



Universidad Tecnológica Metropolitana

Departamento de Informática y Computación

Trabajo N°3:
Analítica Predictiva y Prescritiva

PREDICCIÓN DE VIDEOJUEGOS EXITOSOS

Sebastián Celedón - Thiare Guerrero

Docente Ana Moya Beltrán

Asignatura Analítica

Índice

1	Introducción y motivación	2
2	Hipótesis y planteamiento del problema	3
3	Resultados previos	4
3.1	Éxito por género:	5
3.2	Diferencias significativas:	5
4	Metodología	5
4.1	Definición del problema	5
4.2	Selección de datos	6
4.3	Preparación de los datos	6
4.4	Elección del modelo	6
4.5	Entrenamiento del modelo	7
4.6	Pruebas realizadas	8
5	Resultados	9
5.1	Resultados del modelo	10
5.2	Pruebas para evaluar Overfitting	13
6	Análisis prescriptivo	13
6.1	Análisis y validación de resultados	13
6.2	Recomendaciones	17
7	Bibliografía	18

1 Introducción y motivación

Los géneros de videojuegos, como indie y casual, han demostrado tener una influencia notable en la percepción y aceptación de los jugadores. Sin embargo, las preferencias del consumidor y los factores que impulsan la popularidad y rentabilidad de un videojuego no son completamente predecibles debido a la interacción compleja entre variables como la cantidad de reseñas, valoraciones, precios y plataformas disponibles.

Para el análisis predictivo nuestro objetivo es anticipar el éxito de un videojuego en función de datos históricos y patrones identificados, a través de un modelo “Random Forest”, este enfoque estima la probabilidad de que un videojuego sea exitoso basándose en variables clave como el porcentaje de reseñas positivas, la cantidad de reseñas totales, el género y las plataformas. Esto permite identificar tendencias futuras y proyectar resultados, ayudando a planificar estrategias basadas en datos.

Por otra parte, el análisis prescriptivo busca proporcionar recomendaciones prácticas basadas en los resultados del análisis predictivo. Una vez identificados los videojuegos con alta probabilidad de éxito, este enfoque sugiere acciones específicas, como priorizar la promoción de ciertos géneros, desarrollar videojuegos para múltiples plataformas o enfocar recursos en proyectos con mayor rentabilidad proyectada. Así mismo para aquellos juegos con baja probabilidad de ser exitosos se recomendaría no realizar un nuevo volumen de estos.

El dataset utilizado fue generado mediante el mismo Web-scraping anterior en el entorno “Google Colab Python” desde la sección de novedades de Steam, sin embargo, esta vez se recopiló información de los 4236 videojuegos más recientes, el cual contiene variables clave como género, precio, plataformas, fecha de lanzamiento, cantidad de reseñas y porcentaje de reseñas positivas las cuales fueron seleccionadas por su relevancia en la evaluación del éxito de un videojuego, combinando datos cualitativos y cuantitativos para un análisis integral.

Esta elección se justifica por su representatividad de las tendencias actuales del mercado, su diversidad de características y su origen confiable en Steam, una plataforma líder en la industria de videojuegos.

2 Hipótesis y planteamiento del problema

En la industria de los videojuegos, predecir el éxito de un título antes de su lanzamiento es un desafío crucial que puede definir la rentabilidad de un proyecto. Las decisiones relacionadas con el desarrollo, marketing y distribución suelen basarse en experiencias previas o intuición, lo que puede resultar en inversiones no óptimas. Para abordar este problema, es necesario utilizar técnicas de analítica predictiva y prescriptiva que identifiquen patrones clave en los datos históricos para que brinden recomendaciones prácticas y así maximizar el impacto de los videojuegos en el mercado.

El problema predictivo se define como una tarea de clasificación, donde el objetivo es predecir si un videojuego será exitoso. Se considera éxito a los juegos que cumplan con las siguientes condiciones:

- Más de 100,000 reseñas totales y un porcentaje de reseñas positivas igual o superior al 85%,
- O más de 50,000 reseñas y un porcentaje de reseñas positivas superior al 90%,
- O más de 10,000 reseñas y un porcentaje de reseñas positivas superior al 85%.
- Si no se cumple ninguna de estas condiciones, el juego no se considera exitoso.

Esto permite explorar patrones que ayudan a identificar los factores clave que determinan el desempeño en términos de popularidad y aceptación. Cabe mencionar que la definición de estos parámetros proviene de la analítica descriptiva realizada con anterioridad.

Como hipótesis planteada: **“Los géneros más exitosos son Indie y casual, además la cantidad de reseñas es más influyente que el porcentaje de positividad de un juego”.**

→ **Variables que influyen en la predicción:**

1. **Porcentaje de reseñas positivas:** Refleja la percepción general de los jugadores y es un indicador directo de satisfacción. Juegos con una alta tasa de positividad suelen captar una audiencia más amplia y generar mayor confianza en futuros compradores.
2. **Cantidad de reseñas totales:** Es un proxy del alcance del videojuego en el mercado. Los juegos con más reseñas tienden a haber alcanzado una mayor base de usuarios, lo que sugiere popularidad y visibilidad.
3. **Género del videojuego:** Algunos géneros tienen mayor demanda en el mercado. Esto aumenta sus probabilidades de éxito frente a géneros de nicho.

Las predicciones impactan la toma de decisiones al permitir la optimización de recursos, priorizando videojuegos con mayor probabilidad de éxito y evitando inversiones en proyectos con bajo potencial. Además ayudan a reducir riesgos financieros, guiar el diseño de productos más atractivos basados en características exitosas y planificar estrategias de marketing enfocadas en títulos prometedores.

