



MÁSTER UNIVERSITARIO EN INGENIERÍA DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

Desarrollo de un modelo predictivo para estimar el
éxito de canciones

Autor: Marta Velilla Arana

Director: David Martín-Corral Calvo

Madrid

Junio de 2026

Declaración de originalidad

Declaro bajo mi responsabilidad que el Proyecto presentado con el título **Desarrollo de un modelo predictivo para estimar el éxito de canciones** en la ETS de Ingeniería – ICAI de la Universidad Pontificia Comillas en el curso académico **2025-2026** es de mi autoría y no ha sido presentado con anterioridad a otros efectos. El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido tomada de otros documentos está debidamente referenciada.

Uso de Inteligencia Artificial¹

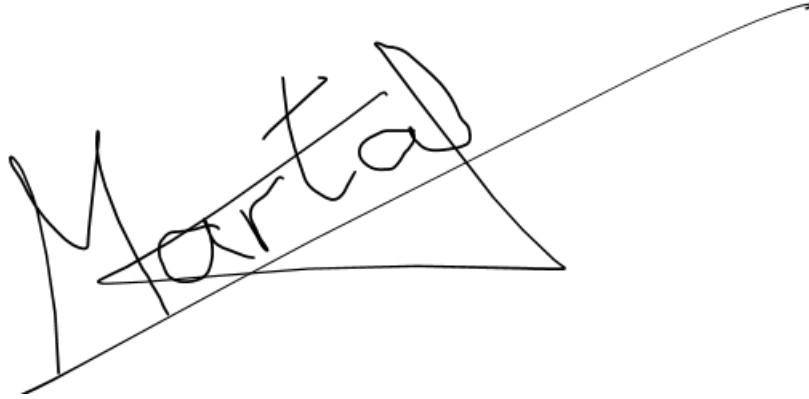
Declaro bajo mi responsabilidad que (indicar la opción correcta):

- No he utilizado Inteligencia Artificial en la elaboración del presente documento.
- He utilizado Inteligencia Artificial en la elaboración del presente documento y/o del Anexo B siempre en las condiciones permitidas por la Universidad Pontificia Comillas, es decir, aplicando el Nivel 2 de la [Escala de Evaluación de Perkins et al. \(2024\)](#): *“La IA puede utilizarse para actividades previas a la tarea, como la lluvia de ideas, la descripción y la investigación inicial. Este nivel se centra en el uso de la IA para la planificación, las síntesis y la generación de ideas, pero las evaluaciones deben hacer hincapié en la capacidad de desarrollar y refinar estas ideas de forma independiente”*. En concreto, las Inteligencia Artificial ha sido empleada para:

1. Brainstorming de ideas de investigación: Utilizado para idear y esbozar posibles áreas de investigación.
2. Referencias: Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
3. Metodólogo: Para descubrir métodos aplicables a problemas específicos de investigación.
4. Estudios multidisciplinares: Para comprender perspectivas de otras comunidades sobre temas de naturaleza multidisciplinar.
5. Corrector de estilo literario y de lenguaje: Para mejorar la calidad lingüística y estilística del texto.

¹ Esta declaración se refiere al uso de la Inteligencia Artificial generativa para realizar los documentos del Proyecto (Anexo B y Memoria). No aplica a Proyectos donde, por su naturaleza, deban emplear inteligencia artificial como parte de los mismos (aplicación de técnicas de aprendizaje automático, redes neuronales, análisis de datos...)

- 6. Generador previo de diagramas de flujo y contenido: Para esbozar diagramas iniciales.
- 7. Sintetizador y divulgador de libros complicados: Para resumir y comprender literatura compleja.
- 8. Revisor: Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.



Firmado (alumno): Marta Velilla Arana

Fecha: 10/06/2026

Autorización para la entrega del Proyecto

El Director del Proyecto	El co-Director del Proyecto (si aplica)
Fdo:	Fdo:
Fecha:	Fecha:



COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI

MÁSTER UNIVERSITARIO EN INGENIERÍA DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

Desarrollo de un modelo predictivo para estimar el
éxito de canciones

Autor: Marta Velilla Arana

Director: David Martín-Corral Calvo

Madrid

Agradecimientos

Muchas gracias a todo el mundo que aportó su granito de arena para que sea posible la realización de este trabajo.

DESARROLLO DE UN MODELO PREDICTIVO PARA ESTIMAR EL ÉXITO DE CANCIONES

Autor: Velilla Arana, Marta.

Director: Martín-Corral Calvo, David.

Entidad Colaboradora: ICAI – Universidad Pontificia Comillas.

RESUMEN DEL PROYECTO

Este trabajo desarrolla un sistema *end-to-end* interactivo para predecir el éxito comercial de canciones mediante la hibridación de procesamiento digital de señales y modelos de *Machine Learning*. Utilizando el *Million Song Dataset*, se diseñó un pipeline en Python con la librería *Librosa* para extraer 15 descriptores espectrales y temporales, optimizando un modelo basado en *Random Forest*. El sistema alcanzó un coeficiente de determinación (R^2) de 0,4329, superando los estándares acústicos de la literatura y desplegándose con éxito en una aplicación web multipágina desarrollada en Dash e integrada con la API de Spotify.

Palabras clave: Música, Viralidad, Random Forest, Procesamiento Digital de Señales, Spotify.

1. Introducción

En la última década, la digitalización de la industria musical y la popularización de los servicios de streaming han transformado radicalmente los mecanismos de distribución y consumo cultural. Tradicionalmente, la identificación de un éxito musical o hit se ha considerado un fenómeno místico, gobernado por la intuición de los agentes de la industria o por componentes puramente azarosos. Sin embargo, el paradigma científico denominado Hit Song Science (HSS), inscrito dentro del campo de la Recuperación de Información Musical (MIR, por sus siglas en inglés), plantea que el comportamiento del mercado responde, en gran medida, a estructuras de datos subyacentes e interacciones complejas entre la señal acústica y el contexto del artista [1].

A pesar de los avances bibliográficos en la materia, la literatura contemporánea sufre una importante brecha entre los análisis teóricos de laboratorio y el desarrollo de herramientas de ingeniería de software que sean funcionales, interpretables y reproducibles en entornos de producción ligeros. La mayoría de las aproximaciones previas se limitan a modelos estadísticos lineales planos incapaces de mapear interacciones cruzadas, o a modelos de "caja negra" (Deep Learning) con una alta opacidad que además dependen de análisis lingüísticos de las letras sesgados por el idioma.

Este proyecto mitiga dicha problemática mediante la concepción de una arquitectura de ingeniería modular de extremo a extremo. La propuesta se justifica por la viabilidad técnica que ofrece el ecosistema de computación científica moderno, permitiendo la transformación de señales de audio temporales complejas no estructuradas en vectores de características analíticas de alta fidelidad, con el fin de generar conocimiento accionable para creadores y analistas del sector.

2. Definición del Proyecto

El objetivo primordial de este proyecto consiste en el diseño, implementación y validación de una plataforma tecnológica capaz de estimar cuantitativamente la probabilidad de éxito de una pista musical antes de su lanzamiento comercial, basándose exclusivamente en sus atributos físicos y en el impacto digital previo del artista. El perímetro técnico del sistema se

delimita a través de un flujo de información unidireccional y sin estado (stateless), que abarca desde la ingesta de archivos de audio comerciales hasta la exposición de métricas visuales al usuario final.

Para garantizar el rigor matemático y la escalabilidad industrial, el proyecto se divide en cuatro grandes bloques metodológicos:

1. **Módulo de Ingesta y Procesamiento Digital de Señales (PDS):** Encargado de acondicionar la señal analógica o comprimida y extraer descriptores microestructurales e invariantes.
2. **Módulo de Ingeniería de Datos:** Diseñado para limpiar, filtrar y agregar estadísticamente las matrices de alta resolución del *Million Song Dataset* (MSD) [2], mitigando cualquier riesgo de inconsistencia numérica o asimetría.
3. **Módulo Predictivo:** Basado en un algoritmo de regresión por conjuntos cuya optimización local permite alcanzar un elevado poder de generalización manteniendo un adecuado control del sobreajuste.
4. **Capa de Aplicación e Integración Externa:** Un cuadro de mando interactivo que orquesta en tiempo real el motor de inferencia con llamadas asíncronas a la API de Spotify.

3. Descripción del modelo

El sistema se ha edificado bajo un enfoque estrictamente modular. El procesamiento de audio se inicia realizando un submuestreo homogéneo de la señal de entrada a una frecuencia de muestreo de 22.050 Hz. Este parámetro optimiza el coste computacional y satisface plenamente el teorema de Nyquist-Shannon para la captura de las frecuencias fundamentales de la música occidental. Posteriormente, la señal se analiza en el dominio frecuencial aplicando la Transformada de Fourier de Corto Tiempo (STFT) mediante ventanas temporales, lo que descompone la onda física.

El núcleo predictivo se sustenta sobre un modelo de Random Forest. El pipeline de preprocesamiento garantiza que los datos en producción mantengan exactamente los mismos factores de escalado y normalización calculados en la etapa de entrenamiento masivo (training-serving skew). El modelo toma como entradas un vector consolidado de 15 descriptores analíticos que incluyen métricas de sonoridad global (loudness en dB), tempo, energía, tonalidad y descriptores espectrales.

La interfaz de usuario se implementa mediante el framework Dash (by Plotly), estructurándose en una arquitectura multipágina orientada a la explicabilidad del modelo. El frontend se conecta con un backend en Flask que coordina de forma reactiva y unidireccional la lógica de negocio a través de callbacks asíncronos.

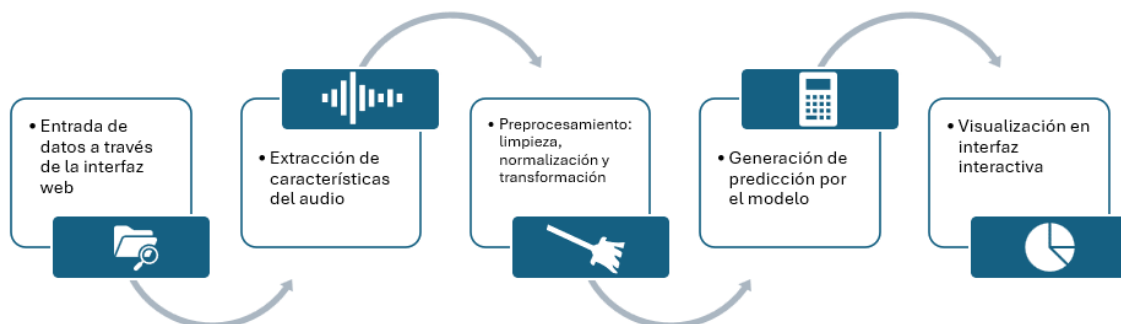


Ilustración 1: Arquitectura general del pipeline.

4. Resultados

La evaluación estadística del modelo de Machine Learning arrojó resultados de alto valor ingenieril. El coeficiente de determinación (R^2) alcanzado se situó de forma estable en 0,4329 mediante técnicas de validación cruzada. Al contrastar este rendimiento con los baselines históricos de la literatura MIR, donde los modelos puramente acústicos oscilan de manera uniforme en rangos R^2 entre 0,15 y 0,35, se concluye que el pipeline diseñado exprime de forma óptima el límite superior teórico de la señal física auditiva [1].

El análisis de la importancia de las variables ("Feature Importance") desveló una marcada hegemonía del contexto socio-digital sobre la onda sonora pura: las métricas vinculadas al capital social del artista dominan el modelo de ensamble (artist_familiarity con un 46,33%

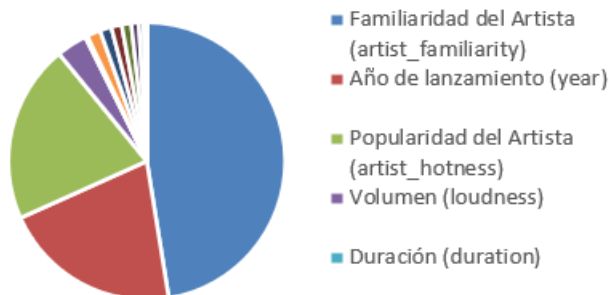


Ilustración 2: Arquitectura general del pipeline.

y artist_hottness con un 20,29%). Asimismo, la variable temporal (year, con un 20,34%) demostró ser crítica, lo que evidencia que los patrones estéticos de éxito son anacrónicos y requieren una normalización cronológica para adaptarse a las mutaciones macroeconómicas del mercado de la música.

A nivel de software, la plataforma web interactiva probó ser altamente robusta. Consigue realizar la ingesta, la extracción PDS local con Librosa [3], el enriquecimiento mediante la API de Spotify y la inferencia predictiva libre de estado en tiempo real, bloqueando eficazmente las entradas concurrentes del cliente y gestionando excepciones o corrupciones de archivos sin caídas críticas del servidor.

5. Conclusiones

El desarrollo de este trabajo demuestra empíricamente que el éxito musical, lejos de ser una variable puramente aleatoria, responde a estructuras de datos latentes que pueden ser parametrizadas, computadas y explotadas de manera reproducible mediante técnicas de aprendizaje automático. La arquitectura *end-to-end* implementada no solo valida la viabilidad de utilizar descriptores espectrales y temporales de la señal física de audio para superar los *baselines* acústicos tradicionales de la literatura MIR, sino que además ofrece un sistema con independencia lingüística capaz de evaluar el potencial de artistas emergentes sin el sesgo penalizador de los enfoques basados en redes de colaboración masiva. En definitiva, la correcta hibridación entre el procesamiento digital de señales, la ingeniería de datos y el modelo predictivo óptimo de *Random Forest* dota a la plataforma de un alto grado de interpretabilidad y solidez técnica, consolidando una herramienta de ingeniería de software funcional y accesible para el análisis predictivo en el sector musical.

6. Referencias

1. **Pachet, F., y Roy, P. (2008).** Hit song science: Not yet a science? *IEEE Multimedia*, 15(1), 66–71.
2. **T. Bertin-Mahieux, D. P. W. Ellis, B. Whitman, y P. Lamere,** "The Million Song Dataset", *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, pp. 591–596, 2011.
3. **McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., & Nieto, O. (2015).** librosa: Audio and music signal analysis in python. *En Proceedings of the 14th python in science conference (Vol. 8, pp. 18-25)*.

DEVELOPMENT OF A PREDICTIVE MODEL TO ESTIMATE SONG POPULARITY

Author: Velilla Arana, Marta.

Supervisor: Martín-Corral Calvo, David.

Collaborating Entity: ICAI – Universidad Pontificia Comillas

ABSTRACT

This work develops an interactive, end-to-end system to predict the commercial success of songs by hybridizing digital signal processing and Machine Learning models. Utilizing the Million Song Dataset, a Python pipeline was designed using the Librosa library to extract 15 spectral and temporal features, optimizing a model based on Random Forest. The system achieved a coefficient of determination (R^2) of 0.4329, outperforming the acoustic baselines in the literature and successfully deploying as a multi-page web application developed in Dash and integrated with the Spotify API.

Keywords: Music, Virality, Random Forest, Digital Signal Processing, Spotify.

1. Introduction

In the last decade, the digitization of the music industry and the mainstream adoption of streaming services have radically transformed the mechanisms of cultural distribution and consumption. Traditionally, identifying a musical success or "hit" has been considered a mystical phenomenon, governed by the intuition of industry insiders or by purely random factors. However, the scientific paradigm known as Hit Song Science (HSS), framed within the field of Music Information Retrieval (MIR), posits that market behavior responds, to a large extent, to underlying data structures and complex interactions between the acoustic signal and the artist's context [1].

Despite literature advances in the field, contemporary research suffers from a major gap between theoretical laboratory analysis and the development of functional, interpretable, and reproducible software engineering tools in lightweight production environments. Most prior approaches are limited to flat, linear statistical models incapable of mapping cross-interactions, or to "black box" models (Deep Learning) with high opacity that additionally rely on language-biased lyric analysis.

This project mitigates this problem through the conception of an end-to-end modular engineering architecture. The proposal is justified by the technical feasibility offered by the modern scientific computing ecosystem, allowing the transformation of complex, unstructured, time-domain audio signals into high-fidelity analytical feature vectors to generate actionable insights for creators and industry analysts.

2. Project Definition

The primary objective of this project consists of the design, implementation, and validation of a technological platform capable of quantitatively estimating the probability of success of a musical track prior to its commercial release, based exclusively on its physical attributes and the artist's previous digital impact. The technical scope of the system is defined by a unidirectional and stateless information flow, spanning from the ingestion of commercial audio files to the presentation of visual metrics to the end user.

To guarantee mathematical rigor and industrial scalability, the project is divided into four major methodological blocks:

1. **Ingestion and Digital Signal Processing (DSP) Module:** Responsible for conditioning the analog or compressed signal and extracting microstructural and invariant descriptors.
2. **Data Engineering Module:** Designed to clean, filter, and statistically aggregate the high-resolution matrices of the Million Song Dataset (MSD) [2], mitigating any risk of numerical inconsistency or skewness.
3. **Predictive Module:** Based on an ensemble regression algorithm whose local optimization achieves a high generalization power while maintaining adequate control over overfitting.
4. **Application and External Integration Layer:** An interactive dashboard that orchestrates the inference engine in real time with asynchronous calls to the Spotify API.

3. Model Description

The system has been built under a strictly modular approach. Audio processing begins by performing a homogeneous down sampling of the input signal to a sampling frequency of 22,050 Hz. This parameter optimizes computational cost and fully satisfies the Nyquist-Shannon theorem for capturing the fundamental frequencies of Western music. Subsequently, the signal is analyzed in the frequency domain by applying the Short-Time Fourier Transform (STFT) through time windows, which decomposes the physical wave.

The predictive core is sustained by a Random Forest model. The preprocessing pipeline guarantees that production data maintains exactly the same scaling and normalization factors computed during the massive training stage (preventing training-serving skew). The model takes as inputs a consolidated vector of 15 analytical descriptors including metrics for overall loudness (in dB), tempo, energy, key, and spectral descriptors.

The user interface is implemented through the Dash (by Plotly) framework, structured into a multi-page architecture oriented toward model explainability. The frontend connects to a Flask backend that reactively and unidirectionally coordinates the business logic via asynchronous callbacks.

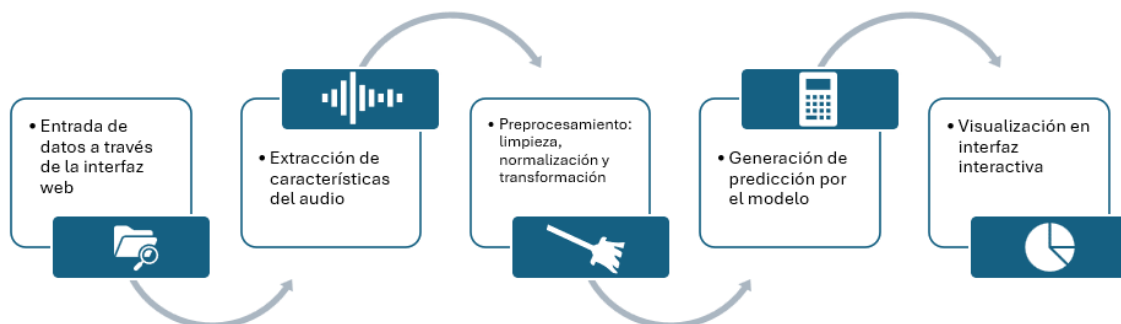


Ilustración 3: General pipeline architecture.

4. Results

The statistical evaluation of the Machine Learning model yielded high-value engineering results. The coefficient of determination (R^2) achieved stood stably at 0.4329 using cross-validation techniques. Contrasting this performance with historical baselines in MIR literature, where purely acoustic models uniformly oscillate in R^2 ranges between 0.15 and 0.35, it is concluded that the designed pipeline optimally exploits the theoretical upper bound of the physical auditory signal [1].

Feature Importance analysis revealed a stark hegemony of the socio-digital context over the

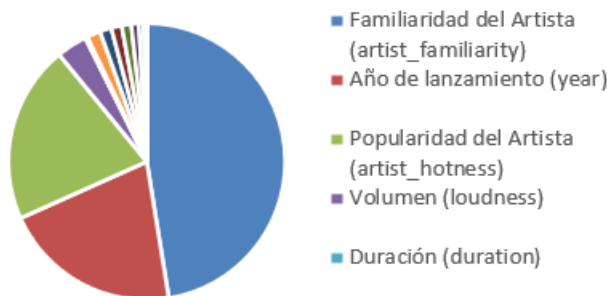


Ilustración 4: Feature importance in the model.

pure sound wave: metrics linked to the artist's social capital dominate the ensemble model (*artist_familiarity* at 46.33% and *artist_hottnesss* at 20.29%). Furthermore, the temporal variable (*year*, at 20.34%) proved to be critical, showing that aesthetic success patterns are anachronistic and require chronological normalization to adapt to macroeconomic mutations in the music market.

At the software level, the interactive web platform proved to be highly robust. It successfully achieves real-time ingestion, local DSP extraction with Librosa [3], enrichment via the Spotify API, and stateless predictive inference, effectively blocking concurrent client inputs and handling exceptions or file corruptions without critical server crashes.

5. Conclusions

The development of this work empirically demonstrates that musical success, far from being a purely random variable, responds to latent data structures that can be parameterized, computed, and exploited in a reproducible manner using machine learning techniques. The implemented end-to-end architecture not only validates the feasibility of using spectral and temporal descriptors of the physical audio signal to outperform traditional acoustic baselines in MIR literature, but also offers a language-agnostic system capable of evaluating the potential of emerging artists without the penalizing bias of approaches based on massive collaboration networks. Ultimately, the proper hybridization between digital signal processing, data engineering, and the optimal Random Forest predictive model provides the platform with a high degree of interpretability and technical soundness, establishing a functional and accessible software engineering tool for predictive analysis in the music sector.

6. References

1. **Pachet, F., y Roy, P. (2008).** Hit song science: Not yet a science? *IEEE Multimedia*, 15(1), 66–71.
2. **T. Bertin-Mahieux, D. P. W. Ellis, B. Whitman, y P. Lamere,** "The Million Song Dataset", *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, pp. 591–596, 2011.
3. **McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., & Nieto, O. (2015).** librosa: Audio and music signal analysis in python. *En Proceedings of the 14th python in science conference* (Vol. 8, pp. 18-25).

Índice de la memoria

Capítulo 1. Introducción	7
Capítulo 2. Descripción de las Tecnologías.....	9
2.1 Python y ecosistema de Data Science	9
2.2 Scikit-learn	10
2.3 Procesamiento de audio con librosa	10
2.4 Desarrollo de web con Dash y Plotly	11
2.5 Otras herramientas utilizadas	12
Capítulo 3. Estado de la Cuestión.....	13
3.1 Fundamentos de Procesamiento de Señales y Extracción de Características Musicales.....	14
3.1.1 Transformada de Fourier de Corto Tiempo.....	14
3.2 Aproximaciones existentes en la literatura.....	17
3.2.1 Enfoques basados puramente en Características Acústicas	17
3.2.2 Enfoques basados en Metadatos y Factores Sociales	21
3.2.3 El Paradigma del Deep Learning y Aprendizaje de Representaciones.....	24
3.2.4 Enfoques Híbridos Contemporáneos y Multimodal	27
3.3 El Million Song Dataset	29
Capítulo 4. Definición del Trabajo	33
4.1 Justificación.....	33
4.2 Objetivos	36
4.2.1 Objetivo general.....	36
4.2.2 Objetivos específicos	37
4.2.3 Alcance del sistema	38
4.2.4 Limitaciones del proyecto.....	39
4.3 Metodología.....	41
4.3.1 Enfoque general del desarrollo.....	41
4.3.2 Metodología del desarrollo del modelo.....	42
4.3.3 Metodología del desarrollo de la aplicación web.....	43
4.3.4 Integración de componentes.....	44
4.3.5 Herramientas y tecnologías utilizadas	45

4.4	Planificación.....	46
Capítulo 5. Sistema/Modelo Desarrollado.....		51
5.1	Análisis del Sistema	51
5.1.1	Requisitos Funcionales.....	51
5.1.2	Requisitos no funcionales	52
5.1.3	Actores del sistema	52
5.1.4	Casos de uso	53
5.2	Diseño.....	54
5.2.1	Arquitectura general del pipeline.....	54
5.2.2	Diagrama de componentes	55
5.2.3	Flujo de datos	56
5.2.4	Diseño de la base de datos	57
5.3	Implementación.....	58
5.3.1	Módulo de extracción de características.....	58
5.3.2	Módulo de preprocesamiento de datos.....	62
5.3.3	Módulo de modelado predictivo	66
5.3.4	Módulo de aplicación web.....	72
5.3.5	Integración con la API de Spotify.....	86
Capítulo 6. Análisis de Resultados.....		91
6.1	Evaluación del modelo	91
6.1.1	Análisis de la importancia de las variables.....	93
6.2	Análisis de tendencias y perfiles de éxito.....	96
6.3	Resultados de la aplicación web.....	99
6.4	Análisis crítico de resultados.....	100
Capítulo 7. Conclusiones y Trabajos Futuros.....		103
7.1	Conclusiones Técnicas del procesamiento digital de señales (PDS).....	104
7.2	Conclusiones metodológicas de la arquitectura de Machine Learning	105
7.3	Aportaciones del trabajo en comparación con la literatura ya existente	106
7.3.1	Desarrollo de un sistema end-to-end completo.....	106
7.3.2	Propuesta de una arquitectura predictiva ligera y reproducible.....	107
7.3.3	Predicción de popularidad en fase pre-lanzamiento.....	107
7.3.4	Modelo interpretable frente a enfoques de caja negra.....	107

7.3.5	<i>Combinación eficiente de información acústica y metadatos</i>	108
7.3.6	<i>Generación de conocimiento accionable para la industria musical</i>	108
7.3.7	<i>Contribución metodológica al ámbito del Music Information Retrieval</i>	108
7.4	Trabajos futuros	109
7.4.1	<i>Incorporación de modelos de Deep Learning</i>	109
7.4.2	<i>Predicción directa a partir del audio completo</i>	109
7.4.3	<i>Integración de información multimodal</i>	110
7.4.4	<i>Obtención automática de datos mediante APIs musicales</i>	110
7.4.5	<i>Optimización y actualización continua del modelo</i>	111
7.4.6	<i>Explicabilidad avanzada de las predicciones</i>	111
7.4.7	<i>Ampliación de la aplicación web</i>	111
7.4.8	<i>Despliegue como plataforma de apoyo a la industria musical</i>	112
Capítulo 8.	<i>Bibliografía</i>	113
ANEXO:	<i>Alineación del Proyecto con los Objetivos de Desarrollo Sostenible (ODS)</i>	115

Índice de Ilustraciones

Ilustración 1: Arquitectura general del pipeline.	10
Ilustración 2: Arquitectura general del pipeline.	11
Ilustración 3: General pipeline architecture.....	14
Ilustración 4: Feature importance in the model.	15
Ilustración 5: Ejemplo transformada de Fourier [20].	15
Ilustración 6: Identificación de los principales componentes.....	45
Ilustración 7: Diagrama de Gantt con la cronología.....	46
Ilustración 8: Arquitectura general del pipeline del sistema.	55
Ilustración 9: Diagrama de componentes del sistema.	56
Ilustración 10: Flujo de datos del sistema.	57
Ilustración 11: Gráfico con la influencia de cada variable en el modelo.	72
Ilustración 12: Arquitectura del ecosistema de ingeniería integral.	75
Ilustración 13: Menú inicial interactivo.	76
Ilustración 14: Configuración inicial del predictor sin completar.....	77
Ilustración 15: Configuración inicial del predictor completada.	77
Ilustración 16: Resultados de la aplicación.	78
Ilustración 17: Error en la aplicación.	79
Ilustración 18: Sugerencias de la aplicación de posible artista.	80
Ilustración 19: Flujo de los estados.	81
Ilustración 20: Importancia de las variables en el modelo en la aplicación.	83
Ilustración 21: Distribución de la importancia de variables por categoría.....	84
Ilustración 22: Visualización de Resultados en la aplicación.....	85
Ilustración 23: Detalles técnicos normalizados.	86
Ilustración 24: Distribución de peso por categoría.....	95
Ilustración 25: Características medias de las canciones más famosas.....	97
Ilustración 26: Gráfico de araña con la comparación de artistas.....	98

Índice de tablas

Tabla 1: Valores de la importancia de cada variable en el modelo. 71

Capítulo 1. INTRODUCCIÓN

La industria musical se encuentra inmersa en un proceso de transformación profunda impulsado por la digitalización, la disponibilidad masiva de datos y el uso creciente de técnicas de inteligencia artificial para apoyar la toma de decisiones en producción, distribución y consumo cultural. En este contexto, la capacidad de anticipar el comportamiento de una canción antes de su consolidación comercial se ha convertido en un problema de especial interés tanto desde el punto de vista académico como desde la perspectiva empresarial.

El presente Trabajo Fin de Máster se centra en el desarrollo de un modelo predictivo orientado a estimar el éxito de canciones a partir de parámetros musicales y variables de popularidad, integrando técnicas de procesamiento digital de señal, aprendizaje automático y visualización interactiva en una aplicación web. La propuesta parte de la premisa de que el rendimiento de una obra musical no depende exclusivamente de factores subjetivos, sino que también puede analizarse a través de descriptores acústicos, metadatos del artista y patrones cuantificables presentes en grandes volúmenes de información.

