

UNIVERSIDAD SAN FRANCISCO DE QUITO

Colegio de Ciencias e Ingenierías

Análisis del Impacto de la Música en el Rendimiento de
los Jugadores de Dota 2 en un Entorno Competitivo

Nahomy Dayanara Varada Salazar

Ingeniería en Ciencias de la Computación

Trabajo de fin de carrera presentado como requisito
para la obtención del título de
Ingeniero en Ciencias de la Computación

Quito, 8 de mayo de 2025

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

HOJA DE CALIFICACIÓN
DE TRABAJO DE FIN DE CARRERA

Análisis del Impacto de la Música en el Rendimiento de
los Jugadores de Dota 2 en un Entorno Competitivo

Nahomy Dayanara Varada Salazar

José David Vega, Ph.D.

Quito, 8 de mayo de 2025

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas. Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombres y apellidos: Nahomy Dayanara Varada Salazar

Código: 00211623

Cédula de identidad: 563431681

Lugar y fecha: Quito, 8 de mayo de 2025

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETHeses>.

UNPUBLISHED DOCUMENT

Note: The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on <http://bit.ly/COPETHeses>.

RESUMEN

Recientemente, la relación entre la música y los procesos cognitivos ha sido objeto de numerosos estudios, evidenciando su impacto en aspectos como la memoria, la concentración y las emociones. Sin embargo, su influencia específica en el rendimiento dentro de videojuegos competitivos, un ámbito que exige altos niveles de concentración, toma de decisiones rápidas y trabajo en equipo coordinado, sigue siendo un campo poco explorado. Este proyecto analiza el impacto de la música en el rendimiento de los jugadores de Dota 2 en un entorno competitivo. Específicamente, se investiga la influencia de diferentes géneros musicales sobre la concentración, el tiempo de reacción, la toma de decisiones y el desempeño general de los jugadores. Para ello, se diseñó un entorno experimental con jugadores divididos en dos grupos: aquellos que juegan con música y aquellos que no.

Los resultados sugieren que la música, especialmente los géneros electrónicos y ambientales, podría tener un impacto positivo en el rendimiento de los jugadores al mejorar métricas clave como el APM (acciones por minuto) y el Performance Score. Estos géneros parecen facilitar la concentración y el flujo durante el juego, lo cual podría mejorar la eficiencia y la toma de decisiones en un entorno competitivo. En contraste, géneros como el metal se asociaron con un aumento en el Tilt Factor, lo que podría indicar una influencia en el control emocional y una mayor propensión a reacciones impulsivas, lo que potencialmente afecta el rendimiento bajo presión.

El análisis de clústeres identificó tres perfiles de jugadores con distintos estilos de juego y preferencias musicales, lo que sugiere que la relación entre la música y el rendimiento podría depender tanto del género musical como de las características individuales de los jugadores. Los perfiles identificados, como los "estratégicos-calmados" y los "agresivos-rápidos", apuntan a una posible interacción entre la música y los estilos de juego, lo que podría permitir la personalización de la música para mejorar el rendimiento en entrenamientos o competiciones futuras.

Estos hallazgos brindan una base para seguir explorando cómo la música podría influir en el rendimiento en contextos competitivos, no solo en videojuegos, sino también en entornos que requieren altos niveles de concentración, toma de decisiones rápidas y control emocional, como en el ámbito laboral o académico.

Palabras clave: Música, rendimiento de jugadores, Dota 2, concentración, toma de decisiones, tiempos de reacción, análisis de clústeres, estilos de juego, Tilt Factor, APM, géneros musicales, eSports.

ABSTRACT

Recently, the relationship between music and cognitive processes has been the subject of numerous studies, highlighting its impact on aspects such as memory, concentration, and emotions. However, its specific influence on performance in competitive video games, which require high levels of concentration, quick decision-making, and coordinated teamwork, remains an underexplored area. This study analyzes the impact of music on the performance of Dota 2 players in a competitive setting. Specifically, it investigates how different music genres influence concentration, reaction time, decision-making, and overall gameplay performance. An experimental environment was designed with players divided into two groups: those who played with music and those who played without.

The results suggest that music, particularly electronic and ambient genres, may positively impact player performance by enhancing key metrics such as actions per minute (APM) and performance scores. These genres appear to facilitate concentration and flow during gameplay, potentially improving efficiency and decision-making in a competitive environment. In contrast, genres such as metal were associated with an increase in the Tilt Factor, indicating an influence on emotional control and a greater propensity for impulsive reactions, which could affect performance under pressure.

Cluster analysis identified three player profiles with distinct gameplay styles and musical preferences, suggesting that the relationship between music and performance may depend on both musical genre and the individual characteristics of players. The identified profiles, such as "strategic-calm" and "aggressive-fast," point to a potential interaction between music and gameplay styles, allowing for music customization to enhance performance in future training or competitions.

These findings provide a basis for further exploration of how music could influence performance in competitive contexts, not only in video games but also in environments requiring high levels of concentration, quick decision-making, and emotional regulation, such as in work or academic settings.

Key words: Music, player performance, Dota 2, concentration, decision-making, reaction times, Tilt Factor, APM, musical genres, eSports.

TABLA DE CONTENIDO

Resumen	4
Abstract.	5
ÍNDICE DE TABLAS	8
ÍNDICE DE TABLAS	8
ÍNDICE DE FIGURAS	9
ÍNDICE DE FIGURAS	9
1.1 Relevancia y justificación	10
1.2 Objetivos generales y específicos	11
1.2.1 Objetivo general.	11
1.2.2 Objetivos específicos.	11
1.3 Alcance y limitaciones	12
1.3.1 Alcance.	12
1.3.2 Limitaciones.	12
2.1 Influencia de la música en la atención y la cognición	14
2.2 Música y rendimiento en videojuegos	15
2.3 Métricas de rendimiento en videojuegos	16
3.1 Contexto experimental	18
3.2 Diseño del estudio	18
3.3 Población y muestra	18
3.3.1 Criterios de selección.	18
3.3.2 Tamaño de la muestra.	19
3.3.3 División experimental	20
3.3.4 Géneros musicales incluidos.	20
3.4 Recolección de datos	20
3.4.1 Scripts en python para la extracción de datos.	21
3.4.2 Datos a recopilar.	21
3.5 Procesamiento y limpieza de datos	22
3.6 Ingeniería de características	23
3.6.1 Creación de nuevas métricas.	23
3.6.2 Selección de variables.	23
3.7 Modelos de machine learning utilizados	24
3.7.1 Clasificación.	24
3.7.2 Regresión.	24
3.7.3 Clustering.	24
3.7.4 Evaluación de modelos.	25
3.8 Herramientas y entorno de trabajo	26
3.8.1 Bibliotecas de Python.	26
4.1 Arquitectura general del sistema	27
4.1.1 Archivos principales de datos.	27
4.1.2 Scripts de adquisición.	28
4.1.3 Problemas encontrados.	28

4.2	Pipeline de procesamiento	28
4.3	Ingeniería de características	29
4.3.1	Métricas creadas.	30
4.3.2	Justificación de las métricas creadas.	31
4.3.3	Técnicas de selección de características.	31
4.4	Entrenamiento y evaluación de modelos	32
4.5	Repositorio GitHub	33
5.1	Resultados de clasificación	34
5.1.1	Conclusión comparativa.	35
5.2	Resultados de regresión	36
5.2.1	Conclusión comparativa.	37
5.3	Resultados de clustering	37
5.4	Discusión e interpretación general	39
6.1	Conclusiones	40
6.2	Trabajo Futuro	40
A	Clasificación de rangos por MMR (Dota 2–Season 4)	44
B	Encuesta Aplicada a Jugadores.	45
C	Código Python	46
C.1	Extracción de datos desde la API de OpenDota	46
C.2	Preprocesamiento y Escalado	47
C.3	Regresión y Métricas	47
C.4	Clustering con KMeans	48
D	Métricas derivadas	49
E	Resultados de la regresión OLS.	50
F	Código de la regresión OLS	51
G	Figuras de Selección de Características.	52
G.1	Figura de Importancia de Características utilizando Información Mutua	52
G.2	Figura de Análisis SHAP de la Importancia de Características	53
H	Estructura del Repositorio.	54
I	Importancia de las Variables en la Predicción del <i>Performance Score</i> para Modelos de Regresión	55
J	Figuras del Análisis de K-Means.	57
K	Distribución de clústeres por género musical	58
K.1	Electrónica	58
K.2	Trap	59
K.3	Pop	60
K.4	Hip-Hop	61
K.5	Latina	62
K.6	Metal	63
K.7	Rock	64

ÍNDICE DE TABLAS

1	Tabla 1. Variables del estudio	19
2	Tabla 2. Porcentaje de géneros musicales en jugadores	20
3	Tabla 3. Datos personales del jugador	21
4	Tabla 4. Datos de comportamiento	21
5	Tabla 5. Métricas de rendimiento del jugador	22
6	Tabla 6. Métricas de evaluación por tipo de modelo	25
7	Tabla 7. Hiperparámetros de los modelos utilizados	32
8	Tabla 8. Relación entre tipo de modelo y variables de entrada-salida	32
9	Tabla 9. Métricas utilizadas para evaluar modelos de clasificación	33
10	Tabla 10. Métricas utilizadas para evaluar modelos de regresión	33
11	Tabla 11. Informe de Clasificación de Regresión Logística sin Música	34
12	Tabla 12. Informe de Clasificación de Random Forest sin Música	34
13	Tabla 13. Informe de Clasificación de Regresión Logística con Música	34
14	Tabla 14. Informe de Clasificación de Random Forest con Música	35
15	Tabla 15. Desempeño de los Modelos de Regresión	36
16	Tabla 16. Resumen general de perfiles de jugadores por clúster	38
17	Tabla 17. Resumen de variables derivadas	49
18	Tabla 18. Resultados de la regresión OLS	50
19	Tabla 19. Resultados de Regresión Lineal para Jugadores que No Escuchan Música	55
20	Tabla 20. Resultados de Random Forest para Jugadores que No Escuchan Música	55
21	Tabla 21. Resultados de Regresión Lineal para Jugadores que Escuchan Música	55
22	Tabla 22. Resultados de Random Forest para Jugadores que Escuchan Música	55
23	Tabla 23. Estadísticas por clúster - Electrónica	58
24	Tabla 24. Estadísticas por clúster - Trap	59
25	Tabla 25. Estadísticas por clúster - Pop	60
26	Tabla 26. Estadísticas por clúster - Hip-Hop	61
27	Tabla 27. Estadísticas por clúster - Música Latina	62
28	Tabla 28. Estadísticas por clúster - Metal	63
29	Tabla 29. Estadísticas por clúster - Rock	64

ÍNDICE DE FIGURAS

1	Diagrama de flujo del sistema	27
2	Tasa de victoria promedio por género musical	35
3	Tilt Factor vs Performance Score por clúster.	38
4	Distribución de MMR por rango en Dota 2 (Season 4). Fuente: Valve / DotaBuff.	44
5	Importancia de características según el análisis de información mutua.	52
6	Análisis SHAP de la importancia de características.	53
7	Método del codo para determinar k	57
8	Silhouette Score para determinar k	57
9	Distribución de clústeres para jugadores que escuchan Electrónica	58
10	Distribución de clústeres para jugadores que escuchan Trap	59
11	Distribución de clústeres para jugadores que escuchan Pop	60
12	Distribución de clústeres para jugadores que escuchan Hip-Hop	61
13	Distribución de clústeres para jugadores que escuchan Música Latina	62
14	Distribución de clústeres para jugadores que escuchan Metal	63
15	Distribución de clústeres para jugadores que escuchan Rock	64

INTRODUCCIÓN

1.1 Relevancia y justificación

En las últimas décadas, la industria de los videojuegos ha experimentado un crecimiento sin precedentes, consolidándose como una de las formas de entretenimiento más influyentes y rentables a nivel global. Los avances en hardware, software, inteligencia artificial (IA) y conectividad han transformado la experiencia de juego, permitiendo mundos virtuales más inmersivos y accesibles para una audiencia diversa. Esta evolución ha llevado a los videojuegos a trascender su propósito inicial de ocio, convirtiéndose en herramientas clave en la educación, la socialización y la competencia profesional (Wolf, 2021).

Actualmente, plataformas en línea facilitan la interacción entre millones de jugadores, fomentando comunidades globales y promoviendo eventos de eSports con premios millonarios. La profesionalización de los videojuegos ha dado lugar a equipos, entrenadores, analistas y patrocinadores que operan dentro de una industria en constante expansión. Su impacto se ha extendido a campos como la salud mental, la rehabilitación y el desarrollo de habilidades cognitivas, evidenciando su relevancia en diversos aspectos de la sociedad contemporánea (Pirker, 2023).

Desde una perspectiva económica, la industria de los videojuegos genera miles de empleos y mueve billones de dólares anualmente, impulsada por la venta de hardware, software, microtransacciones y servicios de streaming. Con la integración de nuevas tecnologías como la Virtual Reality (VR), el cloud gaming y el metaverso, su influencia seguirá en expansión, consolidándose como un pilar clave del entretenimiento y la innovación tecnológica (Wardyga, 2023).

En este contexto, Dota 2 se ha posicionado como uno de los videojuegos más representativos dentro de los eSports. Se trata de un videojuego de estrategia en tiempo real del género Multiplayer Online Battle Arena (MOBA), desarrollado por Valve Corporation. En este juego, dos equipos de cinco jugadores compiten para destruir la estructura central del enemigo, conocida como el Ancestro. Cada jugador elige un héroe con habilidades únicas, lo que requiere estrategias complejas y una alta coordinación en equipo. Con más de 7.5 millones de jugadores activos mensuales y torneos como The International, que han superado los 40 millones de dólares en premios, Dota 2 es uno de los eSports más influyentes a nivel mundial (Thorhauge, 2023).

Paralelamente, el impacto de la música en la cognición y el rendimiento humano ha sido ampliamente estudiado en disciplinas como la psicología y la neurociencia. Numerosos estudios han demostrado que la música influye en procesos como la memoria, la concentración y la toma de decisiones. Sin embargo, su papel en entornos de videojuegos competitivos, donde los jugadores deben procesar información rápidamente, comunicarse eficazmente y mantener precisión en su desempeño, sigue siendo un área poco explorada. Considerando que los eSports exigen altos niveles de concentración, reflejos y coordinación, comprender cómo la música afecta estos factores es crucial para optimizar el rendimiento de los jugadores (Goltz & Sadakata, 2021).

En base a lo anterior, este proyecto busca profundizar en la relación entre la música y el rendimiento en entornos altamente competitivos. Dado que la música ha demostrado influir en diversos procesos cognitivos y emocionales, su impacto en la concentración, la toma de decisiones y el desempeño bajo presión merece una mayor exploración. Este estudio pretende aportar nuevas perspectivas sobre cómo distintos estímulos auditivos pueden afectar la experiencia y eficacia de quienes participan en actividades que requieren atención sostenida y respuestas rápidas.

Los hallazgos no solo podrían contribuir al ámbito de los videojuegos competitivos, sino también generar conocimiento aplicable a otras disciplinas donde la concentración, la velocidad de reacción y la toma de decisiones son clave. Además, estos resultados podrían ser de interés para investigadores en neurociencia, educación y diseño de entornos digitales, abriendo la puerta a futuras investigaciones sobre la influencia de la música en el desempeño humano en diferentes contextos.

1.2 Objetivos generales y específicos

1.2.1 Objetivo general.

Analizar si la música impacta el desempeño competitivo de los jugadores de Dota 2, evaluando su influencia en métricas clave de rendimiento en un entorno competitivo.

1.2.2 Objetivos específicos.

- Examinar cómo diferentes géneros musicales afectan los tiempos de reacción de los jugadores durante el juego.
- Evaluar la influencia de la música en la calidad de la toma de decisiones estratégicas en situaciones de alta presión.
- Comparar el desempeño entre jugadores que escuchan música y aquellos que no, considerando métricas de rendimiento específicas.
- Identificar correlaciones entre géneros musicales y parámetros de rendimiento, tales como acciones por minuto (APM) y tasas de victoria.