3 Resultados previos

En los resultados previos del Trabajo N°2, nuestro estudio empleó técnicas de analítica diagnóstica, como ANOVA y Tukey, para identificar patrones y diferencias entre géneros. Esto permitió evaluar si el éxito en videojuegos se concentra en el género de acción o si otros géneros emergentes como Indie y Casual obtienen mejores resultados.

La hipótesis principal fue: "La mayoría de los juegos exitosos en Steam pertenecen a géneros de acción.". Esta afirmación se basa en el análisis descriptivo previo y en la percepción de que los juegos de acción ofrecen experiencias inmersivas, desafiantes y rejugables, atrayendo a un público amplio. Para verificarlo, se realizó un análisis exhaustivo considerando factores como valoraciones positivas, cantidad de reseñas, precios y géneros de videojuegos.

- Abarcando a los resultados previos como tal:

3.1 Éxito por género:

Los géneros como *Indie* y *Casual* mostraron consistentemente valoraciones positivas más altas en comparación con géneros tradicionales como *Acción* y *Aventura*.

- Estos géneros tienen una desviación estándar baja, lo que indica consistencia en las opiniones.

Generos	mean	std
Indie	88.391304	9.389201
Casual	85.150327	12.966440
Early Access	84.460870	10.831474
Simulation	84.053333	12.544122
Adventure	83.809524	12.578749
Strategy	83.395062	11.789489
Action	82.560074	13.393980
RPG	82.057143	13.678633
Racing	80.657143	15.268081
Sports	79.307692	14.043455
Free To Play	74.366337	14.433103
Massively Multiplayer	74.039604	12.664850

Figura 1. Juegos agrupados por media y desviación estándar

3.2 Diferencias significativas:

El análisis ANOVA y las pruebas de Tukey confirmaron que los géneros *Indie* y *Casual* tienen mayor aceptación que el género *Acción*. Esto refuta la hipótesis inicial.

Resultados de ANOVA: Valor F: 16.01 Valor p: < 0.0001		Grupo 1	Grupo 2	Diferencia de Medias	Valor p ajustado	Se rechaza la hipótesis
Prueba de Levene: 3.51		Action	Adventure	1.6789	0.7153	False
		Adventure	Casual	1.2625	0.9974	False
		Adventure	Massively Multiplayer	-8.8953	0	True
		Action	Free To Play	-8.0709	0	True
		Free To Play	Indie	14.0381	0	True
		Indie	Massively Multiplayer	-14.5766	0	True
		Action	Indie	5.3489	0	True

Figura 2, 3 y 4. Test ANOVA Levene y Tukey.

4 Metodología

4.1 Definición del problema

Se plantea como objetivo predecir si un videojuego será exitoso basado en variables cuantitativas y categóricas, como porcentaje de positividad, número de reseñas, y géneros.

Nombre
Total de Reseñas
Porcentaje de Positividad
Generos

Figura 5. Variables utilizadas.

4.2 Selección de datos

Se utilizó un dataset preprocesado con información sobre **4236** videojuegos, incluyendo métricas clave como porcentaje de positividad y cantidad de reseñas, para establecer la variable objetivo ('Éxito').

Nombre	Precio	Fecha de Lanzamiento	Total de Reseñas	Porcentaje de Positividad	Generos
Counter-Strike 2	Free	Aug 21, 2012	8409009	86	Action,Free To Play
S.T.A.L.K.E.R. 2: Heart of Chernobyl	\$59.99	Nov 20, 2024	32401	78	Action,Adventure,RPG
Call of Duty®: Black Ops 6	\$69.99	Oct 24, 2024	7221	56	Action
Warframe	Free	Mar 25, 2013	599017	87	Action,RPG,Free To Play
THRONE AND LIBERTY	Free	Oct 1, 2024	51975	68	Action,Adventure,Massively Multiplayer,RPG,Free To Play
Farming Simulator 25	\$49.99	Nov 12, 2024	8515	74	Simulation

Figura 6. Dataset videojuegos steam.

4.3 Preparación de los datos

→ **Limpieza de datos:** Se eliminaron columnas irrelevantes y las categóricas se convirtieron en variables dummy para permitir su uso en modelos matemáticos.

Estructura del dataset:
<class 'pandas.core.frame.DataFrame'
RangeIndex: 4237 entries, 0 to 4236
Data columns (total 8 columns):

0	Nombre	4237 non-null	object
1	Precio	4237 non-null	float64
2	Fecha de Lanzamiento	4237 non-null	object
3	Plataformas	4237 non-null	object
4	Total de Reseñas	4237 non-null	int64
5	Porcentaje de Positividad	4237 non-null	int64
6	Generos	4237 non-null	object
7	Éxito	4237 non-null	int64

Figura 7 y 8. Estructura del dataset con su tipo de datos después de la limpieza.

4. 4 Elección del modelo

En base a nuestro dataset, tomamos la decisión de escoger el algoritmo de “Random Forest”, el cual fue elegido por las siguientes razones:

→ Capacidad para manejar conjuntos de datos mixtos: El modelo trabaja con datos numéricos y categóricos gracias a <i>ColumnTransformer</i> , asegurando flexibilidad para diversos tipos de datasets.
→ Robustez frente a datos desbalanceados: Random Forests maneja bien datos desbalanceados, y en este caso, se utilizó sobremuestreo para balancear las clases.
→ Importancia de características: Permite identificar las variables más influyentes mediante el atributo <i>feature_importances_</i> .
→ Predicciones precisas y generalización: Combina múltiples árboles para reducir errores y mejorar la precisión de las predicciones.
→ Interpretabilidad: Aunque menos transparente que un árbol único, es posible analizar árboles individuales para entender las predicciones.