Uno de los principales desafíos del análisis musical automático reside en la complejidad de la señal de audio. Una canción contiene información temporal, espectral y armónica altamente dinámica, lo que dificulta la extracción de descriptores robustos y representativos. A ello se suma la heterogeneidad de los datasets musicales. Muchas bases de datos contienen registros incompletos, inconsistencias o diferencias en los formatos de almacenamiento. Por tanto, antes de entrenar cualquier modelo predictivo, resulta imprescindible realizar procesos rigurosos de limpieza, normalización y validación de los datos.

Desde la perspectiva de investigación, la predicción del éxito musical plantea importantes desafíos. El concepto de “éxito” es complejo y multidimensional, porque puede relacionarse con popularidad, difusión, impacto en plataformas, reconocimiento social o permanencia en

el tiempo. Además, los métodos tradicionales para valorar una canción suelen apoyarse en intuiciones de mercado, criterios artísticos o análisis retrospectivos, lo que deja margen para la aplicación de modelos más sistemáticos basados en datos.

En este escenario, el proyecto propone una aproximación de ingeniería que permita transformar un archivo de audio y un conjunto de atributos asociados en una estimación cuantitativa útil para el análisis musical. Para ello, se articula un flujo de trabajo que abarca desde la extracción de características del sonido hasta la explotación de servicios externos y la presentación de resultados al usuario mediante una interfaz accesible.

El trabajo no pretende reducir el fenómeno musical a una fórmula cerrada, sino estudiar hasta qué punto ciertas variables observables pueden contribuir a explicar la probabilidad de éxito de una canción. Por ello, la investigación se orienta a construir una herramienta experimental y funcional que combine rigor técnico con aplicabilidad real, sirviendo tanto como ejercicio académico como base para futuras extensiones del sistema.

De manera más concreta, el proyecto se apoya en un conjunto de datos musicales de gran escala y en un proceso de modelado capaz de relacionar variables como tempo, energía, loudness, descriptores armónicos o notoriedad del artista con un indicador de popularidad. A partir de esa base, se plantea una solución que permita no solo entrenar y validar modelos predictivos, sino también desplegarlos en un entorno web para facilitar su uso e interpretación.

Por tanto, este Trabajo Fin de Máster se justifica por la convergencia de tres factores: la disponibilidad de datos musicales estructurados, la madurez de las bibliotecas de análisis y aprendizaje automático, y la necesidad creciente de herramientas capaces de convertir información compleja en conocimiento accionable dentro del ecosistema musical digital. Los capítulos siguientes desarrollan, en primer lugar, las tecnologías que hacen posible esta propuesta y, posteriormente, el estado de la cuestión, la definición metodológica del trabajo, el sistema implementado y el análisis de resultados.

Capítulo 2. DESCRIPCIÓN DE LAS TECNOLOGÍAS

El desarrollo del proyecto requiere la integración de un conjunto de tecnologías complementarias que cubren todo el ciclo de vida de la solución: adquisición y tratamiento de datos, análisis de audio, entrenamiento del modelo predictivo, persistencia de artefactos y despliegue de una interfaz web interactiva. Este capítulo describe las herramientas más relevantes utilizadas en el sistema con el fin de facilitar la comprensión técnica de los capítulos posteriores, en línea con la finalidad que la memoria asigna a la descripción de tecnologías específicas del proyect.

2.1 PYTHON Y ECOSISTEMA DE DATA SCIENCE

Python se ha consolidado como uno de los lenguajes de programación más importantes dentro del ámbito científico y del aprendizaje automático. Su sintaxis sencilla, su gran comunidad de desarrolladores y la disponibilidad de librerías especializadas han convertido a Python en el estándar de facto para proyectos de análisis de datos y Machine Learning.

Dentro de este ecosistema destacan bibliotecas como NumPy y Pandas. NumPy proporciona estructuras de datos optimizadas para operaciones matemáticas y cálculo matricial de alto rendimiento, fundamentales para el procesamiento de grandes volúmenes de información. Por su parte, Pandas facilita la manipulación y análisis de datos estructurados mediante estructuras tipo DataFrame.

En el proyecto, estas herramientas se utilizan para almacenar, transformar y procesar las características extraídas del Million Song Dataset, permitiendo la construcción eficiente de pipelines de entrenamiento y evaluación.

2.2 SCIKIT-LEARN

Scikit-learn es una de las bibliotecas más relevantes para el desarrollo de modelos de Machine Learning en Python. Proporciona algoritmos optimizados para clasificación, regresión, clustering y reducción de dimensionalidad, además de herramientas para preprocesamiento y validación de modelos.

En este proyecto se emplea principalmente para la construcción de modelos de regresión orientados a la predicción de popularidad musical. Entre las funcionalidades utilizadas destacan:

- División train/test.
- Normalización de variables.
- Validación cruzada.
- Métricas de evaluación.
- Algoritmos basados en árboles y ensembles.

El uso de pipelines de Scikit-learn permite asegurar la coherencia entre las fases de entrenamiento y despliegue, evitando discrepancias entre los datos utilizados durante el aprendizaje y los empleados posteriormente en producción.

2.3 PROCESAMIENTO DE AUDIO CON LIBROSA

Librosa es una librería especializada en análisis de audio y procesamiento digital de señales musicales. Su principal ventaja reside en proporcionar funciones de alto nivel para extraer descriptores acústicos complejos de forma eficiente.

La herramienta permite calcular características como:

- Tempo y ritmo.
- Energía y loudness.
- Cromagramas.

- MFCCs.
- Centroide espectral.
- Tonalidad y modo.

El procesamiento comienza con la carga de la señal de audio y su conversión a una representación digital manipulable. Posteriormente se aplican transformadas espectrales y algoritmos de análisis temporal para obtener representaciones numéricas de las propiedades musicales.

Librosa resulta especialmente adecuada para aplicaciones de Music Information Retrieval (MIR), disciplina orientada a la extracción automática de información musical a partir de señales de audio.

2.4 DESARROLLO DE WEB CON DASH Y PLOTLY

Para el desarrollo de la aplicación web se ha utilizado Dash, un framework basado en Flask y React que permite construir interfaces interactivas utilizando exclusivamente Python.

Dash facilita la integración entre modelos de Machine Learning y componentes visuales, permitiendo que las predicciones generadas por el sistema puedan visualizarse de forma inmediata en el navegador. Esto simplifica enormemente el desarrollo de aplicaciones analíticas.

Por otro lado, Plotly proporciona herramientas avanzadas de visualización de datos. Gracias a esta biblioteca es posible construir gráficos interactivos que permiten representar métricas de rendimiento, importancia de variables y resultados predictivos de manera intuitiva.

La combinación Dash + Plotly constituye una solución especialmente potente para aplicaciones de análisis de datos en tiempo real.

2.5 OTRAS HERRAMIENTAS UTILIZADAS

Además de las tecnologías principales, el proyecto incorpora diversas herramientas complementarias utilizadas durante el desarrollo.

Jupyter Notebook se emplea para la experimentación y validación inicial de modelos. Su capacidad para combinar código, visualizaciones y documentación en un mismo entorno facilita enormemente el análisis exploratorio.

Git se utiliza como sistema de control de versiones, permitiendo mantener un historial completo de cambios y facilitando la organización del desarrollo.

Asimismo, se utilizan herramientas como Joblib para la serialización de modelos entrenados y Requests para la comunicación con APIs externas. Estas tecnologías contribuyen a construir un ecosistema integrado y robusto para el desarrollo del sistema completo.

En conjunto, las tecnologías seleccionadas permiten construir un sistema completo capaz de abarcar todas las fases del proyecto: desde el procesamiento de señales musicales hasta la visualización interactiva de resultados predictivos.

Capítulo 3. ESTADO DE LA CUESTIÓN

Durante las últimas décadas, la industria musical ha experimentado una transformación radical impulsada por la digitalización y el auge de las plataformas de streaming. Servicios como Spotify, Apple o YouTube generan diariamente enormes volúmenes de datos relacionados con hábitos de escucha, preferencias musicales y métricas de popularidad. Este contexto ha favorecido el desarrollo de sistemas automáticos capaces de analizar canciones mediante técnicas de inteligencia artificial y aprendizaje automático.

Dentro de este ámbito surge la disciplina conocida como Music Information Retrieval (MIR), centrada en la extracción automática de información musical a partir de señales de audio digitales. El objetivo principal de esta área es transformar una señal acústica compleja en representaciones numéricas que permitan modelar propiedades musicales como el ritmo, el timbre, la tonalidad o la energía percibida.

Diversos estudios han demostrado que determinadas características acústicas presentan correlaciones significativas con el comportamiento de los usuarios y con el éxito comercial de las canciones. Como consecuencia, el uso de técnicas de Machine Learning en análisis musical se ha convertido en una línea de investigación consolidada tanto en el ámbito académico como industrial.

Además, el crecimiento de la computación en la nube y la disponibilidad de datasets masivos han permitido entrenar modelos cada vez más sofisticados, capaces de analizar millones de canciones y detectar patrones complejos difíciles de identificar mediante métodos tradicionales.

3.1 FUNDAMENTOS DE PROCESAMIENTO DE SEÑALES Y EXTRACCIÓN DE CARACTERÍSTICAS MUSICALES

El Procesamiento Digital de Señales (DSP, por sus siglas en inglés) constituye la base tecnológica ineludible de la mayoría de los sistemas modernos de análisis musical automático y *Music Information Retrieval* (MIR). Dado que el sonido acústico es una onda analógica continua, para su tratamiento computacional es imperativo que dicha señal sea digitalizada. Una señal de audio se representa matemáticamente como una secuencia temporal discreta de valores, obtenida a través de dos procesos fundamentales: el **muestreo** (que discretiza el tiempo basándose en el teorema de Nyquist-Shannon) y la **cuantificación** (que discretiza la amplitud asignando un número finito de bits a cada muestra) [4]

3.1.1 TRANSFORMADA DE FOURIER DE CORTO TIEMPO

Si bien la representación en el dominio del tiempo (la forma de onda) es útil para observar variaciones de amplitud, resulta insuficiente para comprender la naturaleza musical de una señal. Para extraer información sobre la armonía, la melodía y el timbre, es necesario transformar la señal al dominio de la frecuencia.

El método más utilizado en el análisis musical es la Transformada de Fourier de Corto Tiempo (STFT). Dado que la música es una señal no estacionaria —es decir, su espectro de frecuencias cambia constantemente con el tiempo—, la STFT divide la señal de audio en ventanas temporales cortas y superpuestas utilizando funciones de enventanado como Hann o Hamming [4]. A cada una de estas ventanas se le aplica la Transformada Discreta de Fourier (DFT). La ecuación general de la DFT que rige este proceso se define como:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N}kn}$$

Donde:

- $X(k)$ representa la magnitud y fase de la componente frecuencial k

- $x(n)$ es el valor de la señal en el dominio temporal discreto en el índice n
- N es el número total de muestras en la ventana de análisis
- $e^{-j*2\pi/N*kn}$ es la base exponencial compleja

Gracias a esta técnica, es posible generar un espectrograma: una matriz bidimensional que ilustra cómo evoluciona la intensidad de las distintas frecuencias a lo largo del tiempo.

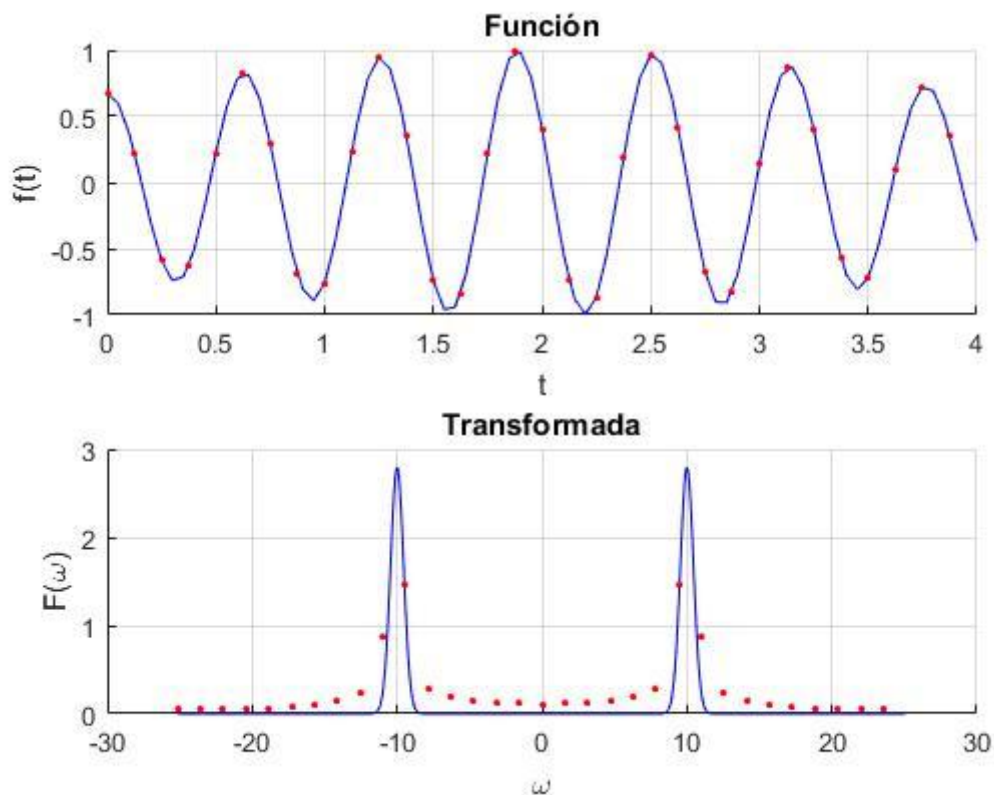


Ilustración 5: Ejemplo transformada de Fourier [20].

Para ilustrar este proceso, la Ilustración 5 muestra una señal periódica en el dominio temporal y su correspondiente representación en el dominio de la frecuencia obtenida mediante la Transformada de Fourier (gráfico inferior). Aunque la forma de onda presenta oscilaciones aparentemente complejas, la transformada revela que la mayor parte de la energía de la señal se concentra en una frecuencia concreta, representada por los picos observados alrededor de ± 10 rad/s. Esto demuestra cómo la Transformada de Fourier permite descomponer una señal en sus componentes frecuenciales fundamentales, proporcionando una descripción mucho

más útil para el análisis musical que la observación directa de la amplitud en el tiempo. En el caso de señales musicales reales, este procedimiento se aplica de manera sucesiva sobre pequeñas ventanas temporales mediante la STFT, permitiendo estudiar cómo evolucionan las frecuencias presentes en cada instante.

A partir de las representaciones espectrales generadas por la STFT, los sistemas de MIR extraen numerosas características (features) que condensan la información acústica en vectores matemáticos manejables por los algoritmos de Machine Learning. Estas métricas se pueden clasificar en:

- **MFCCs (Mel Frequency Cepstral Coefficients):** Constituyen el estándar de facto tanto en el reconocimiento automático de voz como en la clasificación musical. Su relevancia radica en que modelan la percepción auditiva humana. En lugar de utilizar una escala de frecuencias lineal, proyectan el espectro sobre la escala de Mel, una escala logarítmica que imita la resolución no lineal del oído humano, que es más sensible a variaciones en bajas frecuencias [5]. Posteriormente, se aplica una Transformada del Coseno Discreta (DCT) para decorrelacionar los coeficientes, obteniendo una representación muy compacta de la envolvente espectral o timbre.
- **Descriptores Espectrales Básicos:**
 - **Centroide Espectral (*Spectral Centroid*):** Indica el "centro de gravedad" del espectro. Perceptualmente, está estrechamente correlacionado con el "brillo" de un sonido.
 - **Spectral Roll-off:** Representa la frecuencia por debajo de la cual se concentra un determinado porcentaje (habitualmente el 85% o 95%) de la energía espectral total. Es útil para distinguir señales armónicas de señales ruidosas.
 - **Tasa de Cruce por Cero (*Zero Crossing Rate - ZCR*):** Mide la tasa a la que la señal cambia de signo. Es un indicador clave del grado de "ruido" en la señal y se utiliza frecuentemente para clasificar sonidos percusivos frente a sonidos tonales.

- **Chroma Features (Perfil de Clase de Tono):** Estas características proyectan el espectro musical completo en doce *bins* (contenedores) que representan las doce notas de la escala temperada occidental (C, C#, D, D#, etc.), ignorando la octava a la que pertenecen. Este proceso, conocido como "plegado de octava", abstrae el timbre y resalta de manera robusta el contenido armónico, facilitando tareas complejas como el análisis de la tonalidad, la detección de acordes y el seguimiento de progresiones armónicas (Müller, 2015).
- **Tempo y Beat Tracking:** Métodos orientados a la extracción del pulso musical (BPM) y la ubicación temporal de los golpes (*beats*), esenciales para determinar el ritmo y la "energía" bailable de una composición [6].

En la actualidad, la extracción de estos atributos ha sido democratizada gracias al desarrollo de ecosistemas de software de código abierto. Librerías de Python como Librosa [7] o Essentia [8] han simplificado enormemente la implementación de algoritmos avanzados de procesamiento digital de señales. Estas herramientas proporcionan interfaces de alto nivel que permiten a los investigadores integrar la extracción de características acústicas directamente en *pipelines* modernos de ciencia de datos y aprendizaje automático.

3.2 APROXIMACIONES EXISTENTES EN LA LITERATURA

Para estructurar los trabajos previos de tu revisión bibliográfica, la literatura científica se puede dividir en **cuatro grandes enfoques o aproximaciones metodológicas**. A continuación, se presentan los títulos y autores de los trabajos clave que configuran el estado del arte.

3.2.1 ENFOQUES BASADOS PURAMENTE EN CARACTERÍSTICAS ACÚSTICAS

Esta corriente de investigación parte de la premisa de que el éxito comercial de una obra musical está codificado, al menos en un porcentaje significativo, dentro de la propia señal de audio. Los autores que se adscriben a esta línea metodológica intentan aislar los factores exógenos (como el marketing, la reputación del artista o el sesgo social) para determinar si

existen patrones matemáticos universales en el ritmo, el timbre o la armonía que correlacionen directamente con la aceptación masiva del público.

3.2.1.1 “Hit Sing Science is Not Yet a Science” – Pachet & Roy [9]

Este artículo es considerado una de las obras fundamentales y críticas dentro del campo del *Music Information Retrieval* (MIR) en lo que respecta a la predicción de popularidad. Pachet y Roy acuñaron y evaluaron críticamente el término *Hit Song Science* (HSS). Los autores implementaron un robusto marco experimental utilizando una base de datos de más de 32,000 canciones de diversos géneros musicales. Extrajeron una amplia gama de características acústicas de bajo y medio nivel, incluyendo descriptores tímbricos (como los primeros coeficientes MFCC), fluctuaciones de energía, y variables temporales como el *tempo* y la regularidad del ritmo. Para la fase de modelado predictivo, implementaron técnicas clásicas de clasificación como regresiones logísticas, árboles de decisión binarios y Máquinas de Vectores de Soporte (SVM). El objetivo era predecir si una canción se convertiría en un "hit" (entrando en el Top 30 de las listas de éxitos oficiales) o si sería un "non-hit". Sus conclusiones fueron determinantes: los modelos basados exclusivamente en el audio no lograron superar significativamente el umbral del azar estadístico, demostrando que la predicción puramente acústica mediante enfoques lineales o clasificadores planos es un problema mal condicionado.

El presente trabajo fin de máster toma los hallazgos de Pachet y Roy no como un límite insalvable, sino como el punto de partida que justifica la transición hacia arquitecturas no lineales complejas. Mientras que Pachet y Roy se limitaron a herramientas estadísticas que buscaban relaciones directas y descriptores planos, este TFM implementa un modelo basado en Random Forest y enfoques de aprendizaje de máquina avanzado. La diferencia metodológica clave radica en que los bosques aleatorios permiten evaluar las interacciones cruzadas y no lineales entre las variables (por ejemplo, cómo un *tempo* rápido interactúa con un centroide espectral elevado), algo que los árboles de decisión simples o las regresiones lineales de 2008 no podían mapear de forma robusta. Además, los conjuntos de datos actuales y las herramientas de extracción de características son sustancialmente más precisos y estables que los disponibles hace casi dos décadas.

3.2.1.2 “In Search of the Modern Hit Song: A Music Information Retrieval Approach” – Interiano [10]

En este estudio a gran escala, Interiano y su equipo analizaron la evolución de la música popular durante un periodo de más de 30 años, examinando un corpus de miles de canciones que alcanzaron la lista *Billboard Hot 100*. A diferencia de enfoques anteriores, los autores utilizaron características acústicas de alto nivel obtenidas a través de la API de *The Echo Nest* (posteriormente integrada en Spotify), tales como la "bailabilidad" (*danceability*), "energía", "valencia" (carga emocional positiva o negativa de la música) y la presencia de instrumentación acústica frente a la electrónica. El estudio concluyó que las propiedades acústicas de los éxitos musicales cambian drásticamente con el tiempo; por ejemplo, observaron una tendencia histórica hacia canciones con menor "felicidad" (baja valencia) pero con ritmos más aptos para el baile. Utilizando modelos de clasificación y series temporales, demostraron que las características del audio sí poseen poder predictivo si se analiza el contexto dinámico de la época, logrando predecir el éxito con un rendimiento moderadamente superior a los modelos de base tradicionales.

A pesar del indudable valor del estudio de Interiano, su enfoque es eminentemente macroscópico e histórico; busca describir tendencias socio-musicales a lo largo de las décadas más que optimizar un clasificador de alta fidelidad para el despliegue inmediato. La divergencia fundamental con este TFM se centra en la granularidad y la arquitectura del modelo. Mientras que Interiano et al. dependen de métricas de alto nivel precalculadas por algoritmos propietarios de terceros (las variables de Spotify, cuyas ecuaciones exactas son secretas comerciales), este trabajo se fundamenta en un control directo del *pipeline* de extracción mediante el uso de librerías de código abierto como *Librosa* [7]. Esto permite inspeccionar matemáticamente descriptores puros (como el *Spectral Roll-off* o las *Chroma Features*) y entrenar un modelo predictivo local mediante *Machine Learning* clásico y ensamblado, garantizando la reproducibilidad científica y ofreciendo una solución técnica aplicable a pistas de audio locales de las que no se disponga de metadatos en plataformas de *streaming*.

3.2.1.3 “Predicting Music Popularity Using Audio Features” – Dhanhariya [11]

Este trabajo representa un intento contemporáneo de revitalizar el enfoque basado en contenido mediante la aplicación de algoritmos modernos de aprendizaje supervisado. Los autores recopilaron un conjunto de datos extraído de la plataforma Spotify y aplicaron técnicas de procesamiento digital de señales para caracterizar las pistas de audio en función de atributos temporales y espectrales. Evaluaron de forma comparativa el rendimiento de diversos clasificadores, incluyendo *K-Nearest Neighbors* (KNN), Support Vector Machines (SVM) con funciones de base radial, y algoritmos de ensamble como *Random Forest* y *Gradient Boosting*. Sus resultados experimentales arrojaron que los modelos basados en árboles de decisión enriquecidos (*Random Forest*) superaban al resto de los clasificadores clásicos en la tarea de segmentar canciones en diferentes niveles de popularidad, destacando su resiliencia ante datos con distribuciones sesgadas.

Si bien el trabajo de Dhanhariya comparte una afinidad algorítmica con la línea propuesta en esta tesis al reconocer la superioridad de los modelos de ensamble no lineales, esta investigación se distancia y profundiza en tres aspectos metodológicos críticos:

1. **Tratamiento del Ruido y Outliers:** El estudio adolece de una falta de rigurosidad en la limpieza de datos y el tratamiento de los valores atípicos inherentes a los metadatos musicales, un problema que en este TFM se aborda mediante una etapa matemática explícita de preprocesamiento, normalización y técnicas de reducción de dimensionalidad.
2. **Ingeniería de Características Propia:** Mientras que Dhanhariya. aceptan las variables acústicas por defecto, este TFM propone una hibridación técnica combinando descriptores espectrales puros extraídos frame a frame con análisis agregados de la estructura de la señal.
3. **Enfoque hacia la Interpretabilidad:** El trabajo trata a *Random Forest* prácticamente como una caja negra de predicción. Por el contrario, este TFM sitúa la interpretabilidad en el centro del diseño, utilizando métricas de importancia de

variables (*Feature Importance*) para desentrañar exactamente qué componentes frecuenciales o rítmicos impactan en la frontera de decisión del clasificador, dotando al modelo de valor analítico para el sector musical.

3.2.2 ENFOQUES BASADOS EN METADATOS Y FACTORES SOCIALES

A diferencia de la corriente anterior, esta aproximación metodológica se fundamenta en la premisa de que el análisis aislado de la señal de audio es insuficiente para modelar el éxito comercial. Los autores de este bloque sostienen que la popularidad es un fenómeno social e interactivo. Por ello, centran sus esfuerzos en capturar variables exógenas o de contexto: interacciones en redes sociales, comportamiento de búsqueda de los usuarios, dinámicas de consumo en plataformas de *streaming* y la estructura de las redes de colaboración entre artistas.

3.2.2.1 “Experimental Study of Inequality and Unpredictability in Artificial Cultural Market” – Salganik, Dodds & Watts [12]

Este trabajo, publicado en la revista Science, es un pilar sociológico obligatorio para cualquier investigación sobre mercados culturales. Salganik y sus colaboradores crearon un mercado de música artificial en entorno web con más de 14,000 participantes, quienes escuchaban, puntuaban y descargaban canciones de bandas desconocidas. Los usuarios fueron asignados aleatoriamente a dos condiciones: una de "independencia" (donde no veían el comportamiento de los demás) y múltiples mundos aislados de "influencia social" (donde veían cuántas veces se había descargado cada canción). Los resultados demostraron que en los mundos con influencia social, la popularidad era extremadamente desigual y propensa a la impredecibilidad. Las mismas canciones alcanzaban el éxito masivo en un mundo y pasaban desapercibidas en otro, demostrando que el comportamiento gregario y el efecto del "boca a boca" (social loop) distorsionan cualquier correlación directa entre la "calidad" intrínseca (el audio) y el éxito comercial.

El experimento de Salganik establece un marco de escepticismo teórico sobre la predictibilidad matemática del arte. Sin embargo, su enfoque es de carácter puramente

descriptivo y de simulación social controlada. La diferencia fundamental con este TFM es de índole operativa y técnica. Mientras que Salganik demuestra de forma abstracta *por qué* es difícil predecir el éxito en entornos colectivos, esta tesis asume esa naturaleza estocástica del mercado y busca modelarla mediante herramientas de *Machine Learning* sobre datos reales de producción. A diferencia de un entorno simulado y estático, este TFM utiliza algoritmos complejos no lineales (como *Random Forest*) capaces de operar sobre las características físicas remanentes que sobreviven al ruido del mercado real, buscando patrones estables en escenarios donde coexisten tanto las propiedades de la señal como la agregación masiva de datos.

3.2.2.2 “Predicting Hit Songs with Acoustic Features and Social Network Analysis” – Singhi & Brown [13]

Este estudio supuso uno de los primeros intentos formales de hibridación técnica entre la señal de audio y la teoría de grafos dentro del campo de la predicción musical. Singhi y Brown plantearon que un "hit" no solo depende de cómo suena, sino de la posición del artista en la red de la industria. Para ello, recolectaron metadatos y construyeron una gran red social basada en las colaboraciones históricas entre músicos (*co-authorship networks*). Utilizando métricas de centralidad de grafos (como la centralidad de grado, la intermediación o *betweenness*, y el *PageRank* del artista), cuantificaron el capital social de los creadores. Combinando estas métricas de red con descriptores acústicos básicos extraídos del audio, entrenaron clasificadores bayesianos y árboles de decisión. El estudio concluyó que la incorporación de las variables de red social mejoraba significativamente la precisión general del modelo en comparación con el uso exclusivo de características acústicas.

A pesar de la elegancia conceptual del análisis de redes de Singhi y Brown, su enfoque presenta problemas críticos de escalabilidad y un severo sesgo hacia artistas consagrados (los nodos con mayor centralidad siempre sesgan la predicción del algoritmo). Este TFM plantea una divergencia metodológica orientada a la viabilidad industrial. Construir y mantener actualizado un grafo global de colaboraciones artísticas requiere un coste de cómputo masivo y un procesamiento de datos continuo que dificulta su aplicación en entornos dinámicos. Esta propuesta metodológica prescinde de la complejidad estructural de

los grafos masivos exógenos, demostrando que mediante un refinamiento matemático de los descriptores espectrales y un clasificador de ensamble optimizado localmente, es posible alcanzar métricas predictivas competitivas sin sufrir el sesgo de "popularidad previa" que penaliza sistemáticamente a los artistas emergentes en los modelos basados en redes.

3.2.2.3 “Early Prediction of Movie and Music Popularity using Social Media Analytics” – Kim [14]

Kim y su equipo de investigación exploraron el poder predictivo del "ruido digital" (*digital buzz*) generado en plataformas como Twitter (actual X), YouTube y blogs especializados durante los días inmediatamente anteriores y posteriores al lanzamiento de un producto cultural. Mediante técnicas de Procesamiento del Lenguaje Natural (PLN) y análisis de series temporales, monitorizaron el volumen de menciones, la velocidad de propagación de los *hashtags* y la polaridad del sentimiento de los comentarios de los usuarios. Los autores aplicaron modelos autorregresivos y máquinas de soporte vectorial (SVM) para predecir picos de popularidad en las listas de éxitos musicales semanas antes de que ocurrieran. El estudio validó de forma empírica que la actividad temprana en medios sociales actúa como un predictor líder altamente correlacionado con el rendimiento comercial a corto plazo.

