

Factores de contenido asociados al éxito crítico de videojuegos en la plataforma Steam: un análisis interpretativo mediante Canonical Biplot

(Content Factors Associated With the Critical Success of Video Games on Steam: An Interpretative Analysis Using Canonical Biplot)

Ariosto Vicuña Pino, Edwin Pacha-Herrera
Universidad Nacional de Loja

ariosto.vicuna@unl.edu.ec, edwin.paccha@unl.edu.ec

Resumen: La plataforma Steam concentra gran parte de la actividad del mercado de videojuegos para PC, siendo un entorno clave para estudiar factores asociados al desempeño de un juego. Investigaciones recientes relacionan reseñas, actividad de jugadores, precio e idiomas con la popularidad, aunque predominan enfoques predictivos con limitada capacidad interpretativa. El objetivo propuesto es determinar qué factores de contenido se asocian al éxito crítico en la plataforma Steam medido por Metacritic. Se realizó un estudio observacional, correlacional, cuantitativo y transversal, con una muestra no probabilística por cuotas anuales y filtrado por disponibilidad de valoración Metacritic los cuales fueron recolectados por web scraping y analizados bajo CRISP-DM mediante Canonical Biplot. Los resultados muestran que la interacción del jugador (reseñas, jugadores concurrentes, reseñas positivas) es el principal eje diferenciador. Además, en juegos actuales predominan variables de participación, mientras en juegos antiguos adquieren peso los logros, características adicionales y contenido descargable (DLC). El principal aporte es brindar una explicación visual e interpretativa de los factores asociados al éxito crítico de los videojuegos en la plataforma Steam, mostrando que las métricas de interacción del jugador tienen mayor peso que varias características estructurales del producto, y que esta relación cambia entre videojuegos antiguos y actuales.

Palabras clave: Éxito crítico, CRISP-DM, Metacritic.

Abstract: Steam platform accounts for a large share of activity in the PC video game market, making it a key environment for studying the factors associated with game performance. Recent research has linked reviews, player activity, price, and language availability to popularity, although predictive approaches with limited interpretive capacity still predominate. The objective of this study is to determine which content-related factors are associated with critical success on the Steam platform, as measured by Metacritic. This study employed an observational, correlational, quantitative, and cross-sectional design. A non-probability sample was constructed using annual quota sampling and then filtered based on the availability of Metacritic ratings. Data were obtained through web scraping and analyzed under the CRISP-DM methodology using the Canonical Biplot technique. The results show that player interaction variables reviews, concurrent players, and positive reviews constitute the main differentiating axis. In addition, participation-related variables predominate in recent games, whereas achievements, additional features, and Downloadable Content (DLC) carry greater weight in older games. The main contribution of this study is to provide a visual and interpretive explanation of the factors associated with the critical success of video games on the Steam platform, showing that player interaction metrics carry greater weight than several structural product characteristics, and that this relationship changes between older and more recent games.

Keywords: Critical success, CRISP-DM, Metacritic

1. INTRODUCCIÓN

La industria de los videojuegos se ha consolidado como un componente estratégico de la economía digital, tanto por su expansión sostenida como por la creciente complejidad de las interacciones que articula entre plataformas, usuarios y desarrolladores. En este contexto, Steam ocupa una posición relevante como ecosistema de distribución, consumo e interacción para videojuegos de PC, al integrar en un mismo entorno funciones comerciales, sociales e informacionales. Su importancia para la investigación académica radica en que concentra un volumen significativo de datos sobre comportamiento del jugador, recepción del producto y dinámicas de participación, lo que la convierte en un espacio empírico especialmente adecuado para examinar factores asociados al desempeño de un videojuego [1], [2]. A ello se suma que la plataforma emplea análisis de datos masivos para segmentar perfiles de usuarios y diseñar servicios personalizados orientados a reforzar la permanencia y el compromiso dentro del sistema [3]. En conjunto, estas características convierten a Steam en una infraestructura analítica privilegiada para el estudio del mercado de los videojuegos.

Sobre esta base, la literatura reciente ha abordado distintas dimensiones del desempeño de los videojuegos en Steam, principalmente desde perspectivas centradas en la popularidad, la interacción del usuario y la predicción del comportamiento de consumo. En esta línea, [5] aplicó modelos probabilísticos para anticipar la interacción de los jugadores a partir de datos conductuales, evidenciando la capacidad de la plataforma para identificar patrones y proyectar niveles de participación futura. De igual manera, [1] demostraron que variables como el precio, los idiomas soportados, la fecha de lanzamiento y los géneros contribuyen a predecir la popularidad de un videojuego en Steam, mientras que [2] identificaron que los géneros, los precios, las valoraciones y la reputación del publicador influyen en la preferencia de los jugadores y en la visibilidad de los títulos dentro del catálogo. A su vez, estudios centrados en reseñas y jugabilidad han mostrado que la información generada en la plataforma permite examinar tanto la experiencia del usuario como la calidad percibida del videojuego [3], [4]. En conjunto, estos antecedentes han consolidado una agenda de investigación en torno a los determinantes del éxito de los videojuegos en entornos de plataforma.

Sin embargo, el desarrollo de esta agenda ha estado marcado por un predominio de enfoques predictivos orientados a explicar popularidad, actividad del jugador o desempeño comercial, relegando a un segundo plano la comprensión del éxito crítico desde marcos interpretativos. Esta limitación es relevante, ya que la valoración crítica de un videojuego no puede reducirse únicamente a indicadores de mercado o de consumo, dado que también involucra atributos vinculados con la experiencia de juego, el diseño, la reputación y diversas señales de calidad percibida [6], [7]. En términos analíticos, ello obliga a distinguir entre los factores que favorecen la visibilidad o la adopción de un título y aquellos que se asocian con su legitimación crítica. Desde esta perspectiva, el Metascore constituye una medida especialmente pertinente para operacionalizar el éxito crítico, en tanto ofrece un indicador agregado, reconocido y comparable entre videojuegos. En efecto, [7] sostienen que el Metascore refleja de forma fiable la apreciación cualitativa de la crítica especializada, mientras que [8] y [9] respaldan su validez como medida empírica del éxito del videojuego y de la reputación de sus desarrolladores.

