

Predicción de Ventas de Videojuegos en América del Norte

Proyecto de Machine Learning -
Resultados y Conclusiones

Problema de Negocio

- Predecir las ventas en América del Norte (NA_Sales) de un videojuego
- Usando características como Año, Plataforma, Género, Publisher, y ventas en otras regiones.
- Objetivo: ayudar a compañías a tomar decisiones de mercado basadas en datos históricos.

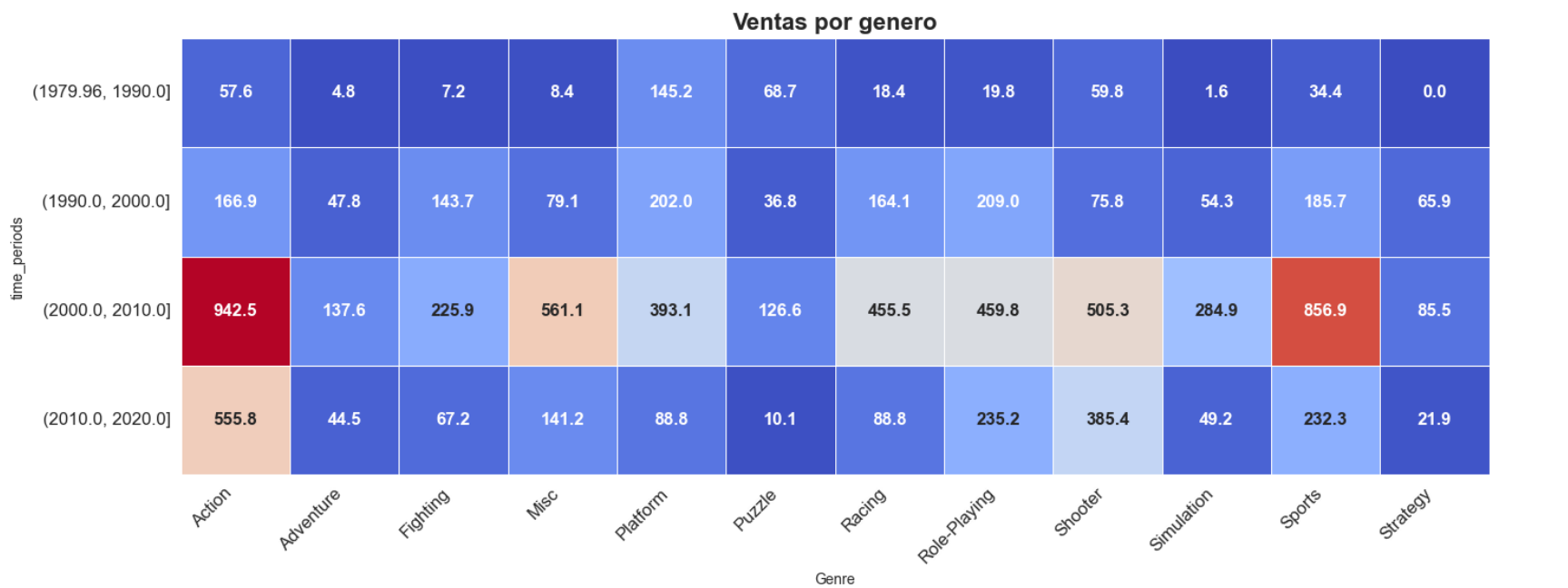
Dataset utilizado

- Fuente: Kaggle - Video Game Sales
- Filas: ~16.000 | Columnas: 11
- Variables categóricas: Platform, Genre, Publisher
- Variables numéricas: Year, ventas por región
- Target: NA_Sales

Análisis Exploratorio de Datos (EDA)