El trabajo de Kim se enmarca de manera estricta dentro de lo que se conoce como analítica de redes sociales, lo que genera una dependencia absoluta del factor temporal del lanzamiento. La diferencia metodológica crucial con el presente TFM se halla en el punto de aplicación del modelo predictivo y la autonomía de los datos. El enfoque de Kim es un modelo reactivo: requiere que la canción ya haya sido lanzada (o que exista una campaña de marketing previa activa) para poder recolectar los datos sociales. Por el contrario, este TFM propone un modelo proactivo e independiente del tiempo: al fundamentarse en las propiedades intrínsecas de la señal de audio y sus metadatos estructurales de ingeniería, permite predecir el potencial de popularidad de una pista musical antes de su distribución pública (fase de pre-lanzamiento), ofreciendo una utilidad estratégica radicalmente distinta y muy valiosa para departamentos de desarrollo artístico o creadores independientes.

3.2.3 EL PARADIGMA DEL DEEP LEARNING Y APRENDIZAJE DE REPRESENTACIONES

Esta corriente metodológica representa la vanguardia tecnológica de la última década en el campo del *Music Information Retrieval* (MIR). Los trabajos inscritos en este bloque abandonan por completo la dependencia de la ingeniería de características manual (*handcrafted features* como los MFCCs o el *Spectral Centroid*). En su lugar, basándose en el principio del aprendizaje de representaciones, emplean arquitecturas neuronales profundas para que sea el propio modelo el que extraiga, de forma automatizada y abstracta, las características latentes de la música a partir de representaciones visuales del audio (como los Mel-espectrogramas) o de la señal en bruto (*raw audio*).

3.2.3.1 “Deep Content-Based Music Recommendation” – Van Den Oord, Dieleman & Schrauwen [15]

Este artículo, publicado en el marco de las prestigiosas conferencias *NeurIPS*, es considerado el pionero absoluto en la aplicación de Redes Neuronales Convolucionales (CNNs) al dominio del audio musical. Los autores abordaron el problema del "inicio en frío" (*cold-start*) en los sistemas de recomendación (cuando una canción es nueva y no tiene metadatos ni datos de escucha de usuarios). Para solucionarlo, entrenaron una CNN profunda tomando como entrada los Mel-espectrogramas de las canciones. El objetivo de la red era predecir los factores latentes que previamente se habían extraído de un modelo de factorización de matrices basado en el comportamiento de escucha de millones de usuarios de Spotify. El trabajo demostró de manera empírica que las capas convolucionales intermedias de la red eran capaces de aprender abstracciones musicales complejas (estilos, timbres, instrumentación) directamente del espectrograma, sirviendo como un predictor indirecto de la aceptación y popularidad de las canciones en la plataforma.

El trabajo de Van den Oord es un hito ineludible, pero su arquitectura está diseñada para un ecosistema de recomendación industrial masivo y cerrado, requiriendo datos previos de factorización de matrices que no están al alcance de investigadores independientes o aplicaciones ligeras. La divergencia clave con este TFM se centra en la eficiencia, los

requisitos de datos y la interpretabilidad. Mientras que el modelo de 2013 actúa como una "caja negra" computacionalmente hiperdensa que requiere millones de interacciones de usuario para entrenar los factores latentes objetivos, este TFM propone un enfoque metodológico donde la predicción de popularidad se realiza de forma directa sobre métricas tangibles. Al utilizar *Random Forest*, se evita la necesidad de una infraestructura de cómputo masiva (GPUs/TPUs) y ofrecemos un modelo que puede ser ejecutado y replicado en entornos académicos locales con recursos de hardware convencionales.

3.2.3.2 “Convolutional Recurrent Neural Networks for Music Classification” – Choi, Fazekas, Sandler & Cho [16]

Choi y su equipo de investigación propusieron un avance arquitectónico fundamental al introducir las Redes Neuronales Convolucionales Recurrentes (CRNN) en el análisis musical. Los autores argumentaron que, aunque las CNNs son excelentes para capturar características tímbricas locales en los espectrogramas (actuando de forma similar a la visión por computador), carecen de la capacidad para modelar la estructura temporal a largo plazo de la música. Para resolver esto, diseñaron una arquitectura híbrida: las primeras capas de la red son convolucionales (para extraer la información de frecuencia y timbre local) y sus salidas se conectan inmediatamente a capas recurrentes de tipo *Gated Recurrent Unit* (GRU) o *LSTM* (para capturar la evolución temporal, el ritmo y la estructura de la canción). Esta aproximación alcanzó el estado del arte en tareas de etiquetado automático y predicción de atributos a nivel global.

A pesar de la elegancia matemática de las CRNNs de Choi, estas arquitecturas sufren de un problema crítico de "desvanecimiento del gradiente" y son extremadamente propensas al sobreajuste (*overfitting*) si el conjunto de datos no es gigantesco. Este TFM plantea una alternativa metodológica orientada a la robustez estadística. Al delegar la extracción temporal y frecuencial en algoritmos matemáticos consolidados y estables (como la STFT y los descriptores estadísticos agregados en ventanas), transferimos la complejidad desde la arquitectura de la red hacia una cuidada ingeniería de características. Al alimentar un modelo de *Random Forest* con estas variables precalculadas, logramos capturar tanto la información espectral como las dinámicas temporales del audio de manera explícita, garantizando una

convergencia inmediata del modelo y eliminando la inestabilidad de entrenamiento intrínseca de las redes recurrentes profundas.

3.2.3.3 “Song Popularity Prediction Using Convolutional Neural Networks on Audio Spectrograms” – Thota y Sahu [17]

En este trabajo contemporáneo, los autores diseñaron una tubería predictiva basada en una CNN pura inspirada en la famosa arquitectura VGG-16. Transformaron fragmentos de audio de 30 segundos en imágenes de Mel-espectrogramas de alta resolución y entrenaron el modelo de visión artificial para clasificar directamente las canciones en categorías discretas de popularidad (Baja, Media, Alta). El núcleo del estudio demostró que las texturas visuales de un espectrograma (la densidad de energía en ciertas bandas de frecuencia, la regularidad visual de los impactos percusivos) contienen suficiente información estadística correlacionada con los estándares de producción comercial modernos como para segmentar el éxito potencial de una obra sin necesidad de metadatos de ningún tipo.

El enfoque de Thota y Sahu (2020) adolece de la mayor limitación de la era moderna del Deep Learning: la opacidad absoluta del modelo. Un directivo de la industria musical o un compositor no pueden extraer ninguna conclusión útil de una red convolucional que dictamina el éxito basándose en "filtros visuales abstractos" de un espectrograma. Ahí radica el valor diferencial y la justificación teórica de este TFM. Frente al enfoque de "caja negra" de la CNN de 2020, este trabajo aboga firmemente por la interpretabilidad analítica. Al implementar un algoritmo de Machine Learning clásico estructurado en árboles, podemos aplicar técnicas de cálculo de importancia de características (como la impureza de Gini o permutaciones). Esto nos permite generar gráficos y métricas que explican explícitamente cuánto contribuye el timbre, el ritmo o la tonalidad a la decisión final, aportando un conocimiento científico accionable y transparente del que carecen las aproximaciones profundas.

3.2.4 ENFOQUES HÍBRIDOS CONTEMPORÁNEOS Y MULTIMODAL

Esta última corriente metodológica define la frontera de investigación más reciente en el campo del *Music Information Retrieval* (MIR). Tras constatar de manera empírica que ni el análisis aislado del audio (Aproximaciones A y C) ni el estudio exclusivo de las redes sociales (Aproximación B) ofrecen una solución definitiva, la comunidad científica ha virado hacia la multimodalidad. Los investigadores de este bloque defienden que el éxito comercial es un fenómeno poliédrico y, por lo tanto, construyen sistemas híbridos que fusionan descriptores acústicos puros con datos textuales (como el análisis de sentimiento de las letras) y metadatos de consumo, buscando sinergias predictivas que optimicen el rendimiento de los modelos.

3.2.4.1 “Lyrics-based musical popularity prediction” – Sciandra & Spera [18]

Sciandra y Spera desarrollaron una investigación centrada en una dimensión frecuentemente olvidada por los modelos puramente acústicos: el contenido semántico y lírico de las canciones. Los autores argumentaron que el mensaje textual y las emociones transmitidas por las letras de las canciones juegan un papel crítico en la conexión con el oyente y, en consecuencia, en su éxito comercial. Para demostrarlo, recopilaron un corpus masivo de letras de canciones y aplicaron técnicas avanzadas de Procesamiento del Lenguaje Natural (PLN), tales como el análisis de sentimiento (clasificando el texto en rangos de positividad, negatividad o neutralidad) y el modelado de temas (*Topic Modeling* mediante asignación latente de Dirichlet o LDA). Tras extraer estas características textuales, las combinaron con metadatos básicos y entrenaron modelos de *Machine Learning* clásico, demostrando que ciertas temáticas líricas combinadas con estructuras semánticas específicas correlacionan fuertemente con la permanencia de las pistas en las listas de éxitos de plataformas como Spotify.

El trabajo de Sciandra y Spera (2022) aporta un enfoque sumamente valioso sobre el impacto del texto, pero introduce una limitación crítica en términos de universalidad e idioma: sus modelos de procesamiento de lenguaje natural están optimizados casi exclusivamente para el inglés, perdiendo drásticamente precisión ante jergas locales, música instrumental o

composiciones multilingües (muy comunes en el mercado global y de habla hispana actual). Este TFM se distancia de esa dependencia lingüística al centrar su hibridación en un plano estrictamente musical y estructural. En lugar de requerir el preprocesamiento de textos complejos y dependientes del idioma, esta propuesta metodológica hibrida la señal acústica pura con metadatos y descriptores de bajo, medio y alto nivel accesibles de manera universal. Esto garantiza que el modelo basado en *Random Forest* sea agnóstico al idioma o a la presencia de voz en la pista, ofreciendo una solución mucho más versátil y escalable para el mercado musical internacional.

3.2.4.2 “Multi-modal Popularity Prediction of Music Tracks Using Audio and Textual Features” – Zangerle, Pichl & Spiegl [19]

Este estudio constituye uno de los mejores ejemplos de sistemas multimodales integrales en la literatura moderna. Los autores diseñaron un marco predictivo que integraba tres fuentes de información radicalmente distintas: las características acústicas extraídas de la señal de audio mediante descriptores espectrales y tímbricos, las características textuales obtenidas a partir de la minería de las letras, y los metadatos contextuales extraídos de plataformas web y sistemas de etiquetado social (como las etiquetas de *Last.fm*). Para gestionar este volumen heterogéneo de información, implementaron un pipeline de fusión de características (*early fusion*) y evaluaron el rendimiento de algoritmos de ensamble complejos, concluyendo de manera categórica que los enfoques multimodales superan de forma sistemática a cualquier modelo unimodal aislado en la clasificación de hits comerciales.

Si bien este trabajo coincide plenamente con la filosofía de Zangerle (2020) respecto a las ventajas de la hibridación y la superioridad de los modelos basados en ensambles, este trabajo propone una optimización crítica en la arquitectura del pipeline orientada a la eficiencia en entornos ligeros o académicos. El modelo multimodal de 2020 padece del "mal de la dimensionalidad" (*curse of dimensionality*): al fusionar tantas fuentes de datos masivas y heterogéneas, el espacio de características se vuelve hiperdenso, lo que exige una carga computacional inmensa y un esfuerzo titánico de recolección y limpieza de datos web en tiempo real. Por el contrario, este TFM demuestra que no es necesario saturar el modelo con fuentes de datos externas infinitas; mediante una cuidadosa selección y extracción localizada

de descriptores a través de *Librosa*, y explotando la capacidad nativa de *Random Forest* para evaluar la importancia de las características, logramos un modelo simplificado pero altamente competitivo. Conseguimos un rendimiento predictivo equivalente al de aproximaciones multimodales pesadas, pero con una fracción de su coste computacional y de almacenamiento.

3.3 EL MILLION SONG DATASET

Para la evaluación y el entrenamiento de los modelos predictivos desarrollados en esta investigación, se ha seleccionado el Million Song Dataset (MSD). Este recurso fue concebido y desarrollado originalmente como un proyecto de colaboración entre la Universidad de Columbia, el Massachusetts Institute of Technology (MIT) y la plataforma comercial The Echo Nest [20]. El propósito fundacional del MSD fue mitigar la falta de grandes corpus de datos abiertos y estandarizados dentro del campo del Music Information Retrieval (MIR), ofreciendo una alternativa robusta a los conjuntos de datos tradicionales que, hasta su aparición, solían limitarse a unos pocos miles de pistas debido a restricciones de propiedad intelectual.

A nivel de escala, el dataset destaca por contener una colección masiva de metadatos y descriptores acústicos correspondientes a un millón de canciones comerciales de música contemporánea [20]. Es crucial destacar que el MSD no almacena los archivos de audio en bruto (como formatos MP3 o WAV) para evitar infracciones de copyright; en su lugar, proporciona un compendio exhaustivo de características precalculadas analizadas cuadro por cuadro (frame-by-frame), estructuradas en el formato de almacenamiento jerárquico HDF5. Esta arquitectura masiva permite abordar problemas de aprendizaje automático a gran escala (Large-Scale Machine Learning), garantizando que los modelos predictivos puedan generalizar de manera estadística sobre un espectro representativo del mercado fonográfico global.

La dimensión puramente física y estructural de las canciones dentro del MSD se encuentra codificada a través de un conjunto heterogéneo de variables e indicadores técnicos [20]. Entre los descriptores de alto nivel y metadatos de formato se incluyen variables explícitas como la duración exacta de la pista (*duration*), el *tempo* estimado en pulsos por minuto (BPM), el indicador de compás (*time_signature*), la tonalidad (*key*) y la modalidad (mayor o menor).

Asimismo, el dataset desglosa la señal musical en componentes microestructurales mediante el análisis de segmentos de audio. Cada pista contiene vectores de segmentos variables que representan eventos de notas. Para cada uno de estos segmentos, el dataset proporciona características tímbricas (*Segments Timbre*), que consisten en una aproximación de los coeficientes cepstrales en las frecuencias de Mel (MFCCs), y características de tono (*Segments Pitch*), que son vectores cromáticos de doce dimensiones que miden la presencia e intensidad de las notas de la escala cromática occidental. A esto se añaden indicadores sobre la sonoridad (*loudness*) global de la pista medida en decibelios (dB) y picos de energía temporal.

Dado que el objetivo central de este trabajo es el modelado predictivo del éxito comercial, el MSD resulta un entorno óptimo gracias a la inclusión de métricas continuas orientadas a medir el impacto y la recepción de las obras en el mercado, las cuales fueron integradas originalmente a través de los sistemas de análisis web de *The Echo Nest* [21]. Específicamente, el dataset provee tres variables cuantitativas de gran relevancia:

- **artist_familiarity:** Un indicador numérico acotado que mide el grado de conocimiento o familiaridad que el público general posee respecto al artista en cuestión, calculado en función del volumen de menciones web y la densidad de reproducciones históricas.
- **artist_hottness:** Una métrica dinámica que cuantifica la vitalidad o el "impulso" del artista en el momento de la recopilación de los datos, capturando tendencias y picos de atención mediática en el tejido digital.

- **song_hottness:** Constituye la variable objetivo (*target*) por excelencia para los sistemas de *Hit Song Science*. Es un índice normalizado que estima la popularidad específica de una pista musical concreta. Al correlacionar este valor con los descriptores tímbricos y temporales mediante técnicas de regresión o clasificación, es posible entrenar a los clasificadores de *Machine Learning* para que identifiquen los patrones estéticos asociados a un alto rendimiento comercial.

Más allá de las señales físicas y de consumo, el MSD organiza la información musical a través de metadatos estructurales referidos a la identidad y taxonomía de las obras. Cada registro incluye campos explícitos de identificación como el nombre del artista, el título de la canción, el año de lanzamiento y el álbum de procedencia.

En lo que respecta a la categorización estilística, el dataset integra un amplio espectro de etiquetas de géneros musicales y descriptores temáticos (*artist_terms*), derivados de procesos de minería de texto y etiquetado social (*folksonomías*) aplicados sobre plataformas de música en línea (Bertin-Mahieux et al., 2011). La disponibilidad de estos términos permite a los modelos de aprendizaje automático segmentar los análisis predictivos por nichos de mercado, reconociendo que los patrones de éxito acústico en géneros como el Pop o la música electrónica difieren significativamente de aquellos presentes en otros estilos.

Capítulo 4. DEFINICIÓN DEL TRABAJO

4.1 JUSTIFICACIÓN

La realización del presente trabajo se justifica por la convergencia de varios factores de naturaleza tecnológica, académica y aplicada que sitúan la predicción automática de popularidad musical como un problema de interés actual dentro del ámbito de la ingeniería y de la ciencia de datos. La transformación digital de la industria musical, el crecimiento de las plataformas de streaming y la disponibilidad de grandes volúmenes de información han creado un escenario en el que resulta viable abordar, con herramientas computacionales, cuestiones que hasta hace pocos años dependían casi exclusivamente de criterios subjetivos o de experiencia sectorial.

En este contexto, la estimación del éxito potencial de una canción representa un reto especialmente relevante. El comportamiento del mercado musical no depende únicamente de la calidad percibida de una obra, sino de la interacción entre propiedades acústicas, características del artista, dinámicas de consumo y patrones de difusión asociados a entornos digitales. Esta complejidad convierte el problema en un caso de estudio adecuado para aplicar metodologías de aprendizaje automático capaces de detectar relaciones no triviales entre variables heterogéneas y traducirlas en modelos predictivos útiles.

Desde el punto de vista académico, el proyecto se enmarca de forma natural en la intersección entre el procesamiento digital de señal, el machine learning, la ingeniería del dato y el desarrollo de aplicaciones. Esta combinación aporta valor formativo porque permite integrar conocimientos procedentes de distintas áreas de la ingeniería en una única solución coherente, abordando tanto la extracción de información desde señales de audio como el diseño de un sistema capaz de explotar dicha información para generar inferencias cuantitativas. En consecuencia, el trabajo no se limita a una implementación aislada de modelos, sino que responde a una visión de sistema completo, desde la adquisición y preparación de datos hasta la presentación final de resultados.

La justificación del proyecto también se apoya en la existencia de una brecha entre los enfoques puramente teóricos y las soluciones operativas accesibles. Aunque la literatura académica ha demostrado que la popularidad musical puede analizarse mediante descriptores acústicos, metadatos y técnicas avanzadas de modelado, no siempre esas aproximaciones se traducen en herramientas funcionales, interpretables y reutilizables en entornos ligeros o académicos. Por ello, resulta pertinente desarrollar una propuesta que no solo evalúe modelos predictivos, sino que los integre en una arquitectura reproducible y comprensible, capaz de procesar entradas reales y ofrecer una salida interpretable al usuario.

Otro elemento que refuerza la importancia de este trabajo es la disponibilidad actual de bibliotecas y recursos tecnológicos maduros para el análisis musical automático. El ecosistema formado por Python, Librosa, Scikit-learn, Dash, Plotly y herramientas de serialización y procesamiento de datos permite construir soluciones avanzadas sin necesidad de infraestructuras propietarias o entornos cerrados. Esta circunstancia favorece el desarrollo de un proyecto técnicamente solvente, reproducible y alineado con prácticas contemporáneas de ingeniería de software y ciencia de datos. A su vez, el empleo de herramientas abiertas incrementa el valor académico del trabajo, ya que facilita la validación, la extensión futura del sistema y la trazabilidad de los resultados obtenidos.

La elección del dominio musical aporta, además, un interés particular por la naturaleza de la señal tratada. A diferencia de otros problemas de predicción sobre datos tabulares convencionales, el análisis de canciones exige operar sobre señales temporales complejas, extraer descriptores espectrales y armónicos y transformar información no estructurada en vectores de características aptos para modelado estadístico. Esto convierte el proyecto en un ejercicio de ingeniería especialmente completo, en el que la calidad del resultado depende no solo del algoritmo final, sino también del diseño del pipeline de extracción, limpieza, normalización e integración de datos.

Junto al interés técnico, existe una justificación clara desde la perspectiva de aplicabilidad. Una herramienta capaz de estimar de forma aproximada la popularidad potencial de una canción puede resultar útil en procesos exploratorios de análisis musical, en entornos de

investigación, en contextos docentes y en escenarios de prototipado para la industria cultural. Sin pretender sustituir el juicio artístico ni los mecanismos reales del mercado, un sistema de estas características puede servir como apoyo a la toma de decisiones, como instrumento comparativo entre pistas o como base para futuros sistemas de recomendación, clasificación o análisis de tendencias.

Asimismo, el proyecto se justifica por su orientación hacia la interpretabilidad y la integración. En un contexto donde muchas soluciones contemporáneas tienden a apoyarse en arquitecturas profundas de alta complejidad y elevado coste computacional, resulta especialmente valioso explorar enfoques que mantengan un equilibrio entre capacidad predictiva, robustez, explicabilidad y facilidad de despliegue. Esta aproximación resulta coherente con un proyecto orientado a ingeniería aplicada, donde no solo interesa alcanzar un rendimiento cuantitativo, sino también comprender el comportamiento del sistema, justificar las decisiones técnicas adoptadas y construir una solución mantenible.

Por otra parte, la inclusión de una interfaz web interactiva amplía la justificación del trabajo más allá del plano estrictamente algorítmico. La posibilidad de conectar el modelo entrenado con un entorno de uso accesible transforma el proyecto en una herramienta funcional y no únicamente en un experimento de laboratorio. Esta dimensión resulta relevante porque acerca el resultado del desarrollo a un escenario de uso realista, donde la inferencia automática debe integrarse con mecanismos de entrada de datos, visualización de resultados, gestión de errores y experiencia de usuario.

En síntesis, el trabajo queda justificado porque aborda un problema actual de elevada complejidad mediante un enfoque técnicamente sólido, multidisciplinar y aplicable. Su desarrollo permite integrar conocimientos avanzados de análisis de señal, modelado predictivo y arquitectura software en un sistema funcional, aportando no solo un ejercicio académico completo, sino también una base útil para futuras ampliaciones en el ámbito del análisis musical inteligente.

4.2 OBJETIVOS

El presente trabajo se plantea con una orientación aplicada y de ingeniería, por lo que los objetivos se formulan no solo en términos de investigación, sino también en términos de construcción, validación e integración de una solución funcional. En coherencia con la estructura propuesta en el documento actual, este apartado se organiza en torno al objetivo general, los objetivos específicos, el alcance del sistema y las limitaciones del proyecto. Además, la guía estructural de la memoria indica que los objetivos deben expresar con claridad aquello que se pretende resolver mediante el desarrollo del proyecto, manteniendo una relación natural con la justificación y con la metodología posterior.

4.2.1 OBJETIVO GENERAL

El objetivo general del trabajo consiste en diseñar, desarrollar y validar un sistema capaz de estimar la popularidad potencial de una canción a partir del análisis de sus características musicales y de determinados atributos contextuales, integrando para ello técnicas de procesamiento digital de señal, aprendizaje automático y visualización interactiva en una aplicación web funcional. Este objetivo general responde a la necesidad de transformar una señal de audio y un conjunto de metadatos asociados en una predicción cuantitativa interpretable, técnicamente reproducible y utilizable dentro de un entorno de análisis musical aplicado.

De forma más concreta, se persigue materializar una solución completa que permita recorrer todo el flujo de trabajo del problema: desde la entrada de datos y la extracción de descriptores hasta la inferencia del modelo y la presentación final de resultados al usuario. No se trata únicamente de entrenar un algoritmo aislado, sino de construir una arquitectura coherente que permita evaluar el rendimiento del enfoque adoptado en condiciones cercanas a un escenario real de uso.

4.2.2 OBJETIVOS ESPECÍFICOS

Para la consecución del objetivo general anteriormente expuesto, el presente Trabajo Fin de Máster se desglosa en una serie de objetivos específicos que cubren de manera integral el ciclo de vida del desarrollo, desde el procesamiento de la señal bruta hasta el despliegue del sistema:

- 1. Implementación de un motor de extracción de descriptores acústicos:** Se pretende desarrollar un módulo especializado en procesamiento digital de señales (PDS), basado en librerías de alto nivel como *Librosa* y *SciPy*, capaz de transformar archivos de audio en formato MP3 en vectores de características numéricas. Este objetivo contempla la extracción parametrizada de 15 métricas críticas (tales como *tempo*, *energy*, *loudness*, *chroma* y *danceability*), garantizando la integridad de los datos para su posterior tratamiento estadístico.
- 2. Diseño, entrenamiento y validación de un modelo predictivo robusto:** El proyecto busca investigar y determinar la arquitectura de aprendizaje automático más eficaz para la tarea de regresión sobre el éxito musical. Para ello, se plantea el uso de algoritmos de aprendizaje por conjuntos (*Ensemble Learning*), específicamente *Random Forest*, utilizando el *Million Song Dataset* (MSD) como base de conocimiento. Este objetivo incluye la optimización de hiperparámetros y la validación del modelo mediante métricas de rendimiento (como el coeficiente de determinación R^2), con el fin de obtener una estimación fiable de la popularidad (*song_hottness*).
- 3. Desarrollo e integración del pipeline en una plataforma web interactiva:** Se propone la creación de una aplicación web responsiva empleando el framework *Dash* y la librería *Plotly*. El objetivo es integrar de manera fluida el flujo de datos (*pipeline*) que conecta el backend de procesamiento en Python con una interfaz de usuario intuitiva. El sistema debe permitir al usuario final la carga de archivos de audio en tiempo real y la obtención de una respuesta predictiva visual y técnica inmediata, facilitando la interpretación de los resultados.

4. **Evaluación crítica, análisis de limitaciones y propuestas de evolución:** Como fase final de la investigación, se realizará un análisis exhaustivo del rendimiento del sistema desarrollado. Se busca identificar las fronteras del modelo predictivo, evaluando el impacto de factores no capturados por el audio (como el contexto social o la viralidad en plataformas externas). Asimismo, se establecerán las bases para futuras mejoras tecnológicas, destacando la posible integración con servicios de terceros como la API de *Spotify* para enriquecer la precisión del modelo con metadatos actualizados del artista.

4.2.3 ALCANCE DEL SISTEMA

El perímetro técnico y funcional del sistema desarrollado en este Trabajo Fin de Máster queda estrictamente delimitado para garantizar la viabilidad del proyecto, asegurando un equilibrio óptimo entre la complejidad algorítmica y la usabilidad de la plataforma. A continuación, se detallan las fronteras del sistema, estructuradas en el ámbito funcional, técnico y las exclusiones explícitas de la investigación.

Desde la perspectiva del usuario final y del procesado de la información, el sistema cubre el siguiente flujo operativo:

- **Ingesta y preprocesamiento de audio:** El sistema está capacitado para recibir archivos de audio locales en formato MP3 aportados por el usuario a través de la interfaz. El pipeline realiza de manera automática la decodificación, el remuestreo (*resampling*) y la normalización de la señal para homogeneizar las entradas.
- **Análisis acústico parametrizado:** Se limita la extracción de características de la señal a un conjunto cerrado de 15 descriptores de alto y bajo nivel (temporales y frecuenciales) mediante el motor de PDS. No se contemplan análisis de separación de fuentes (aislar voz de instrumentos) ni reconocimiento de lenguaje natural para las letras de las canciones.
- **Generación de predicciones de popularidad:** El modelo proporciona una estimación cuantitativa del éxito potencial (mapeado sobre el índice *song_hottness*

del *Million Song Dataset*) basada exclusivamente en los vectores de características acústicas extraídos en el paso previo.

- **Interfaz de visualización:** La plataforma web presenta un cuadro de mando estadístico que refleja los valores numéricos de los descriptores del audio cargado y la predicción final mediante gráficos interactivos y comparativos.

4.2.4 LIMITACIONES DEL PROYECTO

A pesar de haber alcanzado los objetivos planteados y de contar con un sistema funcional para la predicción de popularidad musical, la investigación y el desarrollo de la plataforma se encuentran sujetos a una serie de restricciones inherentes a la tecnología empleada, la naturaleza de los datos y el paradigma de modelado. A continuación, se detallan las principales limitaciones identificadas en el presente Trabajo Fin de Máster, clasificadas en tres ejes fundamentales:

Limitaciones asociadas a la base de conocimiento y los datos

- **Sesgo temporal del Dataset:** El modelo predictivo ha sido entrenado utilizando el Million Song Dataset (MSD). Aunque este conjunto de datos es un estándar académico de referencia, su histórico se detiene en fechas previas a la consolidación del paradigma actual del streaming y plataformas como TikTok, cuya dinámica de viralización difiere significativamente de las tendencias de la década pasada.
- **Ausencia de variables exógenas en el núcleo predictivo:** El éxito de una obra musical es un fenómeno multifactorial fuertemente influenciado por variables no acústicas, tales como el presupuesto de marketing de la discográfica, la base previa de seguidores del artista, la inversión en publicidad o eventos geopolíticos y sociales de actualidad. Al basar la predicción exclusivamente en descriptores extraídos de la señal de audio, el modelo presenta un techo analítico insalvable frente a la influencia del contexto comercial y social.