La literatura presenta un relativo descuido del éxito crítico como objeto de estudio, sino también en la naturaleza de las estrategias analíticas empleadas para abordarlo. De forma general, los trabajos existentes identifican variables con capacidad predictiva sobre determinados resultados, pero ofrecen una comprensión limitada de cómo dichas variables se articulan relacionamente dentro de una estructura multivariante. Así, [4] mostraron que las reseñas de jugadores permitieron evaluar dimensiones relevantes de la jugabilidad; [10] encontraron que las reseñas de pares se asocian con incrementos significativos en la actividad de juego; [11] evidenció que el éxito alcanzado en Steam puede proyectarse hacia otros entornos digitales, como Kickstarter; y [12] analizaron factores asociados a calificaciones altas, destacando la relevancia

de la jugabilidad y de la coherencia técnica. No obstante, estos aportes no abordan de manera central una representación conjunta, visual e interpretativa de la relación entre factores de contenido y éxito crítico. En consecuencia, persiste una brecha teórica y metodológica en torno a la comprensión estructural de los elementos que distinguen a los videojuegos con distinta valoración crítica.

A partir de este vacío, el presente estudio propone una aproximación orientada no solo a identificar asociaciones, sino también a representarlas e interpretarlas de manera integrada. Para ello, se recurre al Canonical Biplot, una técnica particularmente adecuada para este problema, dado que permite representar simultáneamente variables, observaciones y separación entre grupos en un espacio de baja dimensión, favoreciendo una lectura estructural del fenómeno analizado [13]. Su principal fortaleza frente a enfoques de baja interpretabilidad reside en su capacidad para hacer visibles patrones de proximidad, contraste y diferenciación entre categorías, posibilitando una comprensión más precisa de la organización relacional de las variables involucradas en el éxito crítico.

En concordancia con lo anterior, el objetivo general de este artículo es determinar los factores de contenido asociados al éxito crítico de los videojuegos en la plataforma Steam mediante la técnica Canonical Biplot. De este objetivo se derivan tres propósitos específicos: construir una base de datos integrada a partir de las plataformas Steam y Metacritic, analizar las relaciones entre variables de contenido y puntuación crítica, y comparar la estructura de asociación entre videojuegos antiguos y actuales. Asimismo, se plantea como expectativa analítica que las métricas de interacción del jugador y de recepción del mercado presentarán una asociación más intensa con el éxito crítico en videojuegos actuales, mientras que en títulos antiguos algunos atributos estructurales del producto podrían adquirir una relevancia relativa mayor.

La contribución del estudio se sitúa en dos planos complementarios. En el plano sustantivo, aporta evidencia sobre la configuración de los factores asociados al éxito crítico de los videojuegos en Steam. En el plano metodológico, introduce una lectura multivariante, visual e interpretativa que desplaza el análisis desde enfoques predominantemente predictivos hacia una comprensión más estructural de las relaciones entre variables de contenido, recepción del usuario y valoración crítica. Sobre esta base, el artículo se organiza en cinco secciones adicionales: estado del arte, metodología, resultados, discusión y conclusiones.

2. ESTADO DEL ARTE

La discusión dominante en la investigación reciente sobre videojuegos digitales ha estado marcada por tres grandes ejes: la predicción de popularidad o éxito, el análisis de reseñas y comportamiento del jugador, y la búsqueda de modelos explicativos que vinculen atributos del producto con desempeño en plataforma. Steam aparece de manera recurrente como entorno empírico privilegiado por la abundancia de metadatos, la trazabilidad de las reseñas y la posibilidad de observar dinámicas de participación a gran escala [1], [2].

Existen trabajos que modelan el éxito de un videojuego desde una perspectiva predictiva. [1] analizaron la popularidad temprana de juegos en la plataforma Steam y encontraron que variables como precio, idiomas, fecha de lanzamiento y géneros aportan información útil para estimar el número de jugadores. El valor de este estudio radica en mostrar que los metadatos del producto contienen señales tempranas de desempeño; sin embargo, su foco se concentra en popularidad y no en éxito crítico. En una línea parecida, [2] estudiaron el catálogo de la plataforma Steam y concluyeron que algunos factores que influyen en la popularidad del juego son los géneros, los precios, las valoraciones y la reputación del publicador. Este trabajo es útil porque amplía el conjunto de variables relevantes dentro del ecosistema Steam, aunque mantiene una lógica orientada a preferencia y recomendación, más que a interpretación crítica del producto.

También existen estudios que conectan el desempeño de la plataforma Steam con resultados externos. [11] mostró que el éxito de videojuegos en Steam incrementa en 24,1 % el monto recaudado en campañas de Kickstarter, evidenciando que las señales de éxito dentro de una plataforma pueden trasladarse a otros mercados digitales. Este hallazgo es importante porque sugiere que el rendimiento en la plataforma Steam opera como un mecanismo reputacional; sin embargo, sigue siendo una aproximación centrada en efectos causales y económicos más que en estructuras interpretativas de variables.

Por otra parte, existen estudios que priorizan el papel de las reseñas, la jugabilidad y la interacción del usuario. [4] propusieron un enfoque de análisis de jugabilidad basado en reseñas de jugadores, mostrando que la minería de opiniones permite identificar fortalezas y debilidades de un videojuego en términos de calidad percibida. Este estudio resulta especialmente relevante porque desplaza la evaluación desde atributos puramente técnicos hacia la experiencia reportada por los usuarios. En el mismo sentido, [10] encontraron que la creación de una reseña por parte de un amigo se asocia con una tasa significativamente mayor de juego posterior. Este hallazgo refuerza la idea de que las reseñas no solo reflejan experiencia, sino que también modifican comportamiento. A su vez, [7] mostraron que las reseñas de usuarios y críticos no producen efectos idénticos sobre el desempeño de un videojuego, lo que sugiere que ambas fuentes de evaluación aportan señales diferentes dentro del mercado.

Además, [6] argumentan que el éxito del videojuego depende de la interacción entre cualidades de búsqueda y de experiencia. Su trabajo aporta un matiz importante: la recepción positiva de un juego no proviene de un solo atributo, sino de una combinación de señales previas al consumo y experiencias efectivas durante el uso. En un plano cercano, [14] subrayan que la lógica de publicación en la plataforma Steam también se relaciona con componentes de metajuego, interacción prolongada y compromiso del usuario.

En conjunto, el estado del arte permite afirmar que existe cierto consenso en torno a que el desempeño de un videojuego depende de una combinación de factores estructurales, señales de recepción del público y mecanismos de reputación. Sin embargo, persiste un problema no resuelto: la mayoría de los estudios analizan estas variables desde modelos predictivos, análisis parciales de reseñas o aproximaciones centradas en ventas y popularidad. Lo que aún no ha sido suficientemente abordado es una lectura interpretativa, simultánea y visual de la relación entre factores de contenido y éxito crítico.