1.3 Alcance y limitaciones

1.3.1 Alcance.

1. Evaluación del impacto de la música en el rendimiento competitivo de Dota 2, centrada en los tiempos de reacción, toma de decisiones y rendimiento general en un entorno altamente competitivo. La investigación se llevó a cabo durante un torneo local de Dota 2 en una universidad de Ibarra, lo que permitió observar a jugadores en condiciones reales de competencia.
2. Análisis cuantitativo basado en datos de partidas reales. El estudio utilizó datos de rendimiento de partidas y encuestas post-partida para recolectar información sobre la influencia de la música en el rendimiento. Los datos fueron obtenidos a través de la API de Dota 2 y encuestas, que incluyeron detalles sobre si los jugadores escucharon música, qué género y cuánto tiempo estuvieron expuestos.
3. Relevancia en el campo de los eSports y la neurociencia cognitiva. Los hallazgos podrían ser útiles para jugadores, entrenadores e investigadores en psicología cognitiva, neurociencia y diseño de videojuegos. El estudio contribuirá a comprender cómo los estímulos externos, como la música, afectan la toma de decisiones y la concentración en entornos competitivos.
4. Contribución al diseño de videojuegos y entrenamiento de jugadores. Los resultados podrían servir como base para diseñar videojuegos que integren música de manera estratégica para potenciar el rendimiento. También ofrecerán a entrenadores de eSports información valiosa para optimizar entrenamientos utilizando música como herramienta.
5. Desarrollo de herramientas y metodologías personalizadas. Se desarrollarán herramientas para el análisis de datos y la medición de variables relacionadas con la música y su impacto en el rendimiento. Estas herramientas podrían sentar las bases para futuras investigaciones en el campo de los videojuegos y su relación con la música.
6. Investigación aplicada a la mejora del rendimiento en eSports. Los resultados pueden proporcionar recomendaciones prácticas a jugadores y entrenadores, integrando música de forma estratégica en su entrenamiento para mejorar enfoque y toma de decisiones. La música podría convertirse en una herramienta para optimizar el rendimiento competitivo.
7. Posibles aplicaciones en otros juegos y áreas. Aunque el foco principal será Dota 2, los hallazgos podrían extenderse a otros juegos competitivos, proporcionando una visión sobre cómo la música influye en jugadores en diferentes contextos de eSports y otros entornos de alta presión.

1.3.2 Limitaciones.

1. Variabilidad individual. Factores como la experiencia previa del jugador, preferencias musicales y estados emocionales podrían influir en los resultados.

2. Muestra restringida. El estudio se centrará en jugadores de Dota 2, sin considerar otros juegos competitivos.
3. Condiciones controladas limitadas. No se pudo controlar completamente el entorno en el que los jugadores experimentaron la música, incluyendo factores como el tipo de audífonos y el volumen, lo que podría generar variaciones en los resultados.
4. Foco en jugadores de rol core. El estudio se centrará en jugadores de rol *core*, limitando la aplicabilidad de los resultados a otros roles dentro del juego (por ejemplo, *support*).
5. Número de partidas por jugador. El estudio se centrará en un único partido por jugador, lo que limita la capacidad de evaluar el rendimiento a lo largo de múltiples partidas. Esto podría afectar la consistencia de los resultados, ya que el rendimiento de un jugador puede variar significativamente entre diferentes partidas.
6. Tamaño del dataset limitado. Aunque se está ampliando el conjunto de datos, el análisis se basa en una muestra restringida, lo que podría influir en la generalización de los resultados. Un conjunto de datos más grande permitiría comparaciones más robustas.
7. Falta de estudios previos en el área. Actualmente, no existen investigaciones que investiguen específicamente el impacto de la música en el rendimiento competitivo de Dota 2. Esto representa una oportunidad para generar nuevos conocimientos, pero también una limitación, ya que no hay literatura existente para guiar el diseño del estudio o comparar resultados.

ESTADO DEL ARTE

El impacto de la música en el rendimiento humano ha sido objeto de estudio durante décadas en campos como la psicología cognitiva, la neurociencia y las ciencias del deporte. La música ha demostrado tener un efecto significativo en diversos aspectos del comportamiento y las funciones cognitivas, tales como la memoria, la atención, la toma de decisiones y el estado emocional. Sin embargo, su influencia en el contexto de los videojuegos competitivos, como Dota 2, es un área emergente y poco explorada.

A continuación, se presenta una revisión detallada de estudios previos que investigan la relación entre la música y diversos procesos cognitivos, así como su impacto en el rendimiento en videojuegos competitivos. Además, se abordarán las principales métricas de rendimiento en los videojuegos, con especial énfasis en Dota 2.

2.1 Influencia de la música en la atención y la cognición

Investigaciones previas han demostrado que la música puede influir significativamente en el procesamiento cognitivo, la atención y la memoria. Se ha subrayado que la música tiene un impacto positivo en el aprendizaje y la atención, especialmente en tareas repetitivas y en contextos educativos. Estos estudios sugieren que ciertos géneros musicales pueden mejorar la concentración y la memoria, facilitando el desempeño en actividades cognitivas complejas. Sin embargo, el impacto varía según las características específicas de la tarea y el tipo de música que se escuche (FARUQ & ÁZMI, 2024).

En un contexto más amplio, diversas investigaciones han enriquecido la comprensión sobre el impacto de la música en la toma de decisiones en situaciones que requieren un alto nivel de enfoque. Un estudio realizado en 2022 revela que la música no solo mejora la concentración, sino que también influye en la toma de decisiones arriesgadas, lo que puede tener aplicaciones interesantes en videojuegos de estrategia como Dota 2, donde las decisiones rápidas y precisas son clave para el rendimiento (Buelow et al., 2022).

Varios estudios indican que los efectos de la música pueden variar dependiendo del tipo de tarea. En tareas que demandan un alto grado de concentración, como la resolución de problemas complejos, la música puede ser tanto beneficiosa como perjudicial. Se sugiere que la música instrumental o sin letra es más efectiva para mantener el enfoque constante, mientras que otros géneros, como la música con letras, pueden ser más distractivos y afectar negativamente el rendimiento cognitivo. Además, géneros como la música clásica o ambiental tienden a tener efectos favorables en tareas que requieren calma y control emocional. En contraste, la música con ritmos rápidos puede aumentar el estrés, aunque también puede mejorar la motivación y el rendimiento en tareas de alta intensidad (Llanga Vargas & Insuasti Cárdenas, 2019).

El caso de Dota 2 y otros videojuegos competitivos ofrece un entorno único para estudiar estos efectos. En Dota 2, donde las decisiones rápidas y la estrategia son cruciales, los jugadores deben gestionar su atención, concentración y estado emocional bajo presión. La música podría influir en la percepción de la dificultad del juego y en la capacidad para tomar decisiones

acertadas en tiempo real. Esto abre una nueva área de investigación que explora los efectos de la música sobre la gestión emocional, la toma de decisiones bajo presión y el rendimiento general en el juego (Liu et al., 2016).

2.2 Música y rendimiento en videojuegos

La influencia de la música en el rendimiento dentro de los videojuegos ha sido objeto de numerosos estudios. Aunque gran parte de la investigación se ha centrado en aspectos de entretenimiento y experiencia del jugador, su impacto en el rendimiento también ha captado la atención de académicos. Los videojuegos, especialmente aquellos que requieren alta interacción y toma de decisiones rápidas, presentan un entorno ideal para analizar cómo la música puede afectar el rendimiento en situaciones de alta concentración y estrés (Beres, 2024).

Un estudio destacado en este campo, realizado en 2016, analiza cómo la música influye en la adaptación emocional de los jugadores en juegos afectivos. La investigación revela que la música desempeña un papel crucial en la creación del "flujo", un estado de intensa concentración que permite a los jugadores sumergirse completamente en la experiencia del juego. Se ha observado que la música con ritmos rápidos puede mejorar la respuesta inmediata de los jugadores, incentivando su participación y motivación. Sin embargo, en sesiones prolongadas, este tipo de música puede generar estrés cognitivo, afectando negativamente el rendimiento, especialmente en juegos que requieren atención constante y toma de decisiones bajo presión (Bontchev, 2016).

Por lo tanto, el equilibrio entre la estimulación que proporciona la música y el estrés que puede generar es fundamental para crear una experiencia de juego óptima. Aunque la música rápida y enérgica puede ser estimulante, también puede interferir con la capacidad de los jugadores para tomar decisiones precisas durante períodos prolongados. Este hallazgo destaca la importancia de ajustar la música según el ritmo del juego y la naturaleza de la tarea que el jugador debe ejecutar, un efecto que podría ser aún más relevante en videojuegos competitivos como Dota 2, donde los jugadores necesitan mantenerse enfocados durante largos períodos de tiempo (Ünal et al., 2012).

En un enfoque similar, se investiga cómo los factores externos, como la música, afectan el rendimiento en los videojuegos mediante el análisis de métricas de juego. En este estudio, se exploró cómo diferentes géneros musicales influyen en métricas clave, como los tiempos de reacción y la precisión en juegos de rally. Se encontró que la música puede impactar directamente estas métricas: los géneros con ritmos rápidos tienden a acelerar los tiempos de reacción, mientras que los géneros más suaves pueden mejorar la precisión al reducir el estrés y la ansiedad del jugador. Este estudio demuestra que la música no solo influye en la experiencia emocional del jugador, sino también en su rendimiento físico y cognitivo, aspectos cruciales en juegos de acción y estrategia como Dota 2 (Vallejos & Capa, 2010).

Además, se sugiere que los diseñadores de videojuegos pueden utilizar la música de manera estratégica para influir en el rendimiento de los jugadores, ajustando las selecciones musicales según las fases del juego. Por ejemplo, en momentos de alta tensión, una música calmada podría mejorar la precisión de los jugadores al tomar decisiones, mientras que durante las

fases de acción rápida, una música más energética podría potenciar los tiempos de reacción y la motivación (Salas et al., 2010).

El impacto de la música en la toma de decisiones y la precisión también ha sido explorado en otros estudios relacionados con los videojuegos. Guardini y Maninetti destacan cómo la música puede alterar el enfoque de los jugadores, afectando tanto su capacidad para tomar decisiones rápidas como su rendimiento general. En situaciones de alta presión, como en los enfrentamientos dentro de Dota 2, los jugadores deben tomar decisiones estratégicas rápidamente, y la música puede ayudar o perjudicar dependiendo de su naturaleza (Guardini & Maninetti, 2013).

2.3 Métricas de rendimiento en videojuegos

Las métricas de rendimiento en eSports, como el APM y las tasas de victoria, son fundamentales para evaluar el desempeño de los jugadores en videojuegos competitivos. Estas métricas permiten a analistas y entrenadores evaluar la eficiencia, rapidez y precisión de los jugadores en tiempo real. Sin embargo, el análisis de estas métricas debe considerar factores adicionales que pueden influir en el rendimiento, tales como los estímulos ambientales, incluyendo la música y los efectos sonoros, que han mostrado tener un impacto significativo en la capacidad de concentración, la coordinación en equipo y la toma de decisiones rápidas (Usca et al., 2024).

Un estudio reciente, realizado en 2021, explora cómo las métricas de juego, como el APM, se utilizan para analizar el rendimiento de los jugadores en eSports. Se destaca que factores externos, como la música y los sonidos, pueden afectar estas métricas. Por ejemplo, en juegos de ritmo rápido como Dota 2, donde los jugadores deben tomar decisiones rápidas y coordinarse con su equipo, la música con ritmos rápidos y los efectos sonoros pueden influir en la velocidad de las respuestas de los jugadores. El APM, que mide la cantidad de acciones realizadas por minuto, puede verse alterado por estímulos como la música. Algunos jugadores podrían experimentar un aumento en la velocidad de las acciones debido a la motivación o excitación que genera la música, mientras que otros podrían verse afectados negativamente si la música se percibe como distractora o estresante (Beres et al., 2021).

Además, es crucial que las métricas de rendimiento en videojuegos no se analicen de manera aislada, sino en el contexto del entorno completo del jugador, incluyendo factores psicológicos como el nivel de concentración, la motivación y el estrés. La música, como parte de este entorno, tiene el potencial de mejorar el enfoque y la eficiencia del jugador si se ajusta adecuadamente al tipo de juego y a la fase de la partida (Drachen & Canossa, 2009b).

En videojuegos competitivos como Dota 2, investigaciones recientes sugieren que es fundamental explorar cómo la música puede influir en parámetros clave del desempeño, tales como la atención, la coordinación en equipo y el rendimiento general. Por ejemplo, un estudio reciente compara los procesos de atención entre jugadores y no jugadores de videojuegos de disparos, destacando que los jugadores experimentan diferencias significativas en la forma en que manejan la atención y las distracciones durante el juego. Aunque este estudio se centra en juegos de disparos, sus hallazgos pueden extrapolarse a Dota 2 y otros videojuegos com-

petitivos, donde la capacidad de mantener la atención y coordinarse con el equipo es crucial para el éxito (Holm, 2023).

La música puede desempeñar un papel importante en la modulación de la atención de los jugadores. En un juego como Dota 2, donde la multitarea y la toma de decisiones rápidas son esenciales, la música podría ayudar a los jugadores a concentrarse mejor, especialmente en situaciones de alto estrés, o inducir distracción si no está bien seleccionada. Además, la música puede influir en la comunicación y la coordinación entre los miembros del equipo, dos elementos cruciales en los videojuegos cooperativos como Dota 2. Un ambiente sonoro adecuado puede fortalecer la cohesión del equipo y mejorar el rendimiento colectivo (Ángel Aristizábal et al., 2023).

El impacto de la música en el rendimiento general es una dimensión importante en los eSports. En Dota 2, el rendimiento de un jugador no solo se mide por sus habilidades individuales, sino también por su capacidad para trabajar de manera efectiva con su equipo. La música puede afectar tanto el estado de ánimo como la psicología del jugador, lo que a su vez puede influir en la toma de decisiones y en la forma en que interactúan con sus compañeros de equipo. Un estudio sobre la relación entre los factores emocionales y la toma de decisiones en eSports podría arrojar más luz sobre cómo la música puede afectar las decisiones estratégicas de los jugadores durante las partidas (Keith et al., 2016).

METODOLOGÍA

3.1 Contexto experimental

El experimento se realizó en un torneo competitivo de Dota 2 en una universidad local de Ibarra, con el objetivo de evaluar el impacto de la música en el rendimiento de los jugadores.

Los participantes seleccionados tenían un MMR (Matchmaking Rating) promedio entre 3850 y 5420 puntos, lo que corresponde a los rangos de habilidad "Ancient" a "Divine" según la calibración oficial del MMR en Dota 2 (ver figura 4), que se encuentra en el Apéndice A. Este rango fue elegido para garantizar que los jugadores tuvieran una comprensión adecuada de las mecánicas del juego y las estrategias colaborativas. Las partidas se llevaron a cabo en un formato estándar de 5 contra 5, con roles asignados (Carry, Midlaner, Offlaner, Soft Support y Hard Support), y tuvieron una duración promedio de 30 a 45 minutos, representativa del rango de habilidad estudiado.

3.2 Diseño del estudio

El presente estudio se estructuró bajo un enfoque cuantitativo, de tipo experimental y comparativo. Se adoptó un diseño de grupos paralelos, en el cual se manipuló la variable independiente de exposición a música para observar sus efectos sobre distintas métricas de rendimiento en jugadores de Dota 2. Esta aproximación permitió identificar relaciones causales y diferencias significativas entre los grupos de intervención y control (Turner, 2020).

La investigación se enmarca dentro de los estudios aplicados y utiliza un diseño experimental controlado. Al centrarse en la comparación directa entre jugadores expuestos a música y aquellos no expuestos, este enfoque también adquiere un carácter comparativo. Permite evaluar el impacto de distintos géneros musicales y niveles de exposición sobre variables de desempeño cognitivo y conductual (Hedrick et al., 1993).

3.3 Población y muestra

La población objetivo de este estudio está compuesta por jugadores activos de Dota 2 que cuentan con experiencia comprobada en partidas clasificatorias (ranked). Se seleccionaron participantes que cumplen con criterios específicos, tales como un rango de habilidad determinado y un número mínimo de partidas jugadas, para asegurar la homogeneidad del experimento y la validez de los resultados.