4. 5 Entrenamiento del modelo

- **Creación de la variable objetivo:** Se definió 'Éxito' como los videojuegos con más de un 90% de positividad. Para aquellos juegos que poseen más de 5,000 reseñas, la exigencia de positividad disminuye a un 84%. Esto se debe a que, a mayor cantidad de reseñas, el índice de positividad tiende a bajar, según lo **analizado en la página web de Steam**.
- **División de los datos:** El dataset se separó en conjuntos de entrenamiento (80% - 3389 datos) y prueba (20% - 848 datos), utilizando estratificación para mantener la proporción de la variable objetivo.
- **Codificación de datos categóricos:** Codificación de datos categóricos: Las variables categóricas como 'Géneros' y 'Plataformas' se transformaron en variables dummy (binarias) mediante *OneHotEncoder*, para que el modelo pudiera procesarlas.
- **Balanceo de clases:** Balanceo de clases: Dado que había más videojuegos no exitosos que exitosos, se utilizó sobremuestreo en la clase minoritaria ('Éxito'), generando un conjunto balanceado de entrenamiento.

Distribución de la variable 'Éxito': Éxito 0 2150 1 2087	--- Preparando datos para el modelo --- Datos de entrenamiento: 3389 Datos de prueba: 848
---	---

Figura 9. Balanceo de datos con sobremuestreo y datos de entrenamiento y prueba.

→ **Entrenamiento de Random Forest:** Se utilizó el modelo configurado con un estado aleatorio (`random_state=42`) para garantizar la reproducibilidad.

- Durante el entrenamiento:
 - El modelo combinó múltiples árboles de decisión, cada uno construido a partir de un subconjunto aleatorio de las características.
 - Estos árboles trabajaron en conjunto para tomar decisiones basadas en patrones presentes en los datos balanceados.

→ **Optimización interna del modelo:** Los Random Forests seleccionan aleatoriamente características y datos para reducir el riesgo de sobreajuste y mejorar la generalización.

- Cada árbol evaluó divisiones como:
 - **Porcentaje de positividad.**
 - **Cantidad de reseñas.**
 - **Géneros más influyentes.**

→ **Predicciones iniciales:** Después de entrenar el modelo, se aplicaron los datos de prueba para generar predicciones (`y_pred`), clasificando videojuegos como Éxito o No Éxito.

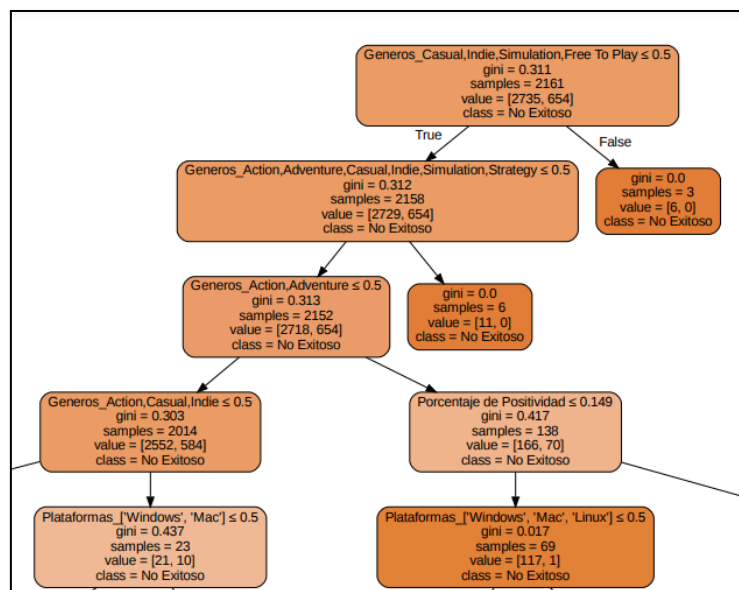


Figura 10. Inicio del árbol de decisión del modelo.

→ El árbol de decisión utiliza características clave del dataset, como género, reseñas positivas, precio y plataformas, para dividir los datos en pasos jerárquicos que clasifican un videojuego como "Exitoso" o "No exitoso". Cada nodo evalúa un umbral específico, refinando la predicción hasta los nodos hoja.

4. 6 Pruebas realizadas

Esta sección describe y contextualiza las pruebas aplicadas (visibles con sus resultados en la sección **5. Resultados**) con el fin de evaluar el desempeño del modelo y garantizar su robustez.

Estas pruebas son esenciales para validar su capacidad predictiva, identificar posibles problemas como el sobreajuste y asegurar su generalización a nuevos datos.

A continuación, se presentan las pruebas y su propósito.

→ Métricas de evaluación:

- **Precisión:** Mide la proporción de predicciones correctas respecto al total.
- **Recall:** Evalúa la capacidad del modelo para identificar correctamente los juegos exitosos.
- **F1-Score:** Combina precisión y recall, siendo ideal para datasets desbalanceados.
- **Curva ROC y AUC:** Permiten analizar la capacidad del modelo para distinguir entre clases exitosas y no exitosas.
- **Log Loss:** Ya que para Random Forest no encaja muy bien el RMSE (el cual se sugiere para este trabajo), decidimos aplicar Log Loss, el cual mide qué tan buenas son las probabilidades predichas por el modelo para clasificar correctamente las clases (Exitoso o No exitoso). Penaliza más las predicciones que asignan una probabilidad baja a la clase correcta, ayudando a evaluar la calidad de las predicciones probabilísticas del modelo.
- **Matriz de confusión:** Se utilizó para analizar los aciertos y errores del modelo en las clasificaciones, distinguiendo entre juegos exitosos y no exitosos. Proporciona una visión detallada del rendimiento, especialmente en el caso de errores específicos (falsos positivos y falsos negativos).