Esta brecha es relevante porque el éxito crítico no equivale exactamente a la popularidad ni al desempeño comercial. Además, la literatura revisada no muestra una aplicación consolidada del Canonical Biplot para estudiar, en un mismo espacio analítico, atributos de contenido, variables de participación del jugador y categorías de valoración crítica. Dado que el CVA biplot permite representar grupos, medias, variables y casos en un mismo plano, constituye una alternativa pertinente para responder ese vacío [13].

3. METODOLOGÍA

Se usó la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) que es considerada robusta, flexible y adaptable para proyectos contemporáneos de ciencia de datos por su capacidad para organizar el ciclo completo del proyecto y, al mismo tiempo, permitir ajustes según el dominio de aplicación, los requisitos del negocio y la complejidad técnica del problema [15], [16]. CRISP-DM se describe mediante seis fases principales: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue. Estas fases no se ejecutan necesariamente de forma lineal, sino que conforman un proceso cíclico con retroalimentación entre etapas, especialmente entre preparación, modelado y evaluación [17]. CRISP-DM mantiene una orientación fuerte hacia el valor práctico del análisis, ya que los

modelos deben responder a necesidades reales del contexto y no limitarse a alcanzar métricas técnicas aceptables [16].

En la fase de comprensión del negocio se concretó que, en el contexto de los videojuegos de la plataforma Steam, el éxito crítico puede llevar a la determinación de las características de éxito de los videojuegos desde un punto de vista no comercial. Tomando en cuenta que el éxito crítico se refiere al reconocimiento y la valoración positiva que un videojuego recibe por parte de la crítica especializada. Las variables que se consideraron se organizaron en cuatro bloques: factores de contenido, factores de éxito, factores de evaluación del jugador y factores del producto. La Tabla 1 presenta las categorías y variables a ser extraídos desde la plataforma Steam (<https://store.steampowered.com>).

Tabla 1. Lista de las variables contenidas en la plataforma Steam.

Categoría	Variables	Interpretación
Factores de contenido	achievements_count	Logros: Representan profundidad, progreso y rejugabilidad del juego.
	features_count	Características: Determinan complejidad estructural y diversidad de la experiencia de juego.
	dlc_count	Contenido descargable: Reflejan soporte continuo y extensión del ciclo de vida del producto.
	languages_count	Idiomas soportados: Mide accesibilidad internacional y alcance global del juego.
Factores de éxito	metacritic_score	Puntuación crítica: Indicador global de calidad crítica del videojuego.
	reviews_count	Popularidad en la plataforma Steam: Refleja aceptación del mercado y difusión del producto.
	peak_players	Máximo de jugadores conectados: Mide intensidad de uso y adopción del videojuego en la plataforma.
Factores de evaluación del Jugador	positive_reviews_percent	Evaluación positiva: Representa satisfacción global de los usuarios.
Factores del producto	price	Precio: Determina barrera económica de acceso al producto.
	genres_count	Géneros: Indica diversidad temática del juego.
	early_access	Acceso anticipado: Indica lanzamiento en fase temprana de desarrollo.
	is_free	Juego gratuito: Define modelo de monetización del producto.

La población de referencia estuvo conformada por videojuegos publicados en la plataforma Steam en la sección “Lo mejor del año”. La muestra se construyó usando una selección no probabilística por cuotas anuales cien registros por año dentro del periodo 2022 - 2025 y aplicando un filtrado por disponibilidad de valoración Metacritic. Como resultado del proceso de depuración, la muestra final quedó conformada por 116 videojuegos. Para la extracción, se realizó un proceso de web scraping, que es una técnica para extraer información de páginas web de forma

automática usando scripts mediante cuatro pasos básicos: acceder a una página, leer su contenido HTML, identificar los datos deseados variables y guardarlos en un formato útil. El lenguaje de programación Python proporcionó las facilidades necesarias para su implementación. El proceso para obtener los valores de las variables descritas en la Tabla 1, tiene varias etapas porque no todos los valores de las variables se encuentran en la misma página. El proceso se ejecutó obteniendo como producto un dataset por año. A continuación, la Tabla 2 describe el desarrollo del web scraping.

Tabla 2. Procedimiento para obtener los valores de las variables desde el sitio de Steam.

Web Scraping	Descripción
Se extrajo la lista de juegos desde la plataforma Steam Best of 202x x= [2, 3, 4, 5]	Se intentó obtener hasta 100 juegos (si la página los expone) desde https://store.steampowered.com/charts/bestofyear/202x Se extrajeron las siguientes columnas: steam_id, nombre, url, image_url desde múltiples fuentes dentro de cada página. Se guardó en un archivo de Excel.
Se obtuvo las puntuaciones de Metacritic de los juegos de la plataforma Steam	Se construyó la URL automáticamente mediante: " https://store.steampowered.com/app/{steam_id}/ " Se extrajeron las siguientes columnas relevantes: release_year, price, genres_count, achievements_count, is_free, languages_count, metacritic_score, early_access, review_score, url Se guardó en un archivo de Excel.
Se adquirió las columnas reviews_count y positive_reviews_percent	Se usó Steam App Reviews API oficial aplicado a la dirección: " https://store.steampowered.com/appreviews/{steam_id}/ " No se utilizó SteamDB ni SteamCharts. Se guardó en un archivo de Excel.
Se consiguió las columnas dlc_count, features_count y peak_players	Se construyó la URL automáticamente mediante: " https://store.steampowered.com/app/{steam_id}/?l=english " Se guardó en un archivo de Excel.

Estudios recientes continúan mostrando que la preparación de datos contiene fases decisivas porque condicionan la confiabilidad del resultado analítico final [17], [18]. El proceso de web scraping dejó cuatrocientos registros, de los cuales se seleccionaron ciento dieciséis registros que cumplieron con tener una valoración de Metacritic (metacritic_score). También se eliminaron todas las columnas no numéricas (name, review_score, developer, publisher, url) o aquellas de ningún valor para el análisis (steam_id). De esta manera, se prepararon los datos para aplicar la técnica de Canonical Biplot.

Se determinó si las variables numéricas necesitan una transformación logarítmica, de acuerdo con su característica. Las variables que no se transformaron porque sus escalas son comparables se mencionan a continuación: positive_reviews_percent, price, genres_count, is_free, early_access, metacritic_score.

Las restantes variables numéricas se sometieron a una transformación logarítmica necesaria porque sus valores vienen de procesos de conteo o con distribuciones de cola larga, como ocurre con métricas de popularidad en plataformas digitales. La transformación logarítmica reduce la asimetría de la distribución, estabiliza la varianza y evita que valores extremos dominen el análisis

factorial. Las variables para transformar son: reviews_count, peak_players, dlc_count, achievements_count, features_count, languages_count.