3.3.1 Criterios de selección.

Para garantizar la validez interna del estudio y la homogeneidad de los participantes, se establecieron criterios de selección específicos. Estos criterios permiten controlar la variabilidad en las métricas de rendimiento y asegurar que los jugadores sean aptos para participar en el experimento, evitando sesgos derivados de diferencias en habilidades o experiencia. A continuación, se describen los criterios establecidos:

- **Roles:** Se seleccionaron jugadores que desempeñan roles core"(mid, carry o offlane), ya que estas posiciones tienen un impacto directo y cuantificable en las métricas de rendimiento del juego.
- **MMR:** Se incluyeron jugadores con un MMR adecuado (entre 3850 y 5420) para garantizar un nivel de habilidad intermedio, lo cual permite evaluar el impacto de la música sin que diferencias extremas de habilidad interfieran con los resultados.
- **Actividad:** Se eligieron jugadores que han participado recientemente en partidas rankeadas (mínimo 50 partidas) para indicar que están activamente involucrados en el juego, asegurando que estén familiarizados con la actual meta y mecánicas, lo que garantiza un nivel de competencia adecuado en el entorno competitivo.

Estos criterios son fundamentales para el estudio y se relacionan directamente con las variables que se detallan en la Tabla 1, donde se presentan las variables del estudio y su clasificación como independientes y dependientes.

Tabla 1. Variables del estudio

Tipo de variable	Descripción
Independientes	<ul style="list-style-type: none"> ▪ Exposición a música (sí / no) ▪ Género musical escuchado ▪ Duración de la exposición a música
Dependientes	<ul style="list-style-type: none"> ▪ Performance Score ▪ Resultado de la partida (WinOrLose) ▪ Acciones por minuto (APM) ▪ Oro por minuto (GPM) ▪ Experiencia por minuto (XPM) ▪ Kills (asesinatos) ▪ Deaths (muertes) ▪ Assists (asistencias) ▪ Tilt Factor ▪ Engagement Level ▪ Clutch Factor

3.3.2 Tamaño de la muestra.

La muestra total utilizada en el estudio fue de 495 observaciones, obtenidas mediante muestreo intencional, seleccionando únicamente jugadores que cumplieran con los criterios de experiencia, nivel y rol. Esta muestra fue dividida equitativamente en dos grupos experimentales paralelos para permitir la comparación controlada entre condiciones con y sin música.

En el grupo con música, se eligieron solo aquellos géneros que representaban los más altos

porcentajes de preferencias entre los jugadores, resultando en un total de siete géneros seleccionados. Esto asegura que la experiencia musical se base en las preferencias más comunes dentro de la comunidad analizada.

3.3.3 División experimental

- **Grupo A** – Jugadores con música: participaron en partidas escuchando música a través de auriculares.
- **Grupo B** – Jugadores sin música: participaron en las mismas condiciones técnicas, pero sin escuchar música.

Ambos grupos participaron en partidas bajo condiciones parcialmente controladas, lo que facilitó una evaluación estandarizada del rendimiento. La duración promedio de las sesiones de juego, de 30 a 45 minutos, permitió analizar el impacto de la música en las variables cognitivas y conductuales.

3.3.4 Géneros musicales incluidos.

Los jugadores del grupo experimental escucharon listas de reproducción asociadas a los géneros musicales más populares entre la comunidad analizada. La selección de estos géneros se basó en los resultados agregados de encuestas previas realizadas a una muestra representativa de jugadores, eligiendo únicamente aquellos que representaban los más altos porcentajes de preferencias. Esto resultó en un total de siete géneros seleccionados, asegurando que la experiencia musical reflejara las preferencias más comunes dentro de la comunidad. Los géneros más frecuentes se encuentran definidos en la Tabla 2.

Tabla 2. Porcentaje de géneros musicales en jugadores

Género Musical	Porcentaje (%)
Electronic	15.38
Hip-Hop	13.77
Latin Music	10.12
Metal	14.17
Pop	15.79
Rock	10.53
Trap/Rap	8.91

Estos géneros fueron elegidos por su alta prevalencia en las respuestas y para mantener la naturalidad del entorno competitivo. El formato de los formularios utilizados se encuentra en el Apéndice B, y los resultados están disponibles en la carpeta `data/Surveys` del repositorio GitHub del proyecto.

3.4 Recolección de datos

La recolección de datos para este estudio se llevó a cabo mediante una combinación de fuentes automatizadas y autodeclaradas, utilizando herramientas tecnológicas orientadas a

la obtención eficiente y estructurada de información relevante.

- **Application Programming Interface (API) de Dota 2:** La recopilación automatizada de datos se realizó a través de la OpenDota API¹, la cual proporciona información detallada sobre partidas, estadísticas de jugadores y eventos relevantes en tiempo real.
- **Instrumento de auto-reporte (cuestionario musical):** Se elaboró y aplicó un cuestionario personalizado con el objetivo de recopilar información cualitativa sobre la presencia de música y sus características durante las sesiones de juego. Este instrumento fue administrado de forma inmediata al finalizar cada partida, y su formato completo se presenta en el Apéndice B.

3.4.1 Scripts en python para la extracción de datos.

Para facilitar la recolección estructurada de datos sobre el rendimiento de los jugadores, se desarrolló un script en Python que automatiza la conexión con la API pública de OpenDota. Este script toma una lista de identificadores de partidas (Match IDs) desde un archivo Excel, realiza solicitudes HTTP a la API para recuperar información detallada de cada partida y extrae métricas relevantes por jugador, tales como GPM, XPM, APM y oro total, entre otros. Posteriormente, los datos son almacenados en un archivo Excel con validación de duplicados.

Para obtener más información sobre el funcionamiento de estos scripts, consulte la sección **Implementación del Experimento**².

3.4.2 Datos a recopilar.

Durante cada partida, se recopilan las métricas que permiten evaluar el desempeño y comportamiento de los jugadores. Estas métricas se presentan en las Tablas 3-5.

Tabla 3. Datos personales del jugador

Dato	Descripción
ID de jugador	Identificación única de cada jugador.
Nivel de habilidad (MMR)	Comparación justa entre jugadores.
Medalla	Medalla asociada con el número de MMR.
Rol	Rol del jugador en la partida.
GPMvsAVG	Comparación del GPM con el promedio del héroe.
XPMvsAVG	Comparación del XPM con el promedio del héroe.

Tabla 4. Datos de comportamiento

Dato	Descripción
Pushing	Puntuación de empuje de líneas y torres (0-10).
Fighting	Puntuación de combate (0-10).
Farming	Puntuación de farreo (0-10).

¹La documentación oficial de la API está disponible en <https://docs.opendota.com/>

²Específicamente, en la subsección 4.1.2 **Scripts de adquisición**.

Tabla 5. Métricas de rendimiento del jugador

Dato	Descripción
WinOrLose	Resultado de la partida (victoria o derrota).
Game Duration	Duración de la partida en minutos.
Kills	Número de enemigos derrotados.
Deaths	Número de muertes del jugador.
Assists	Asistencias proporcionadas.
APM	Número de acciones por minuto.
GPM	Dinero ganado por minuto.
XPM	Experiencia obtenida por minuto.
Final Net Worth	Valor neto final del jugador.

3.5 Procesamiento y limpieza de datos

El preprocesamiento de datos es fundamental para garantizar la calidad y consistencia antes de aplicar técnicas de modelado. Este proceso asegura que los modelos de machine learning no se vean afectados por errores, escalas inconsistentes o ruido, y ayuda a evitar sesgos y sobreajuste durante el entrenamiento. A continuación, se describen los procedimientos implementados:

- **Tratamiento de valores faltantes:** Se identificaron valores nulos en columnas numéricas como APM, GPM y XPM. Para su tratamiento, se utilizó la mediana de cada variable, dada su robustez frente a valores atípicos. Según el Análisis Exploratorio de Datos (EDA), la mediana es preferible cuando los datos no presentan una distribución normal (Imtiaz & Shah, 2008).
- **Codificación de variables categóricas:** La variable *Género musical* fue transformada mediante codificación *One-Hot*, creando una columna binaria por cada género en el conjunto de datos. Para los casos sin música, se asignó un vector de ceros, asegurando una interpretación neutral. Esta estrategia previene la codificación errónea y evita imponer un orden jerárquico artificial entre los géneros (Sahin, 2023).
- **Eliminación de valores atípicos:** Se utilizó el rango intercuartílico (IQR) para detectar y eliminar valores extremos en variables como *Final Net Worth*, *Kills*, *Deaths*, *Assists*, *APM*, *GPM* y *XPM*. Esta técnica identifica observaciones fuera del rango $[Q1 - 1,5 \times IQR, Q3 + 1,5 \times IQR]$, mitigando el impacto de outliers, especialmente en distribuciones asimétricas comunes en datasets de videojuegos (Tukey, 1977).
- **Normalización de variables:** Las variables continuas fueron escaladas utilizando *StandardScaler* de Scikit-learn, transformando los datos para que tengan media cero y desviación estándar uno. Esto mejora la comparabilidad entre métricas de distinta escala y facilita la convergencia de algoritmos de optimización basados en gradiente (Lorenzo, 2024).
- **Conversión de variables categóricas simples:** Variables como *WinOrLose*, así como los conteos de acciones (*Kills*, *Deaths*, *Assists*), fueron convertidas al tipo entero (`int`) para facilitar su procesamiento por los modelos de clasificación. Esta conversión

garantiza que se traten como clases discretas válidas, mejorando la eficiencia computacional y asegurando compatibilidad con librerías como Scikit-learn (Pedregosa et al., 2011).

- **Balanceo del conjunto de datos:** Se implementó un balanceo para evitar sesgos en los modelos de clasificación, detectando un desbalance entre las clases con música "sin música". Se aplicó remuestreo con reposición (upsampling) sobre la clase minoritaria, manteniendo todas las observaciones originales y replicando ejemplos de la clase minoritaria. Esta estrategia previene que los modelos aprendan patrones sesgados (Chawla et al., 2002).

El resultado final del preprocesamiento se consolidó en el archivo `preprocessed/_data.xlsx`, validado manualmente y con scripts automatizados para asegurar su consistencia. Este archivo está disponible en el repositorio de GitHub del proyecto, en la carpeta `Data/Clean/`, y se utilizó como fuente principal para la etapa de ingeniería de características y modelado.

3.6 Ingeniería de características

La ingeniería de características es un paso fundamental en la preparación de datos, ya que permite identificar y crear variables que brindan una comprensión más profunda del comportamiento de los jugadores en Dota 2. En este proyecto, se desarrollaron nuevas métricas y se implementaron diversas técnicas de selección de variables para mejorar el rendimiento predictivo.

3.6.1 Creación de nuevas métricas.

Se diseñaron varias métricas para capturar diferentes aspectos del rendimiento de los jugadores, entre las cuales se incluyen:

- **Performance Score (PS):** Evalúa el rendimiento global de un jugador, considerando Kills, Assists, Deaths, GPM, XPM y el resultado de la partida.
- **Clutch Factor (CF):** Mide la capacidad del jugador para tener un impacto decisivo en momentos clave de la partida.
- **Tilt Factor (TF):** Cuantifica el nivel de frustración o desequilibrio emocional del jugador a lo largo del juego.
- **Engagement Level (EL):** Valora la participación activa del jugador en peleas y en la ofensiva del equipo.

3.6.2 Selección de variables.

Para seleccionar las características más relevantes para el análisis del rendimiento, se utilizaron tres técnicas de selección de características:

- **SHAP (SHapley Additive exPlanations):** Aplica la teoría de juegos para calcular la contribución marginal de cada característica al resultado del modelo.

- **Información Mutua (Mutual Info):** Mide la dependencia entre las variables sin requerir una relación lineal, permitiendo identificar correlaciones no lineales.
- **Importancia por Permutación (Permutation Importance):** Evalúa la importancia de las variables en un modelo entrenado, determinando cómo la permutación de las características afecta el rendimiento del modelo.

3.7 Modelos de machine learning utilizados

Esta investigación implementó un conjunto de modelos de aprendizaje automático supervisado y no supervisado con el objetivo de analizar el impacto de la música y sus géneros en el desempeño competitivo de jugadores de Dota 2. Los modelos fueron clasificados en tres categorías: clasificación, regresión y clustering, y evaluados utilizando métricas cuantitativas específicas.

3.7.1 Clasificación.

Se utilizaron modelos de clasificación para predecir la variable binaria *WinOrLose*, que indica si un jugador ganó o perdió una partida. Los modelos se seleccionaron por su capacidad de manejar múltiples variables e interpretar patrones de comportamiento:

- **Regresión Logística:** Este modelo se utilizó como base debido a su simplicidad y capacidad interpretativa.
- **Random Forest Classifier:** Fue seleccionado por su robustez frente a relaciones no lineales y su habilidad para manejar múltiples características sin requerir una normalización estricta.

3.7.2 Regresión.

La predicción del *Performance Score*, una métrica compuesta destinada a medir el desempeño competitivo, se llevó a cabo mediante modelos de regresión. Estos modelos facilitaron la cuantificación del impacto relativo de cada variable independiente en el rendimiento general del jugador:

- **Regresión Lineal Múltiple:** Para establecer relaciones lineales entre características y su impacto en el rendimiento.
- **Random Forest Regressor:** Para capturar relaciones no lineales entre variables como *Clutch_Factor*, *Tilt_Factor* y la variable dependiente.

3.7.3 Clustering.

Se aplicó el algoritmo K-Means para agrupar a los jugadores en función de su estilo de juego, utilizando las siguientes características:

- APM

- GPM
- XPM
- Clutch_Factor
- Tilt_Factor
- Performance Score

El número óptimo de clústeres se determinó mediante:

- Método del Codo
- Índice de Silueta

El número óptimo de clústeres se determinó mediante:

- Método del Codo (Elbow Method): Este método ayuda a identificar el punto donde la adición de más clústeres no proporciona una mejora significativa en la varianza explicada, permitiendo un equilibrio entre simplicidad y precisión (Cui et al., 2020).
- Índice de Silueta (Silhouette Index): Este índice evalúa la calidad de la agrupación al medir la cohesión y separación de los clústeres, asegurando que los jugadores dentro de cada clúster sean similares entre sí y distintos de otros clústeres (Shutaywi & Kachouie, 2021).

Al utilizar ambos métodos, se obtiene una evaluación más robusta y confiable de la estructura de los clústeres. El Método del Codo proporciona una visualización clara del punto de inflexión en la varianza explicada, lo que ayuda a seleccionar un número adecuado de clústeres sin complicar innecesariamente el modelo. Por otro lado, el Índice de Silueta ofrece una medida cuantitativa de la calidad de la agrupación, asegurando que los clústeres sean coherentes internamente y distintos entre sí. Juntos, estos métodos permiten una toma de decisiones informada y fundamentada en la optimización del análisis de los estilos de juego (Saputra et al., 2020).

3.7.4 Evaluación de modelos.

La evaluación de los modelos de machine learning se llevó a cabo utilizando las siguientes métricas, como se detalla en la Tabla 6.

Tabla 6. Métricas de evaluación por tipo de modelo

Tipo de Modelo	Métricas de Evaluación
Clasificación	Precisión, Recall, F1-Score
Regresión	R^2 (Coeficiente de Determinación), MSE (Error Cuadrático Medio)
Clustering	Inercia (WCSS), Índice de Silueta, Número de Clústeres

Estas métricas permitieron comparar de manera objetiva el rendimiento y la fiabilidad de los modelos aplicados a distintas tareas. Al utilizar métricas específicas para cada tipo de

modelo, se garantiza una evaluación más precisa y adecuada a las características de cada técnica analítica.

3.8 Herramientas y entorno de trabajo

Para el desarrollo del proyecto se utilizaron diversas herramientas de software y bibliotecas que permitieron la adquisición, procesamiento, análisis y visualización de los datos de forma eficiente y reproducible:

- **Python 3.10:** Lenguaje de programación principal para la manipulación de datos y modelado (Hunt, 2019).
- **Google Colab:** Entorno de desarrollo en la nube para la ejecución interactiva de código y colaboración (Bisong, 2019).
- **Microsoft Excel (.xlsx):** Para el almacenamiento estructurado de datos recolectados y procesados.
- **GitHub:** Plataforma de control de versiones utilizada para almacenar el código fuente del proyecto, incluyendo scripts de procesamiento, notebooks y documentación.