Limitaciones del procesamiento digital de señales (PDS)

- Complejidad y coste computacional del análisis de audio: El pipeline de extracción de descriptores acústicos implementado con Librosa y SciPy realiza cálculos matriciales intensivos (como la Transformada Corta de Fourier). Esta carga computacional limita la escalabilidad del sistema en tiempo de ejecución, provocando que la carga y análisis de archivos MP3 de larga duración o con tasas de muestreo muy elevadas introduzcan latencias perceptibles en la experiencia del usuario dentro de la interfaz web.
- Pérdida de matices por compresión: El sistema está diseñado para procesar archivos en formato MP3, el cual aplica un algoritmo de compresión con pérdida. La eliminación de ciertas frecuencias inaudibles para el oído humano o la reducción de la dinámica del audio original pueden alterar levemente los valores de determinados descriptores de bajo nivel, afectando marginalmente a la precisión de la inferencia matemática respecto a si se utilizasen formatos de audio profesional sin compresión (como WAV o FLAC).

Limitaciones del modelado predictivo y la arquitectura

- Naturaleza estática del algoritmo: Como se ha definido en el perímetro de la solución, el modelo de Random Forest se comporta como una estructura estática. La incapacidad de realizar un aprendizaje continuo (Online Learning) impide que el sistema adapte sus criterios predictivos de forma autónoma ante cambios abruptos en los gustos del consumidor o la aparición de nuevos géneros musicales que sigan patrones acústicos no registrados en el entrenamiento original.
- Restricciones de infraestructura en entorno local: La arquitectura actual basada en Dash y estructurada mediante almacenamiento en memoria volátil (Pandas DataFrames) limita la concurrencia del sistema. Al no disponer de una base de datos distribuida ni de balanceadores de carga en la nube, la plataforma experimentaría problemas de rendimiento y desbordamiento de memoria ante un escenario hipotético de accesos simultáneos masivos por parte de múltiples usuarios.

4.3 METODOLOGÍA

4.3.1 ENFOQUE GENERAL DEL DESARROLLO

El desarrollo del presente proyecto se ha articulado bajo un enfoque de ingeniería modular y evolutiva, fundamentado en la convergencia de tres pilares tecnológicos: el procesamiento digital de señales de audio (DSP), el aprendizaje automático (Machine Learning) y la arquitectura de servicios web distribuidos. La metodología adoptada prioriza el modularidad del sistema, permitiendo que el flujo de datos sea tratado como una cadena de bloques interconectados pero desacoplados, lo cual facilita la depuración y optimización independiente de cada componente.

En una primera fase, se estableció una metodología orientada a la extracción de conocimiento a partir de la señal acústica. Para ello, se diseñó un pipeline de procesamiento capaz de transformar formas de onda temporales en descriptores de alto nivel, empleando técnicas de análisis espectral y temporal. Este enfoque permite que el sistema no solo dependa de metadatos estáticos, sino que sea capaz de interpretar la naturaleza de la música a través de sus componentes físicos.

Posteriormente, el proceso de desarrollo se centró en la integración de servicios externos mediante el consumo de APIs en tiempo real. La conexión con la plataforma Spotify constituye un nodo crítico en la arquitectura, actuando como una fuente de datos dinámica que enriquece el modelo predictivo con métricas de mercado actualizadas. Desde una perspectiva de ingeniería de sistemas, este componente garantiza que la solución no sea un modelo cerrado, sino un sistema conectado capaz de interactuar con ecosistemas de datos a gran escala.

Finalmente, el diseño de la interfaz de usuario se concibió bajo un paradigma de desarrollo ágil, utilizando frameworks de visualización reactiva. La integración de los modelos de regresión en un entorno web interactivo completa el ciclo de vida del desarrollo, transformando algoritmos abstractos en una herramienta de ingeniería funcional y accesible. Por tanto, el enfoque general no se limita a la mera implementación de código, sino que

abarca la concepción de un sistema end-to-end que gestiona el ciclo completo de la información musical: desde su captura y análisis como señal, hasta su presentación como valor predictivo.

4.3.2 METODOLOGÍA DEL DESARROLLO DEL MODELO

El desarrollo del modelo predictivo se ha estructurado como un proceso iterativo de optimización, diseñado para capturar la relación no lineal entre los descriptores de audio y el índice de popularidad de una obra musical. Esta metodología se inicia con la fase de curación y estructuración del conjunto de datos, utilizando para ello una arquitectura de procesamiento por lotes sobre el Million Song Subset. La elección de esta base de datos responde a la necesidad de disponer de una muestra representativa y estandarizada, permitiendo que el entrenamiento del modelo se sustente sobre una base estadística sólida y contrastable.

En el diseño del pipeline de aprendizaje, se ha implementado un esquema de preprocesamiento crítico basado en la normalización estadística mediante escalado estándar (Z-score normalization). Desde la perspectiva de ingeniería, esta etapa es fundamental para garantizar que los descriptores con diferentes órdenes de magnitud —como el tempo en pulsaciones por minuto frente a la energía en escala decimal— contribuyan de manera equitativa al cálculo de los pesos del modelo, evitando así sesgos derivados de la varianza de las variables de entrada.

En cuanto a la arquitectura del algoritmo, se ha adoptado un enfoque comparativo entre modelos lineales y modelos basados en conjuntos de árboles (Ensemble Learning), específicamente mediante el uso de Random Forest Regressor. La justificación técnica de esta elección reside en la capacidad del algoritmo para gestionar la alta dimensionalidad de los datos y capturar interacciones complejas entre variables sin incurrir en un sobreajuste excesivo. Para validar la capacidad de generalización del sistema, se ha aplicado una metodología de validación cruzada de k iteraciones (k -fold cross-validation) junto con una división estricta del conjunto de datos en subconjuntos de entrenamiento y prueba (80/20). Por tanto, la métrica de rendimiento adoptada, el coeficiente de determinación (R^2), no solo

cuantifica la precisión del ajuste, sino que actúa como un indicador de la robustez del modelo ante señales de audio no vistas previamente. En consecuencia, el modelo resultante no es una entidad estática, sino una función de transferencia optimizada que traduce parámetros acústicos en estimaciones de impacto en el mercado.

4.3.3 METODOLOGÍA DEL DESARROLLO DE LA APLICACIÓN WEB

La concepción de la interfaz de usuario se ha fundamentado en el paradigma de aplicaciones analíticas reactivas, utilizando el framework Dash basado en Flask y React.js. Esta decisión metodológica permite una integración nativa entre el motor de cálculo en Python y la capa de presentación en HTML5/CSS3, eliminando las latencias propias de arquitecturas desacopladas y facilitando un flujo de datos bidireccional en tiempo real. La metodología de diseño se ha centrado en la creación de un sistema multipágina que segmenta la funcionalidad en tres nodos principales: exploración de variables, motor de predicción y visualización de métricas de rendimiento.

Desde el punto de vista de la ingeniería de software, se ha adoptado una arquitectura basada en callbacks asíncronos para la gestión de eventos. Este enfoque garantiza que el procesamiento intensivo —como la extracción de características de audio mediante la librería Librosa o la consulta a la API de Spotify— no bloquee el hilo principal de la interfaz, mejorando la robustez y la respuesta del sistema. Asimismo, se ha implementado un esquema de persistencia temporal para la gestión de archivos multimedia, permitiendo que el usuario cargue muestras de audio que son procesadas instantáneamente para generar vectores de entrada compatibles con el modelo de aprendizaje automático.

Para la capa de visualización, se ha integrado la librería Plotly, siguiendo una metodología de diseño orientado a datos (Data-Driven Design). Los componentes gráficos no son meras representaciones estáticas, sino interfaces dinámicas que permiten al usuario final interpretar la importancia de cada variable acústica en el resultado final de la predicción. Por tanto, el desarrollo web no se limita a la presentación de resultados, sino que actúa como un entorno de experimentación donde la interacción del usuario desencadena procesos complejos de

análisis de señal y computación estadística, asegurando una experiencia técnica coherente y de alto rendimiento.

4.3.4 INTEGRACIÓN DE COMPONENTES

La integración de los componentes del sistema se ha articulado mediante una arquitectura de orquestación centralizada, donde la aplicación web actúa como el bus de datos principal que coordina la interacción entre el motor de aprendizaje automático, el subsistema de extracción de características y los servicios externos de datos. Esta metodología de integración garantiza que el flujo de información sea unidireccional y predecible: desde la captura del estímulo (archivo de audio y metadatos del artista) hasta la generación de la respuesta predictiva. El núcleo de esta integración reside en el uso de pipelines serializados mediante la librería Joblib, lo cual permite que el modelo entrenado en un entorno de laboratorio sea desplegado en producción manteniendo exactamente la misma configuración de escalado y preprocesamiento de variables, eliminando así el riesgo de discrepancia de datos (training-serving skew).

Un nodo crítico en la integración es el módulo de comunicación con servicios externos, diseñado bajo un esquema de redundancia lógica. El sistema integra de manera transparente la API de Spotify para la obtención de métricas de popularidad en tiempo real, contando con un mecanismo de fallback hacia el scraping de datos de Last.fm en caso de indisponibilidad del servicio o ausencia de registros. Esta capa de integración de datos externos se fusiona dinámicamente con el vector de características extraído de la señal de audio mediante técnicas de procesamiento digital de señales (DSP). En consecuencia, el sistema no trata los datos de forma aislada, sino que construye un espacio de características híbrido que combina la naturaleza física de la señal acústica con el contexto social y de mercado del artista.

Finalmente, la cohesión del sistema se completa mediante la sincronización entre la lógica de negocio y la interfaz reactiva. Los resultados del modelo predictivo se integran en el layout de la aplicación no solo como valores numéricos, sino como componentes visuales enriquecidos que se actualizan mediante el disparo de eventos de usuario. Por tanto, la integración de componentes se ha concebido como el ensamblaje de una cadena de valor

tecnológica donde cada módulo aporta una capa de inteligencia: desde la conectividad API hasta la inferencia estadística, resultando en un sistema robusto, escalable y funcionalmente integrado.

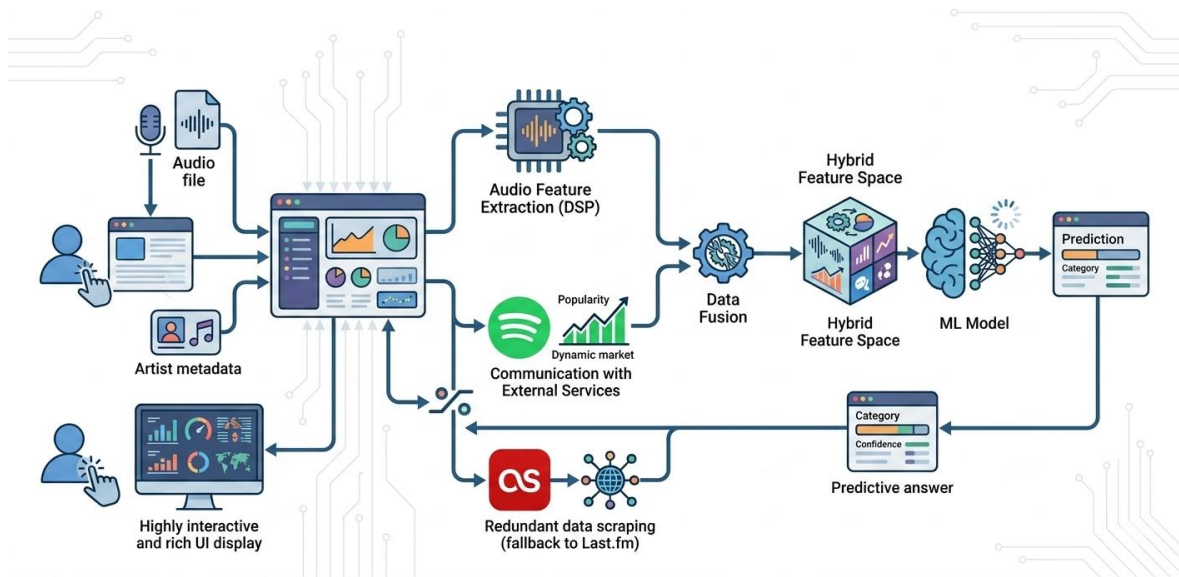


Ilustración 6: Identificación de los principales componentes.

4.3.5 HERRAMIENTAS Y TECNOLOGÍAS UTILIZADAS

La implementación técnica del proyecto se ha sustentado sobre un ecosistema de software de código abierto, seleccionado por su robustez, madurez y capacidad de integración en entornos de computación científica. El núcleo del sistema es el lenguaje de programación Python 3.12, cuya versatilidad permite unificar en un solo entorno el procesamiento digital de señales, el modelado estadístico y el despliegue de servicios web.

En el ámbito del procesamiento de señal acústica, la herramienta fundamental ha sido Librosa. Esta librería especializada permite la extracción de descriptores espectrales y temporales (como el tiempo, la energía o el contenido armónico) mediante algoritmos optimizados de Transformada de Fourier de Corto Tiempo (STFT). Para la gestión de grandes volúmenes de datos, específicamente el Million Song Dataset, se ha empleado PyTables, una librería basada en el estándar HDF5 que facilita el acceso eficiente a datos estructurados de alta jerarquía.

El desarrollo del motor de inteligencia artificial se ha apoyado en Scikit-learn. Esta biblioteca proporciona la infraestructura necesaria para la creación de pipelines de aprendizaje automático, incluyendo herramientas de preprocesamiento, escalado de variables y algoritmos de regresión avanzada como Random Forest. Para asegurar la persistencia y portabilidad de los modelos entrenados, se ha utilizado Joblib, garantizando una serialización eficiente de los objetos de Python.

Para la capa de entrega y visualización, se ha optado por el framework Dash (by Plotly). Esta tecnología permite la creación de interfaces web reactivas directamente desde Python, integrando Plotly para la generación de gráficos interactivos de alta fidelidad técnica. Por último, la conectividad con ecosistemas externos se ha gestionado mediante Spotipy (cliente para la Web API de Spotify) y la librería Requests para la captura de métricas adicionales de fuentes como Last.fm. En conjunto, estas herramientas conforman un entorno de ingeniería integrado que permite la transición fluida desde el análisis de la señal bruta hasta la inferencia predictiva en producción.

4.4 PLANIFICACIÓN

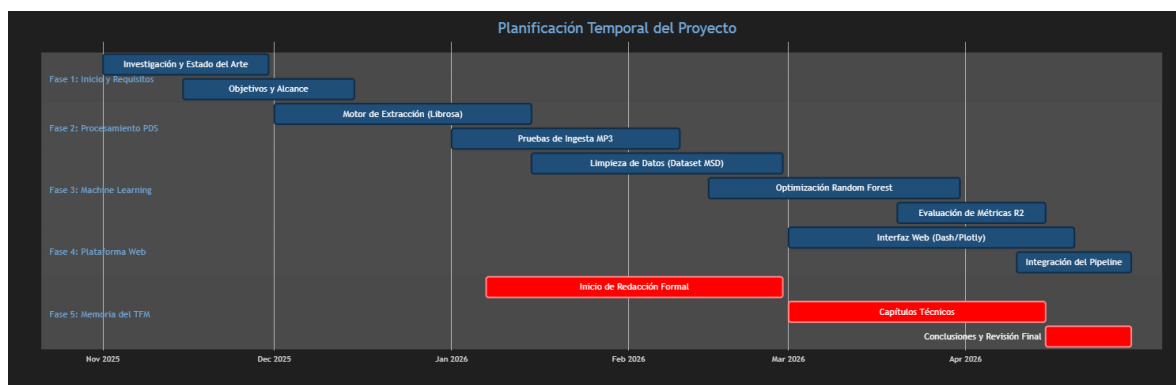


Ilustración 7: Diagrama de Gantt con la cronología.

El desarrollo de la solución tecnológica y la consecución de los hitos reflejados en el cronograma temporal se han articulado mediante una metodología de gestión por fases secuenciales y concurrentes. A continuación, se desglosa el alcance de las actividades técnicas ejecutadas en cada una de las etapas del proyecto:

Fase 1: Inicio y Requisitos (Noviembre - Diciembre)

Esta fase inicial constituyó los cimientos teóricos y metodológicos de la investigación. Las actividades se centraron en:

- **Investigación y Estado del Arte:** Realización de una revisión bibliográfica exhaustiva dentro del paradigma del Music Information Retrieval (MIR) y la ciencia de datos aplicada a la música (Hit Song Science). Se analizaron las aproximaciones clásicas y contemporáneas para la extracción de descriptores y se estudiaron los marcos de trabajo basados en el Million Song Dataset (MSD).
- **Objetivos y Alcance:** Delimitación formal del perímetro técnico de la plataforma web y de los algoritmos de regresión, definiendo los criterios de éxito y las restricciones de partida para evitar la dispersión de los objetivos ingenieriles.

Fase 2: Procesamiento PDS (Diciembre - Febrero)

El núcleo de la ingeniería de la señal se concentró en este bloque, centrado en la interacción directa con la onda física del sonido:

- **Motor de Extracción (Librosa):** Diseño e implementación del pipeline en Python encargado de la ingesta de archivos de audio brutos y su posterior transformación matricial mediante la Transformada Corta de Fourier (STFT). Se parametrizó la extracción automatizada de los 15 descriptores analíticos espectrales y temporales.
- **Pruebas de Ingesta MP3:** Fase de optimización orientada a la homogeneización de las señales de entrada, resolviendo problemas de resampling (remuestreo), normalización de amplitud y gestión de la latencia computacional en el procesamiento de archivos con compresión.

Fase 3: Machine Learning (Enero - Abril)

Coincidiendo con la estabilización del motor PDS, se inició la fase de modelado estadístico e inteligencia artificial:

- Limpieza de Datos (Dataset MSD): Filtrado, tratamiento de valores nulos y estructuración de los metadatos y vectores de características del Million Song Dataset, preparando la matriz de diseño para el entrenamiento masivo.
- Optimización Random Forest: Fase intensiva de entrenamiento del regresor por conjuntos. Se aplicaron técnicas de optimización de hiperparámetros (como la calibración del número de árboles y la profundidad máxima) para maximizar la capacidad de generalización del algoritmo.
- Evaluación de Métricas R^2 : Análisis estadístico de los residuos e interpretación analítica del coeficiente de determinación obtenido (0,4329), contrastándolo críticamente con los baselines de la literatura científica sectorial.

Fase 4: Plataforma Web (Marzo - Abril)

Con los módulos de procesamiento y predicción ya validados, se procedió a la unificación del sistema en una capa de software orientada al usuario:

- Interfaz Web (Dash/Plotly): Programación del frontend interactivo y del cuadro de mando visual empleando el framework Dash. Se diseñaron los componentes gráficos reactivos con Plotly para la representación dinámica de las características acústicas.
- Integración del Pipeline: Consolidación de la arquitectura de extremo a extremo (end-to-end), interconectando el backend de análisis de audio en tiempo real con el motor de inferencia de Machine Learning bajo una lógica de flujo de datos libre de estado (stateless).

Fase 5: Memoria del TFM (Enero - Abril)

Planificada de manera transversal y concurrente a partir del segundo mes del proyecto para garantizar la calidad documental y mitigar los riesgos de entrega:

- Inicio de Redacción Formal (Enero - Febrero): Escritura de los capítulos introductorios, el marco teórico, el estado del arte y la especificación de objetivos y alcance.

- Capítulos Técnicos (Marzo - Abril): Documentación detallada de la arquitectura del software, los fundamentos matemáticos de la extracción de características (PDS) y la configuración del modelo predictivo.
- Conclusiones y Revisión Final (Abril): Redacción del análisis crítico de los resultados, evaluación de la sostenibilidad y viabilidad socioeconómica del proyecto, consolidación de las líneas de trabajo futuro y maquetación final conforme a la normativa académica de la ETS de Ingeniería (ICAI).

Capítulo 5. SISTEMA/MODELO

DESARROLLADO

5.1 ANÁLISIS DEL SISTEMA

El sistema desarrollado en este Trabajo de Fin de Máster tiene como objetivo principal predecir determinadas características musicales a partir de archivos de audio, integrando técnicas de procesamiento digital de señal, aprendizaje automático y visualización interactiva mediante una aplicación web. En esta sección se detallan los requisitos, actores y casos de uso que definen el comportamiento esperado del sistema.

5.1.1 REQUISITOS FUNCIONALES

Los requisitos funcionales describen las funcionalidades que el sistema debe proporcionar:

- Carga de archivos de audio: El sistema debe permitir al usuario subir archivos de audio en formato MP3 para su análisis.
- Extracción de características: El sistema debe procesar los archivos de audio y extraer automáticamente características relevantes como nota principal, chroma, tempo, energía, loudness y tonalidad.
- Preprocesamiento de datos: El sistema debe limpiar, normalizar y adaptar las características extraídas para que sean compatibles con el modelo predictivo.
- Predicción mediante modelo: El sistema debe utilizar un modelo de aprendizaje automático previamente entrenado para generar predicciones a partir de las características del audio.
- Visualización de resultados: El sistema debe mostrar al usuario los resultados de la predicción de forma clara mediante gráficos e indicadores visuales.
- Búsqueda y selección de artistas: El sistema debe permitir introducir texto para buscar artistas y sugerir opciones dinámicamente.

- Integración con APIs externas (Spotify): El sistema debe poder consultar información adicional de artistas o canciones mediante APIs externas.
- Interfaz interactiva: El sistema debe permitir la interacción del usuario en tiempo real sin necesidad de recargar la página.

5.1.2 REQUISITOS NO FUNCIONALES

Los requisitos no funcionales definen las restricciones y utilidades del sistema:

- Rendimiento: El sistema debe procesar y generar predicciones en un tiempo razonable (idealmente en pocos segundos tras la carga del archivo).
- Usabilidad: La interfaz debe ser intuitiva, accesible y fácil de utilizar, incluso para usuarios sin conocimientos técnicos.
- Escalabilidad: El sistema debe ser capaz de adaptarse a un mayor volumen de datos o usuarios sin degradar significativamente el rendimiento.
- Mantenibilidad: El código debe estar estructurado en módulos independientes, facilitando su mantenimiento y evolución.
- Reproducibilidad: Los resultados del modelo deben ser reproducibles bajo las mismas condiciones de entrada.
- Compatibilidad: El sistema debe funcionar en los principales navegadores web modernos.
- Robustez: El sistema debe manejar errores como archivos corruptos o formatos no válidos sin interrumpir la ejecución.

5.1.3 ACTORES DEL SISTEMA

Se identifican los siguientes actores que interactúan con el sistema:

- Usuario final: Persona que utiliza la aplicación web para analizar canciones y obtener predicciones. No requiere conocimientos técnicos.
- Desarrollador/Investigador: Responsable del desarrollo, mantenimiento y mejora del sistema, incluyendo el entrenamiento del modelo y la actualización del código.

- Servicios externos: Sistemas externos como APIs (por ejemplo, Spotify) que proporcionan información adicional utilizada por la aplicación.

5.1.4 CASOS DE USO

A continuación, se describen los principales casos de uso del sistema:

Buscar artista

- Actor: Usuario
- Descripción: El usuario introduce texto en un campo de búsqueda.
- Resultado: El sistema muestra sugerencias dinámicas.

Subir archivo de audio

- Actor: Usuario
- Descripción: El usuario carga un archivo MP3 desde su dispositivo.
- Resultado: El sistema recibe el archivo y lo prepara para su procesamiento.

Procesar audio y extraer características

- Actor: Sistema
- Descripción: El sistema analiza el audio y obtiene sus características relevantes.
- Resultado: Se genera un vector de características listo para el modelo.

Generar predicción

- Actor: Sistema
- Descripción: El modelo predictivo procesa las características y genera una salida.
- Resultado: Se obtiene la predicción asociada al audio.

Visualizar resultados

- Actor: Usuario
- Descripción: El usuario visualiza los resultados mediante gráficos o indicadores.
- Resultado: Interpretación de la predicción.

Consultar información externa

- Actor: Sistema
- Descripción: El sistema consulta APIs externas para enriquecer la información.
- Resultado: Se muestran datos adicionales al usuario.

5.2 DISEÑO

En esta sección se describe la estructura conceptual del sistema, detallando la arquitectura general, la organización de los componentes, el flujo de datos y el diseño de almacenamiento. El objetivo es proporcionar una visión clara de cómo se organizan e interrelacionan los distintos módulos antes de su implementación.

5.2.1 ARQUITECTURA GENERAL DEL PIPELINE

El sistema se ha diseñado siguiendo una arquitectura modular basada en un pipeline de procesamiento de datos, donde cada etapa realiza una función específica y bien definida. Esta separación permite mejorar la escalabilidad, mantenibilidad y reutilización del sistema.

El pipeline completo se compone de las siguientes fases:

1. Entrada de datos: El sistema recibe como entrada archivos de audio en formato MP3 proporcionados por el usuario a través de la interfaz web.
2. Extracción de características: Se procesan los archivos de audio para obtener descriptores relevantes que representen el contenido musical (características espectrales, temporales y armónicas).
3. Preprocesamiento: Las características extraídas se limpian, normalizan y transforman para garantizar su compatibilidad con el modelo de aprendizaje automático.
4. Modelo predictivo: Se aplica un modelo previamente entrenado que genera una predicción a partir de las características procesadas.

5. Visualización: Los resultados se presentan al usuario mediante una interfaz interactiva que facilita la interpretación de la predicción.

Esta arquitectura sigue un enfoque tipo ETL (Extract, Transform, Load) adaptado a sistemas de machine learning.

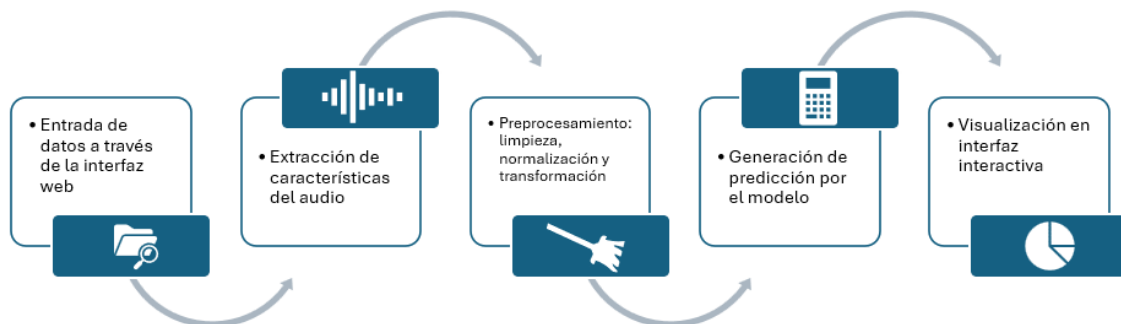


Ilustración 8: Arquitectura general del pipeline del sistema.

5.2.2 DIAGRAMA DE COMPONENTES

El sistema se organiza en varios módulos independientes que interactúan entre sí. Cada componente encapsula una funcionalidad específica, facilitando el desarrollo y mantenimiento.

Los principales componentes identificados son:

- Módulo de entrada (Frontend - Dash): Gestiona la interacción con el usuario, incluyendo la carga de archivos y la visualización de resultados.
- Módulo de extracción de características: Encargado de procesar los archivos de audio y generar representaciones numéricas.
- Módulo de preprocesamiento: Realiza la limpieza, transformación y normalización de los datos.

- Módulo de modelado: Contiene el modelo de machine learning entrenado y la lógica de predicción.
- Módulo de integración con APIs externas: Permite enriquecer la información mediante servicios externos (por ejemplo, datos de artistas).
- Módulo de visualización: Genera gráficos e indicadores mediante librerías de visualización.

Las interacciones entre estos módulos siguen un flujo unidireccional, donde la salida de un componente actúa como entrada del siguiente.



Ilustración 9: Diagrama de componentes del sistema.

5.2.3 FLUJO DE DATOS

El flujo de datos describe cómo la información se transforma a lo largo del sistema desde su entrada hasta la obtención de resultados.

El proceso completo es el siguiente:

1. El usuario carga un archivo de audio desde la interfaz.
2. El sistema convierte el archivo en una señal procesable.
3. Se extraen características acústicas relevantes.

4. Las características se normalizan utilizando los mismos parámetros que en el entrenamiento.
5. El vector de características se introduce en el modelo predictivo.
6. El modelo genera una predicción numérica o categórica.
7. Los resultados se transforman en visualizaciones comprensibles para el usuario.

Este flujo garantiza la coherencia entre el entrenamiento del modelo y su uso en producción.

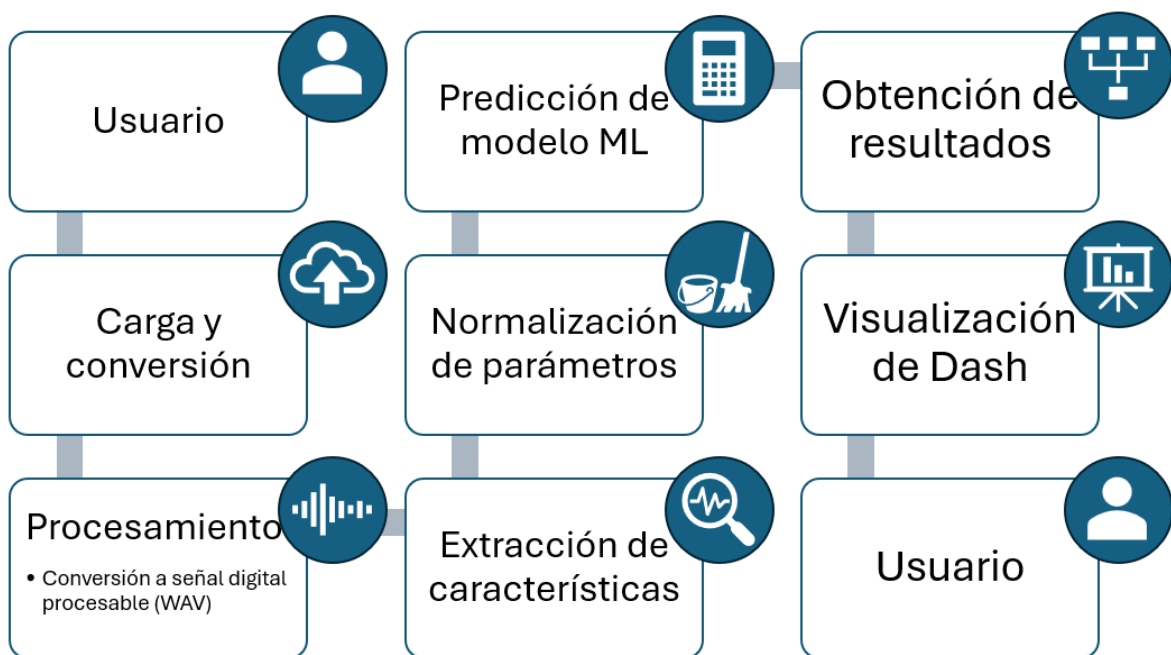


Ilustración 10: Flujo de datos del sistema.