La transformación utilizada se expresa en la Ecuación 1

$$x' = \log(x + 1) \quad (1)$$

donde:

- x representa el valor original de la variable
- x' corresponde al valor transformado.

El término +1 se incorporó para evitar problemas cuando la variable toma el valor cero. La Tabla 3 muestra todas las variables que serán parte del estudio. A las variables que se aplicó una transformación logarítmica se le agregó la notación _log1p.

Tabla 3. Variables que intervienen en el gráfico de Canonical Biplot.

Variable original	Variable final	Descripción
achievements_count	achievements_count_log1p	Número total de logros por juego
features_count	features_count_log1p	Cantidad de funcionalidades o modos disponibles
dlc_count	dlc_count_log1p	Número de expansiones o contenido adicional
languages_count	languages_count_log1p	Número total de idiomas disponibles
reviews_count	reviews_count_log1p	Volumen total de reseñas de jugadores
peak_players	peak_players_log1p	Máximo histórico de jugadores concurrentes
positive_reviews_percent	positive_reviews_percent	Porcentaje de reseñas positivas
price	price	Precio de venta del videojuego
genres_count	genres_count	Número de géneros asociados
early_access	early_access	Estado de desarrollo anticipado
is_free	is_free	Indica si el juego es gratuito

Finalmente, todas las variables fueron estandarizadas mediante Z-score antes del análisis con Canonical Biplot para evitar diferencias de escala entre variables.

En la fase de modelado del análisis de las características de los videojuegos de la plataforma Steam se requirió abordar simultáneamente múltiples dimensiones del producto, incluyendo características estructurales del juego, indicadores de popularidad en plataformas digitales y evaluaciones críticas. En este contexto, el Canonical Biplot constituye una herramienta estadística adecuada para analizar sistemas multivariantes donde los individuos (videojuegos) se describen mediante múltiples variables cuantitativas y, al mismo tiempo, pueden organizarse en grupos definidos por una variable categórica, como los niveles de puntuación crítica en Metacritic.

Desde una perspectiva metodológica, una ventaja central del Canonical Biplot es su capacidad para identificar qué variables contribuyen a la diferenciación entre grupos dentro de un conjunto de datos multivariantes. Canonical Biplot se basa en una descomposición de la matriz de datos

que maximiza la separación entre grupos previamente definidos, de forma similar a los enfoques basados en análisis discriminante multivariante. De acuerdo con [13], el Canonical Biplot permite examinar de forma conjunta la separación entre grupos, la orientación de las variables y la calidad de representación de los casos, lo que lo vuelve especialmente útil cuando el interés de investigación es explicativo antes que puramente predictivo.

Las evaluaciones de críticos influyen en el comportamiento de compra y en el desempeño comercial de los videojuegos, y que estas relaciones presentan estructuras no lineales y dependientes del contexto del mercado [7]. Estas variables forman un sistema complejo de relaciones que requiere métodos multivariantes para ser analizado de manera adecuada. El Canonical Biplot va a proporcionó un análisis simultáneo de las variables relacionadas con contenido del juego (`achievements_count`, `dlc_count`, `features_count`, `languages_count`), indicadores de popularidad en la plataforma (`reviews_count`, `peak_players`, `positive_reviews_percent`) y características del producto (`price`, `genres_count`, `early_access`, `is_free`). La representación conjunta de estas variables en el espacio factorial facilita la identificación de los factores que diferencian videojuegos con distintos niveles de puntuación crítica, permitiendo interpretar el sistema de variables de forma integrada. En este estudio se estimaron tres representaciones: un canonical biplot global, uno para videojuegos antiguos y otro para videojuegos actuales.

La investigación contemporánea resalta que esta evaluación debe ir más allá de la precisión predictiva e incorporar consideraciones sobre robustez, confiabilidad, transparencia y aplicabilidad en escenarios reales [17]. La fase de evaluación tuvo como propósito determinar si los modelos desarrollados permitieron responder al problema de investigación planteado: identificar las características estructurales y de popularidad asociadas con videojuegos con alta puntuación crítica. En este estudio, la evaluación se centró en verificar si el Canonical Biplot permitió: Identificar patrones multivariantes en las características de los videojuegos, diferenciar grupos de juegos según su puntuación en Metacritic y determinar qué variables contribuyen a explicar el éxito crítico de los juegos.

La evaluación de la varianza explicada en la Ecuación 2 analizó el porcentaje de varianza explicada por los primeros ejes factoriales:

$$\text{Varianza explicada} = \frac{\lambda_1 + \lambda_2}{\sum \lambda_i} \quad (2)$$

Donde:

λ son los valores propios asociados a los componentes canónicos. Un porcentaje alto indica que el plano factorial representa adecuadamente la estructura de los datos.

La evaluación de la separación entre grupos de videojuegos verificó si el modelo permitió distinguir entre videojuegos con diferentes niveles de puntuación crítica. Los grupos evaluados fueron: Metacritic alto, Metacritic medio y Metacritic bajo. La evaluación se realizó observando la separación espacial entre grupos en el Canonical Biplot (Tabla 4), la orientación de las variables respecto a los grupos y la concentración o dispersión de los videojuegos en el espacio factorial.

Tabla 4. Grupos de separación espacial en el Canonical Biplot.

Variables de popularidad	Variables estructurales del juego	Variables del producto
<code>reviews_count</code>	<code>achievements_count</code>	<code>price</code>
<code>peak_players</code>	<code>dlc_count</code>	<code>genres_count</code>
<code>positive_reviews_percent</code>	<code>features_count</code>	<code>early_access</code>
	<code>languages_count</code>	<code>is_free</code>

La evaluación de la contribución de las variables analizó la longitud y dirección de los vectores en el Biplot para identificar qué variables influyen en la separación de los videojuegos. Las variables con vectores más largos y alineadas con determinados grupos indican mayor influencia en la estructura multivariante.

En la evaluación comparativa entre videojuegos actuales y antiguos se evaluaron tres configuraciones del modelo: Videojuegos actuales, videojuegos antiguos y conjunto completo. Esto permitió analizar si las variables que caracterizan el éxito del juego se mantienen estables o cambian según el periodo analizado. La comparación se realizó mediante: La estructura del plano factorial, distribución de los grupos y la orientación de los vectores de variables.

La evaluación final del modelo implicó considerarlo adecuado cuando: El plano factorial explica una proporción significativa de la variabilidad de los datos, existe separación observable entre grupos de puntuación, las variables presentan interpretaciones coherentes con el comportamiento del mercado de videojuegos y los resultados son consistentes entre los diferentes conjuntos de datos analizados.