3.8.1 Bibliotecas de Python.

- **Pandas:** Para manipulación y limpieza de datos tabulares.
- **NumPy:** Para operaciones matemáticas y manejo de arrays.
- **Scikit-learn:** Para implementación de modelos de clasificación, regresión y evaluación.
- **Matplotlib y Seaborn:** Para visualizaciones estadísticas y exploratorias.
- **SHAP:** Para interpretación de modelos y análisis de importancia de características.
- **Imbalanced-learn:** Para técnicas de balanceo de clases.

Estas herramientas facilitaron un flujo de trabajo ágil y replicable, adecuado para el tratamiento de grandes volúmenes de datos y técnicas avanzadas de aprendizaje automático.

IMPLEMENTACIÓN DEL EXPERIMENTO

4.1 Arquitectura general del sistema

El sistema implementado para el análisis del rendimiento de jugadores de Dota 2 sigue una arquitectura modular y estructurada, organizada en distintas etapas que abarcan desde la recolección de datos hasta su visualización e interpretación.

A continuación, se presenta un diagrama que ilustra el flujo general de los componentes del sistema (ver Figura 1):

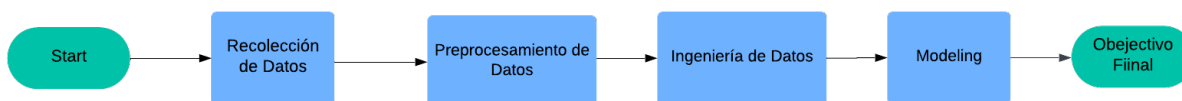


Figura 1: Diagrama de flujo del sistema

Este flujo permite la extracción automatizada de partidas a través de la API oficial de Dota 2, su almacenamiento estructurado en archivos Excel, el procesamiento mediante scripts personalizados en Python y, finalmente, la aplicación de técnicas de machine learning y visualización para la interpretación de los datos recolectados.

4.1.1 Archivos principales de datos.

La gestión de datos se organiza en hojas de cálculo .xlsx, lo que facilita un análisis estadístico accesible. Durante el desarrollo del experimento, se creó una estructura de almacenamiento basada en archivos, permitiendo un procesamiento ágil y flexible de volúmenes variables de datos. Esto optimiza el uso de scripts de Python y herramientas de análisis en Google Colab. Se utilizan cuatro archivos clave a lo largo del proceso:

- **Data_with_music.xlsx**: Contiene el conjunto de datos crudos con todas las observaciones recolectadas a partir de partidas de Dota 2, incluyendo indicadores de rendimiento, duración de partida, actividad del jugador y datos musicales.
- **preprocessed_data.xlsx**: Representa el conjunto de datos limpio, normalizado y balanceado después del tratamiento de valores atípicos, codificación de variables categóricas y escalado de variables continuas.
- **selected_features.xlsx**: Contiene únicamente las variables más significativas para el análisis, seleccionadas mediante técnicas de aprendizaje automático y de ingeniería de características.
- **Popular_Survey_Responses_Music_Dota2.xlsx**: Resultados de encuestas, con la recopilación de información de los géneros musicales más populares entre los jugadores.

4.1.2 Scripts de adquisición.

Para automatizar la obtención y estructuración de datos, se desarrollaron scripts en Python que interactúan con la API oficial de Dota 2. Además, se integraron datos musicales utilizando el archivo **Popular_Survey_Responses_Music_Dota2.xlsx**, el cual recopila información sobre los géneros más escuchados durante las partidas, la duración de escucha y otras preferencias relevantes. Cada entrada fue asociada con las estadísticas de juego mediante el *Match ID* y el *Player ID*, lo que permitió vincular el contexto musical a jugadores específicos dentro de partidas determinadas.

Las principales tareas realizadas por los scripts fueron:

- Asociación de registros musicales (género, duración) con partidas y jugadores mediante el *Match ID* y el *Player ID*.
- Extracción de estadísticas por partida y jugador desde la API.
- Generación automática de archivos .xlsx para su posterior análisis.

Para cumplir con el límite de velocidad de la API (60 solicitudes por minuto), se implementaron pausas programadas dentro de los scripts. Un fragmento del código se encuentra en el Apéndice C.1, y el script completo está disponible en el repositorio del proyecto, en la carpeta `src/DataCollect.py`.

4.1.3 Problemas encontrados.

Durante la implementación del experimento, surgieron algunos desafíos:

- **Registros duplicados:** Algunas partidas se descargaron más de una vez debido a fallas en la conexión con la API, lo cual se resolvió mediante deduplicación usando el ID de partida.
- **Datos inconsistentes:** Algunas partidas contenían valores atípicos o incompletos, especialmente en jugadores con desconexiones o partidas anómalas.
- **Datos faltantes:** Se identificaron campos con valores nulos en variables como Final Net Worth o Kills, que fueron tratados durante la fase de preprocesamiento (ver sección 3.5),

Este enfoque basado en archivos permite un manejo más transparente de los datos y facilita la reproducibilidad de los resultados del experimento.

4.2 Pipeline de procesamiento

El *pipeline* de procesamiento de datos consistió en una secuencia lógica de pasos diseñados para preparar el conjunto de datos de forma eficiente, reduciendo el riesgo de errores y maximizando la calidad del input utilizado para el modelado predictivo. A continuación, se detallan las acciones implementadas:

- **Limpieza inicial:** Se identificaron y trataron los valores nulos mediante imputación con la mediana, decisión tomada por su robustez ante valores extremos y su idoneidad en distribuciones no normales.
- **Transformación de variables:** La variable categórica `Género musical` fue convertida utilizando codificación `One-Hot`, evitando la introducción de relaciones ordinales artificiales entre géneros. Para los casos sin música, se aplicó una codificación con vectores de ceros, manteniendo la coherencia semántica del conjunto de datos.
- **Normalización:** Se optó por la normalización mediante `StandardScaler` para asegurar que todas las variables cuantitativas se encontraran en una escala comparable. Esta decisión fue crucial para mejorar el rendimiento de algoritmos como la regresión logística y reducir la varianza durante el entrenamiento.
- **Eliminación de valores atípicos:** El método del rango intercuartílico (IQR) fue preferido frente a métodos basados en desviación estándar, ya que no se ve afectado por colas largas ni distribuciones asimétricas. En el contexto de videojuegos, métricas como `Net Worth` o `APM` presentan variabilidad extrema debido a partidas anómalas (como `stomps` o `remontadas`)³, lo que justifica el uso de IQR.
- **Conversión de tipos:** Variables como `WinOrLose`, `Kills`, `Deaths` y `Assists` fueron convertidas explícitamente al tipo entero para asegurar compatibilidad con los algoritmos de clasificación y evitar errores de tipo.
- **Balaneo del dataset:** Se identificó un desbalance entre los casos “con música” y “sin música”. Para evitar sesgos de predicción hacia la clase mayoritaria, se aplicó `upsampling` sobre la clase minoritaria, replicando sus observaciones con reposición. Esta técnica fue elegida por su efectividad sin pérdida de información original.

El resultado final fue almacenado en el archivo `preprocessed_data.xlsx`, el cual fue validado manualmente y con scripts automatizados. Este archivo se encuentra disponible en el repositorio del proyecto, dentro de la carpeta `Data/Clean/`, y fue utilizado como base para la ingeniería de características y el modelado.

Un fragmento del código utilizado para este procesamiento se encuentra disponible en el Apéndice C.2, y el código completo está en el repositorio de GitHub del proyecto, en la carpeta `Src`.

4.3 Ingeniería de características

Con el objetivo de identificar las variables más relevantes para el análisis del rendimiento en partidas con o sin música, se aplicaron técnicas avanzadas de **ingeniería de características** y **selección supervisada de variables**. Este proceso permitió enriquecer el dataset con nuevas métricas derivadas del comportamiento del jugador, facilitando así el análisis predictivo del rendimiento general y sus fluctuaciones en diferentes escenarios musicales.

³Un "stomp" se refiere a una partida en la que un equipo es completamente dominado por el otro, mientras que una `remontada`.^{es} cuando un equipo logra recuperar la ventaja y ganar después de estar en desventaja.

4.3.1 Métricas creadas.

Las métricas fueron diseñadas a partir de estadísticas recolectadas a través de la API de Dota 2. Las fórmulas se definieron mediante pruebas de correlación y modelos de regresión lineal para evaluar el peso de cada componente en los resultados finales (**WinOrLose** y **Performance Score**). Para un resumen completo y sus interpretaciones, consulte la Tabla 17 en el Apéndice D.

Los resultados completos del modelo de regresión lineal, incluyendo los pesos de las variables, se presentan en la Tabla 18 en el Apéndice E. Asimismo, el código utilizado para generar los resultados se encuentra en el Apéndice F.

- **Performance Score:** Representa el rendimiento global del jugador. Se calcula ponderando kills, assists, muertes, economía y resultado de la partida:

$$PS = 0,8 \cdot \text{Kills} + 1,2 \cdot \text{Assists} - 1,5 \cdot \text{Deaths} + \frac{\text{GPM}}{3} + \frac{\text{XPM}}{10} + 5 \cdot \text{WinOrLose}$$

Interpretación: Valores altos indican un rendimiento global excelente; valores bajos sugieren baja participación o mal resultado.

- **Clutch Factor (CF):** Mide el rendimiento en situaciones críticas.

$$CF = \frac{0,8 \cdot \text{Kills} + 1,2 \cdot \text{Assists}}{1,5 \cdot \text{Deaths} + 1}$$

Interpretación: Valores altos significan que el jugador fue decisivo bajo presión. Valores bajos reflejan falta de impacto.

- **Engagement Level (EL):** Representa la participación ofensiva del jugador.

$$EL = 0,8 \cdot \text{Kills} + 1,2 \cdot \text{Assists}$$

Interpretación: Un valor alto indica involucramiento constante en peleas. Bajo sugiere poca presencia en el juego.

- **Tilt Factor (TF):** Estima el nivel de frustración o desbalance emocional durante la partida.

$$TF = 0,8 \cdot \text{Kills} + 1,2 \cdot \text{Assists} - 1,5 \cdot \text{Deaths}$$

Interpretación: Valores negativos indican tendencia al tilt o frustración. Positivos reflejan confianza y estabilidad.

Origen de las Ecuaciones: Las ecuaciones para las métricas se basan en investigaciones previas sobre rendimiento en e-sports, las cuales proporcionaron un marco teórico sólido (Álvarez et al., 2021). A partir de un análisis exhaustivo de los factores que influyen en el rendimiento del jugador, se llevaron a cabo análisis estadísticos detallados, incluyendo correlaciones y modelos de regresión lineal. Esto permitió ajustar los pesos asignados a cada componente, asegurando que reflejan con precisión su importancia. Por ejemplo, se demostró que los kills y assists tienen un impacto significativo en el rendimiento global, lo que justificó su mayor peso en el *Performance Score* (Díaz et al., 2024).

4.3.2 Justificación de las métricas creadas.

Las métricas derivadas se desarrollaron para complementar la interpretación del rendimiento de los jugadores más allá de variables tradicionales como Kills, Deaths o APM, ofreciendo una perspectiva más amplia del rendimiento en Dota 2. Algunas métricas clave son:

- **Claridad de propósito:** Cada métrica captura aspectos específicos del comportamiento del jugador, como consistencia, respuesta bajo presión o frustración (Drachen & Canossa, 2009a).
- **Fundamento teórico:** Métricas como el *Tilt Factor* se basan en estudios de psicología del rendimiento, donde los jugadores muestran variabilidad emocional ante eventos negativos (Cregan et al., 2025). Otras, como el *Clutch Factor*, reflejan la capacidad de sobresalir en momentos críticos (Poulus et al., 2022).
- **Interpretabilidad de resultados:** La inclusión de estas métricas mejora la interpretación no solo de si un jugador ganó, sino también de *por qué* y bajo qué condiciones se alcanzó ese resultado (Hooshyar et al., 2018), proporcionando así una visión más profunda del rendimiento del jugador.

Además, abordan limitaciones de variables existentes:

- Kills y Deaths no siempre reflejan impacto en equipo (Kleinman et al., 2022).
- APM no captura calidad de decisiones, solo cantidad (Sharpe et al., 2022).
- Win/Loss es binario y no expresa desempeño relativo (Varga et al., 2024).

4.3.3 Técnicas de selección de características.

Para evaluar el impacto de cada componente en las variables dependientes `WinOrLose` y `Performance Score`, se aplicaron tres técnicas de selección de características a un conjunto de métricas derivadas. Cada método proporciona una perspectiva única sobre la importancia de las variables mediante enfoques estadísticos distintos, mejorando la comprensión y robustez de los hallazgos, así como permitiendo la validación cruzada. A continuación se describen las técnicas utilizadas:

- **Información Mutua (Mutual Information):** Esta técnica mide la dependencia entre variables sin requerir una relación lineal. Se utilizó `mutual_info_regression` de Scikit-learn para identificar variables con mayor información compartida con `Performance Score`, facilitando la selección de correlaciones no lineales, especialmente para el *Tilt Factor* (Cardona & VELÁSQUEZ, 2006). Las características más relevantes identificadas son: **Tilt Factor**, **Clutch Factor**, **XPM**, **GPM** y **APM**, cuyos resultados se presentan en el apéndice G.1
- **Importancia por Permutación (Permutation Importance):** Se entrenó un modelo de Random Forest y se evaluó la disminución del rendimiento al permutar los valores de cada característica, determinando así la importancia relativa de cada variable (Hooker et al., 2021).

- **SHAP (SHapley Additive exPlanations)**: Este método calcula la contribución marginal de cada característica al resultado del modelo, aplicado al clasificador Random Forest para obtener interpretaciones precisas y transparentes sobre el impacto de cada variable en las predicciones (Li et al., 2024). Este análisis revela que **Clutch Factor**, **WinOrLose**, **Kills** y **Assists** tienen un alto impacto en la predicción del *Performance Score*, cuyos resultados se muestran en el apéndice G.2.

Las características seleccionadas se documentaron en el archivo *selected_features.xlsx*, que incluye variables originales y derivadas, organizadas alfabéticamente para facilitar el análisis y asegurar la trazabilidad en las etapas de modelado y visualización.

4.4 Entrenamiento y evaluación de modelos

Para los modelos de clasificación, como la Regresión Logística y el Random Forest Classifier, se empleó validación cruzada para evaluar su desempeño. Los modelos de regresión, como la Regresión Lineal y el Random Forest Regressor, fueron ajustados con hiperparámetros predefinidos y también evaluados mediante validación cruzada para asegurar una adecuada generalización. Todos los modelos se entrenaron utilizando la biblioteca Scikit-learn en Python.

A continuación, se presentan los modelos utilizados junto con la biblioteca correspondiente y los hiperparámetros configurados para su entrenamiento en la Tabla 7.

Tabla 7. Hiperparámetros de los modelos utilizados

Modelo	Librería	Hiperparámetros
Regresión Logística	Scikit-learn	C=10, penalty='l2', solver='liblinear'
Random Forest Classifier	Scikit-learn	n_estimators=300, min_samples_split=2, max_depth=10
Regresión Lineal	Scikit-learn	Default
Random Forest Regressor	Scikit-learn	n_estimators=200, max_depth=None, max_features='sqrt'

La Tabla 8 muestra la relación entre los diferentes modelos de machine learning y las variables de entrada utilizadas para predecir los resultados.

Tabla 8. Relación entre tipo de modelo y variables de entrada-salida

Modelo	Tipo	Entrada (Features)	Salida
Regresión Logística	Clasificación	APM, GPM, XPM, Clutch Factor, Tilt Factor, Engagement Level	WinOrLose (0 o 1)
Random Forest Classifier	Clasificación	Igual que arriba	WinOrLose (0 o 1)
Regresión Lineal	Regresión	Igual que arriba	Performance Score
Random Forest Regressor	Regresión	Igual que arriba	Performance Score

Para medir el desempeño de los modelos, se utilizaron diversas métricas según el tipo de tarea (clasificación o regresión). En la Tabla 9 se presentan las métricas empleadas para los modelos de clasificación.