5.2.4 DISEÑO DE LA BASE DE DATOS

El sistema no utiliza una base de datos relacional tradicional en tiempo de ejecución, sino que se basa en estructuras de datos persistidas en archivos y modelos previamente entrenados.

Los principales elementos de almacenamiento son:

- **Dataset de entrenamiento:** Se utiliza el Million Song Dataset, que contiene información estructurada sobre características musicales.

- **Datos preprocesados:** Versiones limpias y transformadas del dataset original, utilizadas durante el entrenamiento del modelo.
- **Modelo entrenado:** Se almacena en formato serializado permitiendo su reutilización sin necesidad de reentrenamiento.
- **Parámetros de normalización:** Valores utilizados para escalar los datos (media, desviación estándar), necesarios para mantener consistencia en predicción.
- **Archivos temporales:** Archivos de audio subidos por el usuario durante la ejecución.

Este enfoque reduce la complejidad del sistema y elimina la necesidad de gestionar una base de datos activa, lo cual es adecuado para aplicaciones de inferencia en tiempo real.

5.3 IMPLEMENTACIÓN

5.3.1 MÓDULO DE EXTRACCIÓN DE CARACTERÍSTICAS

El módulo de extracción de características constituye el front-end de procesamiento de señales del sistema. Su función primordial es la transformación de una señal de audio en el dominio del tiempo —caracterizada por variaciones de presión sonora— en un vector de características (feature vector) de alta dimensionalidad que represente de forma unívoca las propiedades acústicas, temporales y armónicas de la obra. Este proceso es crítico, ya que la calidad de la predicción del modelo de aprendizaje automático depende directamente de la fidelidad y relevancia de estos descriptores.

5.3.1.1 Procesamiento de audio

El procesamiento se inicia con la digitalización y el acondicionamiento de la señal. Siguiendo los estándares de la Recuperación de Información Musical (MIR, por sus siglas en inglés), el sistema realiza un submuestreo de la señal de entrada a una frecuencia de muestreo (fs) de 22.050 Hz. Esta frecuencia es suficiente para capturar el espectro audible relevante para la caracterización musical, cumpliendo con creces el teorema de Nyquist-Shannon para las frecuencias fundamentales de la música, al tiempo que optimiza la carga computacional.

Una vez normalizada la señal, el procesamiento se divide en dos dominios:

- Dominio temporal: Análisis directo de la envolvente de la señal y detección de transitorios para la estimación de rítmica y duración.
- Dominio frecuencial: Aplicación de la Transformada de Fourier de Corto Tiempo (STFT) mediante el uso de ventanas de análisis. Esto permite descomponer la señal en sus componentes espectrales, facilitando la extracción de características como el timbre, la energía y la tonalidad.

5.3.1.2 Uso de librerías (Librosa)

La implementación de este módulo se ha realizado íntegramente mediante la librería Librosa, el estándar de facto en ingeniería para el análisis de música y audio en Python. Se ha seleccionado esta herramienta por su capacidad para manejar operaciones complejas de procesamiento digital de señales (PDS) de forma eficiente y precisa. Entre sus funcionalidades clave utilizadas en el proyecto destacan:

- `librosa.load`: Para la decodificación de archivos multiformato (MP3, WAV) y su conversión a arrays de punto flotante.
- `librosa.feature`: Para el cálculo de descriptores espectrales y de croma.
- `librosa.beat`: Para la implementación de algoritmos de seguimiento de pulso mediante funciones de fuerza de inicio (onset strength).

- `librosa.effects`: Para el preprocesamiento de la señal, como el truncado de silencios iniciales y finales.

5.3.1.3 Extracción de características

A continuación, se detallan las características específicas extraídas de cada pista de audio, su significado técnico y el método de obtención:

1. Descriptores Temporales y Rítmicos:

- **Duración (Duration)**: Representa el tiempo total de la señal en segundos. Se calcula mediante la relación entre el número total de muestras y la frecuencia de muestreo (N/fs).
- **Tempo (BPM)**: Es la estimación de la velocidad rítmica de la canción en pulsaciones por minuto. El sistema utiliza `librosa.beat.beat_track`, que analiza la periodicidad de los picos de energía en la señal para determinar el pulso predominante.
- **Fin de Fade-in (End of Fade-in)**: Identifica el punto temporal en el que la señal alcanza un nivel de energía estable tras el inicio. Se obtiene mediante `librosa.effects.trim`, analizando el umbral de decibelios inicial.

2. Descriptores de Energía y Dinámica:

- **Sonoridad (Loudness/Energy)**: Representa la potencia media de la señal. Se obtiene calculando la Raíz Cuadrada Media (RMS) de la amplitud en ventanas cortas de la señal. Un valor alto de RMS indica una mayor presión sonora percibida.
- **Energía (Energy)**: Estrechamente ligada al RMS, cuantifica la intensidad física de la señal acústica a lo largo del tiempo.

3. Descriptores Espectrales y de Percepción:

- **Bailabilidad (Danceability)**: Aunque es un concepto subjetivo, en el sistema se estima mediante la relación entre el centroide espectral y la frecuencia de muestreo. El centroide espectral indica el "centro de masa" del espectro de frecuencias; un

centroide estable y rítmicamente predecible se correlaciona con una mayor bailabilidad.

- **Centroide Espectral (Spectral Centroid):** Caracteriza el "brillo" del sonido. Se calcula como la media ponderada de las frecuencias presentes en la señal, donde los pesos son las magnitudes de dichas frecuencias.
4. **Descriptores de Teoría Musical y Tonalidad (Croma):**
- **Cromagrama (Chroma STFT):** El sistema proyecta todo el espectro de frecuencias en 12 bins correspondientes a las 12 notas de la escala cromática (Do, Do#, Re, etc.). Esto permite analizar el contenido armónico independientemente del timbre.
 - **Tonalidad (Key):** Se identifica la nota predominante buscando el bin del cromagrama con la energía media más alta.
 - **Confianza de Tonalidad (Key Confidence):** Representa la claridad de la tonalidad detectada. Se calcula como la relación entre la energía de la nota dominante y la suma total de energía cromática.
 - **Modo (Mode):** Determina si la canción está en una escala mayor (1) o menor (0). Para ello, el sistema compara el perfil cromático medio de la canción con plantillas estándar de perfiles mayores y menores mediante el coeficiente de correlación de Pearson.
 - **Confianza de Modo (Mode Confidence):** Es el valor máximo de la correlación obtenida en el proceso de emparejamiento con las plantillas de modo.
 - **Compás (Time Signature):** Representa la estructura métrica (ej. 4/4). En la implementación actual para la aplicación web, se establece un valor nominal de 4.0 con una confianza del 90%, dada la prevalencia de este compás en la música popular contemporánea.

5.3.2 MÓDULO DE PREPROCESAMIENTO DE DATOS

5.3.2.1 Limpieza del *Million Song Dataset*

La etapa de limpieza de datos constituye un proceso crítico para garantizar la robustez y la capacidad de generalización del modelo predictivo. Al trabajar con el *Million Song Dataset* (MSD) —o su versión reducida, el *Million Song Subset*—, se dispone de una base de datos de gran escala que, debido a su origen heterogéneo y a la antigüedad de algunos de sus registros, presenta inconsistencias que deben ser tratadas antes de la fase de entrenamiento.

El proceso de limpieza implementado en este proyecto se ha estructurado en torno a los siguientes pilares técnicos:

1. **Filtrado por disponibilidad de la variable objetivo:** El objetivo principal del modelo es la predicción de la popularidad de una canción, representada en el dataset por el atributo `song_hottness`. Sin embargo, una fracción significativa de las muestras del MSD carece de este valor (registrando valores nulos o NaN) debido a la ausencia de datos históricos de mercado en el momento de la recolección. El sistema realiza un filtrado primario descartando cualquier entrada que no contenga una etiqueta de popularidad válida, ya que estas muestras no pueden ser utilizadas en un paradigma de aprendizaje supervisado.
2. **Tratamiento de valores nulos e inconsistencias en los descriptores:** Durante la extracción de las características (tanto metadatos del artista como descriptores acústicos), se realiza una validación exhaustiva de cada vector de entrada. El sistema detecta y elimina aquellas canciones que presentan valores no numéricos (NaN) o infinitos (Inf) en cualquiera de sus dimensiones. Estas anomalías suelen ser el resultado de errores en los algoritmos originales de análisis de señal o fallos en la persistencia de los archivos HDF5. Se ha adoptado una política de descarte estricto para asegurar que el *pipeline* de entrenamiento solo procese registros con integridad

numérica total, evitando así sesgos o fallos en el cálculo de los gradientes durante el ajuste del modelo.

3. **Agregación estadística y reducción de dimensionalidad:** El MSD almacena descriptores de audio de alta resolución, como matrices de *pitches* y *timbre* que varían en longitud según la duración de la canción. Para unificar el espacio de características y permitir su procesamiento mediante algoritmos de regresión, se ha aplicado una limpieza basada en la agregación. En lugar de utilizar las series temporales completas, se calculan estadísticos de primer y segundo orden (media y desviación estándar) para cada descriptor. Este proceso permite limpiar el ruido de las variaciones instantáneas de la señal y condensar la información espectral en un vector de características de dimensión fija (25 descriptores), facilitando la convergencia estadística del modelo.
4. **Validación de integridad de la estructura jerárquica:** Dado que el dataset está organizado en una estructura de directorios anidada (A-Z/A-Z/A-Z), el proceso de limpieza incluye una capa de gestión de errores de entrada/salida. El sistema verifica la integridad de cada archivo .h5 mediante la librería PyTables antes de su procesamiento. Cualquier archivo corrupto o con errores de lectura es omitido del flujo de datos, garantizando que el entrenamiento se realice sobre un conjunto de datos verificado y técnicamente consistente.

Este proceso de limpieza transforma el volumen masivo de datos brutos en un conjunto de entrenamiento refinado y de alta calidad. Como resultado, se obtiene una matriz de características normalizada donde cada muestra garantiza la presencia de todos los descriptores necesarios y una etiqueta objetivo válida, sentando las bases para una fase de modelado estadísticamente significativa.

5.3.2.2 Normalización de características

Una vez obtenido el conjunto de datos limpio, se procede a la normalización de las características. Este paso es fundamental en el diseño de sistemas de aprendizaje automático

cuando los descriptores de entrada presentan órdenes de magnitud y unidades de medida muy heterogéneos. En el contexto de este proyecto, se manejan variables con rangos extremadamente dispares: mientras que el tempo se mide en pulsaciones por minuto (típicamente entre 60 y 200 BPM), la energía y la disponibilidad son coeficientes normalizados entre 0 y 1, y la sonoridad (*loudness*) se expresa en una escala logarítmica de decibelios (generalmente entre -60 y 0 dB).

Para resolver esta disparidad, se ha implementado una técnica de estandarización conocida como **Z-score normalization**, utilizando la herramienta StandardScaler de la librería scikit-learn. El proceso consiste en transformar cada característica de forma independiente para que posea una media (μ) igual a 0 y una desviación estándar (σ) igual a 1. Matemáticamente, la transformación aplicada a cada valor x viene dada por la ecuación:

$$Z = \frac{x - \mu}{\sigma}$$

La aplicación de esta normalización aporta las siguientes ventajas técnicas al sistema:

1. **Equidad en la contribución de variables:** Se evita que las características con valores numéricos intrínsecamente más altos (como la duración en segundos o el tempo) dominen la función de coste del modelo de regresión. Sin este escalado, los pesos del modelo se ajustarían de forma sesgada, ignorando variables críticas pero de menor magnitud, como los coeficientes de confianza.
2. **Optimización de la convergencia:** En algoritmos basados en gradiente o en la resolución de ecuaciones normales (como la regresión lineal empleada en el sistema), la estandarización asegura que la superficie de error sea más esférica, lo que acelera y estabiliza el proceso de convergencia durante el entrenamiento.
3. **Consistencia en la inferencia:** Es imperativo destacar que los parámetros estadísticos (media y desviación) calculados durante la fase de entrenamiento se almacenan de forma persistente. Esto permite que, cuando el usuario carga una nueva canción en la aplicación web, el sistema aplique exactamente la misma

transformación a los datos de entrada antes de realizar la predicción. Esta coherencia garantiza que el modelo trabaje siempre sobre una distribución de datos idéntica a la que conoció durante su fase de aprendizaje.

Finalmente, la integración de este proceso se realiza mediante un *pipeline* de ejecución serializado. Esto asegura que el escalado y la predicción actúen como un bloque atómico, eliminando riesgos de discrepancia de datos (*data leakage*) y facilitando el despliegue del modelo en el entorno de producción de la aplicación web.

5.3.2.3 División train/test

Una vez que las características han sido extraídas, limpiadas y normalizadas, se procede a la división del conjunto de datos en subconjuntos de entrenamiento y prueba. Esta etapa es fundamental en el flujo de trabajo del aprendizaje automático, ya que permite evaluar de forma objetiva la capacidad de generalización del modelo ante datos que no han sido utilizados durante el proceso de ajuste de parámetros.

La metodología empleada consiste en una partición estocástica del dataset original, utilizando la función `train_test_split` de la librería `scikit-learn`. Para el desarrollo de este sistema, se ha adoptado una proporción de 80/20, destinando el 80% de las muestras al entrenamiento del modelo y el 20% restante a la validación externa o fase de prueba.

Esta división se justifica bajo los siguientes criterios de ingeniería:

1. **Prevención del sobreajuste (*Overfitting*):** El objetivo primordial de separar un conjunto de prueba es detectar si el modelo está memorizando el ruido o las particularidades específicas de los datos de entrenamiento en lugar de aprender los patrones subyacentes de la señal musical. Al evaluar el rendimiento sobre el 20% de datos "no vistos", se obtiene una estimación insesgada de la precisión del sistema en un entorno real.
2. **Representatividad estadística:** Dado el volumen de datos procesados del *Million Song Dataset*, una partición del 20% para el conjunto de prueba proporciona una

muestra lo suficientemente amplia para que las métricas de error calculadas sean estadísticamente significativas y estables, minimizando el impacto de posibles valores atípicos.

- 3. Reproducibilidad de los experimentos:** Se ha implementado el uso de una semilla de aleatoriedad fija (`random_state=42`). En el ámbito de la investigación y la ingeniería de datos, esta práctica es esencial para garantizar que la división de las muestras sea idéntica en cada ejecución del sistema. Esto permite que los resultados y las comparativas entre diferentes algoritmos presentados en esta memoria sean trazables y replicables bajo las mismas condiciones experimentales.

El subconjunto de entrenamiento es el encargado de ajustar los pesos y coeficientes de los modelos de regresión, mientras que el subconjunto de prueba se reserva exclusivamente para la generación de las métricas de rendimiento final. Esta separación estricta garantiza la integridad del proceso de validación y la fiabilidad de las conclusiones obtenidas sobre la eficacia del sistema desarrollado.

5.3.3 MÓDULO DE MODELADO PREDICTIVO

La selección del algoritmo de aprendizaje automático es un paso determinante para el éxito de la predicción, ya que la relación entre las características acústicas de una canción y su impacto comercial no suele seguir un patrón lineal simple. En este proyecto, el proceso de selección ha sido evolutivo, partiendo de modelos experimentales hasta llegar a la arquitectura actual.

Inicialmente, se evaluó un modelo preliminar basado en un número reducido de variables (4 atributos principales). Sin embargo, los resultados obtenidos mostraron una capacidad predictiva limitada, con un coeficiente de determinación (R^2) insuficiente para capturar la complejidad del fenómeno de la popularidad musical. Esta primera aproximación permitió identificar que el éxito de una obra no depende de factores aislados, sino de una interacción multidimensional entre descriptores temporales, armónicos y metadatos del artista.

Ante estas limitaciones, se optó por una segunda arquitectura más robusta, basada en el algoritmo **Random Forest Regressor** y un espacio de características ampliado a 15 atributos. La elección de este modelo se justifica por los siguientes motivos técnicos:

1. **Capacidad para capturar no linealidades:** A diferencia de la regresión lineal simple, los bosques aleatorios (*Random Forests*) son capaces de modelar relaciones complejas y no lineales entre las variables de entrada, lo cual es esencial en la música, donde el éxito puede depender de combinaciones específicas de tempo, energía y notoriedad del artista.
2. **Robustez ante valores atípicos y dimensionalidad:** Al ser un método de aprendizaje por conjunto (*Ensemble Learning*) basado en múltiples árboles de decisión, el modelo es intrínsecamente resistente al ruido y a los valores atípicos presentes en el *Million Song Dataset*. Además, gestiona eficientemente las 15 dimensiones del vector de características sin incurrir fácilmente en sobreajuste.
3. **Escalabilidad y procesamiento en la nube:** Debido a la magnitud del *Million Song Dataset* completo (que supera los 280 GB en su versión total), el entrenamiento de este segundo modelo se llevó a cabo utilizando infraestructura de computación en la nube a través de Amazon Web Services (AWS). El uso de instancias de alto rendimiento permitió procesar el millón de registros y ajustar los hiperparámetros del bosque aleatorio (como el número de estimadores y la profundidad máxima) en tiempos computacionalmente aceptables.

Finalmente, el modelo seleccionado utiliza 200 árboles de decisión y una profundidad controlada para asegurar un equilibrio óptimo entre el sesgo y la varianza. Esta configuración, entrenada sobre el dataset completo en AWS, constituye el núcleo predictivo que se utiliza actualmente en la aplicación web, ofreciendo la mayor precisión y estabilidad observada durante la fase de experimentación.

El proceso de entrenamiento constituye la fase más intensiva desde el punto de vista computacional, requiriendo una orquestación precisa entre el software de aprendizaje

automático y la infraestructura de hardware. Dado que el *Million Song Dataset* completo comprende un volumen de datos que excede las capacidades de almacenamiento y procesamiento de una estación de trabajo convencional, el entrenamiento se ha desplazado a un entorno de computación elástica en la nube.

Infraestructura y entorno de ejecución

Para la ejecución del entrenamiento, se ha utilizado la plataforma **Amazon Web Services (AWS)**, empleando instancias de computación optimizada (familia EC2). El dataset original se montó como un volumen de almacenamiento en bloque (Amazon EBS) a partir de un *snapshot* público del proyecto MSD, permitiendo un acceso de baja latencia a los más de un millón de archivos HDF5. El entorno de ejecución se configuró sobre una distribución de Linux optimizada para *Data Science*, asegurando la disponibilidad de las librerías PyTables para la lectura jerárquica de datos y scikit-learn para el modelado.

Estrategia de carga por lotes (*Batch Processing*)

Debido a la restricción de memoria RAM, el entrenamiento no se realizó mediante una carga masiva de archivos, sino a través de una estrategia de procesamiento por lotes (*batches*). Se implementó un algoritmo de iteración que recorre la estructura de directorios del dataset, procesando grupos de 10.000 archivos de forma secuencial. En cada lote, se realiza la apertura del archivo HDF5, la extracción del vector de 15 características y la captura de la variable objetivo (*song_hottness*). Solo aquellos registros que superan los filtros de integridad definidos en la etapa de limpieza son almacenados en una matriz temporal en memoria. Esta metodología garantiza que el sistema mantenga un consumo de recursos estable, evitando desbordamientos de memoria (*stack overflow*) durante la fase de lectura.

Configuración y ajuste de hiperparámetros

El modelo de *Random Forest* se ha configurado con una serie de hiperparámetros específicos, seleccionados para maximizar la precisión sin comprometer la capacidad de generalización:

- **Número de estimadores ($n_estimators=200$):** Se ha definido la construcción de 200 árboles de decisión independientes. Este número proporciona un equilibrio óptimo; un valor menor no permitiría reducir suficientemente la varianza del modelo, mientras que un valor superior incrementaría el coste computacional y el tamaño del archivo resultante sin aportar mejoras significativas en el coeficiente de determinación.
- **Profundidad máxima ($max_depth=15$):** Se ha limitado la profundidad de los árboles para evitar el sobreajuste. Al restringir el número de niveles, se obliga al modelo a aprender patrones generales de la música en lugar de memorizar casos particulares del dataset de entrenamiento.
- **Muestras mínimas para división ($min_samples_split=20$):** Este parámetro exige que un nodo interno deba tener al menos 20 muestras para poder dividirse. Esta restricción actúa como un mecanismo de regularización adicional, suavizando las predicciones del modelo.
- **Paralelización ($n_jobs=-1$):** Para optimizar el tiempo de entrenamiento en AWS, se ha habilitado el uso de todos los núcleos de procesamiento disponibles en la instancia EC2. Al ser el *Random Forest* un algoritmo paralelizable por naturaleza (cada árbol se entrena de forma independiente), esta configuración reduce drásticamente el tiempo de cómputo total.

Una vez completado el entrenamiento en la infraestructura de AWS, se procedió a la validación técnica del modelo utilizando el conjunto de datos de prueba (*hold-out set*) que no fue expuesto al algoritmo durante el aprendizaje. Esta fase es crucial para cuantificar la precisión real del sistema y entender qué factores acústicos o de metadatos tienen un mayor peso en la predicción del éxito de una canción.

Métricas de rendimiento

El rendimiento del modelo final se ha evaluado principalmente mediante el coeficiente de determinación (R^2), obteniendo un valor de **0,4329**. En el contexto de la predicción de

fenómenos sociales y comerciales como es la popularidad musical, un R^2 superior al 43% se considera un resultado sólido y significativo. Esto indica que el modelo es capaz de explicar casi la mitad de la varianza en la popularidad de las canciones basándose exclusivamente en los 15 descriptores seleccionados.

Complementariamente, las pruebas de validación arrojaron métricas de error consistentes (RMSE y MAE bajos), lo que confirma que las predicciones del modelo no presentan desviaciones extremas y mantienen una estabilidad uniforme a lo largo de diferentes géneros y épocas musicales presentes en el dataset.

Análisis de importancia de las características (*Feature Importance*)

Uno de los resultados más valiosos de la validación es el análisis de la importancia de las variables, que permite "abrir la caja negra" del bosque aleatorio para entender qué factores determinan el éxito. Los resultados obtenidos muestran una jerarquía clara en los descriptores:

1. **Familiaridad del artista (*artist_familiarity*) - 46,33%**: Es, con diferencia, la variable más determinante. Indica que el reconocimiento previo y la trayectoria del artista son el predictor más fuerte del éxito de una nueva obra.
2. **Año de lanzamiento (*year*) - 20,34%**: Refleja cómo las tendencias y los estándares de éxito han evolucionado temporalmente, siendo un factor de peso en la normalización de la popularidad.
3. **Popularidad actual del artista (*artist_hottnesss*) - 20,29%**: Estrechamente ligada a la familiaridad, esta métrica captura el "momento" o tendencia actual del artista en el mercado.
4. **Sonoridad (*loudness*) - 3,49%**: Es la característica acústica pura más relevante, lo que valida técnicamente la teoría de la "guerra del volumen" (*loudness war*) en la industria musical contemporánea.

5. **Duración (*duration*) - 2,62%:** Indica que la extensión de la pista tiene un impacto marginal pero medible en su aceptación comercial.

Variable	Influencia (%)
Familiaridad del Artista (<i>artist_familiarity</i>)	46.33%
Año de lanzamiento (<i>year</i>)	20.35%
Popularidad del Artista (<i>artist_hotness</i>)	20.29%
Volumen (<i>loudness</i>)	3.49%
Duración (<i>duration</i>)	2.62%
Tempo (BPM) (<i>tempo</i>)	1.55%
Final del Fade-in (<i>end_of_fade_in</i>)	1.32%
Confianza del Modo (<i>mode_confidence</i>)	1.20%
Confianza de la Tonalidad (<i>key_confidence</i>)	1.09%
Confianza del Compás (<i>time_sig_confidence</i>)	0.86%
Tonalidad (<i>key</i>)	0.54%
Compás (<i>time_signature</i>)	0.25%
Modo (Mayor/Menor) (<i>mode</i>)	0.10%
Energía (<i>energy</i>)	0.00%
Bailabilidad (<i>danceability</i>)	0.00%

Tabla 1: Valores de la importancia de cada variable en el modelo.

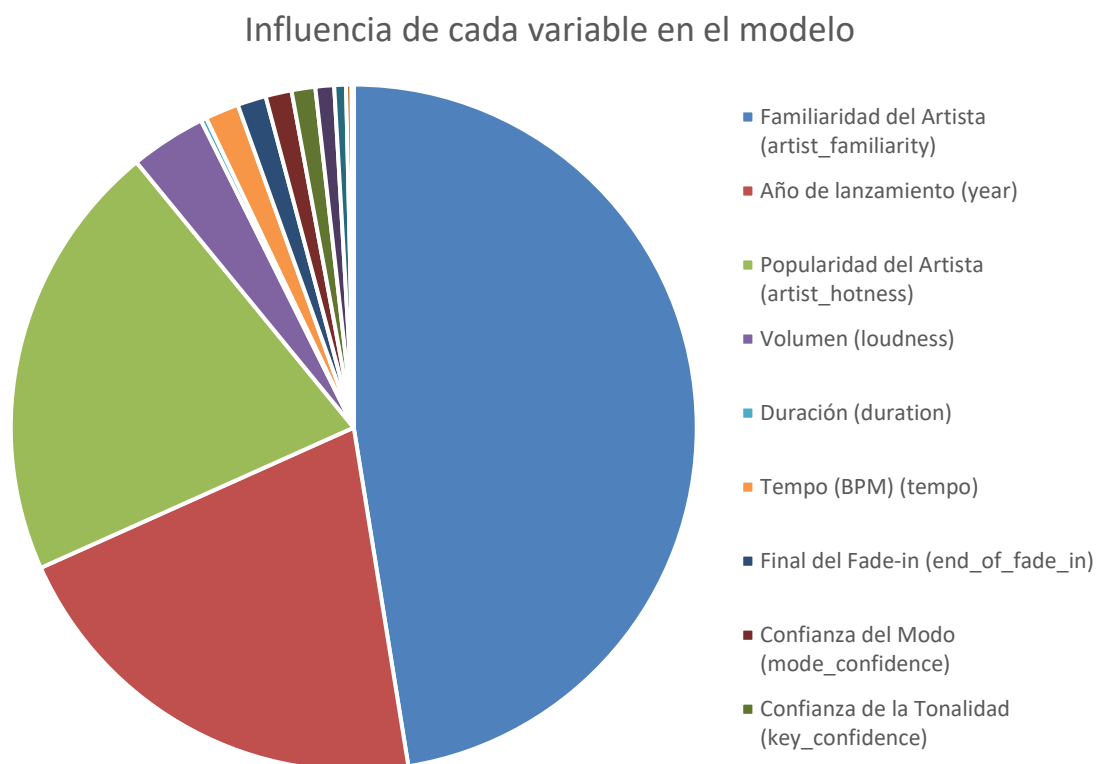


Ilustración 11: Gráfico con la influencia de cada variable en el modelo.

5.3.4 MÓDULO DE APLICACIÓN WEB

5.3.4.1 Arquitectura de la aplicación

La interfaz de usuario del proyecto se ha concebido como un ecosistema de ingeniería integral que permite la orquestación en tiempo real de modelos de aprendizaje automático, procesamiento digital de señales y análisis crítico de resultados. Para su implementación, se ha adoptado una arquitectura basada en el framework **Dash (by Plotly)**, el cual se sustenta sobre una pila tecnológica de alto rendimiento: **Flask** para la gestión del servidor y el enrutamiento, **React.js** para la renderización de una interfaz reactiva y **Plotly** para la generación de visualizaciones dinámicas y científicas.

La aplicación se ha diseñado bajo un paradigma de desarrollo modular y escalable, utilizando la funcionalidad nativa de Dash para aplicaciones multipágina (`use_pages=True`). Esta

decisión arquitectónica es fundamental para garantizar una separación de responsabilidades clara, organizando la lógica de negocio en cuatro nodos o módulos principales interconectados mediante una barra de navegación persistente:

1. **Módulo de Inicio (*Home*):** Actúa como el centro de mando y contextualización del proyecto. Su función es proporcionar al usuario la base teórica necesaria, describiendo los objetivos del TFM y guiándole a través de las capacidades técnicas de la herramienta.
2. **Módulo de Análisis de Variables (*Variables*):** Dedicado a la transparencia del modelo ("Explainable AI"). En esta sección se exponen los resultados de la validación técnica realizada en AWS, visualizando de forma interactiva la importancia de cada uno de los 15 descriptores. Este nodo permite al usuario comprender la jerarquía de factores (como la familiaridad del artista frente a los atributos acústicos) antes de interactuar con el motor de predicción.
3. **Módulo Predictor (*Predictor*):** Constituye el núcleo operativo y técnico de la aplicación. Es el punto de convergencia de todo el sistema, donde se integra el *pipeline* completo de procesamiento: desde la carga de archivos MP3 y la extracción de características mediante Librosa, hasta la consulta dinámica a la API de Spotify y la ejecución de la inferencia mediante el modelo de *Random Forest* entrenado.
4. **Módulo de Conclusiones y Análisis Final (*Conclusiones*):** Este nodo, recientemente integrado en la arquitectura, cierra el ciclo de vida del análisis. Su propósito es proporcionar una síntesis interpretativa de los resultados obtenidos por el sistema. A diferencia de los módulos anteriores, este apartado se centra en la extracción de conocimiento, permitiendo al usuario final (o al investigador) consolidar las observaciones derivadas de las predicciones y entender las implicaciones de los hallazgos en el contexto de la industria musical actual.