→ **Validación cruzada:** Se utilizó un esquema de validación cruzada de 10 folds para asegurar que el modelo generalice correctamente a datos no vistos. Este método divide el conjunto de datos en 10 partes, utilizando una para prueba y las otras nueve para entrenamiento, rotando este proceso hasta completar los folds.

→ Pruebas de sobreajuste:

- **Curvas de aprendizaje:** Analizan la diferencia entre las métricas de entrenamiento y validación para identificar posibles problemas de sobreajuste.
- **Curvas de ganancias acumuladas:** Permiten evaluar si el modelo presenta un comportamiento consistente y predecible frente a nuevos datos.

5 Resultados

El análisis del modelo de predicción de éxito de videojuegos se basó en las métricas de **precisión, recall, F1-Score, curva ROC y AUC, validación cruzada, y log loss**, las cuales permiten evaluar tanto su desempeño general como su capacidad de generalización. A continuación, se detallan los resultados obtenidos.

5.1 Resultados del modelo

→ **Desempeño del modelo:** El modelo alcanzó una precisión del **98.94%**, lo que implica que hizo predicciones correctas para la mayoría de las observaciones en el conjunto de prueba.

- **Clase 0 (No Exitoso):** Precisión de **0.99** y **recall** de **1.00**, lo que significa que clasificó correctamente casi todos los juegos no exitosos.
- **Clase 1 (Exitoso):** Precisión de **1.00** y **recall** de **0.94**, identificando correctamente la mayoría de los juegos exitosos.

Accuracy del modelo: 98.94 %					
Reporte de Clasificación:					
	precision	recall	f1-score	support	
0	0.99	1.00	0.99	687	
1	1.00	0.94	0.97	161	
accuracy			0.99	848	
macro avg	0.99	0.97	0.98	848	
weighted avg	0.99	0.99	0.99	848	

Figura 11. Reporte de clasificación.

→ **Matriz de confusión:** El modelo tiene una alta capacidad para predecir correctamente tanto los juegos exitosos como los no exitosos, con muy pocos errores (solo 9 juegos exitosos mal clasificados).

Valor real "No Exitoso":

- 687 juegos correctamente clasificados como No exitoso.
- 0 juegos mal clasificados como Exitosos.

Valor real "Exitoso":

- 9 juegos mal clasificados como No exitosos.
- 152 juegos correctamente clasificados como exitosos.

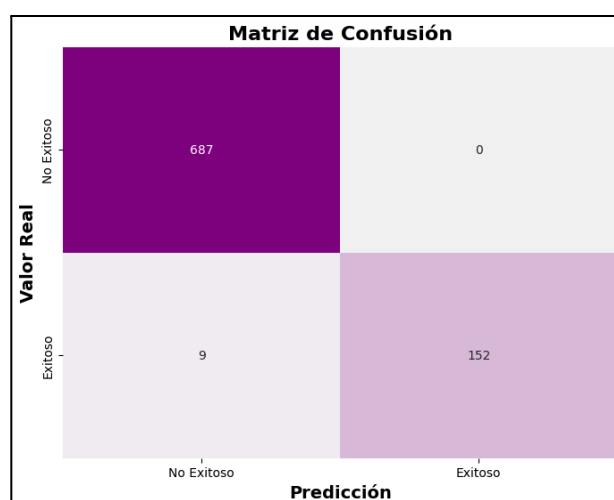


Figura 12. Matriz de confusión.

→ Curva ROC y AUC:

El modelo presentó una “curva ROC” con un AUC igual a **1.0**, confirmando que clasifica muy bien entre clases positivas y negativas en el conjunto de prueba. Pese a la diferencia de 9 juegos mal clasificados, esta es mínima en comparación con la cantidad total de juegos clasificados correctamente (848 muestras en total), lo que significa que el modelo sigue teniendo un rendimiento sobresaliente.

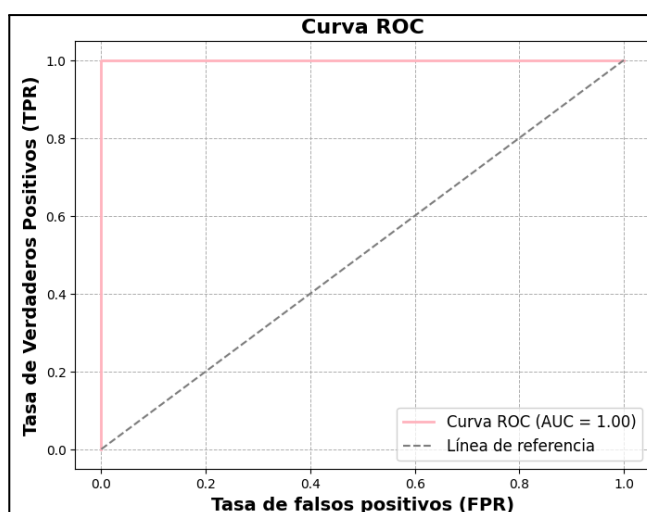


Figura 13. Curva ROC y AUC.

→ Log Loss:

El valor de Log Loss fue extremadamente bajo, lo que indica que el modelo produce predicciones probabilísticas muy precisas.

El Log Loss de **0.184** refuerza que el modelo Random Forest no solo clasifica con precisión, sino que también asigna probabilidades bien calibradas. Esto respalda la confiabilidad del modelo para identificar videojuegos exitosos basándose en las características seleccionadas.

