

FACULTAD DE MATEMÁTICAS E INGENIERIAS

MODELADO Y SIMULACIÓN I

Proyecto Final (Éxito de un Videojuego)

Ruby Alexandra Tovar Palma - 614 232700

Santiago Steven Sanchez Barbosa - 614 232708

Docente: Julián Orlando Jiménez Cárdenas

Fundación Universitaria Konrad Lorenz
Bogotá D.C.
21 de marzo de 2025

Índice

1	Introducción	2
2	Objetivo general	2
2.1	Objetivos específicos	2
3	Requisitos de datos	3
4	Definición de Variables	3
4.1	Definición de Variables Independientes y Dependientes	4
4.1.1	Variables Independientes	4
4.1.2	Variables Dependientes	4
4.2	Uso de Regresiones para la Predicción de Datos	5
4.2.1	Métodos de Regresión	5
4.3	Cálculo del Éxito	6
5	Plan de implementación	6
5.1	Técnicas de procesamiento y limpieza de datos	6
5.2	Bibliotecas	7
5.3	Métodos para la validación y prueba del modelo	7
6	Resultados esperados	7
6.1	Tipo de información o soluciones esperadas	7
6.2	Visualizaciones o representaciones potenciales	8
6.3	Aplicación de los resultados al problema real	8
7	Referencias	8

1. Introducción

El mundo de los videojuegos durante las ultimas décadas ha sido un mercado en constante crecimiento ya que cuenta con lo mejor de todos los sectores del entretenimiento y ademas con la evolución de los dispositivos tecnológicos, los cuales cada vez son más avanzados y con mayor capacidad para lograr ejecutar videojuegos de mas calibre y mas calidad, es normal el querer entrar en este mercado lo antes posible. Sin embargo, este nicho tiene múltiples factores a tener en cuenta para lograr entrar y, sobre todo, triunfar con el lanzamiento de algún título. Siendo tan así que, incluso, empresas ya posicionadas en el sector tienen dificultades para poder replicar éxitos anteriores o en general entender que es lo que quiere su audiencia; por tales razones es de suma importancia tener a la mano una visión estadística y matemática de que hace a u juego exitoso ser, justamente, exitoso.

Cada videojuego que ha sido exitoso y a su vez se ha convertido en un referente para la industria tiene un matiz y una característica que lo hace único y especial frente a la comunidad amante del gaming. Para destacar no es necesario reinventar la rueda, a veces basta con trabajar bajo un modelo que cuente con características fundamentales que permita tener una cierta parte de la atención de la audiencia ya ganada. Tales características van desde lo mas intrínseco del juego hasta algunos factores externos al mismo, como lo son el genero, los gráficos, la fecha de lanzamiento, el precio y muchas otras variables de las cuales hablaremos mas adelante.

El modelo matemático del cual se entrara en profundidad en las siguientes paginas mostrara una forma de predecir de una manera objetiva mediante el uso de datos reales las probabilidades que tiene un nuevo videojuego de ser un éxito según la tendencia de los 100 videojuegos mejor valorados tanto por la critica como por la audiencia en general. El modelo no solo busca poder ser una herramienta que permita conocer el éxito de los videojuegos próximos a ser lanzados, sino que incluso poder predecir si la idea de un nuevo desarrollo que ni siquiera ha comenzado pueda ser viable para el estudio que lo lleve a cabo.

2. Objetivo general

Desarrollar un modelo matemático basado en datos reales que permita predecir la probabilidad de éxito de un videojuego antes de su lanzamiento, considerando tanto tendencias del mercado como factores clave identificados en los 100 videojuegos mejor valorados por la crítica y la audiencia.

2.1. Objetivos específicos

1. **Analizar los factores determinantes del éxito** de un videojuego, incluyendo género, gráficos, fecha de lanzamiento, precio y otros elementos internos y externos.
2. **Construir un modelo predictivo** basado en tendencias estadísticas y matemáticas extraídas de los videojuegos mejor valorados en la industria.
3. **Evaluar la viabilidad de nuevas ideas de desarrollo** antes de que inicie la producción, con el fin de optimizar recursos y minimizar riesgos financieros.
4. **Proporcionar una herramienta de apoyo** para estudios de videojuegos que permita tomar decisiones informadas sobre el diseño y comercialización de nuevos títulos.

5. **Comparar los resultados del modelo con casos reales** para validar su precisión y mejorar su capacidad de predicción a lo largo del tiempo.

3. Requisitos de datos

Para el desarrollo del modelo matemático se harán uso de dos bases de datos pertenecientes a dos medios muy confiables, dichas fuentes de información son:

1. **SteamDB:** Es una base de datos no oficial pero confiable que extrae información directamente de Steam, la principal plataforma de distribución de videojuegos en PC. Ofrece datos detallados sobre precios, número de jugadores, reseñas y tendencias de ventas en tiempo real, siendo una herramienta clave para analizar el éxito y la popularidad de los juegos.
2. **IGDB (Internet Game Database):** Es una base de datos de videojuegos respaldada por Twitch, lo que garantiza la calidad y precisión de sus datos. Proporciona información estructurada sobre géneros, estudios desarrolladores, fechas de lanzamiento y puntuaciones de críticos y jugadores, siendo una referencia esencial para el análisis de tendencias en la industria del gaming.