Tabla 9. Métricas utilizadas para evaluar modelos de clasificación

Métrica	Descripción
Accuracy	Proporción de predicciones correctas sobre el total de predicciones realizadas.
Precision	Proporción de verdaderos positivos sobre el total de predicciones positivas.
Recall	Proporción de verdaderos positivos sobre el total de casos reales positivos.
F1-Score	Media armónica entre Precision y Recall, útil en conjuntos de datos desbalanceados.
Matriz de Confusión	Tabla que muestra verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos.

La Tabla 10 presenta las métricas utilizadas para evaluar los modelos de regresión, las cuales son fundamentales para entender el rendimiento y la efectividad de los modelos en la predicción de valores.

Tabla 10. Métricas utilizadas para evaluar modelos de regresión

Métrica	Descripción
Coefficiente de determinación (R^2)	Mide la proporción de la varianza de la variable dependiente explicada por el modelo.
Error cuadrático medio (MSE)	Media de los cuadrados de los errores, indica el grado de precisión del modelo.
Gráfico de Predicción vs Valor real	Visualización que permite comparar los valores predichos con los valores reales.

4.5 Repositorio GitHub

El código fuente, los datos, notebooks y documentación del proyecto se encuentran organizados en un repositorio público en GitHub. La estructura del repositorio se detalla en el Apéndice H⁴.

⁴Repositorio: *Dota2 Performance Music* en GitHub: <https://github.com/N3pthys/Dota2-Performance-Music.git>

RESULTADOS Y DISCUSIÓN

Este capítulo presenta los hallazgos de los modelos de clasificación, regresión y clustering aplicados a los datos de jugadores de Dota 2. Se evaluó el efecto de la música y los géneros musicales en el rendimiento competitivo, utilizando métricas como `WinOrLose`, `Performance Score`, `APM`, `Clutch Factor` y `Tilt Factor` para capturar diferentes dimensiones del desempeño en las partidas.

5.1 Resultados de clasificación

Se entrenaron dos modelos supervisados —Regresión Logística y Random Forest— para predecir si un jugador ganaría o perdería una partida (`WinOrLose`). El conjunto de datos se segmentó en dos grupos: con música y sin música durante las partidas.

En el grupo de jugadores que no escuchaban música, los modelos mostraron los resultados que se presentan en la Tabla 11 y Tabla 12.

Tabla 11. Informe de Clasificación de Regresión Logística sin Música

Clase	Precisión	Recall	F1-Score	Soporte
0 (Lose)	0.81	0.93	0.86	27
1 (Win)	0.89	0.74	0.81	23
Exactitud	0.84			
Promedio macro	0.85	0.83	0.84	50
Promedio ponderado	0.85	0.84	0.84	50

Tabla 12. Informe de Clasificación de Random Forest sin Música

Clase	Precisión	Recall	F1-Score	Soporte
0 (Lose)	0.86	0.93	0.89	27
1 (Win)	0.90	0.83	0.86	23
Exactitud	0.88			
Promedio macro	0.88	0.88	0.88	50
Promedio ponderado	0.88	0.88	0.88	50

Para el grupo de jugadores que escucharon música, ambos modelos mantuvieron buena precisión, aunque el rendimiento fue diferente. En la Tabla 13 y Tabla 14 se presentan los resultados.

Tabla 13. Informe de Clasificación de Regresión Logística con Música

Clase	Precisión	Recall	F1-Score	Soporte
0 (Lose)	0.62	0.86	0.72	21
1 (Win)	0.86	0.62	0.72	29
Exactitud	0.72			
Promedio macro	0.74	0.74	0.72	50
Promedio ponderado	0.76	0.72	0.72	50

Tabla 14. Informe de Clasificación de Random Forest con Música

Clase	Precisión	Recall	F1-Score	Soporte
0 (Lose)	0.83	0.95	0.89	21
1 (Win)	0.96	0.86	0.91	29
Exactitud	0.90			
Promedio macro	0.90	0.91	0.90	50
Promedio ponderado	0.91	0.90	0.90	50

El análisis por género musical se presenta en la siguiente figura 2, que muestra la tasa de victoria promedio de los jugadores según el género musical que escuchan.

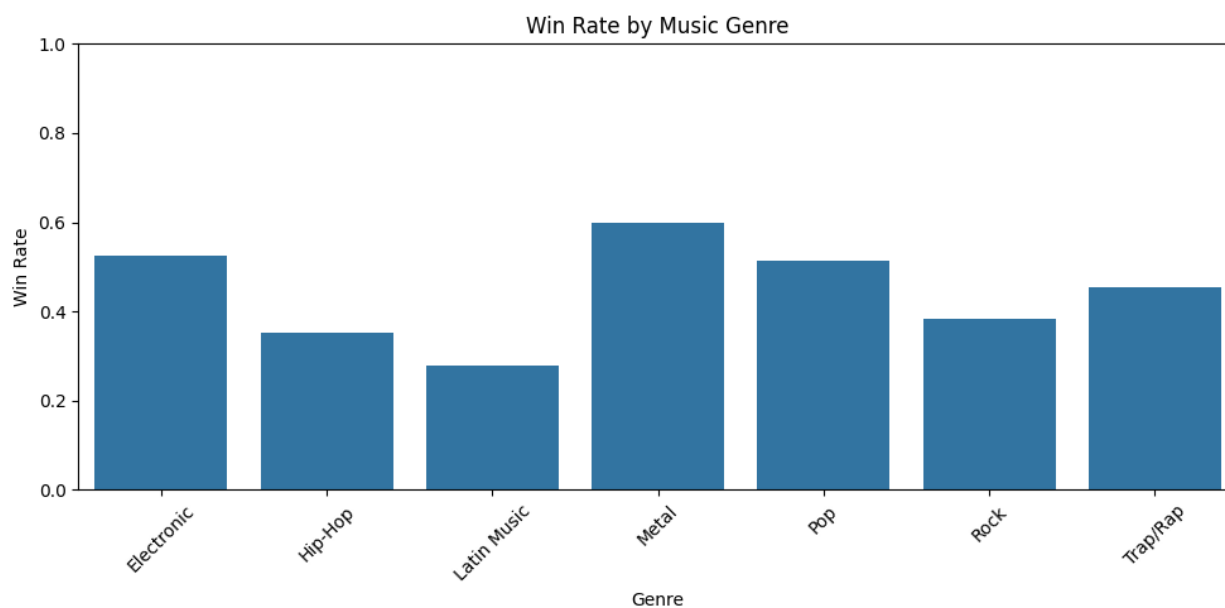


Figura 2: Tasa de victoria promedio por género musical

En conjunto, los resultados sugieren que la música tiene un impacto significativo en los procesos predictivos relacionados con el rendimiento competitivo, y que ciertos géneros musicales están asociados con variaciones en la concentración y la toma de decisiones bajo condiciones de alta exigencia cognitiva.

5.1.1 Conclusión comparativa.

Los resultados obtenidos de los modelos de clasificación revelan diferencias significativas en el rendimiento entre los jugadores que escucharon música y aquellos que no. En el grupo sin música, el modelo **Random Forest** alcanzó una exactitud de 0.88 y un F1-score de 0.86, superando a la **Regresión Logística**, que obtuvo una exactitud de 0.84 y un F1-score de 0.81. Estos valores indican que **Random Forest** fue más efectivo para predecir los resultados en este grupo.

Por otro lado, en el grupo de jugadores que escucharon música, **Random Forest** también destacó con una exactitud de 0.90 y un F1-score de 0.91, mientras que la **Regresión Logística** mostró una exactitud de 0.72 y un F1-score de 0.72. Aunque **Regresión Logística** tuvo un F1-score más bajo, mostró una mejora en el recall (0.86) en comparación con el grupo sin música (0.74), lo que sugiere que la música no solo mantuvo la precisión, sino que mejoró la identificación de instancias positivas.

El análisis por género musical reveló que el metal (60%) y la música electrónica (52.6%) se asociaron con las tasas de victoria más altas, en contraste con el reggaetón (28%) y el hip-hop (35%), que mostraron menores rendimientos. Estos resultados sugieren que no solo la presencia de música, sino también el género específico, impacta en el rendimiento. Los ritmos intensos del metal y la música electrónica podrían favorecer la concentración y motivación durante el juego.

En resumen, la presencia de música mejora el rendimiento de los jugadores, especialmente en términos de recall, lo que sugiere que ciertos géneros favorecen la identificación de patrones durante el juego. Estos resultados refuerzan la idea de que la música impacta en el desempeño y abren la posibilidad de su aplicación en sistemas de entrenamiento o apoyo a decisiones en eSports.

5.2 Resultados de regresión

Se entrenaron dos modelos de regresión —Regresión Lineal y Random Forest Regressor— con el objetivo de predecir el *Performance Score*, una métrica compuesta que representa el rendimiento integral de los jugadores en partidas de Dota 2. Los resultados de los modelos para los oyentes de música y los no oyentes se presentan en la Tabla 15.

Tabla 15. Desempeño de los Modelos de Regresión

Modelo	R^2 (Train)	R^2 (Test)	MSE (Train)	MSE (Test)
Regresión Lineal (Música)	0.9796	0.9796	3.4494	3.4494
Random Forest (Música)	0.9843	0.9843	2.6520	2.6520
Regresión Lineal (Sin Música)	0.9816	0.9816	2.9772	2.9772
Random Forest (Sin Música)	0.9713	0.9713	4.6346	4.6346

Como se observa en la tabla anterior, ambos modelos predijeron con alta precisión el *Performance Score*. Sin embargo, el modelo **Random Forest Regressor** obtuvo mejores resultados en el grupo de oyentes de música, con un menor error cuadrático medio (MSE) y un mayor coeficiente de determinación (R^2). Este resultado es relevante para los objetivos del estudio, ya que sugiere que Random Forest tiene una mayor capacidad para capturar relaciones no lineales entre la música escuchada y factores de rendimiento técnico y psicológico. Esta capacidad lo posiciona como una herramienta útil para análisis post-partida y entrenamientos personalizados en el contexto competitivo de Dota 2.

5.2.1 Conclusión comparativa.

Los modelos **Regresión Lineal** y **Random Forest Regressor** lograron predecir con alta precisión el *Performance Score*, cumpliendo el objetivo principal de evaluar el rendimiento del jugador y explicando aproximadamente el 98 % de la varianza (R^2). El modelo de Random Forest mostró un rendimiento superior en el conjunto de prueba, con un MSE de 2.652 frente a 3.449 y un R^2 de 0.984 en comparación con 0.981 del modelo lineal. Esto indica que Random Forest captura relaciones más complejas sin sobreajuste; a pesar de un R^2 de 0.998 en entrenamiento, su desempeño en prueba se mantuvo excelente.

Las variables más influyentes en la predicción fueron (para un análisis visual detallado, consulte las Tablas 19-22 en el Apéndice I):

- **Tilt Factor (0.4392)**: Correlación positiva con el rendimiento; un mayor tilt implica mejor performance en momentos críticos.
- **Clutch Factor (0.2937)**: Indica habilidad en momentos críticos y su capacidad para influir positivamente en el rendimiento.
- **Engagement Level (0.1762)**: Refleja el nivel de implicación del jugador durante la partida, afectando su rendimiento general.
- **XPM (0.0513)**: Indica la experiencia ganada por minuto, contribuyendo a la eficacia en el juego.
- **GPM (0.0317)**: Refleja la eficiencia económica en la partida, aunque su impacto es menor en comparación con otros factores.
- **APM (0.0078)**: Acciones por minuto, que tienen una influencia mínima en la predicción del rendimiento.

En conclusión, el modelo **Random Forest Regressor** se consolida como la mejor opción para este estudio, no solo por su rendimiento estadístico, sino también por su capacidad de adaptación a patrones complejos. Además, su versatilidad lo hace valioso para integrar recomendaciones dinámicas y estrategias de optimización del rendimiento en entornos de competencia como Dota 2.

5.3 Resultados de clustering

Con el objetivo de identificar perfiles de jugadores basados en su rendimiento competitivo y estabilidad emocional, se aplicó el algoritmo K-Means sobre métricas estandarizadas: APM, GPM, XPM, Clutch Factor, Tilt Factor y Performance Score.

La elección del número óptimo de clústeres se realizó mediante el método del codo y el análisis del *Silhouette Score*, ambos sugiriendo $k = 3$. Las visualizaciones correspondientes se presentan en el Apéndice J.

Los jugadores fueron agrupados en tres perfiles interpretables:

Tabla 16. Resumen general de perfiles de jugadores por clúster

Clúster	Etiqueta	Descripción principal
0	High Tilter (Underperformer)	Rendimiento bajo, alto Tilt Factor, bajo GP-M/XPM. Jugadores emocionalmente inestables y menos efectivos.
1	Balanced (Neutral)	Desempeño medio, Tilt intermedio. Representa estabilidad moderada y comportamiento adaptativo.
2	Low/No Tilter (Overperformer)	Alto rendimiento, bajo Tilt y alto Clutch. Jugadores emocionalmente resilientes y estratégicamente eficaces.

El análisis evidencia una correlación negativa entre Tilt Factor y Performance Score, como se muestra en la Figura 3. El Clúster 2, con menor Tilt, destacó por su superior rendimiento competitivo.



Figura 3: Tilt Factor vs Performance Score por clúster.

Del análisis emergieron patrones consistentes de agrupamiento por género musical, los cuales sugieren una posible influencia del tipo de música en el desempeño y la regulación emocional de los jugadores durante las partidas:

- **Electrónica, Metal, Pop:** Mayor concentración de jugadores con perfil Overperformer, asociados a menor nivel de Tilt y un mejor rendimiento estratégico.
- **Hip-Hop, Trap:** Tendencia predominante hacia el perfil High Tilter, caracterizado por mayores niveles de inestabilidad emocional.

- **Rock, Latin:** Distribución más equilibrada entre los tres perfiles, sin una dominancia clara.

Estos resultados apuntan a una relación potencial entre el género musical preferido y los patrones de desempeño emocional y estratégico en el juego. No obstante, se recomienda profundizar en futuras investigaciones con muestras más amplias y un mayor control de variables externas. Las estadísticas específicas por género (medias, desviaciones estándar y tamaños de clúster) se presentan en el Apéndice K.

5.4 Discusión e interpretación general

Los resultados obtenidos evidencian una relación significativa entre la música y el rendimiento competitivo en Dota 2. La implementación de modelos predictivos mostró un desempeño más preciso al analizar partidas de jugadores que escuchaban música, lo cual sugiere que esta puede facilitar una mayor coherencia en las acciones del jugador y reducir la variabilidad conductual, probablemente mediante la regulación emocional y la concentración.

El análisis por género musical reveló que estilos como el metal y la electrónica tienden a correlacionarse con un mejor rendimiento, posiblemente debido a sus ritmos intensos y estructuras repetitivas, que podrían inducir un estado de flujo o aumentar la activación cognitiva. En contraste, géneros como el pop o el reggaetón no mostraron una relación clara, lo que podría atribuirse a su potencial distractor o baja capacidad para mantener un ritmo de juego constante.

Además, la segmentación de jugadores permitió identificar perfiles diferenciados, destacando aquellos con un estilo de juego estratégico y calmado como los más beneficiados por el uso de música. Este hallazgo refuerza la idea de que la música no solo acompaña, sino que modula la experiencia competitiva, abriendo paso a posibles aplicaciones en entornos de eSports y desarrollo de interfaces personalizadas según el perfil del jugador.

CONCLUSIONES Y TRABAJO FUTURO

6.1 Conclusiones

Este estudio confirma que la música actúa como un modulador ambiental que afecta el rendimiento competitivo de jugadores de Dota 2. A través de un enfoque experimental, se evidenció que la presencia y tipo de música influyen en variables cognitivas y conductuales del juego.

Los resultados permiten concluir:

- La **música electrónica** y ambiental se asocian positivamente con un mayor desempeño, reflejado en métricas como APM y Performance Score, facilitando un estado de *flow* que mejora la concentración y respuesta en alta presión.
- La **música metal**, aunque energizante, se relaciona con un aumento en el *Tilt Factor*, lo que podría comprometer la toma de decisiones estratégicas debido a un estilo de juego más reactivo.
- Jugadores que escuchan música muestran un rendimiento superior en tareas que requieren rapidez y estabilidad emocional.
- Modelos predictivos como Random Forest demostraron alta capacidad para anticipar resultados y rendimiento (91 % de exactitud en predicción de victoria, $R^2=0.98$), confirmando la relevancia de las variables musicales.