Log Loss del modelo: 0.184

Figura 14. Resultado Log Loss

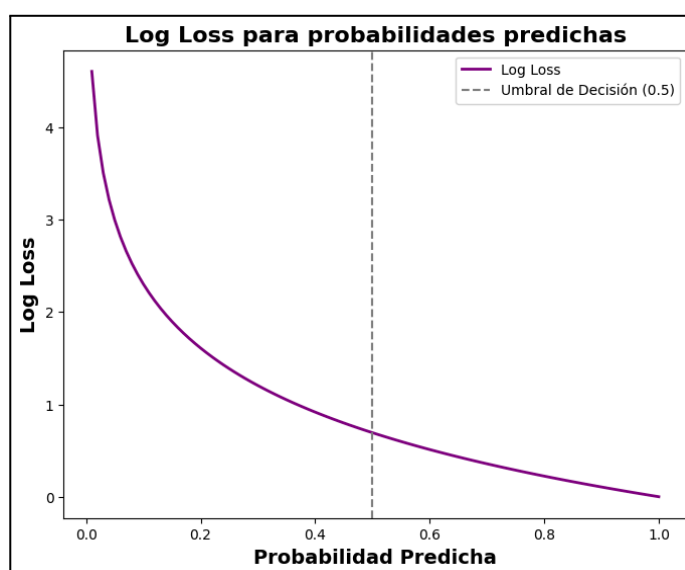


Figura 15. Curva Log Loss para probabilidades predichas.

→ Validación cruzada:

El modelo muestra un rendimiento sobresaliente en la validación cruzada, con puntuaciones entre 97.64% y 100% en los 10 folds. El promedio de precisión es 99.11%, y la desviación estándar de 0.93% indica una baja variabilidad. Esto confirma que el modelo es consistente y confiable, con alta precisión y poca fluctuación en su capacidad para generalizar.

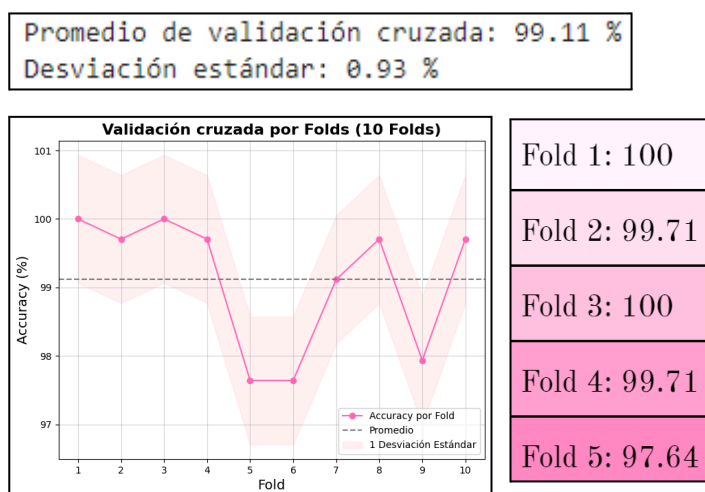
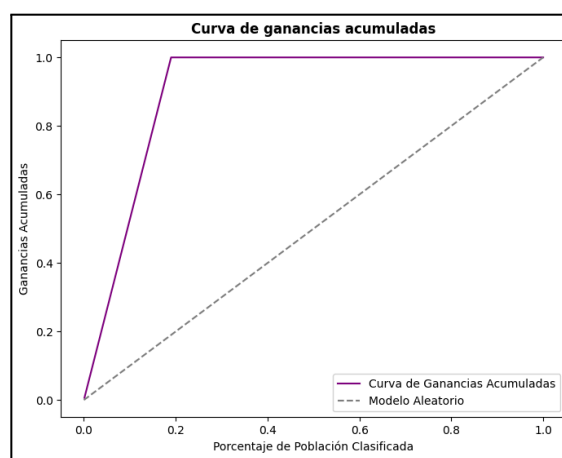
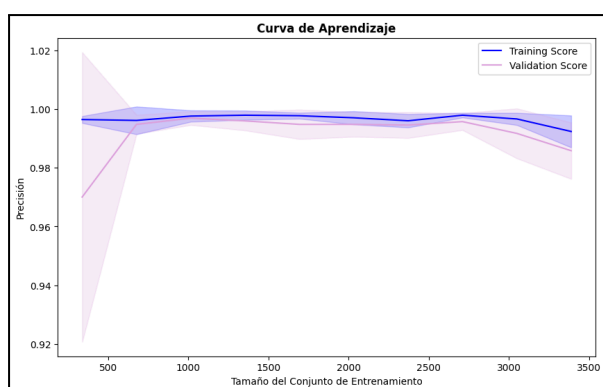


Figura 16. Validación cruzada y 5 primeros Folds.

5.2 Pruebas para evaluar Overfitting

Los resultados obtenidos muestran un rendimiento **casi perfecto** y muy consistente, por lo que se realizaron **3 pruebas** para confirmar la ausencia de sobreajuste (**overfitting**):



Accuracy en el conjunto de prueba: 98.94 %
Promedio de validación cruzada: 98.64 %
Desviación estándar en validación cruzada: 1.18 %

Figura 17, 18 y 19. Curva de aprendizaje, Curva de ganancias acumuladas y promedio de validación cruzada.

