%20%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-bias>

[3] Valdés, B. (2022). *Machine Learning*. Recuperado 9 de septiembre de 2022, de [https://docs.google.com/document/d/1WOv6P6BzoFV0x5bN2PW\\_1MwP4JLdw\\_To/edit](https://docs.google.com/document/d/1WOv6P6BzoFV0x5bN2PW_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/videogamesales>