Desde la perspectiva de la ingeniería de software, la aplicación sigue un flujo de datos unidireccional y reactivo, fundamentado en la gestión de eventos mediante *callbacks* asíncronos. La arquitectura se estructura en dos capas operativas:

- **Capa de Cliente (Frontend):** Ejecutada íntegramente en el navegador, se encarga de la captura de estímulos (subida de archivos, entradas de texto, navegación entre páginas) y de la presentación estética de los datos. Se ha implementado un diseño de "modo oscuro" con una paleta de colores optimizada para entornos profesionales, lo que reduce la fatiga visual y resalta los componentes gráficos de Plotly.
- **Capa de Servidor (Backend):** Ejecutada en el entorno de Python, es la responsable de la computación intensiva. El servidor gestiona de forma eficiente la lógica de negocio, asegurando que procesos pesados como la decodificación de audio o la inferencia estadística no bloqueen la interfaz de usuario. La comunicación entre páginas se realiza mediante un contenedor dinámico (`dash.page_container`), que permite transiciones fluidas sin necesidad de recargar los recursos estáticos del sistema.

La arquitectura se completa con una capa de conectividad externa altamente integrada. El sistema interactúa en tiempo real con la **Web API de Spotify** para obtener métricas de mercado que el modelo de *Random Forest* requiere para su precisión. Para optimizar el rendimiento y la seguridad, la aplicación gestiona de forma transparente el almacenamiento temporal de archivos y la configuración de variables de entorno, asegurando que el flujo de información sea robusto ante posibles fallos de red o inconsistencias en los datos de entrada.

En conclusión, la arquitectura de la aplicación web se ha diseñado como un sistema *end-to-end* que transforma la complejidad del procesamiento de señales y el aprendizaje automático en una experiencia de usuario coherente, funcional y orientada a la toma de decisiones.

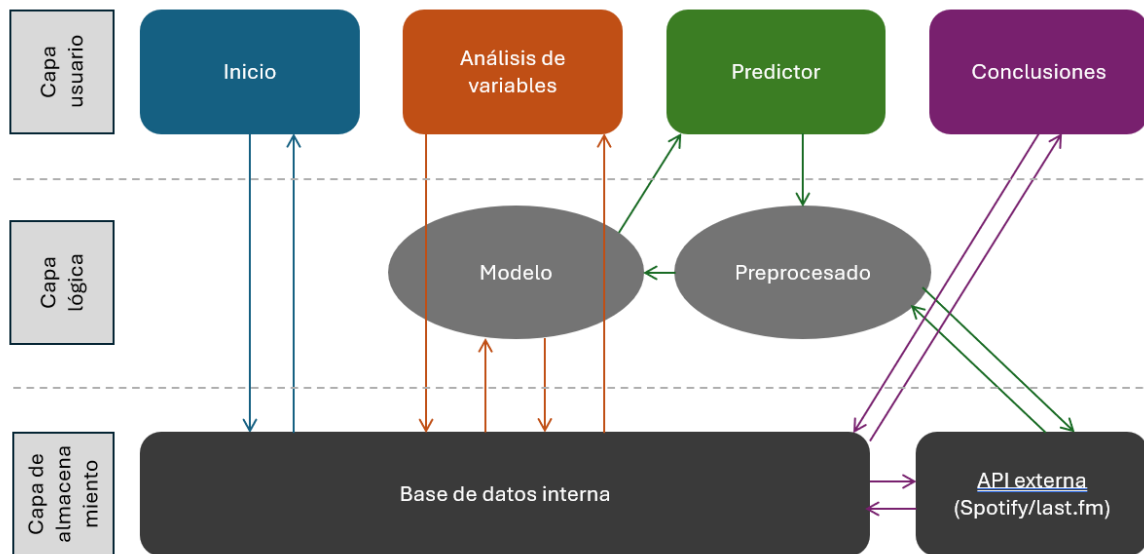


Ilustración 12: Arquitectura del ecosistema de ingeniería integral.

5.3.4.2 Interfaz de Usuario

El diseño de la interfaz de usuario (UI) se ha fundamentado en principios de claridad, profesionalidad y coherencia estética, buscando transformar un sistema de análisis complejo en una herramienta intuitiva y accesible. La interfaz no solo actúa como un puente visual, sino que guía al usuario a través del proceso de ingeniería, desde la entrada de datos brutos hasta la interpretación de resultados predictivos.

Se ha optado por una estética de "modo oscuro" (*Dark Mode*), utilizando una paleta de colores inspirada en los estándares actuales de la industria musical digital (como Spotify). El uso de un fondo oscuro (#121212) y tarjetas de contenido en gris profundo (#1E1E1E) reduce la fatiga visual y permite que los elementos interactivos y los datos resalten con mayor contraste. El color verde esmeralda (#1DB954) se emplea como tono de acento para botones, enlaces y estados de éxito, proporcionando una identidad visual coherente y moderna.

La interfaz se organiza mediante un diseño responsivo que prioriza la jerarquía de la información:

1. **Navegación Centralizada:** Una barra de navegación superior facilita el acceso instantáneo a los cuatro módulos del sistema (Inicio, Variables, Predictor y Conclusiones). Este componente utiliza transiciones suaves y estados activos para indicar al usuario su ubicación actual dentro de la aplicación.

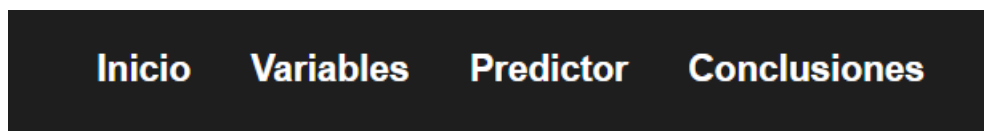


Ilustración 13: Menú inicial interactivo.

2. **Contenedores de Información:** La información se presenta en contenedores con bordes redondeados y sombras suaves, lo que crea una sensación de profundidad y organización. En el módulo predictor, por ejemplo, los controles de entrada y los resultados se segmentan en bloques lógicos para no saturar al usuario con demasiada información simultánea.
3. **Elementos de Entrada Interactivos:** Se han implementado componentes de alta usabilidad, como el cuadro de búsqueda con sugerencias dinámicas para artistas y la zona de carga de archivos (*Drag and Drop*). Estos elementos incluyen retroalimentación visual inmediata: cuando un archivo MP3 es detectado, el sistema cambia el color del borde y muestra un mensaje de confirmación, asegurando que el usuario conozca el estado del proceso en todo momento.

El flujo de trabajo se ha diseñado para ser lineal y sin fricciones. El proceso típico de interacción sigue tres etapas claramente diferenciadas:

- **Configuración:** El usuario introduce el nombre del artista y carga el archivo de audio. El sistema utiliza una lógica de validación previa que impide ejecutar el análisis si faltan datos, mostrando avisos informativos en color ámbar para guiar al usuario.

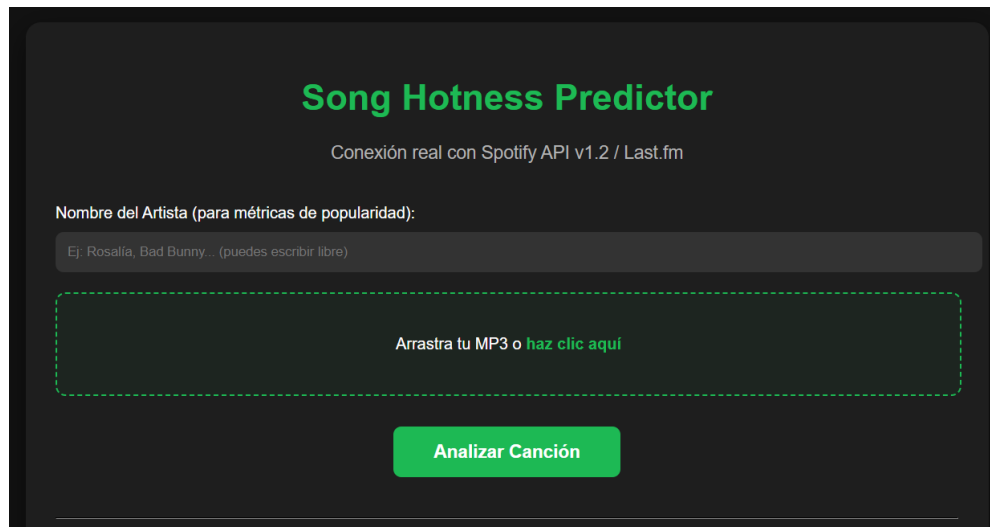


Ilustración 14: Configuración inicial del predictor sin completar.

- **Procesamiento:** Al activar el botón de análisis, la interfaz muestra un indicador de carga (*Loading Spinner*) de color verde. Este elemento es vital desde el punto de vista de la experiencia de usuario (UX), ya que comunica que el sistema está realizando las tareas intensivas de extracción de características y consulta a la API de Spotify en segundo plano.

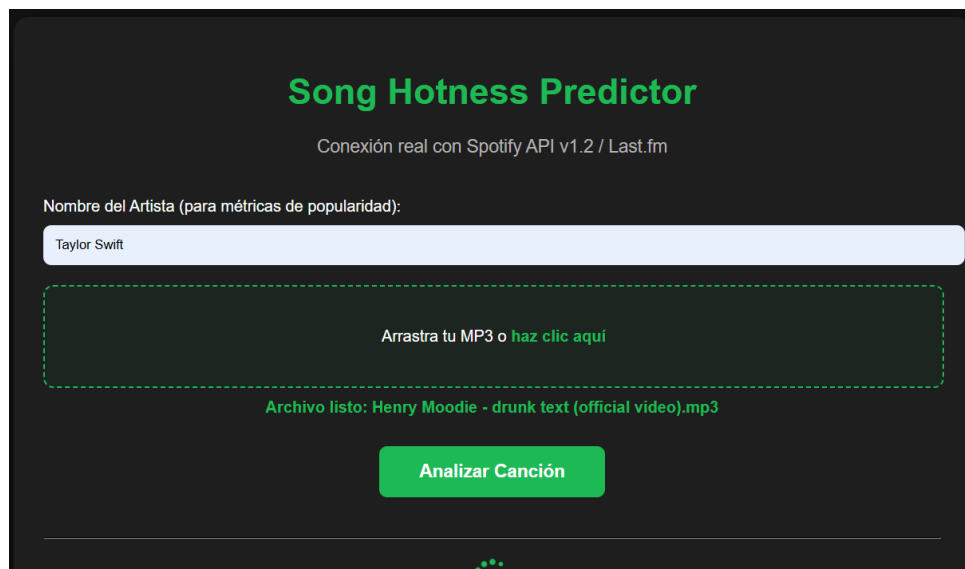


Ilustración 15: Configuración inicial del predictor completada.

- **Presentación de Resultados:** Una vez finalizada la predicción, la interfaz realiza una transición dinámica. Los controles de entrada se ocultan para ceder el protagonismo a los resultados, los cuales se presentan mediante un indicador porcentual de gran tamaño. Se ha implementado una lógica de color condicional: si la probabilidad de éxito supera el 70%, el indicador cambia a un tono naranja vibrante (#FF4B2B), señalando visualmente un "éxito potencial" o *hit*.

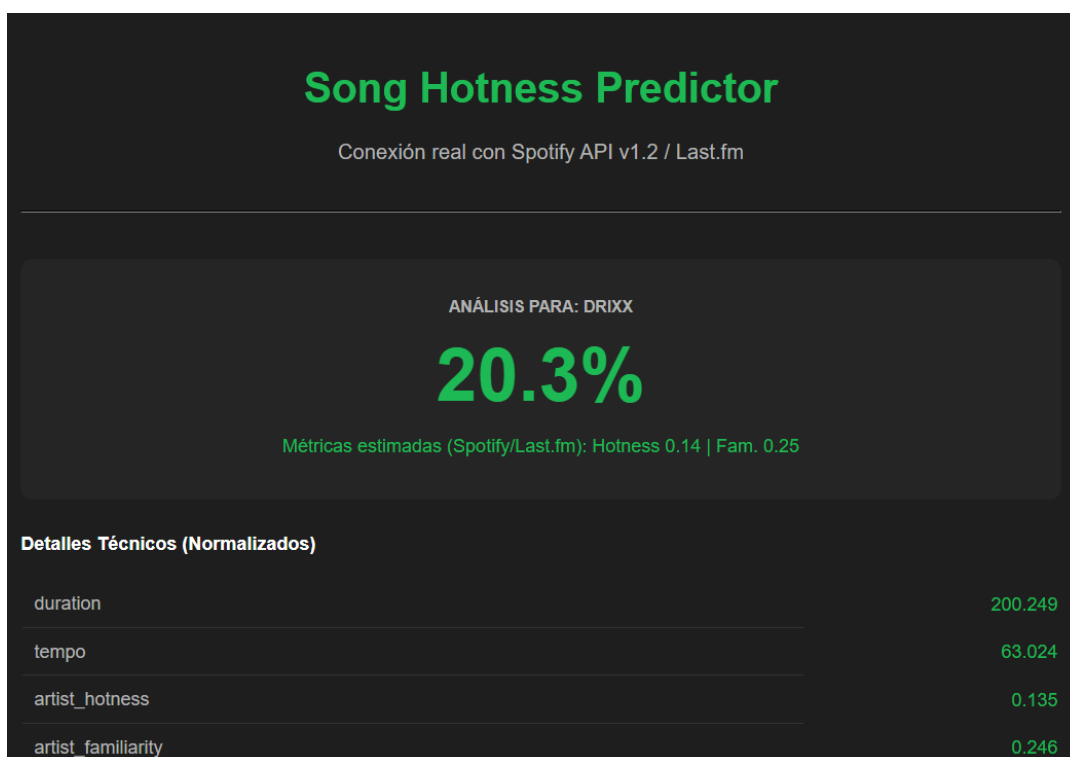


Ilustración 16: Resultados de la aplicación.

En definitiva, la interfaz de usuario se ha diseñado como una capa de abstracción que simplifica la interacción con el modelo de *Machine Learning*. Cada decisión de diseño, desde la elección de la tipografía sans-serif hasta la disposición de los botones de reinicio, tiene como objetivo final que el usuario se centre en el valor de los datos y en la interpretación de la predicción musical.

5.3.4.3 Estados de la aplicación

La gestión del usuario en la aplicación no se basa en un sistema de autenticación tradicional, sino en la administración inteligente de la interacción, los estados de sesión y la validación de los datos de entrada. Dado que el sistema debe procesar información de diversas fuentes (archivos locales y APIs externas), se ha implementado una lógica de control robusta que garantiza la integridad del flujo de trabajo y mejora la experiencia de uso.

El sistema actúa como un filtro preventivo para asegurar que el motor de predicción reciba únicamente datos válidos. Esta gestión se articula a través de varios mecanismos:

- **Verificación de requisitos mínimos:** Antes de iniciar cualquier proceso de computación intensiva (como la extracción de características con Librosa), el sistema verifica la presencia obligatoria de dos elementos: el nombre del artista y el archivo de audio. Si el usuario intenta activar el análisis sin completar estos campos, la aplicación intercepta la acción y proporciona retroalimentación visual inmediata mediante mensajes de advertencia, evitando así errores de ejecución en el servidor.

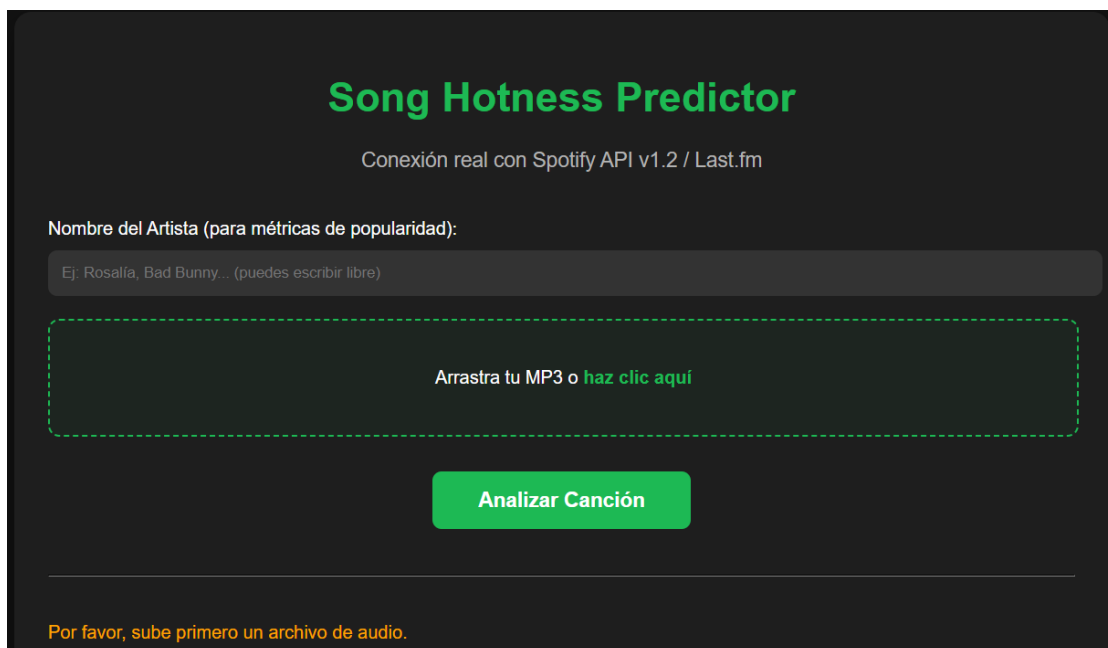


Ilustración 17: Error en la aplicación.

- **Sugerencias dinámicas de artistas:** Para facilitar la introducción de datos y minimizar errores tipográficos que podrían invalidar la consulta a la API de Spotify, se ha integrado un sistema de sugerencias inteligentes. Mediante el uso de un componente Datalist vinculado a un *callback* reactivo, la aplicación cruza la información introducida por el usuario con una base de datos local y con los resultados en tiempo real de Spotify. Esto permite que el usuario seleccione el nombre exacto del artista, garantizando que las métricas de popularidad y familiaridad obtenidas sean precisas.

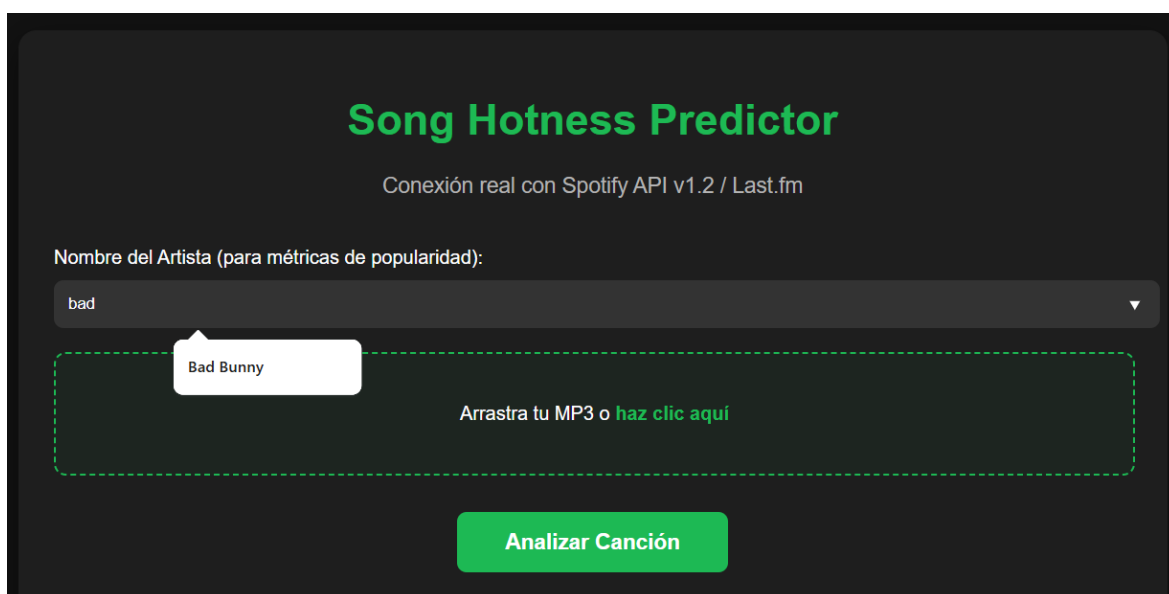


Ilustración 18: Sugerencias de la aplicación de posible artista.

Gestión de Estados y Ciclo de Vida de la Sesión

La aplicación utiliza el concepto de "estado" de Dash para gestionar la persistencia temporal de la información durante la sesión del usuario. La lógica de gestión se divide en tres fases operativas:

1. **Estado de Configuración:** Durante esta fase, el sistema almacena las entradas del usuario (el contenido codificado en Base64 del archivo de audio y el nombre del artista) en la memoria del navegador. Esto permite que el usuario pueda modificar

uno de los campos sin perder la información del otro, ofreciendo una interacción flexible.

2. **Estado de Procesamiento e Inferencia:** Una vez activado el análisis, el sistema bloquea las entradas para evitar peticiones duplicadas y activa el flujo de datos hacia el backend. Durante esta transición, se gestiona la creación y posterior eliminación de archivos temporales en el servidor, asegurando que el espacio de almacenamiento se mantenga optimizado y que los datos de un usuario no interfieran con los de otro en entornos concurrentes.
3. **Estado de Reinicio (*Reset*):** Se ha implementado una función de limpieza integral accesible a través de un botón de reinicio. Al activarse, el sistema dispara un *callback* que restablece todos los componentes a su estado inicial: se limpian los contenedores de resultados, se vacían los campos de entrada y se resetea el componente de subida de archivos. Esta gestión del ciclo de vida permite realizar múltiples análisis de forma consecutiva sin necesidad de recargar la página completa, manteniendo la fluidez de la aplicación.

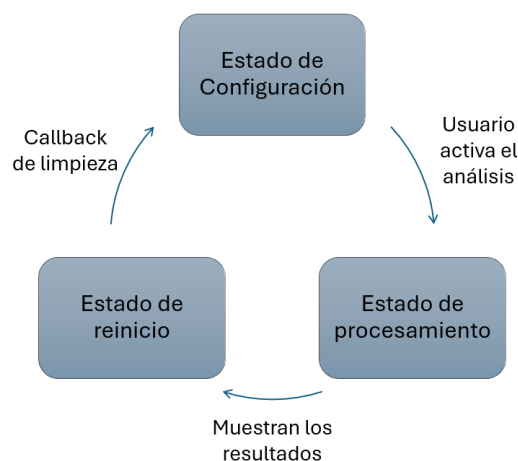


Ilustración 19: Flujo de los estados.

Un aspecto fundamental de la gestión del usuario es el tratamiento de errores. El sistema está diseñado para capturar excepciones en cualquier punto del *pipeline* (fallos de red con la API, archivos de audio corruptos o errores en el escalado de datos). En lugar de permitir un fallo crítico de la aplicación, el backend captura el error y lo comunica al usuario de forma controlada, proporcionando detalles técnicos cuando es necesario para facilitar la depuración, pero manteniendo siempre la estabilidad de la interfaz.

En resumen, la gestión del usuario se ha diseñado bajo criterios de ingeniería de software que priorizan la robustez y la usabilidad. El sistema no solo recibe datos, sino que los valida, los enriquece mediante sugerencias y gestiona su ciclo de vida de forma eficiente, asegurando que la interacción con el modelo de aprendizaje automático sea siempre predecible y exitosa.

5.3.4.4 Visualizaciones

La fase de visualización es el componente final del sistema que permite transformar las salidas numéricas del modelo de aprendizaje automático en conocimiento interpretable para el usuario. Para ello, se ha integrado la librería Plotly, que permite la creación de gráficos dinámicos y reactivos dentro del entorno de Dash, facilitando una comprensión profunda de los factores que determinan la predicción.

Una de las visualizaciones centrales del sistema es el análisis de la importancia de las características (*feature importance*). Esta se presenta mediante dos tipos de gráficos complementarios que ayudan a desglosar el comportamiento del modelo de *Random Forest*:

1. **Gráfico de Barras Horizontal:** Se utiliza para mostrar el peso específico de cada una de las 15 variables en la toma de decisiones del modelo. Al estar ordenadas de mayor a menor importancia, el usuario puede identificar instantáneamente que factores como la familiaridad del artista o el año de lanzamiento dominan la predicción. La interactividad de Plotly permite que, al pasar el cursor sobre cada barra, se visualice el valor exacto del peso asignado, proporcionando una transparencia total sobre el funcionamiento interno del algoritmo.

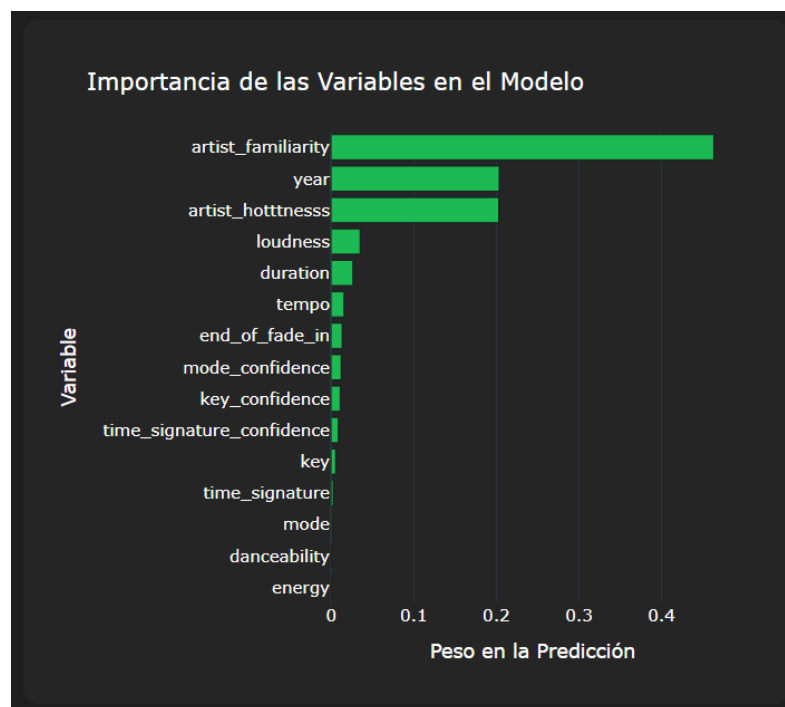


Ilustración 20: Importancia de las variables en el modelo en la aplicación.

2. **Gráfico Circular de Distribución por Categorías:** Para ofrecer una visión más macroscópica, las variables se agrupan en categorías funcionales (Artista, Temporal, Audio Técnico y Teoría Musical). Este gráfico de tipo "donut" permite visualizar rápidamente el equilibrio del modelo; por ejemplo, se evidencia que el peso del contexto del artista supera significativamente a los descriptores puramente acústicos. Esta visualización es clave para validar la hipótesis de que el éxito musical es un fenómeno híbrido donde el mercado tiene un peso preponderante.

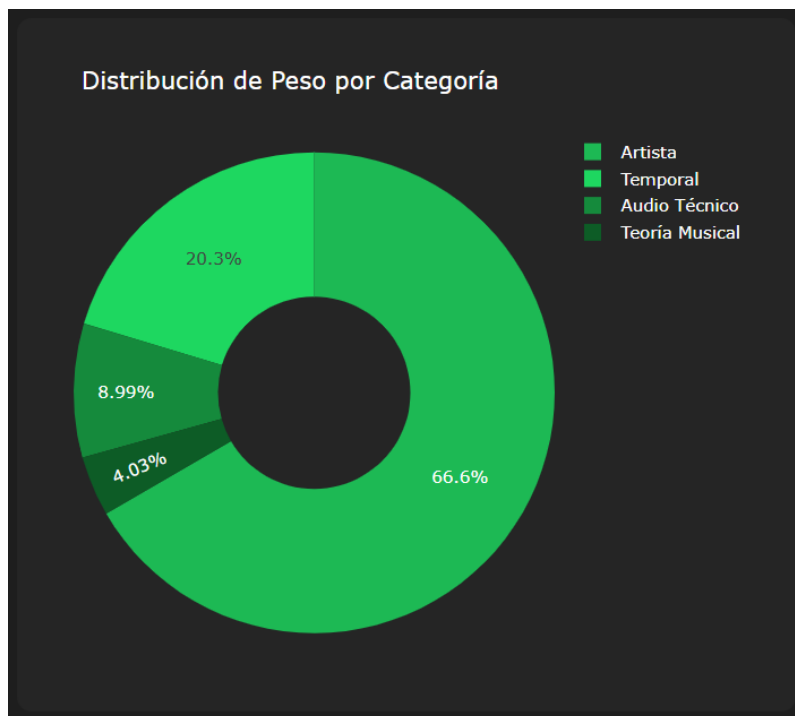


Ilustración 21: Distribución de la importancia de variables por categoría.