La fase de despliegue en este estudio se realizó mediante la generación y análisis de tres representaciones gráficas basadas en Canonical Biplot, que según [13] y [19] permitió visualizar relaciones multivariadas entre variables, combinando precisión analítica y capacidad explicativa. La técnica se basa en el Análisis de Correlación Canónica (ACC), cuyo objetivo es encontrar las combinaciones lineales de variables canónicas de cada conjunto que tienen la máxima correlación entre sí [20]. El modelo desarrollado se desplegó a través de tres gráficos de Canonical Biplot que representan diferentes configuraciones del conjunto de datos analizado. Los gráficos generados están en la Tabla 5.

Tabla 5. Resultados de la fase de despliegue.

Canonical Biplot	Objetivo	Permite analizar
Videojuegos actuales	Identificar las variables que caracterizan el éxito crítico en el mercado actual.	Relación entre popularidad y evaluación crítica.
		Distribución de los juegos según niveles de Metacritic.
		Variables dominantes en el contexto actual del mercado.
Videojuegos antiguos	Evaluar si las características asociadas al éxito crítico eran diferentes en periodos anteriores.	Cambios estructurales en las variables relevantes.
		Diferencias en la distribución de los juegos.
		Evolución de los factores asociados al éxito.
Conjunto completo de videojuegos	Identificar la estructura global de relaciones entre variables y puntuación crítica.	Visualizar la estructura multivariante general
		Identificar variables dominantes en el conjunto total de datos
		Comparar la posición relativa de videojuegos actuales y antiguos.

Como resultado del proyecto se generaron los siguientes artefactos analíticos:

1. Dataset limpio y estructurado para análisis multivariante.
2. Scripts reproducibles para la generación del modelo.
3. Tres representaciones gráficas basadas en Canonical Biplot.
4. Interpretación analítica de los patrones observados.

Además, se usó Python con Visual Studio Code para realizar el proceso de web scraping y R Statistics junto con el paquete MultBiplotR creado por [21] para realizar los gráficos de Canonical Biplot. Con estos elementos es factible reproducir el análisis y facilitar su aplicación en futuros estudios sobre videojuegos.

4. RESULTADOS

4.1 Estructura general de las variables asociadas al éxito crítico

El análisis global mostró que las variables `reviews_count`, `peak_players` y `positive_reviews_percent` se alinearon hacia una misma región del espacio factorial, lo que indicó una asociación positiva dentro de la estructura multivariante del modelo. Los videojuegos clasificados con puntuación crítica alta se concentraron en la dirección de estos vectores, mientras que los títulos con puntuación crítica baja aparecieron en regiones opuestas del plano. Ver Figura 1.

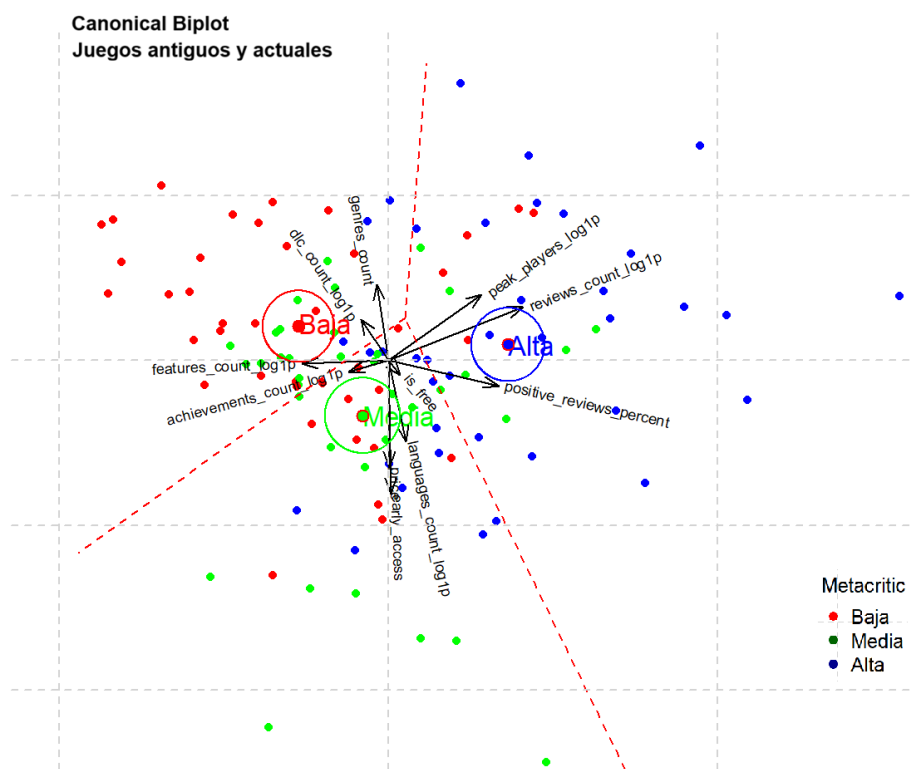


Figura 1. Canonical Biplot aplicado al conjunto de datos entre los años 2022-2025 (videojuegos actuales y antiguos).

En contraste, variables como `features_count`, `achievements_count` y `dlc_count` se proyectaron con mayor cercanía a zonas asociadas a puntuaciones críticas inferiores o intermedias, lo que sugiere una contribución distinta en la organización del espacio multivariante. La estructura discriminante del modelo se concentra principalmente en las dos primeras dimensiones canónicas. La inercia canónica obtenida para la dimensión uno fue 84,36 y 15,63 para la dimensión dos.

4.2 Videojuegos actuales

En el subconjunto de videojuegos actuales se observó una agrupación más clara de los títulos con puntuación crítica alta en torno a `reviews_count`, `peak_players` y `positive_reviews_percent`. Este patrón indicó que, en los juegos recientes, las variables ligadas a participación del usuario y recepción del mercado presentaron mayor peso discriminante en la diferenciación entre niveles de Metacritic (Figura 2).