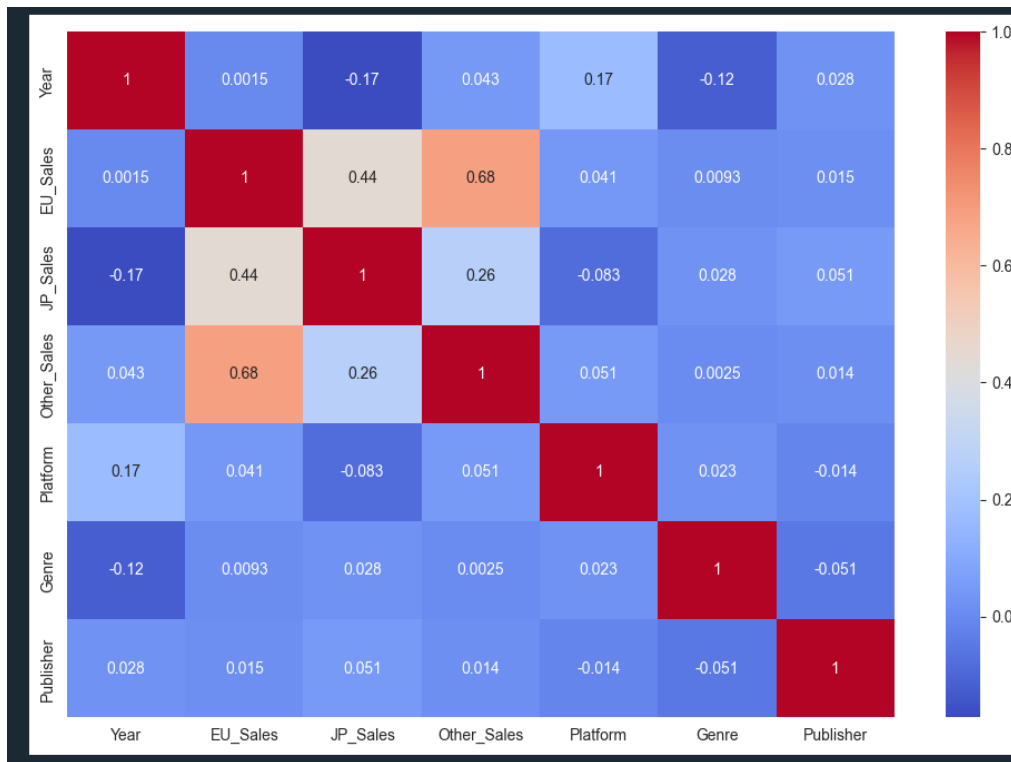
- Se analizaron valores nulos, distribuciones y correlaciones.
- Se eliminaron columnas irrelevantes: Rank, Name, Global_Sales.
- Se observaron correlaciones entre NA_Sales y otras regiones (EU, JP, Other).

Rank		Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	1	Wii Sports	Wii	2006.0	Sports	Nintendo	41.49	29.02	3.77	8.46	82.74
1	2	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24
2	3	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.85	12.88	3.79	3.31	35.82
3	4	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.75	11.01	3.28	2.96	33.00
4	5	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37



Preprocesamiento

- Separación de variables categóricas y numéricas.
- Escalado de variables numéricas con StandardScaler.
- Codificación ordinal para variables categóricas.
- División del dataset en train y test (80/20).



Modelos Probados

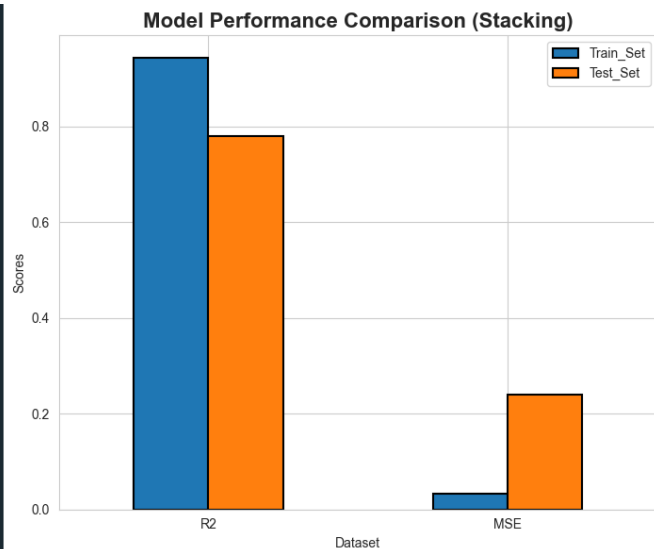
- Regresión Lineal
- K-Nearest Neighbors (KNN)
- Árbol de Decisión
- Modelo Final: Stacking Regressor con Gradient Boosting como meta-modelo

Evaluación del Modelo

- Métricas utilizadas: R^2 (score), MSE (error cuadrático medio)
- Evaluación en conjunto de entrenamiento y prueba
- Validación cruzada para asegurar consistencia
- Visualización de resultados en gráfico de barras + tabla

Resultados

- Modelo final (Stacking + Gradient Boosting):
 - R^2 en entrenamiento: ~ 0.92
 - R^2 en prueba: ~ 0.88
 - Mejora significativa respecto a modelos individuales.



	Train_Set	Test_Set
R2	0.9430360416423594	0.7792002052030316
MSE	0.03265025879511125	0.2403995414050363

Conclusiones y Mejoras Futuras

- El modelo muestra buen rendimiento general.
- Stacking mejora frente a modelos por separado.
- Posibles mejoras: tuning de hiperparámetros, más variables, uso de LightGBM.
- El proyecto demuestra una aplicación práctica del ML en predicción de mercado.