→ Dado los resultados, se concluye que no hay sobreajuste porque las curvas de aprendizaje muestran estabilidad entre las precisiones de entrenamiento y validación, la validación cruzada muestra resultados consistentes con baja desviación estándar, y la curva de ganancias acumuladas indica un buen rendimiento sin comportamiento aleatorio, **todo lo cual sugiere que el modelo generaliza bien.**

→ 6 Análisis prescriptivo

6.1 Análisis y validación de resultados

Este análisis tiene como objetivo proporcionar recomendaciones prácticas y estratégicas basadas en los resultados de las predicciones, ayudando a las partes interesadas (desarrolladores, inversores, etc.) a **tomar decisiones informadas** sobre qué videojuegos deberían ser promovidos o desarrollados.

→ Géneros con mayor probabilidad de éxito

Como primer análisis a realizar, tenemos un gráfico sobre los géneros con mayor cantidad de juegos exitosos según nuestro modelo, en el cual nos indica que los géneros; “Action”, “Indie”, “Adventure” y “Simulation” poseen la mayor cantidad de juegos exitosos en ese mismo orden. Esto refuta parcialmente la hipótesis anterior, ya que aunque los juegos Indie destacan como exitosos, el género Acción supera a Casual en términos de éxito predicho.



Figura 20. Géneros con mayor probabilidad de éxito.

→ Plataformas con mayor probabilidad de éxito

Por otra parte, como era de esperar la plataforma “Windows” se mantiene como la más exitosa, esto mayormente a que todos los videojuegos se ofrecen en ella, sin embargo se observa un cambio en las otras dos, antes “Linux” era considerada como la segunda más famosa y predominante, sin embargo el éxito de los videojuegos que se lanzan en ella no es tanto como lo son en “Mac”, ya que por 90 juegos este último es más exitoso que la plataforma de código abierto, cambiando así por completo nuestro análisis descriptivo.

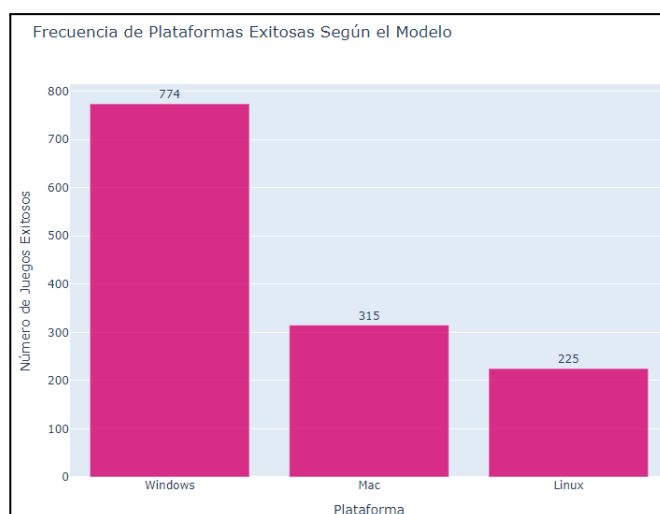


Figura 21. Plataformas con mayor probabilidad de éxito

→ Videojuegos con mayor probabilidad de éxito

Ya pasando al análisis de los videojuegos exitosos según nuestro modelo, se observa que los juegos que se clasifican como exitosos tienen características consistentes que respaldan su éxito. En primer lugar, muchos de estos videojuegos presentan un alto porcentaje de positividad en las reseñas, lo que indica una buena aceptación por parte de la comunidad. Además, para confirmar que realmente son juegos exitosos o no, nos ayudaremos con la página web “[Metacritic](#)” la cual calcula una media de las reseñas de otras páginas o usuarios, dando más importancia a las que considera relevantes y obteniendo una valoración óptima y precisa, siendo una de las principales y más importante empresas de clasificación de videojuegos, películas, entre otros medios de entretenimientos.

Nuestro modelo nos indicó “Grand Theft Auto V” como el más exitoso y si buscamos las valoraciones de él en “Metacritic”, tiene una media de 97% de valoraciones positivas en 66 páginas relacionadas, confirmando así el verdadero éxito del videojuego y que a sido correctamente clasificado como el mejor ya que lleva más de 10 años desde su lanzamiento recibiendo críticas positivas.

Por último, se observa la tendencia de los géneros “Action”, “Indie” y “Adventure”, demostrando la correcta popularidad de estos vista anteriormente.

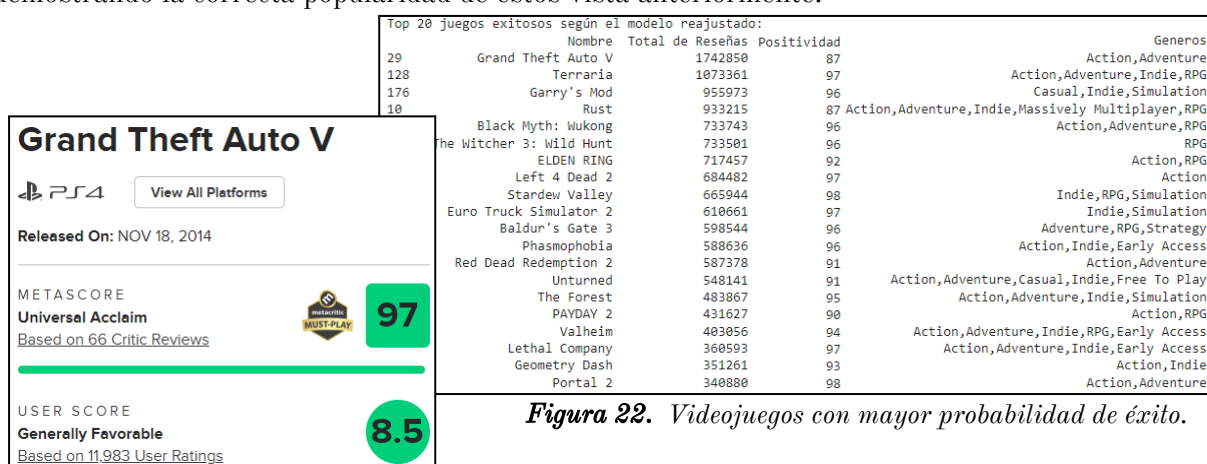


Figura 22. Videojuegos con mayor probabilidad de éxito.