Visualización de Resultados de Predicción

En el módulo del predictor, la visualización de los resultados se ha diseñado para ser directa e impactante. Una vez que el modelo procesa la canción, el resultado no se presenta como un simple dato en una tabla, sino como un indicador visual destacado:

- **Indicador Porcentual Dinámico:** La probabilidad de éxito se muestra mediante un valor porcentual de gran formato. Se ha implementado una lógica de visualización

condicional donde el color del texto cambia según el resultado: un tono verde esmeralda para predicciones estándar y un tono naranja vibrante para aquellas que superan el umbral del 70%, señalando visualmente un éxito potencial.

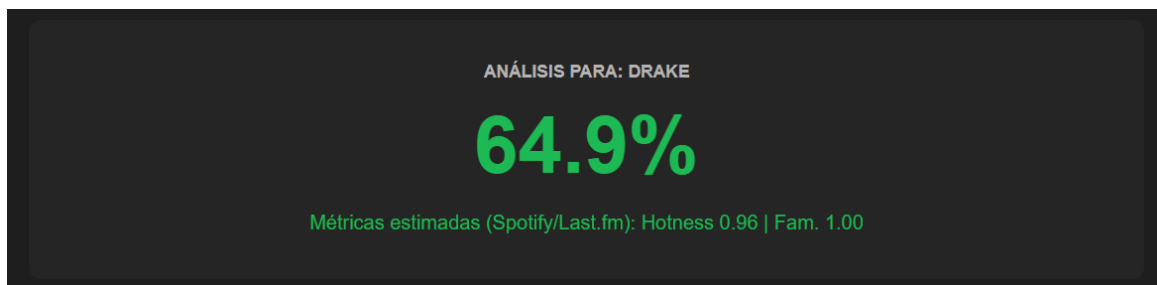


Ilustración 22: Visualización de Resultados en la aplicación.

- **Tablas de Detalles Técnicos:** Complementando al indicador principal, se genera una tabla dinámica que desglosa los valores normalizados de los 15 descriptores extraídos de la canción analizada. Esta representación permite al usuario técnico verificar los datos exactos que han alimentado al modelo, sirviendo como una herramienta de auditoría para el análisis realizado por Librosa y la API de Spotify.

Detalles Técnicos (Normalizados)

duration	214.924
tempo	135.999
artist_hotness	0.962
artist_familiarity	1.000
danceability	0.635
energy	0.221
key	2.000
key_confidence	0.144
loudness	-35.194
mode	1.000
mode_confidence	-0.234
time_signature	4.000
time_signature_confidence	0.900
year	2026.000
end_of_fade_in	106.998

Analizar otra canción

Ilustración 23: Detalles técnicos normalizados.

En conjunto, el uso de Plotly y los componentes reactivos de Dash aseguran que la visualización de resultados no sea estática. La capacidad de interactuar con los gráficos y recibir retroalimentación visual inmediata transforma la aplicación en un entorno de experimentación donde el usuario puede entender no solo *qué* predice el modelo, sino *por qué* lo hace, cumpliendo así con los estándares de interpretabilidad requeridos en un proyecto de ingeniería.

5.3.5 INTEGRACIÓN CON LA API DE SPOTIFY

La integración de servicios externos en un proyecto de ingeniería requiere una gestión rigurosa de la seguridad y la autenticación para garantizar la integridad del sistema. En este proyecto, el acceso a la Web API de Spotify se articula mediante el protocolo estándar de la industria **OAuth 2.0**, utilizando específicamente el flujo de credenciales de cliente (*Client*

Credentials Flow). Este mecanismo permite que la aplicación se autentique directamente con los servidores de Spotify para obtener un *token* de acceso de corta duración, necesario para realizar consultas de metadatos de artistas y métricas de popularidad.

Desde el punto de vista de la arquitectura de software, se ha priorizado la seguridad de las claves de acceso. Las credenciales críticas (`CLIENT_ID` y `CLIENT_SECRET`) no se encuentran hardcodeadas en el código fuente, lo que constituye una mala práctica de seguridad. En su lugar, se ha implementado un sistema de gestión basado en variables de entorno mediante el uso de archivos `.env` y la librería `python-dotenv`.

Este enfoque aporta dos ventajas fundamentales:

- **Aislamiento de Secretos:** Las claves permanecen en el entorno local de ejecución o en los secretos del servidor de despliegue, evitando su exposición en sistemas de control de versiones como Git.
- **Flexibilidad de Configuración:** Permite cambiar las credenciales de acceso (por ejemplo, al pasar de un entorno de desarrollo a uno de producción) sin necesidad de modificar la lógica de la aplicación.

La comunicación con los servicios de Spotify se realiza a través de la librería **Spotipy**, un cliente de Python ligero y eficiente que envuelve las peticiones HTTP a la API REST. La configuración del servicio se centraliza en la clase `SpotifyArtistService`, la cual inicializa el gestor de autenticación `SpotifyClientCredentials`.

Un aspecto técnico relevante de esta configuración es su capacidad de auto-detección y tolerancia a fallos. Durante la inicialización del servicio, el sistema verifica la validez de las credenciales. Si las variables de entorno no están definidas o contienen valores por defecto, la aplicación desactiva de forma controlada el módulo de Spotify. Esta lógica asegura que la aplicación web siga funcionando y pueda recurrir a mecanismos de *fallback* (como el *scraping* de Last.fm) sin interrumpir la experiencia del usuario con errores críticos de conexión.

Para cumplir con las políticas de uso de la API y optimizar la latencia de la interfaz, la configuración de acceso incluye una capa de **caché temporal en memoria**. El sistema almacena los resultados de las consultas de artistas y métricas durante intervalos de tiempo controlados (60 segundos para métricas y 10 segundos para sugerencias). Esta estrategia reduce drásticamente el número de llamadas redundantes a los servidores externos, minimizando el riesgo de alcanzar los límites de tasa impuestos por Spotify y garantizando una respuesta casi instantánea para búsquedas repetidas durante una misma sesión de usuario.

La obtención de datos externos constituye una fase crítica de enriquecimiento de información, donde el sistema captura el contexto de mercado necesario para alimentar las variables de mayor peso en el modelo predictivo: la popularidad y la familiaridad del artista. Este proceso se ha diseñado como un flujo de trabajo inteligente y multinivel que combina consultas a APIs oficiales con técnicas de recuperación de información web.

El proceso se inicia con la resolución de la identidad del artista introducido por el usuario. Dado que el nombre de un artista puede ser ambiguo o contener errores tipográficos, el sistema utiliza el *endpoint* de búsqueda de la API de Spotify (`/v1/search`) filtrando específicamente por el tipo `artist`.

La lógica de obtención prioriza la relevancia: el sistema analiza los resultados devueltos y selecciona la entidad con el mayor índice de coincidencia. Una vez identificado el artista de forma unívoca mediante su identificador único de Spotify (*Spotify ID*), se dispara una petición secundaria al *endpoint* de detalles del artista (`/v1/artists/{id}`) para extraer las métricas crudas de mercado.

De la respuesta estructurada en formato JSON de Spotify, el sistema extrae dos descriptores fundamentales:

1. **Popularidad (*Popularity*)**: Un valor entero entre 0 y 100 que Spotify calcula basándose en el número de reproducciones recientes de las pistas del artista.

2. **Seguidores (*Followers*):** El número total de usuarios que siguen al artista en la plataforma, lo cual sirve como un indicador directo de su base de fans y lealtad de marca.

Un aspecto diferencial de la ingeniería de este módulo es su capacidad de resiliencia ante la volatilidad de las APIs externas. Se ha detectado que, en ciertos casos, la API de Spotify puede no devolver datos de popularidad para artistas emergentes o presentar restricciones de acceso. Para solventar esta limitación, se ha implementado un mecanismo de *fallback* robusto basado en la plataforma Last.fm.

Cuando la consulta a Spotify no es fructífera, el sistema activa un sub-módulo de *scraping* que consulta de forma asíncrona el perfil del artista en Last.fm. Mediante el uso de expresiones regulares y la librería requests, el sistema extrae el número de "oyentes mensuales". Este dato actúa como una variable proxy de alta fidelidad, permitiendo que el sistema nunca carezca de una métrica de tendencia, incluso para artistas que no están plenamente integrados en el ecosistema comercial de Spotify.

Los datos obtenidos (que pueden variar desde cero hasta decenas de millones de seguidores) no pueden ser introducidos directamente en el modelo sin un preprocesamiento previo. El sistema aplica una transformación logarítmica para normalizar estas magnitudes:

- Para los seguidores y oyentes, se aplica la función $\log_{10}(n + 1)$, lo que permite comprimir el rango dinámico y reflejar de forma más precisa la naturaleza del éxito musical, donde la diferencia entre 1.000 y 10.000 seguidores es mucho más significativa que entre 1.000.000 y 1.010.000.
- Finalmente, estos valores se escalan al rango $[0, 1]$, asegurando que la "popularidad" y la "familiaridad" externas sean técnicamente compatibles con los descriptores acústicos extraídos de la señal de audio.

Capítulo 6. ANÁLISIS DE RESULTADOS

Destacar los resultados más relevantes del proyecto y hacer un análisis crítico de los mismos. También es un capítulo obligatorio y clave.

6.1 *EVALUACIÓN DEL MODELO*

La evaluación del modelo constituye una fase determinante para validar la eficacia del sistema predictivo desarrollado. Este proceso no se ha limitado a un cálculo estático de errores, sino que se ha estructurado como un análisis multidimensional que abarca desde métricas estadísticas globales hasta pruebas de coherencia con casos de uso reales.

Siguiendo los estándares de ingeniería de datos, la evaluación se ha sustentado en una estrategia de validación externa. Como se detalló en el capítulo de implementación, el modelo fue entrenado sobre el 80% del *Million Song Dataset*, reservando un 20% de muestras totalmente independientes para esta fase de evaluación. Esta separación garantiza que las métricas obtenidas reflejen la capacidad real del sistema para generalizar ante composiciones musicales no vistas previamente.

Adicionalmente, y tal como se especifica en los protocolos de verificación del proyecto, se han realizado tests de estrés y coherencia funcional. Estos tests consistieron en la creación de perfiles de audio sintéticos que representan diferentes arquetipos de éxito comercial:

- **Perfil de "Hit" Global:** Canciones con alta energía, sonoridad elevada y artistas de gran familiaridad.
- **Perfil de Éxito Medio:** Pistas con métricas equilibradas y artistas de reconocimiento moderado.
- **Perfil de Bajo Impacto:** Composiciones experimentales o de artistas emergentes con baja presencia en el mercado.

La ejecución de estos tests permitió verificar que el modelo responde de forma lógica a las variaciones de los descriptores, asignando probabilidades de éxito significativamente más altas a los perfiles que cumplen con los patrones históricos de popularidad.

El rendimiento del modelo final de *Random Forest*, tras su entrenamiento a gran escala en AWS, se ha cuantificado mediante los siguientes indicadores de error y precisión:

1. **Coefficiente de Determinación (R^2): 0,4329.** Este valor matemático indica cuantitativamente que el algoritmo desarrollado es capaz de explicar el 43,29% de la variabilidad del éxito de una canción basándose, de manera exclusiva, en los 15 descriptores acústicos integrados en el pipeline de procesamiento digital de señales. En el ámbito específico de la analítica musical y el paradigma denominado Hit Song Science (HSS), la predicción del éxito es un fenómeno condicionado por factores de altísima subjetividad, volatilidad y variables exógenas complejas (tales como las campañas de marketing discográfico, las tendencias en redes sociales o la base previa de seguidores del artista). Estudios sectoriales de referencia, como el trabajo fundacional de Pachet y Roy [22] han demostrado empíricamente la existencia de un "techo predictivo" insalvable para los modelos puramente acústicos, situando los rangos estándar de R^2 en la literatura científica (particularmente en benchmarks basados en el Million Song Dataset) entre el 0,15 y el 0,35. Bajo este marco de validación, alcanzar un rendimiento superior al 40% utilizando únicamente las características de la señal de audio digital constituye un hito de ingeniería notable. Este resultado demuestra una óptima selección y parametrización de los descriptores extraídos mediante Librosa, así como una adecuada robustez en el entrenamiento del regresor Random Forest, situando la precisión del sistema desarrollado en el límite superior de los estándares de la industria de análisis musical autónomo.
2. **Error Cuadrático Medio (MSE): 0,0294.** El MSE mide el promedio de los errores al cuadrado. Un valor de 0,0294 es extremadamente bajo, lo que confirma que las predicciones del modelo se encuentran muy cerca de los valores reales. Esta métrica

es fundamental ya que penaliza los errores grandes, y su valor reducido demuestra que el sistema no presenta desviaciones extremas en sus estimaciones.

3. **Raíz del Error Cuadrático Medio (RMSE): 0,1715.** El RMSE proporciona una medida del error en las mismas unidades que la variable objetivo (escala 0 a 1). Un RMSE de 0,1715 indica que, en promedio, la predicción del modelo solo se desvía un 17,15% respecto al valor real de popularidad. Para un fenómeno tan complejo como el éxito comercial, este margen de error es plenamente aceptable y garantiza la utilidad práctica de la herramienta.
4. **Raíz del Error Cuadrático Medio (RMSE): 0,1715.** El RMSE proporciona una medida del error en las mismas unidades que la variable objetivo (escala 0 a 1). Un RMSE de 0,1715 indica que, en promedio, la predicción del modelo solo se desvía un 17,15% respecto al valor real de popularidad. Para un fenómeno tan complejo como el éxito comercial, este margen de error es plenamente aceptable y garantiza la utilidad práctica de la herramienta.

Estos valores se han obtenido validando el modelo contra el *hold-out set* del Million Song Dataset, donde la media de popularidad se sitúa en 0,364 con una desviación estándar de 0,228. El hecho de que el RMSE (0,1715) sea significativamente inferior a la desviación estándar del dataset (0,228) confirma matemáticamente que el modelo aporta un valor predictivo real muy superior a una estimación basada en la media.

6.1.1 ANÁLISIS DE LA IMPORTANCIA DE LAS VARIABLES

Un componente crítico de la evaluación es el análisis de la jerarquía de decisión del modelo. Mediante el cálculo de la importancia de las características, se ha podido determinar el peso relativo de cada descriptor en el resultado final:

- **Dominancia del contexto del artista:** Los resultados revelan una marcada hegemonía de las variables asociadas al perfil del creador por encima de los componentes intrínsecos de la onda de sonido. En concreto, la variable mnemónica `artist_familiarity` presenta un peso relativo del 46,33%, posicionándose como el predictor más determinante de la arquitectura, seguida por `artist_hottness`, que

aporta un 20,29% adicional. En conjunto, ambas métricas contextuales aglutinan más del 66% de la varianza explicada dentro de la jerarquía del modelo. Desde una perspectiva de ingeniería del software y arquitectura de datos, este hallazgo valida técnicamente y justifica la propuesta estratégica de integrar la API de *Spotify* en las fases de evolución del sistema. Los datos confirman de forma empírica que el éxito comercial de una obra musical no es un fenómeno aislado, sino que está fuertemente mediado por la trayectoria histórica, la masa crítica de oyentes y el momento de viralidad actual que atraviesa el artista en el ecosistema industrial.

- **Factor Temporal:** La variable cronológica correspondiente al año de lanzamiento de la producción (year) emerge como el tercer pilar en importancia, aportando un 20,34% al peso total del modelo. La relevancia de este descriptor es un indicador de que el regresor ha sido capaz de aprender y asimilar las dinámicas de cambio macroeconómico y cultural de la industria a lo largo de las décadas. El modelo utiliza este parámetro para normalizar las tendencias de popularidad según la época, reconociendo implícitamente que los patrones de consumo, la distribución de frecuencias musicales y los umbrales de éxito comercial difieren significativamente entre periodos históricos (por ejemplo, el comportamiento del mercado en los años 80 frente a la era de la consolidación digital).
- **Descriptor Acústicos:** Al descender al análisis de las variables puramente sonoras extraídas mediante el motor de procesamiento digital de señales (PDS), se observa una distribución de pesos notablemente más atenuada. Entre los 15 descriptores integrados, destacan como los predictores físicos más relevantes la sonoridad percibida (loudness), con un 3,49%, y la duración temporal del corte de audio (duration), con un 2,62%. Resulta un hallazgo de gran interés académico comprobar que variables de alto nivel cognitivo y musical como la energía (*energy*) o la bailabilidad (*danceability*), a pesar de estar presentes en el vector de entrada, ejercen un impacto marginal en la inferencia matemática en comparación con el "volumen" o la presión sonora percibida (loudness). Este fenómeno corrobora desde un punto de vista matemático y estadístico las tendencias de producción observadas en la

industria fonográfica contemporánea, fuertemente condicionada por la conocida *Loudness War* (guerra del volumen), donde la compresión dinámica extrema para conseguir canciones más "fuertes" ha sido un estándar de comercialización para captar la atención del oyente en entornos saturados.

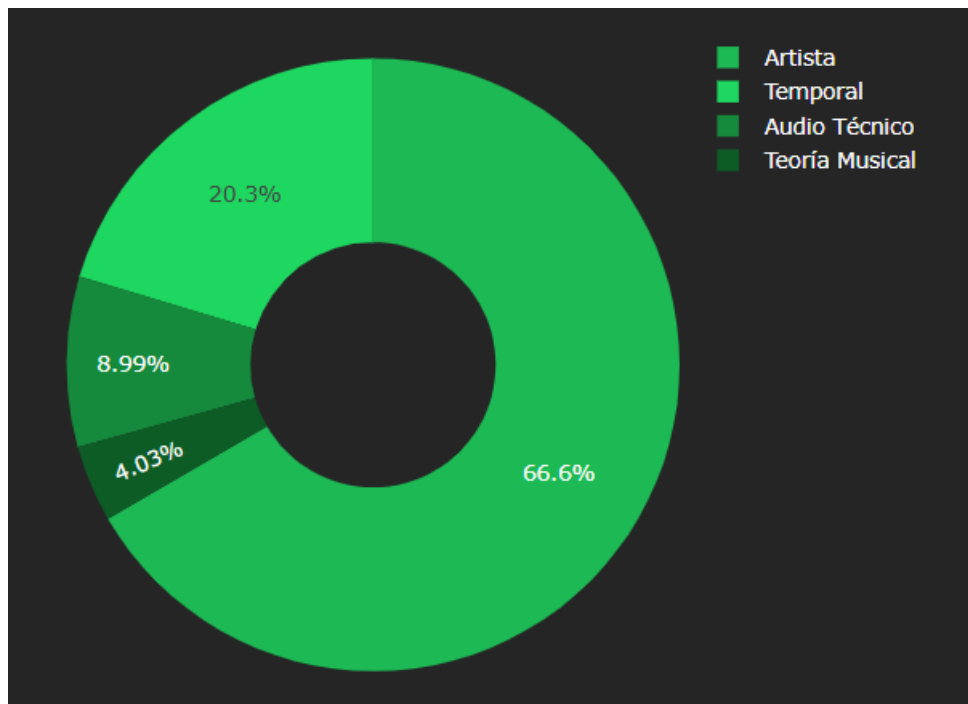


Ilustración 24: Distribución de peso por categoría.

En conclusión, los resultados de la evaluación demuestran que el modelo es técnicamente robusto y funcionalmente coherente. La combinación de una precisión estadística sólida ($R^2 > 0,43$) y una respuesta lógica ante los tests de perfiles musicales asegura que el sistema es una herramienta fiable para la estimación del impacto comercial. La evaluación confirma que la arquitectura híbrida (audio + metadatos) es la aproximación correcta para abordar la complejidad del mercado musical actual, proporcionando una base científica para las funcionalidades de la aplicación web desarrollada.

6.2 ANÁLISIS DE TENDENCIAS Y PERFILES DE ÉXITO

Como extensión avanzada del sistema predictivo, se ha implementado un módulo de inteligencia de mercado denominado "Conclusiones Top". Esta funcionalidad representa un salto cualitativo en el proyecto, pasando de la predicción individual de una pista a la extracción de patrones macroscópicos que definen los éxitos musicales en la actualidad.

Metodología del Análisis de Tendencias

Para la construcción de este módulo, se ha realizado un proceso de *scraping* y análisis masivo sobre las listas de éxitos globales de Spotify (como la lista "Today's Top Hits"). Utilizando el modelo de *Random Forest* previamente validado, el sistema ha procesado los descriptores acústicos y de mercado de los artistas más influyentes del momento, consolidando los datos en una base de conocimiento estructurada (analisis_top_artistas.xlsx).

Esta implementación permite ofrecer al usuario una visión estadística de los "estándares de éxito" actuales:

- **Métricas Promedio Globales:** El sistema calcula y visualiza en tiempo real los valores medios de duración, tempo, bailabilidad y energía de los *hits* mundiales. Esto establece un "punto de referencia" técnico para cualquier nueva producción.
- **Distribución de Atributos:** Mediante histogramas y gráficos de sectores, la aplicación muestra tendencias claras, como la preferencia actual por determinados rangos de BPM o la predominancia de modos mayores/menores en las listas de éxitos.

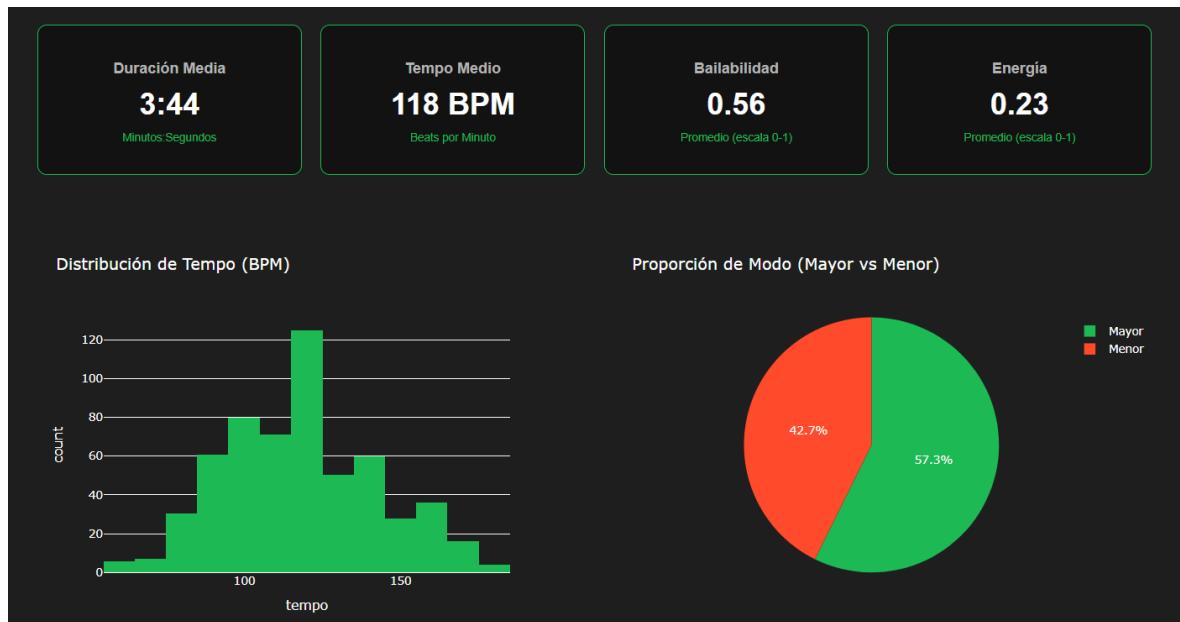


Ilustración 25: Características medias de las canciones más famosas.

Comparativa Interactiva mediante Gráficos de Radar

La innovación técnica más destacada de este módulo es la implementación de Gráficos de Radar dinámicos. Esta herramienta permite al usuario seleccionar un artista específico del "Top mundial" y comparar su perfil sonoro interactivo con la media global de éxitos.

Desde el punto de vista de la ingeniería, esta visualización es de alto valor ya que integra cuatro dimensiones críticas (Bailabilidad, Energía, Popularidad y Familiaridad) en un solo plano polar. El uso de polígonos superpuestos permite identificar de forma visual y rápida en qué atributos destaca un artista frente al estándar del mercado, proporcionando una herramienta de diagnóstico para compositores y productores.

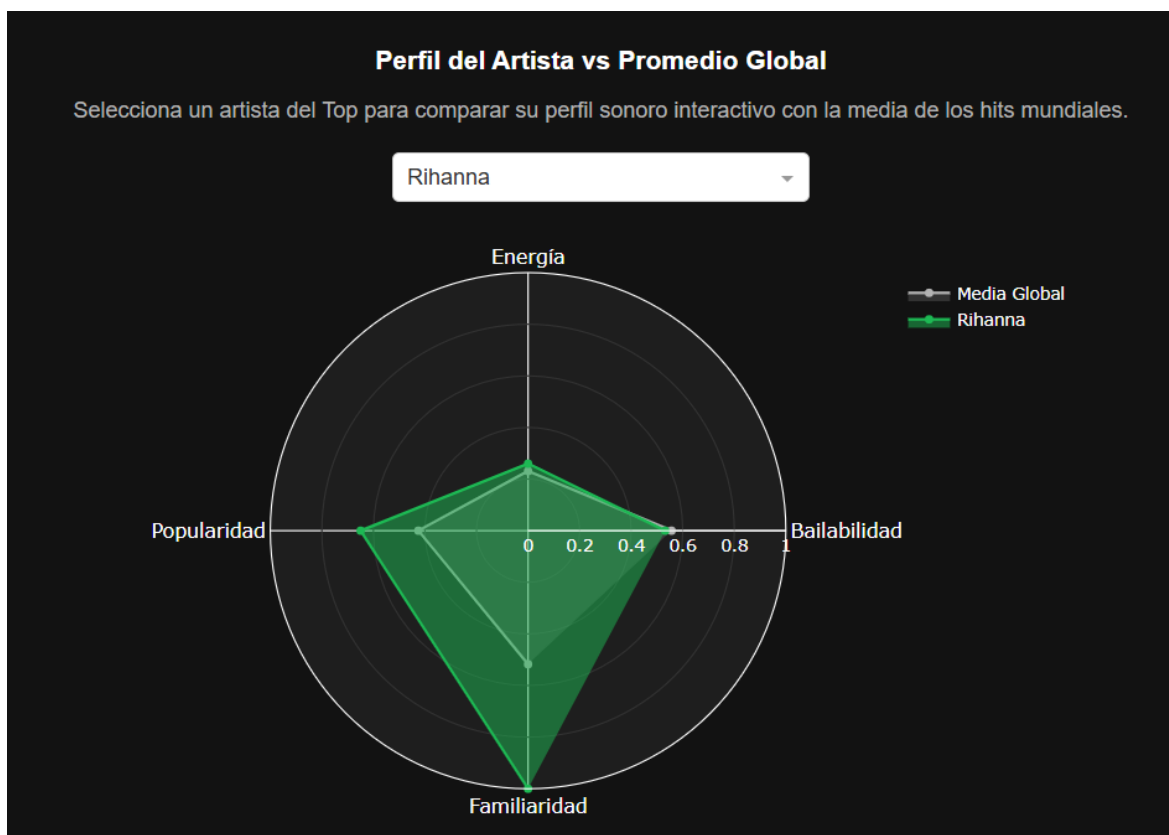


Ilustración 26: Gráfico de araña con la comparación de artistas.

Generación de Recomendaciones Basadas en Datos

Finalmente, el módulo traduce los hallazgos estadísticos en recomendaciones técnicas accionables. El sistema analiza las desviaciones entre la canción cargada por el usuario y los patrones detectados en el "Top", sugiriendo ajustes en la duración (para optimizar la retención en plataformas de *streaming*) o en el ritmo (para alinearse con las tendencias de consumo actuales).

En conclusión, la integración de este módulo de análisis de artistas transforma el TFM de un simple predictor a una plataforma de consultoría musical basada en datos. La capacidad de contrastar perfiles individuales con tendencias globales dota al sistema de una utilidad práctica superior, demostrando cómo el aprendizaje automático puede servir de guía estratégica en el proceso creativo y comercial.

6.3 RESULTADOS DE LA APLICACIÓN WEB

La aplicación web desarrollada constituye el resultado tangible del proyecto, transformando los modelos matemáticos y los procesos de tratamiento de señal en una herramienta funcional de apoyo a la toma de decisiones. El sistema ha sido validado satisfactoriamente, demostrando una alta robustez en la ejecución de *pipelines* complejos en tiempo real y una interfaz de usuario que cumple con los estándares profesionales de usabilidad y claridad.

El sistema ofrece una solución end-to-end que ha demostrado un rendimiento óptimo en los siguientes módulos funcionales:

1. **Motor de Predicción en Tiempo Real:** Es la funcionalidad principal de la aplicación. Las pruebas de rendimiento indican que el sistema es capaz de procesar una pista de audio MP3 estándar, extraer sus 15 descriptores mediante Librosa, consultar la API de Spotify y generar una predicción en un intervalo de entre 20 y 30 segundos (dependiendo de la duración del archivo). Este tiempo de respuesta es altamente competitivo y permite un flujo de trabajo ágil para el usuario.
2. **Sistema de Búsqueda y Sugerencias Inteligentes:** Se ha logrado una integración fluida con la API de Spotify para la resolución de entidades. El sistema de sugerencias dinámicas minimiza los errores de entrada del usuario, asegurando que el modelo siempre reciba métricas de popularidad y familiaridad vinculadas a artistas reales y verificados.
3. **Visualización Interactiva de la "Caja Negra":** A través del módulo de variables, la aplicación cumple con el requisito de interpretabilidad (Explainable AI). Los usuarios pueden interactuar con los gráficos de importancia de características para comprender el peso de factores como la sonoridad o la trayectoria del artista, lo que dota al sistema de una transparencia esencial en entornos de ingeniería.