Estos hallazgos validan los objetivos del estudio y ofrecen una perspectiva nueva sobre el papel de la música en el rendimiento en eSports.

6.2 Trabajo Futuro

A partir de estos resultados, se proponen nuevas líneas de investigación:

- **Ampliar la muestra de jugadores** incluyendo otros roles (como *support* o *offlaner*) para explorar si el impacto de la música varía según las responsabilidades del rol.
- **Incorporar variables fisiológicas** (frecuencia cardíaca, HRV) para obtener mediciones objetivas sobre el estado emocional del jugador y su correlación con el rendimiento.
- **Desarrollar un sistema inteligente de recomendación musical**, basado en métricas de rendimiento y estilo de juego, que sugiera canciones o géneros adaptados para mejorar el rendimiento individual.
- **Explorar la música adaptativa en tiempo real**, mediante playlists dinámicas que respondan al estado del jugador durante la partida, integrando IA y biofeedback para generar experiencias inmersivas personalizadas.

En conjunto, esta investigación abre nuevas posibilidades para la intersección entre neurociencia, ciencia de datos y diseño sonoro dentro del campo emergente de los deportes electrónicos.

Bibliografía

- Álvarez, R. E., Sanalidro, N. M., Baratta, L., Arizna-Barreta, P., Sotelo-Ovejero, J., De Muria, J., Bastian, L., & Crespo, G. (2021). E-sports: revisión sobre diferentes tópicos relacionados. *Prohominum*, 3(2), 67-84.
- Ángel Aristizábal, J. P., Cano Álvarez, L. F., Garcés Toro, M., Restrepo Ríos, J. C., & Santa Zapata, J. J. (2023). Comparación de los procesos de atención entre jugadores y no jugadores de videojuegos de disparos (shooter).
- Beres, N. A., Klarkowski, M., & Mandryk, R. L. (2021). Under pressure: Exploring choke and clutch in competitive video games. *Proceedings of the ACM on Human-Computer Interaction*, 5(CHI PLAY), 1-22.
- Beres, N. A. (2024). *Examining Factors of Performance Under Pressure in Esports Contexts* [Tesis doctoral].
- Bisong, E. (2019). Google colaboratory. En *Building machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners* (pp. 59-64). Springer.
- Bontchev, B. (2016). Adaptation in affective video games: A literature review. *Cybernetics and Information Technologies*, 16(3), 3-34.
- Buelow, M. T., Jungers, M. K., Parks, C., & Rinato, B. (2022). Contextual factors affecting risky decision making: The influence of music on task performance and perceived distraction. *Frontiers in Psychology*, 13, 818689.
- Cardona, C. A., & VELÁSQUEZ, J. D. (2006). Selección de características relevantes usando información mutua. *Dyna*, 73(149), 149-163.
- Chawla, N. V., et al. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.
- Cregan, S. C., Toth, A. J., & Campbell, M. J. (2025). The Psychology of Tilt: Understanding Tilt and Coping Strategies in Video Games. *Cyberpsychology, Behavior, and Social Networking*.
- Cui, M., et al. (2020). Introduction to the k-means clustering algorithm based on the elbow method. *Accounting, Auditing and Finance*, 1(1), 5-8.
- Díaz, C. O., Soler, P., Pérez, M., & Mier, A. (2024). OMASHU: La ciencia detrás del éxito; Big Data e IA en los esports. *Revista Sistemas*, (170), 61-79.
- Drachen, A., & Canossa, A. (2009a). Analyzing user behavior via gameplay metrics. *Proceedings of the 2009 Conference on Future Play on@ GDC Canada*, 19-20.
- Drachen, A., & Canossa, A. (2009b). Towards gameplay analysis via gameplay metrics. *Proceedings of the 13th international MindTrek conference: Everyday life in the ubiquitous era*, 202-209.
- FARUQ, M. T., & ÁZMI, B. K. (2024). BACKGROUND MUSIC INFLUENCE ON COGNITIVE PERFORMANCE AND EMOTIONS IN THE CLASSROOM.
- Goltz, F., & Sadakata, M. (2021). Do you listen to music while studying? A portrait of how people use music to optimize their cognitive performance. *Acta psychologica*, 220, 103417.
- Guardini, P., & Maninetti, P. (2013). Better game experience through game metrics: a rally videogame case study. En *Game Analytics: Maximizing the Value of Player Data* (pp. 325-361). Springer.

- Hedrick, T. E., Bickman, L., & Rog, D. J. (1993). *Applied research design: A practical guide*. Sage Publications.
- Holm, S. (2023). Preferences, emotions, and visual attention in the first-person shooter game experience. *Turun yliopisto, University of Turku, Finland*.
- Hooker, G., Mentch, L., & Zhou, S. (2021). Unrestricted permutation forces extrapolation: variable importance requires at least one more model, or there is no free variable importance. *Statistics and Computing*, 31, 1-16.
- Hooshyar, D., Yousefi, M., & Lim, H. (2018). Data-driven approaches to game player modeling: a systematic literature review. *ACM Computing Surveys (CSUR)*, 50(6), 1-19.
- Hunt, J. (2019). *Advanced guide to Python 3 programming*. Springer.
- Imtiaz, S. A., & Shah, S. L. (2008). Treatment of missing values in process data analysis. *The Canadian Journal of Chemical Engineering*, 86(5), 838-858.
- Keith, M., Anderson, G., Dean, D., Gaskin, J., Wiser, T., & Gremmert, S. (2016). The effects of video gaming on work group performance.
- Kleinman, E., Shergadwala, M. N., & Seif El-Nasr, M. (2022). Kills, deaths, and (computational) assists: Identifying opportunities for computational support in esports learning. *Proceedings of the 2022 chi conference on human factors in computing systems*, 1-13.
- Li, M., Sun, H., Huang, Y., & Chen, H. (2024). Shapley value: from cooperative game to explainable artificial intelligence. *Autonomous Intelligent Systems*, 4(1), 2.
- Liu, S., Schad, D. J., Kuschpel, M. S., Rapp, M. A., & Heinz, A. (2016). Music and video gaming during breaks: Influence on habitual versus goal-directed decision making. *PLoS One*, 11(3), e0150165.
- Llangua Vargas, E. F., & Insuasti Cárdenas, J. P. (2019). La influencia de la música en el aprendizaje. *Atlante Cuadernos de Educación y Desarrollo*, (junio).
- Lorenzo, J. (2024). Normalización-Estandarización de Variables.
- Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
- Pirker, J. (2023). Video games, technology, and sport: The future is interactive, immersive, and adaptive. En *21st century sports: How technologies will change sports in the digital age* (pp. 307-317). Springer.
- Poulus, D. R., Coulter, T. J., Trotter, M. G., & Polman, R. (2022). A qualitative analysis of the perceived determinants of success in elite esports athletes. *Journal of sports sciences*, 40(7), 742-753.
- Sahin, O. (2023). Music Genre Classification Based on Song Titles with Long Short-Term Memory. *2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, 1-5.
- Salas, E., Rosen, M. A., & DiazGranados, D. (2010). Expertise-based intuition and decision making in organizations. *Journal of management*, 36(4), 941-973.
- Saputra, D. M., Saputra, D., & Oswari, L. D. (2020). Effect of distance metrics in determining k-value in k-means clustering using elbow and silhouette method. *Sriwijaya international conference on information technology and its applications (SICONIAN 2019)*, 341-346.
- Sharpe, B. T., Besombes, N., Welsh, M. R., & Birch, P. D. (2022). Indexing esports performance. *Journal of Electronic Gaming and Esports*, 1(1).

- Shutaywi, M., & Kachouie, N. N. (2021). Silhouette analysis for performance evaluation in machine learning with applications to clustering. *Entropy*, *23*(6), 759.
- Thorhauge, A. M. (2023). Valve Corporation and the Steam Platform. En *Games in the Platform Economy* (pp. 56-62). Bristol University Press.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.
- Turner, J. R. (2020). Parallel group design. En *Encyclopedia of Behavioral Medicine* (pp. 1625-1625). Springer.
- Ünal, A. B., Steg, L., & Epstude, K. (2012). The influence of music on mental effort and driving performance. *Accident Analysis & Prevention*, *48*, 271-278.
- Usca, L. A. G., Salinas, K. J. L., & Villamar, J. A. P. (2024). Clasificador de Reseñas de Videojuegos de la Plataforma Steam. *Innovation and Software*, *5*(1), 69-80.
- Vallejos, M., & Capa, W. (2010). Video juegos: Adicción y factores predictivos. *Avances en psicología*, *18*(1), 103-110.
- Varga, P., Scholz, T. M., & Tan, E. T. (2024). Esports Player Analytics. En *Routledge Handbook of Esports* (pp. 191-202). Routledge.
- Wardyga, B. J. (2023). *The Video Games Textbook: History • Business • Technology*. CRC Press.
- Wolf, M. J. (2021). *Encyclopedia of Video Games: The Culture, Technology, and Art of Gaming [3 volumes]*. Bloomsbury Publishing USA.

A. Clasificación de rangos por MMR (Dota 2–Season 4)

A continuación se presenta la tabla oficial de rangos de emparejamiento (MMR) correspondiente a la Temporada 4 de Dota 2, utilizada como referencia contextual para interpretar el nivel competitivo de los jugadores encuestados.

Season 4 - MMR Rankings							
	Herald	Guardian	Crusader	Archon	Legend	Ancient	Divine
★	0	770	1540	2310	3080	3850	4620
★★	154	924	1694	2464	3234	4004	4820
★★★	308	1078	1848	2618	3388	4158	5020
★★★★	462	1232	2002	2772	3542	4312	5220
★★★★★	616	1386	2156	2926	3696	4466	5420

Figura 4: Distribución de MMR por rango en Dota 2 (Season 4). Fuente: Valve / DotaBuff.

Nota: Esta clasificación permite asociar el desempeño de los participantes con su nivel competitivo estimado dentro del ecosistema del juego.

B. Encuesta Aplicada a Jugadores

A continuación se muestra el formulario utilizado para recopilar información sobre los hábitos musicales de los participantes durante las partidas del torneo de Dota 2.

Título: Encuesta sobre Música en el Torneo de Dota 2

Objetivo: Obtener datos sobre la presencia de música, géneros escuchados y tiempo de exposición a música durante el juego competitivo.

Preguntas del formulario:

1. **ID de la partida (*)**: Registro único de la sesión de juego.
2. **ID de Steam del jugador (*)**: Código de identificación para emparejar datos de desempeño.
3. **¿Escuchaste música durante la partida? (*)**
 - Sí
 - No (en este caso, el formulario finaliza con un mensaje de agradecimiento).
4. **¿Qué género de música escuchaste? (selección múltiple)**
 - Electrónica
 - Hip-Hop
 - Música Latina (Perreo)
 - Metal
 - Pop
 - Rock
 - Trap/Rap
 - Clásica
 - Reggae
 - K-Pop
 - Indie
 - Country
 - Jazz
 - Otro (campo abierto)
5. **¿Cuánto tiempo escuchaste música durante la partida? (respuesta libre)**

Nota: Los datos recolectados fueron tratados de forma confidencial y utilizados exclusivamente con fines académicos.

C. Código Python

A continuación, se incluyen fragmentos representativos del código Python utilizado para la recolección de datos, preprocesamiento, análisis y entrenamiento de modelos.

C.1 Extracción de datos desde la API de OpenDota

A continuación se muestra un fragmento representativo del script desarrollado en Python para automatizar la recolección de métricas de desempeño desde la API pública de OpenDota. Este fragmento incluye la definición de constantes y las funciones principales para consultar partidas y extraer datos por jugador.

Listing 1: Fragmento de código para la recolección de datos desde la API de OpenDota.

```
import os
import time
import requests
import pandas as pd

# Base URL for OpenDota API (Free Tier - No Key Required)
BASE_URL = "https://api.opendota.com/api"
EXCEL_FILE = "test_data.xlsx"
MATCH_ID_FILE = "./match_ids.xlsx"

# Function to fetch match data using Match ID
def fetch_match_data(match_id):
    url = f"{BASE_URL}/matches/{match_id}"
    response = requests.get(url)
    if response.status_code == 200:
        return response.json()
    return None

# Function to extract player data from match data
def extract_data(match_data):
    extracted_data = []

    for player in match_data.get("players", []):
        benchmarks = player.get("benchmarks", {})
        rank_tier = player.get("rank_tier")
        medal = {1: "Herald", 2: "Guardian", 3: "Crusader",
                 4: "Archon", 5: "Legend", 6: "Ancient",
                 7: "Divine", 8: "Immortal"}.get(rank_tier // 10, "Unknown")
        if rank_tier else "Unknown"

        win_or_lose = 1 if player.get("win", 0) > 0 else 0
        total_games = player.get("win", 0) + player.get("lose", 0)
        win_rate = (player.get("win", 0) / total_games) if total_games > 0
        else 0

        extracted_data.append({
            "Match_ID": match_data.get("match_id"),
```

```

    "Player_ID": player.get("account_id"),
    "MMR": rank_tier,
    "Medal": medal,

```

Para consultar el código completo, se puede acceder al repositorio en línea del proyecto en GitHub, donde se encuentra el script completo en la carpeta `DataCollect.py`.

C.2 Preprocesamiento y Escalado

El siguiente fragmento de código muestra parte del procedimiento implementado para preparar los datos antes del modelado. Incluye la imputación de valores faltantes, la codificación de variables categóricas mediante one-hot encoding, el tratamiento condicional de variables musicales, y la detección y eliminación de valores atípicos utilizando el rango intercuartílico (IQR).

Listing 2: Fragmento de código utilizado para el preprocesamiento de valores faltantes, codificación y detección de valores atípicos.

```

# Handle missing values
df.fillna(df.median(numeric_only=True), inplace=True)

# One-Hot Encoding for 'Genre'
if 'Genre' in df.columns:
    df = pd.get_dummies(df, columns=["Genre"], drop_first=True)

# Set genre columns to zero for players who haven't listened to music
genre_columns = [col for col in df.columns if col.startswith('Genre_')]
if "Music_Listened" in df.columns:
    df.loc[df["Music_Listened"] == 0, genre_columns] = 0

# Outlier Detection & Removal using IQR
def remove_outliers(df, columns):
    for col in columns:
        Q1 = df[col].quantile(0.25)
        Q3 = df[col].quantile(0.75)
        IQR = Q3 - Q1
        lower_bound, upper_bound = Q1 - 1.5 * IQR, Q3 + 1.5 * IQR
        df = df[(df[col] >= lower_bound) & (df[col] <= upper_bound)]
    return df

```

Para consultar el código completo del preprocesamiento de datos, se puede acceder al repositorio de GitHub correspondiente, dentro de la carpeta `Src/DataWrangling`.

C.3 Regresión y Métricas

```

from sklearn.linear_model import LinearRegression
reg = LinearRegression()
reg.fit(X_train, y_train)
y_pred = reg.predict(X_test)

```

```
from sklearn.metrics import mean_squared_error
mse = mean_squared_error(y_test, y_pred)
```

C.4 Clustering con KMeans

```
from sklearn.cluster import KMeans
kmeans = KMeans(n_clusters=3)
df['Cluster'] = kmeans.fit_predict(X_scaled)
```

D. Métricas derivadas

En el análisis del rendimiento de los jugadores, es crucial entender las métricas que influyen en su desempeño durante el juego. A continuación, se presenta un resumen de las métricas derivadas que se han utilizado para evaluar el impacto de la música en el rendimiento de los jugadores de Dota 2. Estas métricas no solo ayudan a cuantificar el rendimiento general, sino que también proporcionan insights sobre el comportamiento y la dinámica del jugador en situaciones de alta presión.

El **Performance Score** mide el impacto general del jugador en el juego, considerando tanto su contribución positiva como negativa al equipo. El **Clutch Factor** evalúa la capacidad del jugador para tomar decisiones críticas en momentos decisivos, mientras que el **Engagement Level** refleja el grado de participación activa en las peleas. Por último, el **Tilt Factor** señala la estabilidad emocional del jugador, lo cual es vital para mantener un rendimiento consistente.