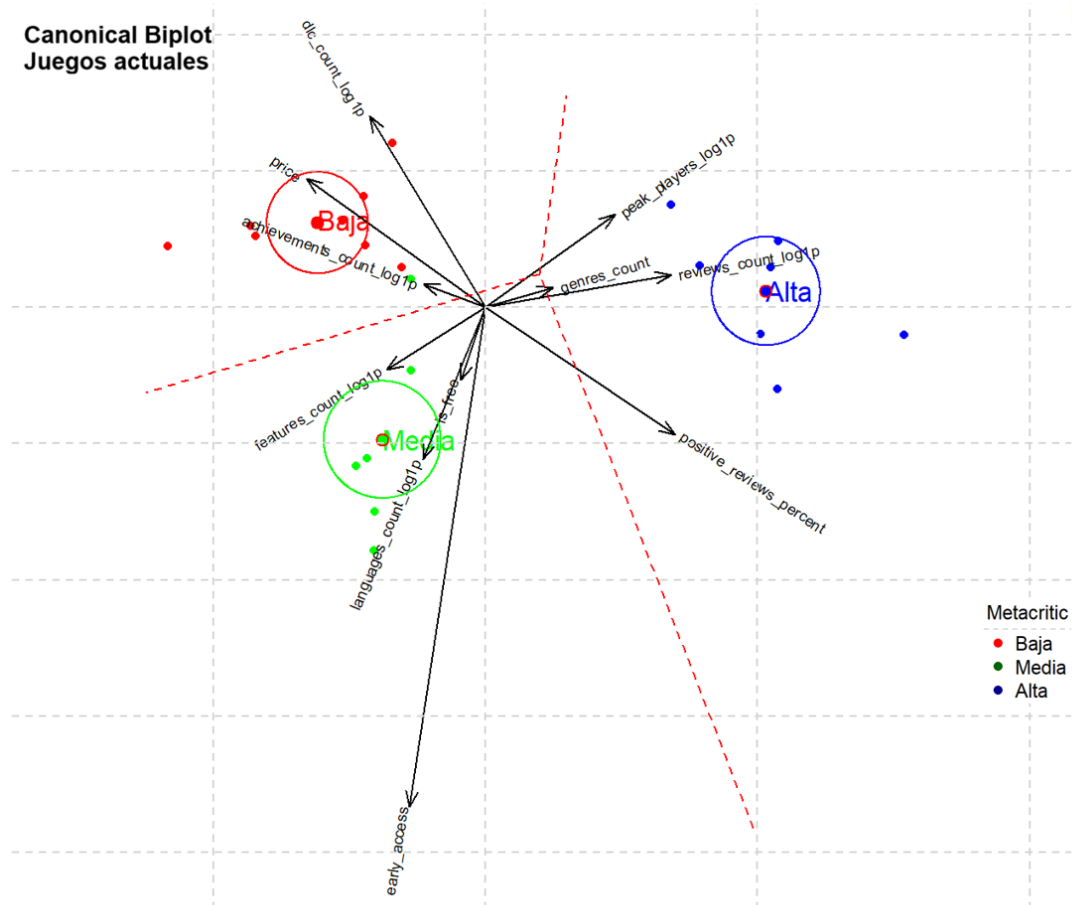


Figura 2. Canonical Biplot aplicado al conjunto de datos del año 2025(videojuegos actuales).

Por su parte, los juegos con puntuación crítica baja tendieron a orientarse hacia variables estructurales como `price` y `achievements_count`, mientras que la categoría media ocupó una posición intermedia cercana a `features_count` y `languages_count`. En este subconjunto, la separación entre grupos es más definida que en el caso de los videojuegos antiguos. En cuanto a la inercia canónica se tuvo 83,68 para la dimensión uno y 16,31 para la dimensión dos.

4.3 Videojuegos antiguos

En los videojuegos antiguos se mantuvo la asociación entre puntuación crítica alta y `positive_reviews_percent`, pero la orientación de `reviews_count` y `peak_players` apareció más dispersa. A diferencia de lo observado en los juegos recientes, en este subconjunto adquirieron mayor peso relativo variables de contenido y complejidad estructural, como `features_count`, `achievements_count` y `dlc_count`. Además, la distribución espacial de los grupos mostró una superposición mayor entre categorías de puntuación crítica, lo que indica una separación menos marcada entre niveles de éxito crítico en comparación con los videojuegos actuales (Figura 3).

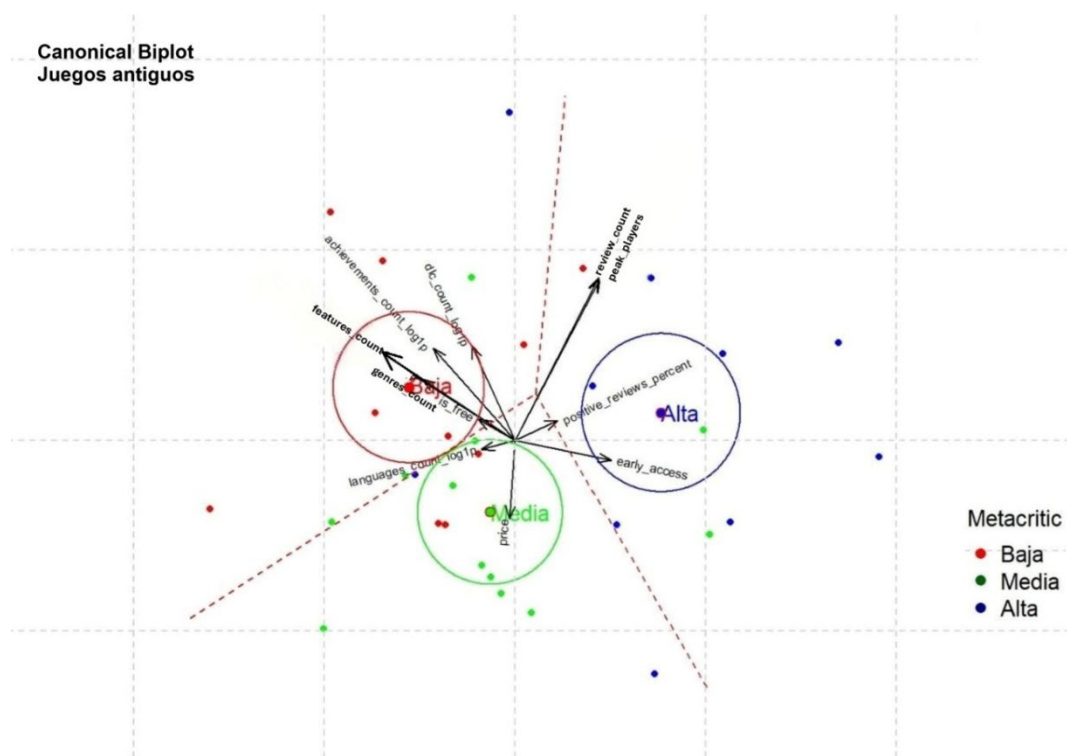


Figura 3. Canonical Biplot aplicado al conjunto de datos del año 2022 (videojuegos antiguos).