Figura 1: SteamDB



Figura 2: IGDB (Internet Game Database)

4. Definición de Variables

Para este caso, se hará uso de un modelo que sea capaz de predecir el éxito de videojuegos a partir de factores de popularidad extraídos tras el análisis de datos históricos de la base de datos SteamDB y de IGDB. Es así como las variables empleadas para el modelo serán:

- Género del videojuego.
- Gráficos.

- Modo de juego.
- Microtransacciones.
- Número de descargas.
- Puntuación.
- Jugadores activos.
- Mes de lanzamiento.
- Tamaño de la desarrolladora.
- Éxito o popularidad de la desarrolladora.
- Precio de compra.

Puesto que existen datos categóricos como el género, gráficos, entre otras variables, se extraerá la repetición de cada dato categórico a manera de un patrón que se repite en juegos exitosos, y se van a mapear los datos categóricos mediante valores numéricos. Por ejemplo, si el género que lidera a los juegos más populares es indie, entonces a este se le asignará un valor en un intervalo $[a, b]$ que valore patrones de popularidad, siendo $b > a$, con b siendo el valor asignado a los juegos de género indie y a el valor asignado al género de juegos menos popular.

4.1. Definición de Variables Independientes y Dependientes

La idea principal es definir una ecuación general que modele el éxito de un videojuego. Primero, se hace un análisis previo de las variables, dado que algunas tienen dependencia de otras, por lo que se definirán variables independientes y dependientes, con las cuales se crearán funciones compuestas.

4.1.1. Variables Independientes

- Género del videojuego (G).
- Modo de juego (M).
- Microtransacciones (Mt).
- Gráficos (Gf).
- Tamaño del estudio (Te).
- Mes del lanzamiento (Lm).

4.1.2. Variables Dependientes

- **Precio** (Pv): Depende de la desarrolladora, el modo de juego, los gráficos y las microtransacciones. $Pv(Te, M, Mt, Gf)$.
- **Wishlist** (W): Indica el interés de los jugadores y depende del género, mes de lanzamiento y precio. $W(G, Lm, Pv)$.

- **Puntuaciones (P):** Relación calidad-precio. $P(W, Pv)$.
- **Número de descargas (D):** Depende de las puntuaciones, el precio y los jugadores activos. $D(Ja)$.
- **Jugadores activos (Ja):** Depende del precio, la calidad y las puntuaciones. $Ja(P)$.

Dado que las funciones son compuestas, para evitar la repetición de variables, se ha simplificado la notación eliminando redundancias. Así, para obtener la expresión con todas las variables necesarias, basta con reemplazar las variables dependientes en las funciones: $Ja(P) \rightarrow Ja(W, Pv) \rightarrow Ja(G, Lm, Pv, Pv) \rightarrow Ja(G, Lm, Pv) \rightarrow Ja(G, Lm, Te, M, Mt, Gf, Ed)$.

Aunque existen muchas variables dependientes, en esta modelación el éxito de un videojuego se determinará principalmente por el número de descargas, ya que este refleja tanto la ganancia económica como el crecimiento de la comunidad, factores clave para considerar un juego exitoso. Sin embargo, el éxito no depende únicamente de este aspecto, sino también de la popularidad sostenida del juego a lo largo del tiempo. Por ello, el éxito también estará influenciado por la cantidad de jugadores activos, las puntuaciones y el precio, ya que estos factores incentivan a más usuarios a descargar el videojuego. Así, el éxito puede representarse mediante la siguiente ecuación:

$$E(D, Ja, P) = aD + bJa + cP + dPv$$

Sin embargo, como el precio ya está involucrado las otras variables dependientes, se puede omitir, simplificando la función.

$$E(D, Ja, P) = aD + bJa + cP \tag{1}$$

4.2. Uso de Regresiones para la Predicción de Datos

Una vez definida la función que modela el éxito, es necesario determinar los valores de a , b y c . Para ello, se emplearán modelos de regresión, que podrán variar entre regresión lineal y logística, dependiendo del comportamiento de las variables dependientes.

4.2.1. Métodos de Regresión

Puesto que tanto SteamDB como IGDB proporcionan suficientes datos históricos, se pueden aplicar métodos de regresión:

1. **Regresión Lineal:** Se analiza si las relaciones entre los datos históricos y las variables independientes son lineales.
2. **Regresión Logística:** Para un manejo de datos con resultados binarios.

3. **Regresión Polinómica:** Si los datos tienen una varianza significativa en su distribución.

Según la distribución de los datos en un sistema de coordenadas cartesianas y su varianza, se aplicarán distintos métodos de regresión a las variables dependientes D , Ja y P . Esto permitirá optimizar el ajuste de datos mediante análisis estadístico y seleccionar el método más adecuado en función del cálculo de la varianza y las distribuciones.