Figura 23. Validación de videojuego en Metacritic.

→ Videojuegos con menor probabilidad de éxito

Ya pasando al análisis de los videojuegos menos exitosos según nuestro modelo, destaca *PUBG: BATTLEGROUNDS*, que cuenta con más de 2.4 millones de reseñas y un porcentaje de positividad de solo 58%. Este dato refleja críticas significativas de su comunidad, posiblemente relacionadas con problemas de rendimiento, balance o monetización dentro del juego.

Además en “Metacritic” tiene una media de 72% en 52 páginas de reseñas, siendo esto un porcentaje muy bajo y confirmando que el juego no fue exitoso, además con las críticas de usuarios obtuvo 4.9/10 puntos con más de dos mil valoraciones, afirmando su mal recibimiento por parte de los jugadores.

Por último, se observa que el género **Action** es recurrente en los juegos menos exitosos, indicando que, a pesar de su popularidad, no siempre garantiza aceptación si no cumple con las expectativas de los usuarios.

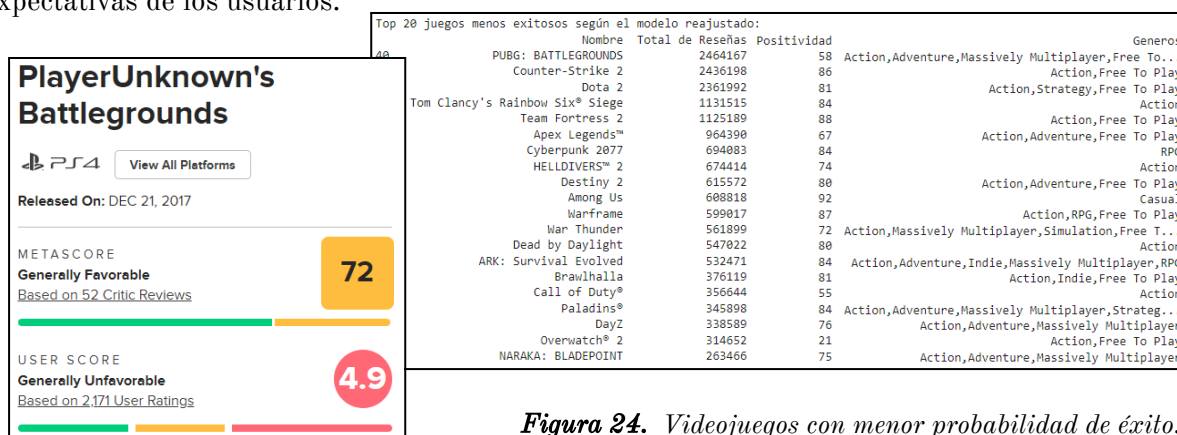


Figura 24. Videojuegos con menor probabilidad de éxito.

Figura 25. Validación de videojuego en Metacritic.

→ Pruebas con videojuegos específicos

Por último, se realizó una prueba en donde se le indicaba un juego en específico al modelo y este nos devolvía si era exitoso o no.

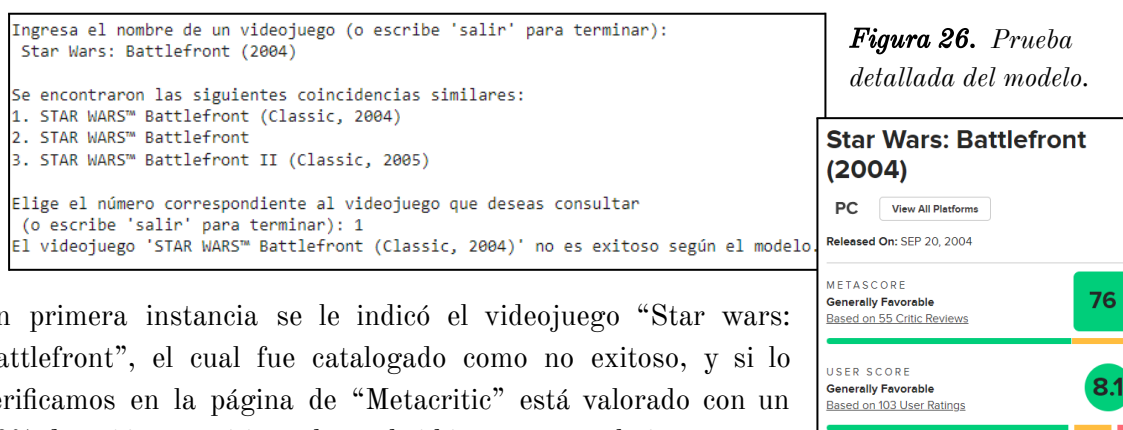


Figura 26. Prueba detallada del modelo.

En primera instancia se le indicó el videojuego “Star wars: Battlefront”, el cual fue catalogado como no exitoso, y si lo verificamos en la página de “Metacritic” está valorado con un 76% de críticas positivas, lo cual si bien no es tan bajo, tampoco se considera como exitoso.