La comparación entre los tres biplots permite afirmar que la estructura del éxito crítico no es idéntica entre videojuegos actuales y antiguos. En los títulos recientes, la diferenciación entre puntuaciones se alinea principalmente con variables de participación y recepción del jugador. En los títulos antiguos, en cambio, se observó una mayor influencia relativa de características internas del producto. También se identificó que variables como `languages_count` y `genres_count` presentan vectores de menor magnitud, lo que indica una contribución discriminante más limitada dentro del espacio factorial. Se obtuvo una inercia canónica de 77,42 y 22,57 para las dimensiones uno y dos, respectivamente.

5. DISCUSIÓN

Los resultados de esta investigación sugieren que el éxito crítico de los videojuegos en la plataforma Steam se organiza, ante todo, alrededor de señales de participación del jugador y recepción del mercado. En particular, el número de reseñas, el máximo histórico de jugadores concurrentes y el porcentaje de reseñas positivas constituyen el núcleo estructural que diferencia a los videojuegos con mejor y peor puntuación crítica. Este hallazgo coincide con lo reportado por [4], quienes mostraron que las reseñas de jugadores permitieron capturar dimensiones significativas de calidad percibida, y con [10], quienes observaron que las reseñas también influyen en la actividad de juego.

Los hallazgos también dialogan con [7], quienes mostraron que las reseñas de usuarios y críticos no tienen un efecto idéntico sobre el desempeño del videojuego. En el presente estudio, la proximidad entre métricas derivadas de usuarios y categorías de alta valoración crítica sugiere que, aunque la crítica profesional y la evaluación del público no sean equivalentes, sí pueden converger dentro de una misma estructura de legitimación del producto. Esto ayuda a explicar por qué variables sociales y de compromiso aparecieron tan cerca del grupo de mejor evaluación crítica.

De forma adicional, los resultados se alinean con la literatura que subraya la relevancia de señales de mercado dentro del ecosistema de Steam. [1] y [2] ya habían mostrado que metadatos como precio, géneros e idiomas ayudan a explicar popularidad. Sin embargo, en este estudio dichas variables no aparecieron como el principal eje discriminante del éxito crítico. Esta diferencia es importante porque sugiere que no es lo mismo explicar popularidad que explicar valoración crítica. En otras palabras, algunas variables útiles para estimar visibilidad o número de jugadores pueden ser menos decisivas cuando el resultado de interés es la recepción crítica agregada.

Otro hallazgo relevante es la diferencia entre videojuegos actuales y antiguos. En los juegos recientes, las métricas de participación del jugador desempeñan un papel más fuerte en la organización del espacio factorial. En los juegos antiguos, por el contrario, adquieren mayor protagonismo variables estructurales del producto, como logros, características adicionales y DLC. Este contraste puede interpretarse como un cambio en la lógica de validación del videojuego dentro del ecosistema digital: mientras que en etapas previas el contenido estructural podía funcionar como señal central de valor, en el contexto actual parecen pesar más las métricas visibles de interacción, comunidad y aceptación del usuario.

Los resultados también encuentran apoyo parcial en [6], quienes argumentan que el éxito del videojuego no depende de un único atributo, sino de la interacción entre cualidades de búsqueda y experiencia. Esa lógica se observa aquí en forma de estructura multivariante: no es una sola variable la que explica la puntuación crítica, sino la articulación entre recepción positiva, actividad de juego y difusión social del producto. Asimismo, [14] habían señalado que la lógica de metajuego y la persistencia de participación pueden ser claves en plataformas como Steam; el presente estudio aporta evidencia visual de esa centralidad dentro del éxito crítico.

Desde el punto de vista metodológico, el trabajo verifica la pertinencia del Canonical Biplot como herramienta de análisis interpretativo. Tal como explican los autores [13], esta técnica permite representar al mismo tiempo grupos, variables y observaciones, lo que facilita la comprensión de sistemas complejos. En este caso, esa representación fue especialmente útil para traducir relaciones estadísticas en patrones visuales legibles para investigación y toma de decisiones.

No obstante, el estudio presenta limitaciones. En primer lugar, la muestra se restringe a videojuegos populares de la plataforma Steam con datos disponibles y puntuación crítica trazable, lo que excluye a videojuegos de menor visibilidad. Esta omisión podría condicionar la profundidad del análisis al determinar las características que diferencian a los videojuegos actuales de los antiguos. En segundo lugar, el diseño es transversal, de modo que no permite establecer causalidad. Finalmente, no se incorporaron variables cualitativas como complejidad narrativa, estética, estrategia de marketing o calidad técnica profunda, que podrían enriquecer la explicación del éxito crítico.

En términos aplicados, los resultados sugieren que desarrolladores y editores deberían prestar especial atención a mecanismos que fortalezcan la participación del jugador, el volumen de reseñas positivas y la sostenibilidad de la comunidad alrededor del producto. Al mismo tiempo, la comparación temporal muestra que el diseño del contenido sigue siendo relevante, aunque su peso relativo puede variar según el contexto histórico del mercado.

6. CONCLUSIONES

Este estudio permitió establecer que el éxito crítico de los videojuegos en la plataforma Steam, medido mediante la puntuación en Metacritic, se relaciona principalmente con variables de interacción del jugador y recepción del mercado. En particular, el volumen de reseñas, el pico histórico de jugadores concurrentes y el porcentaje de reseñas positivas constituyen el principal eje de diferenciación entre videojuegos con distintas categorías de valoración crítica.

Asimismo, el análisis comparativo mostró que esa estructura cambia según la temporalidad del videojuego. En los videojuegos actuales, la participación del usuario y la recepción visible dentro de la plataforma tienen una capacidad discriminante mayor. En los videojuegos antiguos, en cambio, algunas variables estructurales del producto adquieren mayor protagonismo relativo. Esto sugiere que la lógica del éxito crítico en la plataforma Steam no es estática, sino históricamente variable.

Desde una perspectiva metodológica, el Canonical Biplot demostró ser una herramienta útil para estudiar fenómenos complejos en plataformas digitales porque permite comprender no solo qué variables importan, sino también cómo se relacionan dentro de una estructura multivariante interpretable. Esa es, precisamente, una de las principales contribuciones de este artículo frente a la literatura predominantemente predictiva.

Como línea futura, sería pertinente ampliar la muestra a videojuegos con menor visibilidad, incorporar análisis longitudinales y complementar el modelo con minería de texto de reseñas para integrar de manera más explícita dimensiones cualitativas de la experiencia del jugador y de la crítica especializada.