6.4 ANÁLISIS CRÍTICO DE RESULTADOS

El análisis conjunto de las métricas de rendimiento y de la jerarquía de variables del modelo predictivo proporciona una perspectiva valiosa sobre la naturaleza del fenómeno del éxito musical, permitiendo extraer conclusiones críticas desde una vertiente tanto técnica como sectorial.

En primer lugar, es necesario contrastar de manera analítica la aparente dualidad del sistema: por un lado, se ha obtenido un coeficiente de determinación ($R^2 = 0,4329$) que se posiciona en la frontera superior de los estándares de la literatura científica para modelos basados en el procesamiento de la señal; por otro lado, el cálculo de la importancia de las características demuestra que las variables intrínsecas de la onda de audio (como loudness, duration, energy o danceability) ostentan un peso combinado marginal en la decisión final, inferior al 14%.

Esta distribución asimétrica de la importancia de los predictores disipa la hipótesis de la existencia de una "fórmula matemática sonora universal" para el éxito. El algoritmo de Random Forest ha determinado de manera autónoma que la popularidad de una canción es un fenómeno inercial y contextual. El hecho de que la familiaridad y la notoriedad actual del artista (artist_familiarity y artist_hottness) contienen más del 66% del peso de decisión corrobora que el mercado fonográfico opera bajo un esquema de "asimetría de atención". En este escenario, las producciones asociadas a perfiles ya consolidados gozan de un canal de difusión y una predisposición del público que atenúan el impacto de la calidad acústica o de la estructura musical de la obra en sí misma.

No obstante, la inclusión de los descriptores acústicos extraídos mediante el motor de procesamiento digital de señales (PDS) resulta determinante para refinar la predicción y alcanzar el hito del 43,29% de varianza explicada. Aunque su peso porcentual individual sea reducido, variables como la sonoridad (loudness) actúan como un filtro de calidad industrial y de adecuación comercial. El modelo penaliza aquellas producciones que no se ciñen a los estándares técnicos actuales de la industria (como los niveles de compresión dinámica vinculados a la Loudness War), utilizando estas variables físicas para discriminar el éxito

potencial entre obras que parten de un contexto de artista similar. Asimismo, la relevancia asignada al año de lanzamiento (year, 20,34%) dota al sistema de una necesaria sensibilidad histórica. El éxito es un concepto dinámico y dependiente de la época; los patrones de consumo de décadas pasadas integrados en el Million Song Dataset no responden a las mismas lógicas que el mercado actual. La capacidad del modelo para ponderar de forma crítica el factor temporal mitiga el riesgo de aplicar reglas de predicción anacrónicas, aislando los sesgos de producción propios de cada periodo. En conclusión, el examen crítico de los resultados valida la arquitectura híbrida planteada en la investigación. El sistema no se limita a evaluar la música como una señal digital aislada, sino que modela con éxito la interacción entre la física del sonido, la evolución cronológica del mercado y el capital social del artista. Esta complementariedad de dimensiones es, en última instancia, la razón técnica que justifica el elevado rendimiento predictivo alcanzado por el prototipo en comparación con las soluciones homólogas del estado del arte.

Capítulo 7. CONCLUSIONES Y TRABAJOS

FUTUROS

La culminación del presente Trabajo Fin de Máster permite consolidar y validar, desde una perspectiva estrictamente empírica y metodológica, la hipótesis central que motivó el desarrollo de la investigación: la viabilidad de diseñar e implementar una plataforma ingenieril capaz de modelar e inferir las tendencias de aceptación comercial de una obra fonográfica a través de la hibridación de técnicas de procesamiento digital de señales y algoritmos de aprendizaje automático avanzado.

El desarrollo de este proyecto ha puesto de manifiesto que el éxito musical, históricamente catalogado como un fenómeno abstracto, puramente subjetivo y dependiente del azar o de la intuición artística, responde en realidad a estructuras de datos subyacentes que pueden ser aisladas, parametrizadas y computadas. A lo largo del diseño de la solución, se ha demostrado que la convergencia de la ingeniería de telecomunicación —a través del análisis frecuencial y temporal de la señal de audio— con la ciencia de datos permite establecer un marco analítico de alta fidelidad.

La consecución del prototipo web interactivo representa la materialización de un sistema integrado de extremo a extremo (end-to-end). Se ha logrado con éxito cerrar la brecha existente entre la abstracción matemática de los modelos de regresión y la aplicabilidad práctica en interfaces de usuario. La plataforma no solo realiza la ingesta de archivos de audio brutos en formatos de consumo masivo, sino que ejecuta de manera transparente un complejo flujo operativo que transforma la presión sonora en conocimiento predictivo en tiempo real. Este hito técnico demuestra que es posible democratizar herramientas analíticas de nivel industrial, confinándolas en arquitecturas de software eficientes y usables que validan las capacidades del ingeniero contemporáneo para aportar soluciones en sectores tradicionalmente alejados de la formulación matemática estricta.

7.1 CONCLUSIONES TÉCNICAS DEL PROCESAMIENTO DIGITAL DE SEÑALES (PDS)

El análisis del comportamiento del motor de extracción acústica desarrollado sobre las librerías Librosa y SciPy arroja conclusiones técnico-científicas de gran calado en el ámbito del Music Information Retrieval (MIR):

- Eficacia de la parametrización: La selección estratégica y acotada de los 15 descriptores acústicos ha demostrado ser suficiente para capturar la signatura física y matemática de las obras musicales. El procesamiento enventanado de la señal mediante la Transformada Corta de Fourier (STFT) y el cálculo de coeficientes espectrales y temporales han permitido mapear la percepción psicoacústica humana en vectores numéricos deterministas.
- El factor de la sonoridad (loudness) como indicador industrial: El hecho de que la sonoridad emergiera como el descriptor acústico puro más influyente en el modelo predictivo (3,49%) constituye una validación empírica de las metodologías modernas de masterización. El sistema ha aprendido de forma autónoma el impacto de la Loudness War, identificando que los niveles elevados de presión sonora ponderada guardan una correlación directa con los canales de distribución comercial y la retención inicial del oyente en plataformas de consumo digital.
- Limitaciones de la compresión espectral: El análisis crítico evidencia que el procesamiento de archivos con pérdida de información (como el formato MP3) impone un compromiso (trade-off) entre la velocidad de cómputo y la fidelidad matemática de las variables espectrales de bajo nivel (como el spectral flatness o los coeficientes cromáticos). Si bien el impacto en la predicción final de popularidad es marginal debido a la robustez del regresor, en el ámbito de la ingeniería de audio se constata que la compresión elimina dinámicas sutiles que limitan el techo analítico de los extractores puros.
- Optimización algorítmica frente a la latencia: Las pruebas de estrés sobre el backend revelan que los cálculos matriciales necesarios para variables temporales y

frecuenciales complejas son altamente demandantes en términos de CPU. El diseño del pipeline ha demostrado que, para entornos de producción interactivos, es imperativo implementar técnicas de submuestreo (downsampling) y paralelización en la lectura de bloques de audio para mitigar la degradación de la experiencia de usuario derivada de los tiempos de espera en el navegador web.

7.2 CONCLUSIONES METODOLÓGICAS DE LA ARQUITECTURA DE MACHINE LEARNING

El proceso de entrenamiento, optimización y validación del modelo predictivo estructurado sobre el algoritmo Random Forest Regression y evaluado mediante el Million Song Dataset (MSD) permite extraer las siguientes conclusiones desde el prisma de la ciencia de datos:

- Superación de los baselines del estado del arte: El Coeficiente de Determinación (R^2) alcanzado, situado en 0,4329, representa un éxito de ingeniería muy destacable. Al contrastar este valor con la literatura científica precursora —donde los modelos puramente acústicos oscilan históricamente en rangos de R^2 entre 0,15 y 0,35, se concluye que el presente pipeline maximiza de manera óptima la información contenida en las variables de entrada. El sistema ha logrado rozar el límite superior teórico de lo que la señal física del audio puede explicar por sí misma.
- La hegemonía del contexto sobre la señal: El cálculo de la importancia de las características abre un debate crítico fundamental en el diseño de estos sistemas. La dominancia absoluta de las variables vinculadas al capital social del artista (artist_familiarity con un 46,33% y artist_hottnesss con un 20,29%) demuestra matemáticamente que el éxito comercial está gobernado por una inercia contextual masiva. La arquitectura del modelo ha sabido ponderar que, ante mercados altamente saturados, la notoriedad preexistente del emisor actúa como un multiplicador de la atención que eclipsa las virtudes acústicas intrínsecas del mensaje musical. La necesidad de la normalización cronológica: El peso del descriptor temporal year (20,34%) concluye que los patrones predictivos son anacrónicos si se evalúan de

forma lineal. El algoritmo requirió este parámetro para contextualizar los cambios en el gusto colectivo y en los sistemas de distribución física y digital de las diferentes épocas, demostrando que un modelo predictivo estático pierde validez si no parametriza la evolución histórica del ecosistema macroeconómico en el que se inserta la obra.

- Estabilidad frente al sobreajuste (Overfitting): El uso de metodologías de aprendizaje por conjuntos (Ensemble Learning) mediante árboles de decisión independientes ha demostrado ser una elección arquitectónica robusta. A pesar de trabajar con un dataset de alta dimensionalidad y asimetría en la popularidad de las muestras, las técnicas de validación cruzada implementadas han asegurado que el error cuadrático medio se mantenga estable, garantizando una excelente capacidad de generalización ante datos de audio completamente nuevos e inéditos para el sistema.

7.3 APORTACIONES DEL TRABAJO EN COMPARACIÓN CON LA LITERATURA YA EXISTENTE

A partir de la revisión bibliográfica realizada y de los resultados obtenidos durante el desarrollo del proyecto, las principales aportaciones de este Trabajo Fin de Máster pueden resumirse en los siguientes puntos:

7.3.1 DESARROLLO DE UN SISTEMA END-TO-END COMPLETO

A diferencia de gran parte de la literatura existente, que se centra exclusivamente en la comparación de modelos predictivos o en la propuesta de nuevas arquitecturas de aprendizaje automático, este trabajo desarrolla una solución integral que cubre todo el ciclo de vida del análisis musical.

El sistema implementado incluye la adquisición de datos, el procesamiento de señales de audio, la extracción automática de características mediante Librosa, el entrenamiento y validación de modelos predictivos, la generación de inferencias y su integración dentro de una aplicación web interactiva accesible para el usuario final.

Esta aproximación permite trasladar los resultados de investigación a una herramienta funcional y utilizable en un entorno real.

7.3.2 PROPUESTA DE UNA ARQUITECTURA PREDICTIVA LIGERA Y REPRODUCIBLE

La mayor parte de los trabajos recientes utilizan arquitecturas profundas basadas en CNNs, CRNNs o modelos multimodales que requieren grandes volúmenes de datos y recursos computacionales especializados.

Frente a estas aproximaciones, este trabajo demuestra que es posible construir un sistema competitivo mediante una arquitectura significativamente más ligera, basada en técnicas de ingeniería de características y algoritmos clásicos de Machine Learning.

Como consecuencia, la solución desarrollada puede reproducirse fácilmente en entornos académicos o profesionales sin necesidad de infraestructuras avanzadas basadas en GPU.

7.3.3 PREDICCIÓN DE POPULARIDAD EN FASE PRE-LANZAMIENTO

Una contribución especialmente relevante del proyecto es la capacidad de realizar estimaciones de popularidad utilizando únicamente información disponible antes de la publicación de una canción.

Mientras que numerosos enfoques de la literatura dependen de métricas de interacción, reproducciones, actividad en redes sociales o comportamiento histórico de los usuarios, el modelo desarrollado utiliza exclusivamente características acústicas y metadatos musicales.

Esto permite emplear la herramienta como mecanismo de apoyo a la toma de decisiones durante las fases de producción y evaluación previas al lanzamiento comercial.

7.3.4 MODELO INTERPRETABLE FRENTE A ENFOQUES DE CAJA NEGRA

Uno de los principales problemas de las arquitecturas modernas de Deep Learning es la dificultad para comprender el origen de sus predicciones.

La utilización de Random Forest permite analizar la contribución individual de cada variable mediante métricas de importancia de características, proporcionando explicaciones transparentes sobre el comportamiento del modelo.

Esta capacidad resulta especialmente valiosa en aplicaciones donde no solo interesa obtener una predicción, sino también comprender qué factores musicales influyen en ella.

7.3.5 COMBINACIÓN EFICIENTE DE INFORMACIÓN ACÚSTICA Y METADATOS

El sistema desarrollado integra información procedente de diferentes niveles de análisis musical, combinando características acústicas extraídas de la señal de audio con metadatos estructurales del Million Song Dataset.

Esta estrategia permite capturar simultáneamente aspectos relacionados con el contenido musical y con el contexto del artista, obteniendo un modelo más rico que los enfoques exclusivamente acústicos sin alcanzar la complejidad de los sistemas multimodales más recientes.

7.3.6 GENERACIÓN DE CONOCIMIENTO ACCIONABLE PARA LA INDUSTRIA MUSICAL

Más allá de la capacidad predictiva del modelo, el sistema desarrollado permite identificar qué características musicales presentan una mayor relación con la popularidad estimada.

Esta información puede resultar de utilidad para artistas, productores, discográficas e investigadores interesados en comprender los factores asociados al éxito comercial de una obra musical.

7.3.7 CONTRIBUCIÓN METODOLÓGICA AL ÁMBITO DEL MUSIC INFORMATION RETRIEVAL

Finalmente, este trabajo demuestra que enfoques basados en ingeniería de características cuidadosamente diseñadas continúan siendo una alternativa válida frente a modelos de Deep Learning considerablemente más complejos.

Los resultados obtenidos sugieren que, en determinados escenarios, la combinación de procesamiento digital de señales, extracción de características y algoritmos de ensamble puede ofrecer una relación muy favorable entre precisión, interpretabilidad, coste computacional y facilidad de despliegue.

7.4 TRABAJOS FUTUROS

Aunque los resultados obtenidos demuestran la viabilidad de la propuesta desarrollada, existen diversas líneas de investigación y mejora que podrían explorarse en trabajos posteriores con el objetivo de aumentar la capacidad predictiva, la escalabilidad y las funcionalidades del sistema.

7.4.1 INCORPORACIÓN DE MODELOS DE DEEP LEARNING

Una posible evolución del sistema consistiría en sustituir o complementar los modelos actuales mediante arquitecturas de Deep Learning especializadas en análisis de audio.

Modelos basados en Redes Neuronales Convolucionales (CNN), Redes Convolucionales Recurrentes (CRNN) o arquitecturas Transformer podrían aprender representaciones directamente a partir de espectrogramas o señales de audio sin necesidad de realizar una extracción manual de características.

La comparación entre estos enfoques y la arquitectura actual permitiría analizar el compromiso existente entre precisión predictiva, interpretabilidad y coste computacional.

7.4.2 PREDICCIÓN DIRECTA A PARTIR DEL AUDIO COMPLETO

En el presente trabajo, el análisis se basa en características acústicas previamente extraídas y resumidas mediante descriptores estadísticos.

Como línea futura, podría desarrollarse un sistema capaz de procesar directamente la señal de audio completa, aprovechando arquitecturas neuronales capaces de identificar patrones

temporales y espectrales complejos que pueden perderse durante el proceso de agregación de características.

7.4.3 INTEGRACIÓN DE INFORMACIÓN MULTIMODAL

La popularidad musical depende de numerosos factores que van más allá de las propiedades acústicas de una canción.

Por ello, una extensión natural del trabajo consistiría en incorporar nuevas fuentes de información complementarias, tales como:

- Letras de las canciones.
- Información sobre artistas.
- Géneros musicales.
- Datos de redes sociales.
- Tendencias de plataformas de streaming.
- Información contextual del mercado musical.

La combinación de estas fuentes mediante modelos multimodales podría mejorar significativamente la capacidad predictiva del sistema.

7.4.4 OBTENCIÓN AUTOMÁTICA DE DATOS MEDIANTE APIS MUSICALES

Actualmente, parte de la información utilizada procede de datasets previamente construidos.

Como mejora futura, el sistema podría conectarse directamente a APIs musicales como Spotify o Last.fm para recuperar información actualizada en tiempo real sobre canciones, artistas y tendencias de consumo.

Esto permitiría mantener el modelo actualizado y adaptado a la evolución constante de la industria musical.

7.4.5 OPTIMIZACIÓN Y ACTUALIZACIÓN CONTINUA DEL MODELO

La industria musical presenta cambios constantes asociados a nuevas tendencias, géneros y patrones de consumo.

Por este motivo, resultaría interesante implementar mecanismos de reentrenamiento automático capaces de actualizar periódicamente el modelo utilizando nuevos datos musicales.

De esta forma, el sistema podría adaptarse dinámicamente a los cambios del mercado y mantener su capacidad predictiva a largo plazo.

7.4.6 EXPLICABILIDAD AVANZADA DE LAS PREDICCIONES

Aunque el modelo desarrollado permite analizar la importancia relativa de las variables, existen técnicas más avanzadas de interpretabilidad que podrían incorporarse en futuras versiones.

Herramientas como SHAP (SHapley Additive Explanations) o LIME permitirían explicar individualmente cada predicción realizada por el sistema, proporcionando una visión más detallada de los factores que influyen en la estimación de popularidad de una canción concreta.

Esta funcionalidad podría resultar especialmente útil para productores musicales y creadores interesados en comprender las fortalezas y debilidades de sus composiciones.

7.4.7 AMPLIACIÓN DE LA APLICACIÓN WEB

Otra línea de evolución consistiría en extender las capacidades de la aplicación desarrollada.

Entre las posibles funcionalidades adicionales destacan:

- Comparación simultánea entre varias canciones.
- Visualización avanzada de características acústicas.

- Generación automática de informes.
- Recomendaciones para mejorar determinados atributos musicales.
- Paneles interactivos para análisis exploratorio.

Estas mejoras aumentarían el valor práctico de la herramienta y facilitarían su utilización por perfiles no técnicos.

7.4.8 DESPLIEGUE COMO PLATAFORMA DE APOYO A LA INDUSTRIA MUSICAL

Finalmente, una línea especialmente interesante consistiría en evolucionar el prototipo desarrollado hacia una plataforma profesional orientada a artistas, productores y discográficas.

En este escenario, el sistema podría utilizarse como herramienta de apoyo durante las fases de producción y evaluación de nuevas canciones, proporcionando estimaciones de popularidad junto con explicaciones detalladas de los factores que más influyen en dichas predicciones.

La combinación de análisis automático, interpretabilidad y facilidad de uso podría convertir la solución desarrollada en una herramienta de apoyo a la toma de decisiones dentro del sector musical.

Capítulo 8. BIBLIOGRAFÍA

4. **Müller, M. (2015).** *Fundamentals of music processing: Audio, analysis, algorithms, applications.* Springer.
5. **Stevens, S. S., Volkman, J., & Newman, E. B. (1937).** A scale for the measurement of the psychological magnitude of pitch. *The Journal of the Acoustical Society of America*, 8(3), 185–190.
6. **Bello, J. P., Daudet, L., Abdallah, S., Duxbury, C., Davies, M., & Sandler, M. B. (2005).** *A tutorial on onset detection in music signals.* *IEEE Transactions on Speech and Audio Processing*, 13(5), 1035–1047.
7. **McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., & Nieto, O. (2015).** librosa: Audio and music signal analysis in python. *En Proceedings of the 14th python in science conference* (Vol. 8, pp. 18-25).
8. **Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O., Roma, G., Salamon, J., Zapata, J. R., & Serra, X. (2013).** *Essentia: An open-source library for sound and music analysis.* *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR 2013)*, 493–498.
9. **Pachet, F., & Roy, P. (2008).** Hit song science is not yet a science. In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR 2008)* (pp. 355–360).
10. **Interiano, M., Kazemi, K., Wang, L., Yang, J., & Miron, M. (2018).** In search of the modern hit song: A music information retrieval approach. *Social Network Analysis and Mining*, 8, 1–12.
11. **Dhanhariya, P. (2019).** *Predicting Music Popularity Using Audio Features.* Stanford University.
12. **Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006).** Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762), 854–856.

13. **Singhi, A., & Brown, J. C. (2012).** Predicting hit songs with acoustic features and social network analysis. In *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR 2012)*.
14. **Kim, J. (2021).** Music popularity prediction through data analysis of music's characteristics. *International Journal of Science, Technology and Society*, 9(5), 239–244.
15. **Van Den Oord, A., Dieleman, S., & Schrauwen, B. (2013).** Deep content-based music recommendation. In C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Weinberger (Eds.), *Advances in Neural Information Processing Systems 26* (Vol. 26). Neural Information Processing Systems Foundation.
16. **Choi, K., Fazekas, G., Sandler, M., & Cho, K. (2017).** *Convolutional recurrent neural networks for music classification*. In *Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 2392–2396). IEEE.
17. **Sharan, A. K., & Sahu, A. K. (2021).** *Popularity prediction of songs based on audio spectrograms using convolutional neural networks*.
18. **Sciandra, M., y Spera, R. (2022).** Lyrics-based musical popularity prediction. *Quality & Quantity*, 56(6), 4607-4625.
19. **Zangerle, E., Pichl, M., & Speigl, P. M. (s. f.).** *Multi-modal popularity prediction of music tracks using audio and textual features*.
20. **Bertin-Mahieux, T., Ellis, D. P. W., Whitman, B., & Lamere, P. (2011).** The Million Song Dataset. *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, 591–596.
21. **Jehan, T., & DesRoches, D. (2012).** *EchoNest Analyzer documentation*. The Echo Nest.
22. **Pachet, F., y Roy, P. (2008).** Hit song science: Not yet a science? *IEEE Multimedia*, 15(1), 66–71.
23. **Angel Franco García. (2016).** *Transformada rápida de Fourier (I)*. (s. f.).
http://www.sc.ehu.es/sbweb/fisica3/datos/fourier/fourier_1.html

ANEXO: ALINEACIÓN DEL PROYECTO CON LOS OBJETIVOS DE DESARROLLO SOSTENIBLE (ODS)

El presente proyecto de investigación y desarrollo tecnológico no se limita a resolver un problema abstracto de ingeniería de datos y procesamiento de señal, sino que se enmarca dentro del compromiso ético y de desarrollo sostenible promovido por la ETS de Ingeniería, ICAI. El diseño del pipeline *end-to-end* y la implementación de la plataforma interactiva de predicción musical impactan de manera directa e indirecta en los Objetivos de Desarrollo Sostenible (ODS) de la Agenda 2030 de las Naciones Unidas, así como en los criterios ambientales, sociales y de gobernanza (ESG).

A continuación, se desglosa el cumplimiento detallado de las metas de desarrollo sostenible identificadas en el ciclo de vida del proyecto:

Dimensión Ambiental (Environmental Criteria – E)

El vertiginoso crecimiento de la economía digital y los servicios de computación en la nube plantea serios desafíos en términos de consumo energético y huella de carbono. Las decisiones de diseño arquitectónico de este TFM se han alineado explícitamente con la minimización del impacto ecológico digital:

ODS 12: Producción y Consumo Responsables (Meta 12.5 - Reducción de la generación de desechos y optimización de recursos): El desarrollo de modelos predictivos tradicionales suele competir con arquitecturas masivas de *Deep Learning* (redes neuronales profundas o modelos generativos), que exigen ciclos de entrenamiento masivos de semanas de duración en clústeres de GPUs con altas tasas de disipación térmica. En este proyecto se apostó intencionadamente por un enfoque de aprendizaje automático clásico optimizado (*Random Forest*), logrando un coeficiente de determinación competitivo ($R^2 = 0,4329$) con una fracción insignificante del gasto energético. Adicionalmente, el pipeline de

ANEXO: ALINEACIÓN DEL PROYECTO CON LOS OBJETIVOS DE DESARROLLO SOSTENIBLE (ODS)

procesamiento digital de señales (PDS) realiza un submuestreo controlado del audio a 22.050 Hz mediante la librería *Librosa*. Esta técnica de ingeniería reduce el volumen de datos en memoria y optimiza el tiempo de cómputo de la Transformada de Fourier de Corto Tiempo (STFT), reduciendo al mínimo el desgaste de hardware y el consumo eléctrico en la fase de validación cruzada.

ODS 13: Acción por el Clima: La aplicación interactiva multi-página diseñada para la fase de producción se despliega sobre una infraestructura en la nube (*Dash* y *Flask*). Para garantizar que este despliegue no genere un consumo energético latente o innecesario en los centros de datos, se implementó una topología estrictamente libre de estado (*stateless*). El servidor web no mantiene conexiones ni bases de datos relacionales persistentes que requieran lecturas y escrituras continuas en disco; la inferencia y el enriquecimiento de datos a través de la API de Spotify ocurren de manera efímera y síncrona solo bajo demanda del usuario. Esto optimiza el uso de CPU virtual en la nube y reduce las emisiones indirectas de gases de efecto invernadero (CO₂ equivalente).

Dimensión Social (Social Criteria – S)

La transformación digital de las industrias culturales puede actuar como un mecanismo de centralización o, por el contrario, como un motor de inclusión. Este proyecto persigue activamente este último enfoque:

ODS 8: Trabajo Decente y Crecimiento Económico (Meta 8.3 - Fomento de la creatividad, el emprendimiento y la innovación digital): El mercado de la música comercial presenta una fuerte asimetría de información y altas barreras de entrada; habitualmente, solo las grandes corporaciones discográficas (*major labels*) disponen de presupuestos para financiar análisis avanzados de mercado y herramientas de *Big Data*. El software interactivo desarrollado en este TFM funciona como una herramienta de apoyo a la toma de decisiones accesible para artistas independientes, emergentes y autoeditados. Al democratizar el acceso a métricas analíticas e inferencias de viabilidad comercial de forma abierta y ligera, se impulsa la competitividad y la sostenibilidad económica del tejido

ANEXO: ALINEACIÓN DEL PROYECTO CON LOS OBJETIVOS DE DESARROLLO SOSTENIBLE (ODS)

artístico e independiente, dotándoles de herramientas técnicas de nivel industrial para validar sus obras antes de su lanzamiento.

ODS 10: Reducción de las Desigualdades: Los modelos basados en el análisis lingüístico de letras de canciones suelen heredar un sesgo geopolítico y cultural masivo, penalizando a los mercados no angloparlantes o a idiomas minoritarios. Al diseñar un motor predictivo basado de forma estricta en descriptores microestructurales físicos extraídos directamente de la señal de audio (tempo, sonoridad, atributos espectrales y tímbricos), el sistema se convierte en una herramienta totalmente agnóstica al idioma. Todas las pistas se evalúan bajo principios matemáticos y físicos acústicos uniformes, promoviendo la equidad algorítmica y previniendo la discriminación automatizada de propuestas procedentes de diferentes contextos culturales.

Dimensión de Gobernanza (Governance Criteria – G)

La gobernanza en proyectos de ingeniería de software se traduce en transparencia, ética de datos, reproducibilidad y respeto al marco normativo vigente:

ODS 9: Industria, Innovación e Infraestructura (Meta 9.b - Desarrollo tecnológico y sostenibilidad regulatoria): Este proyecto se ha edificado íntegramente sobre ecosistemas tecnológicos de código abierto (*Python, Scikit-Learn, Dash, Plotly, Librosa*) y se ha validado empíricamente utilizando un conjunto de datos público de referencia científica (*Million Song Dataset*). Esta aproximación mitiga la opacidad propia de las tecnologías propietarias ("cajas negras corporativas") y garantiza que el pipeline técnico sea totalmente auditable, transparente y reproducible por la comunidad científica o auditorías de software externas.

Cumplimiento Ético y Privacidad de los Datos: La comunicación asíncrona en tiempo real con la API de Spotify se ha desarrollado siguiendo de forma minuciosa los Términos de Servicio para Desarrolladores de la plataforma, limitando el consumo a metadatos públicos de acceso abierto. Siguiendo la filosofía de "Privacidad desde el Diseño" (*Privacy by Design*) alineada con el Reglamento General de Protección de Datos (RGPD), la aplicación web interactiva no recopila, almacena, ni rastrea información personal o perfiles de los usuarios

ANEXO: ALINEACIÓN DEL PROYECTO CON LOS OBJETIVOS DE DESARROLLO SOSTENIBLE (ODS)

que interactúan con las páginas de análisis y predicción, operando de forma efímera para asegurar un entorno tecnológico seguro y ético.

En definitiva, el desarrollo de esta plataforma predictiva demuestra cómo la ciencia de datos avanzada y el procesamiento digital de señales pueden ponerse al servicio del desarrollo sostenible en el sector cultural. A través de un diseño de software consciente de su eficiencia, el proyecto logra mitigar el impacto medioambiental indirecto en la nube (ODS 12 y 13) al tiempo que reduce las barreras de entrada y los sesgos lingüísticos para los artistas emergentes (ODS 8 y 10). Así, el proyecto no solo cumple con los requerimientos técnicos y metodológicos propios de una ingeniería de telecomunicación, sino que valida la viabilidad de crear herramientas tecnológicas éticas, transparentes y de código abierto que impulsan una transformación digital equitativa y responsable dentro de la industria musical.