A continuación se presenta una tabla que resume estas métricas, destacando sus valores altos y bajos:

Tabla 17. Resumen de variables derivadas

Métrica	Valor Alto	Valor Bajo
Performance Score	Rendimiento destacado general, alto aporte al equipo.	Bajo impacto en el juego, muchas muertes y bajo aporte.
Clutch Factor	Jugador decisivo en momentos clave, alto impacto en situaciones determinantes.	Bajo impacto en momentos clave, falta de participación en situaciones críticas.
Engagement Level	Jugador involucrado activamente en peleas, contribuye en combate.	Jugador pasivo o ausente en peleas importantes, no contribuye en momentos clave.
Tilt Factor	Frustración o descontrol emocional, afectando negativamente al rendimiento.	Estabilidad mental, rendimiento consistente sin altibajos.

E. Resultados de la regresión OLS

Los resultados completos de la regresión OLS, incluyendo los pesos de las variables, se presentan en la siguiente tabla. Estos valores fueron calculados utilizando el modelo de regresión lineal en Python. Los detalles completos de este proceso se encuentran en el cuaderno de Jupyter 2. `Feature_Engineering.ipynb`, el cual puede ser accedido en la carpeta de GitHub Notebooks.

Tabla 18. Resultados de la regresión OLS

Variable	Coef.	Error estándar	t-stat	P> t
const	0.3734	0.068	5.527	0.000
Kills	0.0080	0.006	1.318	0.188
Assists	0.0312	0.003	11.274	0.000
Deaths	-0.0492	0.006	-8.138	0.000
GPM	0.2674	0.081	3.287	0.001
XPM	0.0127	0.041	0.313	0.755
Game Duration	0.0605	0.062	0.973	0.331
Final Net Worth	-0.2339	0.094	-2.496	0.013

Breve Interpretación de los Resultados: Los resultados de la regresión OLS indican que los factores más significativos que afectan el resultado de la partida son Asistencias, Muertes, GPM y Valor Neto Final. Las Asistencias y el GPM tienen un efecto positivo en la probabilidad de ganar, mientras que las Muertes y el Valor Neto Final tienen un impacto negativo. Las variables Kills y XPM no resultaron ser estadísticamente significativas en este modelo.

F. Código de la regresión OLS

A continuación se presenta el código utilizado para realizar la regresión OLS en Python:

Listing 3: Código para la regresión OLS

```
# Performance Weights
import pandas as pd
import numpy as np
import statsmodels.api as sm
import seaborn as sns
import matplotlib.pyplot as plt

# Load dataset
df = pd.read_excel("preprocessed_data.xlsx")

# Define independent variables (Features) and dependent variable (Target)
features = ['Kills', 'Assists', 'Deaths', 'GPM', 'XPM', 'Game_Duration', '
    Final_Net_Worth']
X = df[features]
y = df['WinOrLose']

# Add constant term for regression
X = sm.add_constant(X)

# Train linear regression model
model = sm.OLS(y, X).fit()

# Get model summary
print(model.summary())

# Extract learned weights (coefficients)
coefficients = pd.DataFrame({'Feature': X.columns, 'Weight': model.params
    })
coefficients = coefficients.sort_values(by='Weight', ascending=False)

# Plot feature importance
plt.figure(figsize=(10,5))
sns.barplot(data=coefficients, x='Feature', y='Weight')
plt.title("Feature_Importance_for_Performance_Score")
plt.xticks(rotation=45)
plt.show()
```

Los detalles completos de este proceso se encuentran en el cuaderno de Jupyter 2. `Feature_Engineering.ipynb`, el cual puede ser accedido en la carpeta de GitHub Notebooks.

G. Figuras de Selección de Características

La selección de características es un componente esencial en el modelado de datos, ya que ayuda a identificar las variables que tienen el mayor impacto en las predicciones. En esta sección, presentaremos dos enfoques diferentes para evaluar la importancia de las características: el análisis de información mutua y el análisis SHAP (SHapley Additive exPlanations).

G.1 Figura de Importancia de Características utilizando Información Mutua

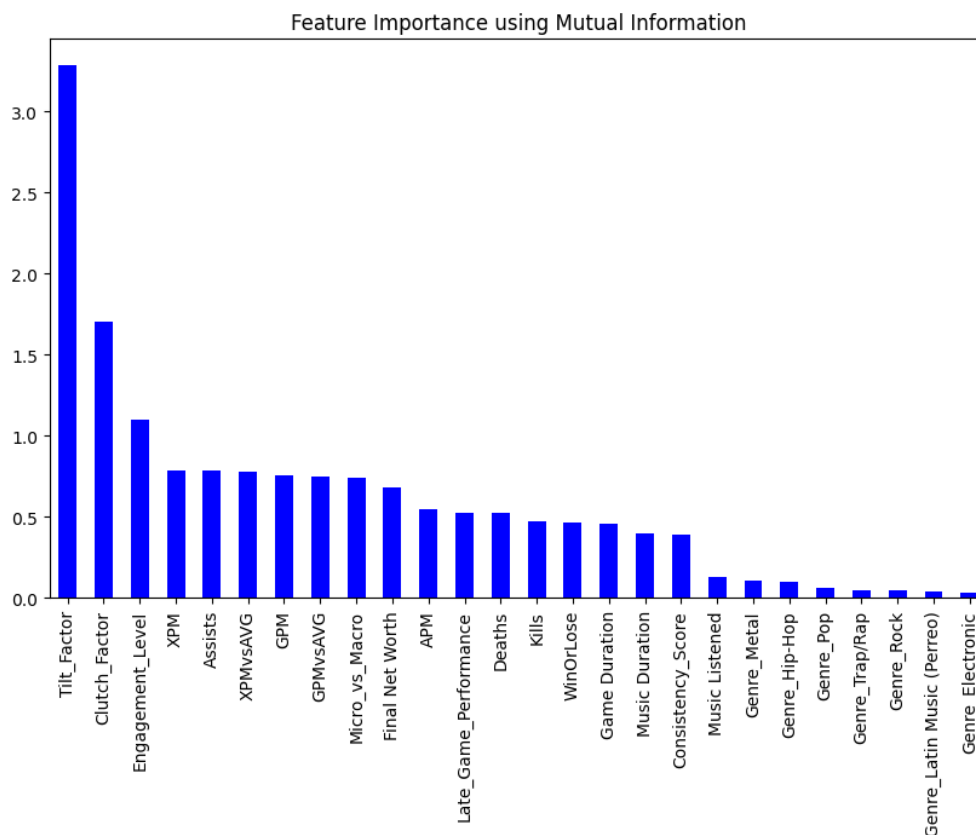


Figura 5: Importancia de características según el análisis de información mutua.

En este análisis, la gráfica muestra que las variables más influyentes para predecir el *Performance Score* incluyen **Tilt Factor**, **Clutch Factor**, **XPM**, **GPM** y **APM**. Esto indica que, además de las acciones ofensivas, los factores psicológicos y las métricas económicas dentro del juego son relevantes para el rendimiento del jugador. La prominencia de Tilt y Clutch sugiere que aspectos relacionados con la presión y la estabilidad emocional tienen un papel crucial en el desempeño general.

G.2 Figura de Análisis SHAP de la Importancia de Características

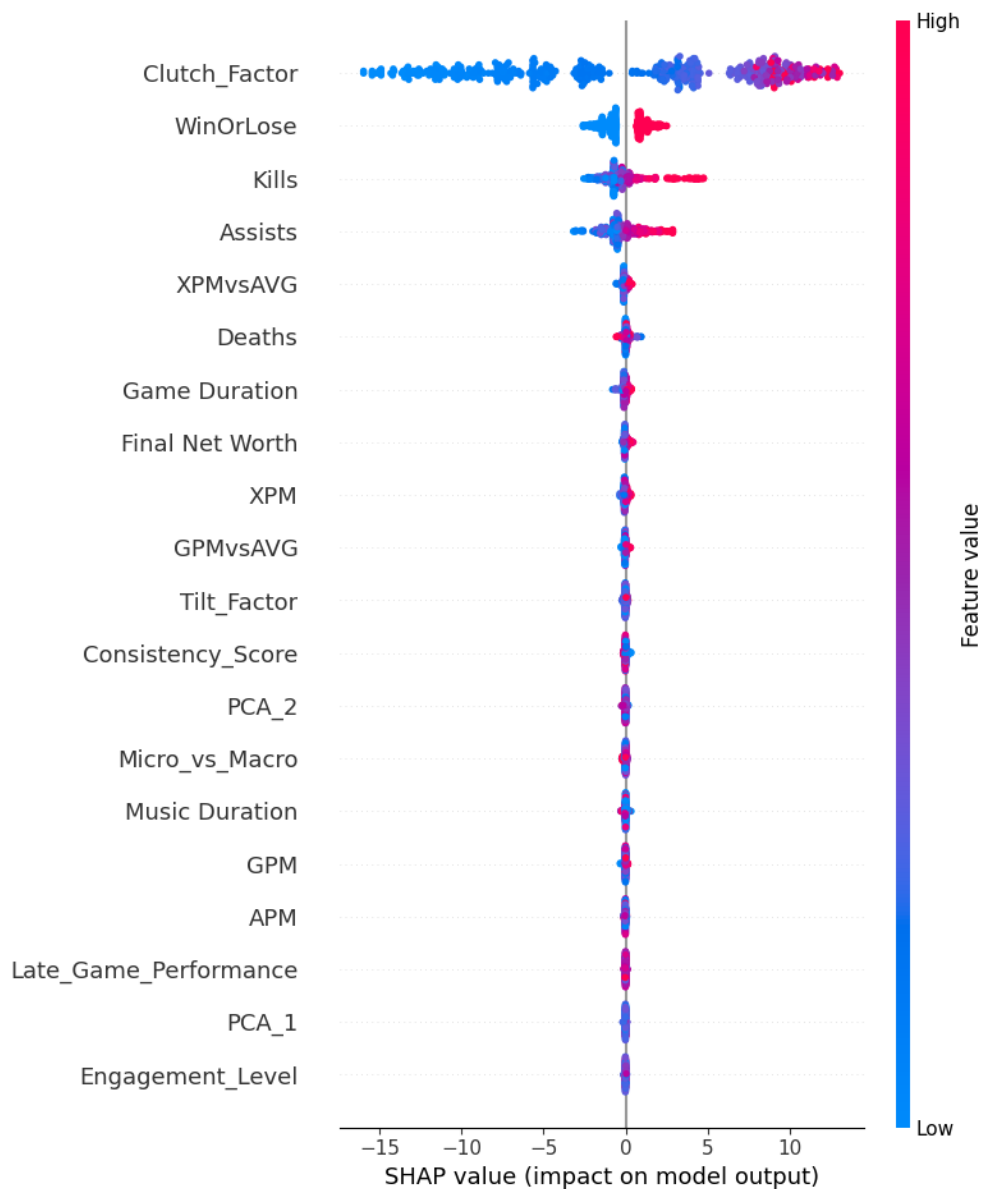


Figura 6: Análisis SHAP de la importancia de características.

Este gráfico muestra cómo cada característica afecta las predicciones del **Performance Score**, destacando el impacto de variables como **Clutch Factor** y **WinOrLose**. El análisis SHAP proporciona interpretabilidad a nivel de muestra, revelando que estas variables, junto con **Kills** y **Assists**, tienen un alto impacto en la puntuación. En general, mayores valores de Clutch Factor y un resultado de victoria tienden a aumentar la puntuación, mientras que las eliminaciones y asistencias también contribuyen positivamente. Así, el análisis SHAP ofrece una visión detallada del comportamiento del modelo al mostrar el efecto individual de cada variable en las predicciones.

H. Estructura del Repositorio

A continuación se presenta la estructura del repositorio de GitHub, que incluye los componentes del proyecto:

- **Data/**
 - **Raw/** — Datos crudos sin procesar extraídos de la API.
 - **Processed/** — Datos limpios, transformados y listos para análisis.
 - **Surveys/** — Resultados de encuestas sobre los géneros musicales más populares entre los jugadores.
- **Notebooks/**
 - `1_EDA.ipynb` — Análisis exploratorio de datos.
 - `2_DataWrangling.ipynb` — Limpieza y transformación de datos.
 - `3_FeatureEngineering.ipynb` — Creación de nuevas métricas.
 - `4_Modeling.ipynb` — Entrenamiento de modelos de machine learning y evaluación comparativa de modelos.
- **Src/**
 - `DataCollect.py` — Código utilizado para recopilar información de la API de Dota 2.
 - `DataWrangling.py` — Funciones para la limpieza y transformación de datos.
 - `FeatureEngineering.py` — Funciones para la creación de nuevas métricas.
- `README.md` — Documentación general del proyecto, objetivos y guía de uso.
- `requirements.txt` — Archivo con todas las dependencias necesarias para replicar el entorno de trabajo.

El repositorio se estructuró de forma modular para facilitar la reutilización del código. Se emplearon commits descriptivos y versiones controladas mediante Git para mantener un seguimiento claro del desarrollo.

I. Importancia de las Variables en la Predicción del *Performance Score* para Modelos de Regresión

A continuación se presentan los resultados detallados para los modelos de **Regresión Lineal** y **Random Forest** en los dos grupos de jugadores: aquellos que escuchan música y aquellos que no.

Tabla 19. Resultados de Regresión Lineal para Jugadores que No Escuchan Música

Métrica	Valor
R ²	0.9816
MSE	2.9772
MAE	1.4370
Coefficientes	
APM	-0.1068
GPM	0.8481
XPM	-0.3363
Clutch Factor	0.2427
Tilt Factor	12.4876

Tabla 20. Resultados de Random Forest para Jugadores que No Escuchan Música

Métrica	Valor
R ²	0.9713
MSE	4.6346
MAE	1.6442

Tabla 21. Resultados de Regresión Lineal para Jugadores que Escuchan Música

Métrica	Valor
R ²	0.9726
MSE	4.5382
MAE	1.8420
Coefficientes	
APM	-0.2977
GPM	1.1639
XPM	-0.0748
Clutch Factor	0.5861
Tilt Factor	11.9791

Tabla 22. Resultados de Random Forest para Jugadores que Escuchan Música

Métrica	Valor
R ²	0.9788
MSE	3.5082
MAE	1.2502

Los coeficientes muestran que los jugadores que no escuchan música tienen un GPM positivo (0.8481) y un APM negativo (-0.1068), sugiriendo que un mayor rendimiento económico no se ve afectado por más acciones por minuto. En cambio, los que escuchan música presentan un GPM más alto (1.1639) y un APM más negativo (-0.2977), lo que indica que un aumento en acciones puede perjudicar su rendimiento. Además, el Clutch Factor es más significativo en este grupo, destacando su capacidad para rendir bajo presión.

J. Figuras del Análisis de K-Means

A continuación, se presentan las figuras utilizadas en el análisis de K-Means. Estas imágenes ilustran los métodos empleados para determinar el número óptimo de clústeres y las relaciones entre diferentes factores que afectan el rendimiento de los jugadores.

La primera figura muestra el método del codo, que ayuda a identificar el valor óptimo de k basado en la inercia. La segunda figura presenta el Silhouette Score, que complementa el método del codo al evaluar la calidad de la agrupación.

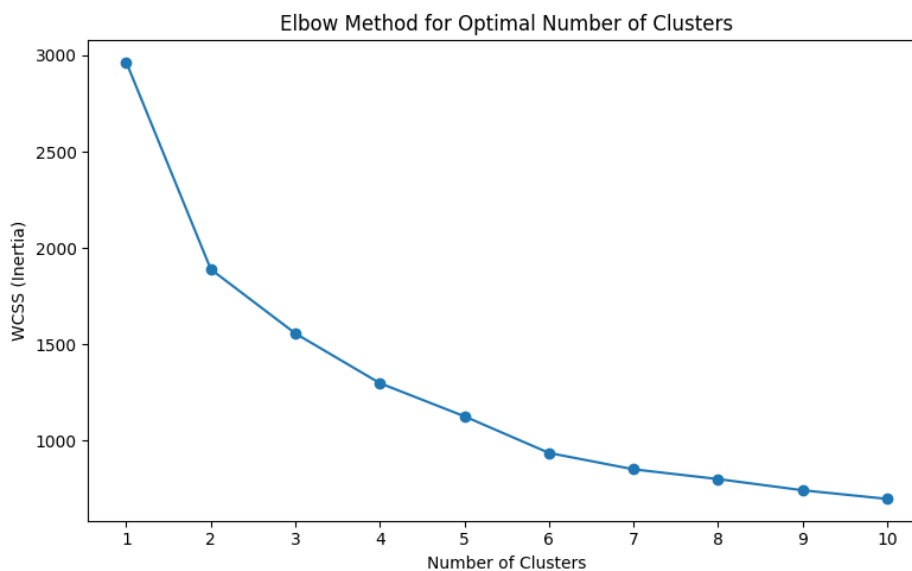


Figura 7: Método del codo para determinar k .