REFERENCIAS

- [1] A. De Luisa, J. Hartman, D. Nabergoj, S. Pahor, M. Rus, B. Stevanoski, J. Demšar, y E. Štrumbelj, "Predicting the popularity of games on Steam," *arXiv*, 2021. [En línea]. Disponible en: <https://doi.org/10.48550/arXiv.2110.02896>
- [2] E. Salkanović, N. Zukorlić, L. Oković, y D. Kečo, "A data analysis of Steam's game catalog and diverse recommendation strategies," *International Journal of Computer Applications*, vol. 186, no. 54, pp. 39-49, dic. 2024. [En línea]. Disponible en: <https://doi.org/10.5120/ijca2024924261>
- [3] H. Zhou, "More games, more playtime," *Interdisciplinary Humanities and Communication Studies*, vol. 1, no. 1, jun. 2023. [En línea]. Disponible en: <https://doi.org/10.61173/2y9xnj78>
- [4] X. Li, Z. Zhang, y K. Stefanidis, "A data-driven approach for video game playability analysis based on players' reviews," *Information*, vol. 12, no. 3, mar. 2021. [En línea]. Disponible en: <https://doi.org/10.3390/info12030129>
- [5] H. Ding, "Modeling user behavior online: A comparison of Pareto/NBD and BG/NBD models," *Highlights in Engineering and Technology*, vol. 61, pp. 128-132, jul. 2023. [En línea]. Disponible en: <https://doi.org/10.54097/hset.v61i.10283>
- [6] S. Heidenreich, F. Handrich, y T. Kraemer, "Flawless victory! Investigating search and experience qualities as antecedent predictors of video game success," *Electronic Markets*, vol. 33, no. 1, 2023. [En línea]. Disponible en: <https://doi.org/10.1007/s12525-023-00647-2>
- [7] T. Kraemer, W. H. Weiger, y S. Heidenreich, "Do all stars shine the same? Investigating the nonlinear effects of user and critic reviews on video game sales," *Journal of Business Research*, vol. 188, 2025. [En línea]. Disponible en: <https://doi.org/10.1016/j.jbusres.2024.115034>
- [8] J. Pfau, M. Debus, J. Juul, E. Lundedal Hammar, A. Canossa, y M. Seif El-Nasr, "Predicting success factors of video game titles and companies," in *Entertainment Computing - ICEC 2022*, B. Göbl, E. Van Der Spek, J. Baalsrud Hauge, y R. McCall, Eds. Cham, Switzerland: Springer, 2022, vol. 13477, pp. 269-282. [En línea]. Disponible en: https://doi.org/10.1007/978-3-031-20212-4_22

- [9] H. Yu, "A comprehensive analysis of game success: Synergy between quality, brand, and market," *Applied and Computational Engineering*, vol. 110, no. 1, pp. 22-28, nov. 2024. [En línea]. Disponible en: <https://doi.org/10.54254/2755-2721/110/2024MELB0101>
- [10] V. Brodschneider, y J. Pirker, "On the influence of reviews on play activity on Steam: A statistical approach," en *2023 IEEE Conference on Games (CoG)*, pp. 1-5, 2023. [En línea]. Disponible en: <https://doi.org/10.1109/CoG57401.2023.10333235>
- [11] D. Chen, "Signals of video game success on Steam: Evidence from Kickstarter campaigns," *Finance Research Letters*, vol. 86, 2025. [En línea]. Disponible en: <https://doi.org/10.1016/j.frl.2025.108995>
- [12] G. C. Ullmann, C. Politowski, Y.-G. Guéhéneuc, y F. Petrillo, "What makes a game high-rated?: Towards factors of video game success," in *Proc. 6th International CSE Workshop on Games and Software Engineering: Engineering Fun, Inspiration, and Motivation*, Pittsburgh, PA, USA, May 2022, pp. 16-23. [En línea]. Disponible en: <https://doi.org/10.1145/3524494.3527628>
- [13] N. le Roux y S. Gardner-Lubbe, "A two-group canonical variate analysis biplot for an optimal display of both means and cases," *Advances in Data Analysis and Classification*, vol. 19, no. 3, pp. 721-748, 2025. [En línea]. Disponible en: <https://doi.org/10.1007/s11634-024-00593-7>
- [14] M. N. Rizani, M. N. A. Khalid, y H. Iida, "Application of meta-gaming concept to the publishing platform: Analysis of the Steam games platform," *Information*, vol. 14, no. 2, pp. 110, 2023. [En línea]. Disponible en: <https://doi.org/10.3390/info14020110>
- [15] A. M. Shimaoka, R. C. Ferreira, y A. Goldman, "The evolution of CRISP-DM for data science: Methods, processes and frameworks," *SBC Reviews on Computer Science*, vol. 4, no. 1, pp. 28-43, 2024. [En línea]. Disponible en: <https://doi.org/10.5753/reviews.2024.3757>
- [16] V. Plotnikova, M. Dumas, y F. P. Milani, "Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements," *Data & Knowledge Engineering*, vol. 139, 2022. [En línea]. Disponible en: <https://doi.org/10.1016/j.datak.2022.102013>
- [17] S. Tripathi, D. Muhr, M. Brunner, H. Jodlbauer, M. Dehmer, y F. Emmert-Streib, "Ensuring the robustness and reliability of data-driven knowledge discovery models in production and manufacturing," *Frontiers in Artificial Intelligence*, vol. 4, 2021. [En línea]. Disponible en: <https://doi.org/10.3389/frai.2021.576892>
- [18] M. Elkabalawy, A. Al-Sakkaf, E. M. Abdelkader, y G. Alfalah, "CRISP-DM-based data-driven approach for building energy prediction utilizing indoor and environmental factors," *Sustainability*, vol. 16, no. 17, 2024. [En línea]. Disponible en: <https://doi.org/10.3390/su16177249>
- [19] C. J. F. Ter Braak, P. Smilauer, y S. Dray, "Algorithms and biplots for double constrained correspondence analysis," *Environmental and Ecological Statistics*, vol. 25, no. 2, pp. 171-197, jun. 2018. [En línea]. Disponible en: <https://doi.org/10.1007/s10651-017-0395-x>
- [20] L. Wells, K. Thurimella, y S. Bacallado, "Regularised canonical correlation analysis: Graphical lasso, biplots and beyond," *arXiv*, nov. 16, 2025. [En línea]. Disponible en: <https://doi.org/10.48550/arXiv.2403.02979>
- [21] J. L. Vicente-Villardón, *MultBiplotR: Multivariate Analysis Using Biplots in R*, ver. 23.11.0. CRAN, 2023. [En línea]. Disponible en: <https://CRAN.R-project.org/package=MultBiplotR>