Figura 27. Valoración Metacritic.

```

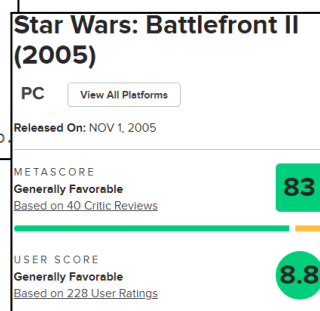
Ingresa el nombre de un videojuego (o escribe 'salir' para terminar):
Star Wars: Battlefront II(2004)

Se encontraron las siguientes coincidencias similares:
1. STAR WARS™ Battlefront (Classic, 2004)
2. STAR WARS™ Battlefront II (Classic, 2005)
3. STAR WARS™ Battlefront

Elige el número correspondiente al videojuego que deseas consultar
(o escribe 'salir' para terminar): 2
El videojuego 'STAR WARS™ Battlefront II (Classic, 2005)' es exitoso según el modelo.

```

Figura 28. Segunda prueba detallada del modelo.



Como segunda prueba se indicó el juego “Star wars: Battlefront 2”, el cual es la segunda entrega del videojuego antes visto, en esta ocasión el modelo lo categorizó como exitoso, y verificando con “Metacritic”, posee un 83% de positividad en 40 sitios, mejorando un 7%.

Figura 29. Valoración Metacritic

6.2 Recomendaciones

1. **Priorizar géneros exitosos y plataformas rentables:** Dado que los géneros Indie, Adventure y Action muestran mayores tasas de éxito, desarrollar videojuegos en estas categorías ofrece una mayor probabilidad de aceptación, sin embargo esto no garantiza su éxito. Asimismo, centrarse en Windows y Mac maximiza el alcance del mercado.

Possible escenario: Invertir en un juego de género Adventure/Action para Windows y Mac podría captar un amplio público y garantizar rentabilidad.

2. **Enfocarse en la calidad de géneros saturados como Acción:** El género Acción puede fracasar si no cumple con las expectativas de los usuarios. Por ello, invertir en una jugabilidad refinada y diferenciación creativa es esencial. Esto se basa en la observación de títulos como PUBG, que, pese a su gran base de usuarios, enfrenta críticas constantes.

Possible escenario: Juegos de Action con innovaciones únicas pueden evitar el desgaste en un mercado competitivo.

3. **Evitar secuelas de juegos mal recibidos:** Según los resultados, juegos con baja positividad y pocas reseñas no justifican una nueva inversión. En su lugar, se recomienda redirigir recursos hacia conceptos frescos y prometedores.

Possible escenario: Cancelar una secuela con baja proyección financiera y desarrollar un título independiente con un concepto original.

4. **Mantener calificaciones altas en plataformas externas:** Los datos confirman que la percepción en sitios como Metacritic influye directamente en la aceptación del público. **Responder a críticas de usuarios y lanzar actualizaciones** puede mejorar significativamente la imagen de un título.

Possible escenario: Una actualización que resuelva problemas técnicos o la implementación de nuevas mecánicas dentro del juego podría convertir un juego mediocre en uno exitoso, atrayendo nuevos usuarios.

7 Conclusión y discusión

El análisis predictivo realizado con el modelo Random Forest permitió identificar patrones fundamentales para comprender el éxito de los videojuegos.

Entre los hallazgos más relevantes se destacan:

1. **Géneros más exitosos:** Los géneros Acción, Indie y Aventura sobresalen en las tasas de éxito, pese a esto, se concluyó que Acción también figura entre los menos exitosos. Esto demuestra que, más allá de la popularidad de un género, **la calidad y la satisfacción del usuario son elementos esenciales.**
2. **Plataformas líderes:** Windows mantiene su liderazgo como la plataforma más exitosa, mientras que, Mac ha superado a Linux en títulos exitosos en comparación a nuestro anterior análisis, lo que no solo refleja cambios en las preferencias del mercado sino también en el buen recibimiento por parte de los jugadores, pese a las restricciones de usuario de este sistema operativo.
3. **Importancia de la fecha de lanzamiento:** La fecha de lanzamiento se posiciona como un factor fundamental para la predicción, ya que permite analizar cómo la popularidad de un juego evoluciona con el tiempo, permitiendo mejorar la precisión del modelo.

Abarcando el análisis prescriptivo, este permite proponer estrategias prácticas como priorizar géneros rentables, innovar en mercados no saturados y evitar secuelas de juegos mal recibidos. Un punto destacable es la necesidad de responder a las críticas de los usuarios y mantener calificaciones altas en plataformas como Metacritic, ya que estas acciones pueden mejorar significativamente la percepción y el rendimiento de un título.

Reflexión final

El modelo no solo demostró ser una herramienta precisa para predecir el éxito, sino que también ofrece un enfoque integral para optimizar recursos, minimizar riesgos y guiar decisiones estratégicas en una industria competitiva. La combinación de análisis predictivo y prescriptivo resalta el potencial de este enfoque para anticiparse a las tendencias y maximizar el impacto en el mercado e industria de los videojuegos.

Bibliografia

1. Steam. (2024). Steam Store dataset. <https://store.steampowered.com>
2. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
3. Friedman, J., Hastie, T., & Tibshirani, R. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
4. Japkowicz, N., & Shah, M. (2011). *Evaluating learning algorithms: A classification perspective*. Cambridge University Press.
5. Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.
6. Microsoft Corporation. (2024). Introduction to machine learning. <https://learn.microsoft.com>
7. Scikit-learn developers. (2024). Scikit-learn: Machine learning in Python. <https://scikit-learn.org/stable>
8. Zhang, H., & Singer, B. H. (2010). *Recursive partitioning and applications*. Springer.
9. Metacritic. (2024). *Metacritic*. <https://www.metacritic.com/>