4.3. Cálculo del Éxito

Una vez determinadas las ecuaciones de D , Ja y P mediante regresiones, se procederá a calcular el éxito del videojuego. Para ello, se pueden emplear tres enfoques diferentes:

1. **Método de mínimos cuadrados:** Utilizando el cálculo de D , Ja y P para la construcción de matrices que permitan modelar la relación entre las variables.
2. **Asignación de pesos a las variables a , b y c :** Estos valores porcentuales determinan la influencia de D , Ja y P en el éxito del videojuego, según los datos históricos.
3. **Aplicación de modelos de regresión:** Dependiendo de la naturaleza de los datos, se empleará regresión logística para datos binarios o regresiones polinomiales y lineales para datos continuos.

5. Plan de implementación

5.1. Técnicas de procesamiento y limpieza de datos

Dado el contexto del problema y la naturaleza de las bases de datos con las cuales haremos el modelo predictivo, hay que tener ciertas pautas para garantizar la calidad de la predicción así como también la coherencia del modelo planteado.

- Se recopilarán datos de fuentes confiables como **SteamDB** e **IGDB**, asegurando que sean relevantes y actualizados. A su vez se hará la debida eliminación de valores nulos y duplicados.
- Conversión de datos categóricos en valores numéricos mediante codificación ordinal o one-hot encoding (ejemplo: asignar valores a géneros de videojuegos según su popularidad).
- Escalado de variables numéricas para mejorar la estabilidad del modelo (normalización o estandarización según convenga).
- Análisis de correlación entre variables para identificar posibles redundancias y optimizar el modelo.

5.2. Bibliotecas

Se hará uso de las siguientes bibliotecas tanto para el manejo como para la gestión de los datos a la hora de implementar el modelo en Python.

- **Pandas y NumPy:** Para manipulación de datos.
- **Matplotlib y Seaborn:** Para visualización de distribuciones y correlaciones.
- **Scikit-learn:** Para aplicar regresiones lineales, polinómicas o logísticas según corresponda.
- **Statsmodels:** Para realizar análisis de regresión y determinar la significancia estadística de las variables.

5.3. Métodos para la validación y prueba del modelo

Para garantizar la precisión y fiabilidad del modelo propuesto, es necesario validarlo utilizando diferentes estrategias. Estas pruebas permiten comparar los resultados del modelo con la realidad y ajustar los parámetros para mejorar su desempeño. A continuación, se presentan los métodos clave para evaluar la efectividad del modelo:

- Usar el modelo con un juego que ya haya salido antes y de esta manera llevar a cabo una comparación entre la realidad y la predicción propuesta por el modelo.
- **Cálculo de métricas de rendimiento según el tipo de regresión:** MSE (Error Cuadrático Medio) y R^2 para regresiones lineales y polinómicas. Y también Accuracy, precisión, recall y F1-score para regresión logística.
- Validación cruzada para evitar sobreajuste y verificar la robustez del modelo.

6. Resultados esperados

El modelo desarrollado permitirá extraer información valiosa sobre los factores que influyen en el éxito de un videojuego. Además, ayudará a comprender el impacto de distintas variables en la popularidad de un título.

6.1. Tipo de información o soluciones esperadas

Uno de los principales objetivos del modelo es identificar qué características tienen mayor impacto en el éxito de un videojuego. Además, se busca evaluar cómo factores como el precio afectan la popularidad del juego, lo que puede ser útil para estudios de mercado y estrategias de lanzamiento.

- Identificación de los factores con mayor impacto en el éxito de un videojuego.
- Evaluación del impacto del precio y otros factores en la popularidad del juego.

6.2. Visualizaciones o representaciones potenciales

Para facilitar la interpretación de los datos y las relaciones entre las variables, se utilizarán diversas herramientas de visualización. Estas representaciones gráficas ayudarán a comprender mejor los patrones que influyen en el éxito de un videojuego.

- **Matriz de correlación** para analizar relaciones entre variables.
- **Gráficos de dispersión** mostrando la relación entre puntuaciones, descargas y éxito.
- **Histogramas y boxplots** para distribución de datos categóricos y numéricos.
- **Gráficos de regresión** para visualizar el ajuste de los modelos a los datos.

6.3. Aplicación de los resultados al problema real

Los hallazgos del modelo podrán ser utilizados en la industria del videojuego para mejorar la toma de decisiones. A partir de los resultados obtenidos, se podrán diseñar estrategias para aumentar las probabilidades de éxito de un título antes de su lanzamiento.

- Recomendaciones para estudios de videojuegos sobre géneros, precios y estrategias de lanzamiento para maximizar el éxito.
- Identificación de patrones en videojuegos exitosos y no exitosos.
- Evaluación de tendencias para prever qué características impulsarán la popularidad de futuros lanzamientos.

7. Referencias

1. SteamDB. (n.d.). *Steam Database*. Retrieved from <https://steamdb.info>
2. IGDB. (n.d.). *Internet Game Database*. Retrieved from <https://www.igdb.com>