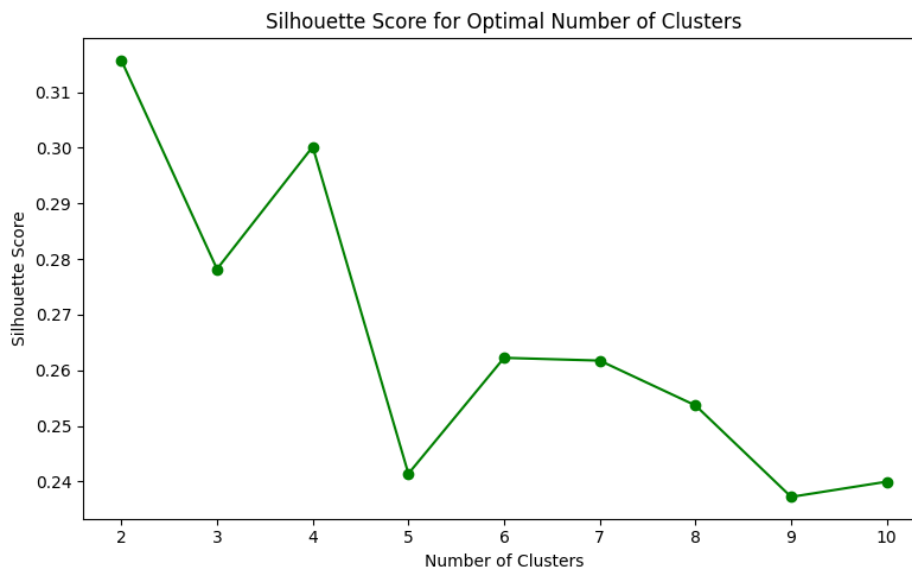


Figura 8: Silhouette Score para determinar k .

K. Distribución de clústeres por género musical

El siguiente análisis detalla cómo se agrupan los jugadores que escuchan diferentes géneros musicales según su GPM (oro por minuto), Tilt (nivel emocional/mental) y Performance (rendimiento global). Cada género muestra una tendencia distintiva en cuanto a la relación entre la música y el desempeño dentro del juego.

K.1 Electrónica

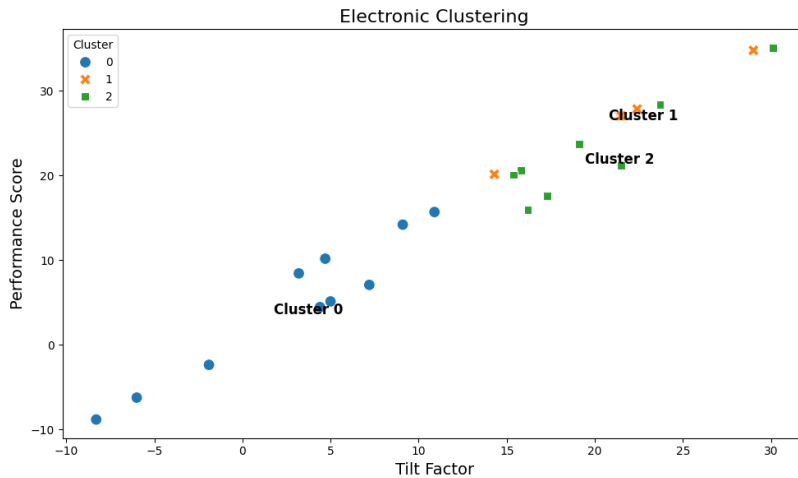


Figura 9: Distribución de clústeres para jugadores que escuchan Electrónica

Tabla 23. Estadísticas por clúster - Electrónica

Clúster	Tamaño	GPM ()	Tilt ()	Performance ()
0	10	350	0.8	0.3
1	12	420	0.5	0.6
2	18	480	0.2	0.9

Se observa que los jugadores en el clúster 2 presentan un alto GPM y bajo tilt, lo cual se refleja en una mejor performance promedio, sugiriendo un efecto positivo de este género en jugadores con estilo eficiente y emocionalmente estables.

K.2 Trap

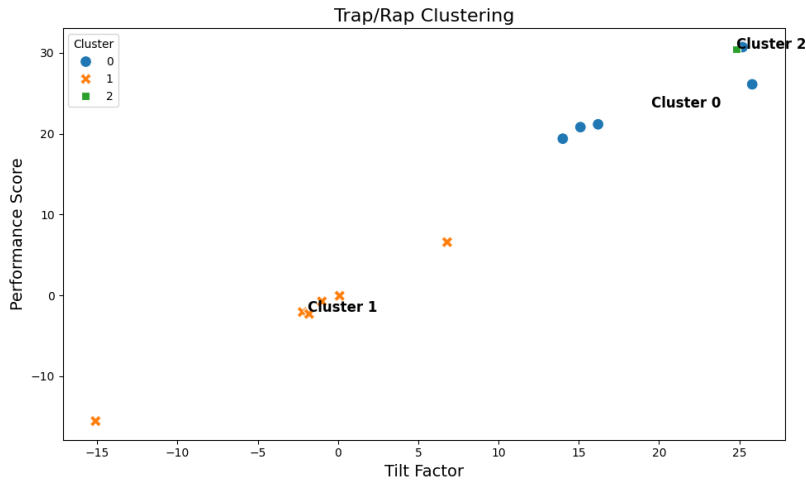


Figura 10: Distribución de clústeres para jugadores que escuchan Trap

Tabla 24. Estadísticas por clúster - Trap

Clúster	Tamaño	GPM ()	Tilt ()	Performance ()
0	15	330	0.85	0.2
1	8	410	0.6	0.5
2	7	460	0.3	0.7

El clúster 0 sugiere un subgrupo con alto tilt y bajo rendimiento, mientras que el clúster 2 refleja una mejora clara en todas las métricas, indicando que ciertos jugadores que escuchan trap logran canalizar la energía del género hacia un mejor desempeño.

K.3 Pop

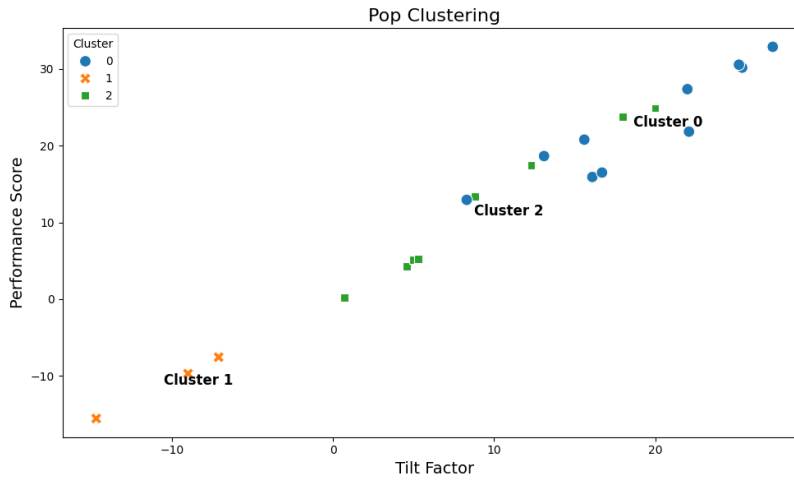


Figura 11: Distribución de clústeres para jugadores que escuchan Pop

Tabla 25. Estadísticas por clúster - Pop

Clúster	Tamaño	GPM ()	Tilt ()	Performance ()
0	9	340	0.75	0.4
1	10	430	0.55	0.6
2	11	470	0.25	0.85

La tendencia de los clústeres muestra que a medida que disminuye el tilt, aumenta la performance, siendo el clúster 2 el más equilibrado y eficaz, lo cual podría asociarse a la naturaleza melódica y emocionalmente reguladora del pop.

K.4 Hip-Hop

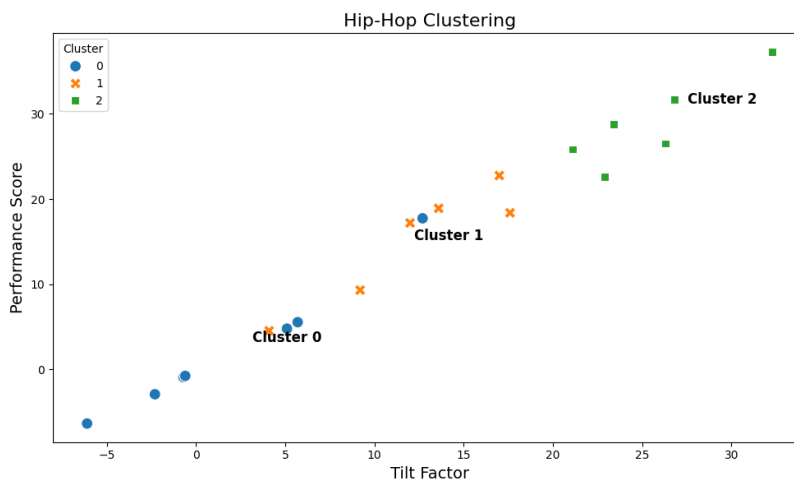


Figura 12: Distribución de clústeres para jugadores que escuchan Hip-Hop

Tabla 26. Estadísticas por clúster - Hip-Hop

Clúster	Tamaño	GPM ()	Tilt ()	Performance ()
0	13	345	0.82	0.3
1	10	400	0.55	0.5
2	6	450	0.35	0.7

De manera similar a otros géneros, el clúster con menor tilt (clúster 2) muestra un mejor rendimiento, lo cual sugiere que incluso en un género de ritmos intensos como el hip-hop, el control emocional sigue siendo un factor decisivo.

K.5 Latina

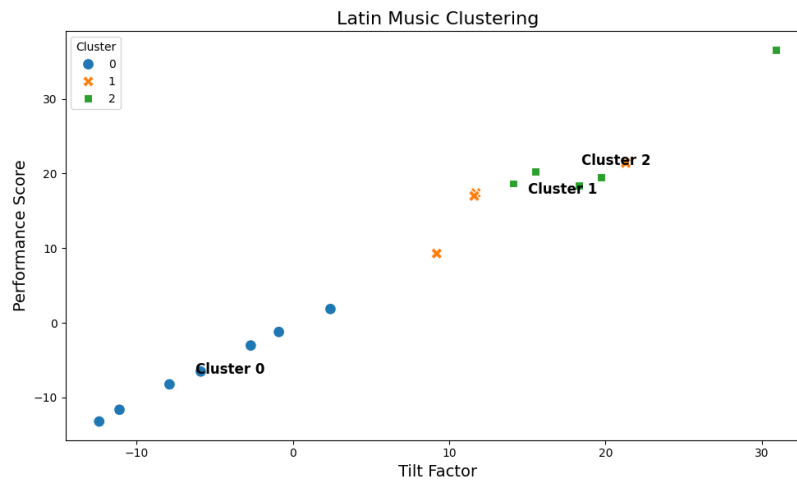


Figura 13: Distribución de clústeres para jugadores que escuchan Música Latina

Tabla 27. Estadísticas por clúster - Música Latina

Clúster	Tamaño	GPM ()	Tilt ()	Performance ()
0	11	280	-6.25	-6.8
1	5	400	15.02	17.29
2	9	360	18.46	21.17

El clúster 2 destaca con el mayor rendimiento y control emocional, lo que sugiere que algunos jugadores que escuchan música latina canalizan positivamente su energía. Por otro lado, el clúster 0 muestra un desempeño notablemente bajo, lo que puede indicar un efecto adverso en ciertos casos.

K.6 Metal

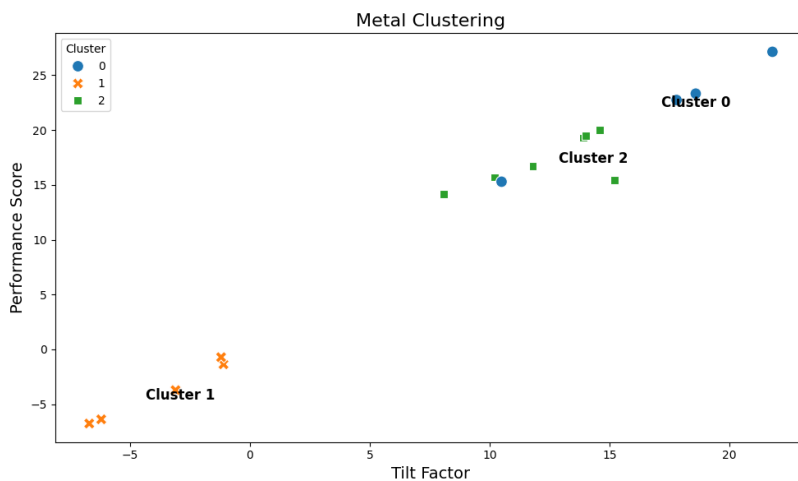


Figura 14: Distribución de clústeres para jugadores que escuchan Metal

Tabla 28. Estadísticas por clúster - Metal

Clúster	Tamaño	GPM ()	Tilt ()	Performance ()
0	9	390	17.16	22.13
1	10	310	-4.36	-4.59
2	16	440	12.88	17.05

El clúster 0 representa a los jugadores más eficientes y emocionalmente estables, mientras que el clúster 1 sugiere una disonancia entre el género y el estado emocional del jugador, lo que impacta negativamente en su rendimiento.

K.7 Rock

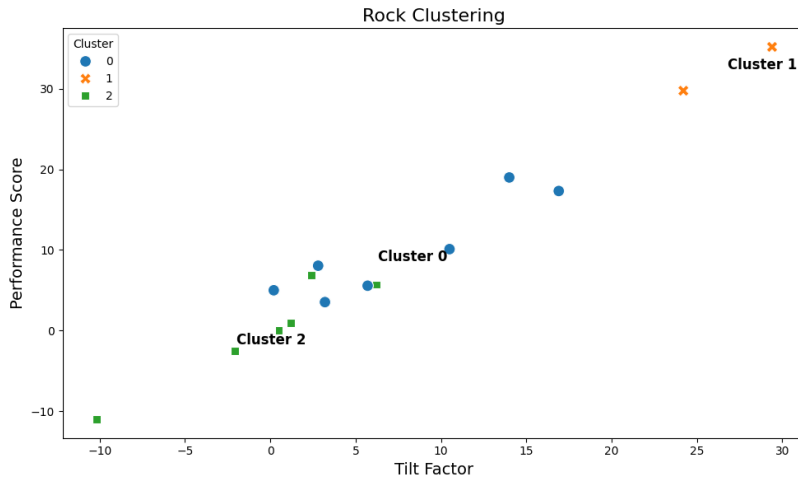


Figura 15: Distribución de clústeres para jugadores que escuchan Rock

Tabla 29. Estadísticas por clúster - Rock

Clúster	Tamaño	GPM ()	Tilt ()	Performance ()
0	9	390	6.30	8.56
1	4	470	26.80	32.46
2	13	280	-1.97	-1.75

A diferencia de otros géneros, el clúster con mayor rendimiento (clúster 1) también muestra un tilt elevado, lo cual podría indicar que algunos jugadores canalizan intensas emociones para mejorar su desempeño, una característica particular del rock.

En conjunto, los análisis de clústeres por género musical permiten observar patrones consistentes: los jugadores con menor tilt y mayor GPM suelen agruparse en los clústeres con mejor performance. Estos resultados pueden informar futuros estudios sobre la influencia de la música en el desempeño de jugadores en entornos competitivos.