

Clasificación Automática de Canciones como Tristes o Alegres basada en Audio Features de Spotify utilizando K-means y Random Forest

Mateo Gutiérrez Melo - Universidad Nacional De Colombia - mgutierrezca@unal.edu.co

25 Mayo de 2023

Este artículo propone un sistema de clasificación automática de sentimientos en canciones utilizando los audio features proporcionados por Spotify. El objetivo es desarrollar una herramienta eficiente y precisa que pueda clasificar las canciones según las emociones que generan. Se implementará un algoritmo de Random Forest de clasificación, utilizando los audio features como variables predictoras y etiquetando previamente los sentimientos mediante el algoritmo de K-means.

Keywords— Random Forest, K-Means, Machine Learning, Spotify, Audio features

1. Introducción

La música ha sido una forma de expresión artística que ha evocado y transmitido una amplia gama de emociones a lo largo de la historia. En la actualidad, la cantidad de música disponible es vasta, lo que presenta un desafío para los oyentes que buscan canciones que se alineen con sus estados de ánimo o emociones particulares. Por lo tanto, surge la necesidad de desarrollar herramientas eficientes y precisas que puedan clasificar las canciones según los sentimientos que estas generan.

Este proyecto propone realizar una clasificación de sentimientos de canciones utilizando los audio features proporcionados por la API de Spotify. La API de Spotify ofrece una amplia variedad de información sobre las canciones, incluyendo información del artista y características acústicas. Estas características incluyen el tempo, la energía, la valencia y la instrumentalidad, entre otras. Por ejemplo, la variable "danceability" describe qué tan adecuada es una pista para bailar en función de una combinación de elementos musicales como el tempo, la estabilidad del ritmo y la fuerza del ritmo.

Para lograr el objetivo de clasificar los sentimientos de las canciones, se implementará un algoritmo de Random Forest de clasificación, una técnica de machine learning efectiva en la clasificación de datos. Así mismo, se utilizarán los audio features mencionados anteriormente como variables predictoras y se etiquetarán previamente los sentimientos mediante el algoritmo de K-means. Para finalizar se entrenará el modelo con un conjunto de datos etiquetados y se evaluará su rendimiento utilizando métricas de evaluación adecuadas.

El resultado de este proyecto podría tener aplicaciones prácticas significativas. Por ejemplo, las plataformas de streaming de música podrían utilizar este sistema de clasificación para recomendar canciones basadas en los sentimientos deseados por el usuario en un momento dado. Además, los usuarios podrían utilizar esta herramienta para descubrir canciones que se ajusten a su estado emocional actual o incluso para explorar nuevos géneros o artistas en función de las emociones que deseen experimentar.

En resumen, este proyecto busca desarrollar un sistema de clasificación de sentimientos de canciones utilizando los audio features proporcionados por la API de Spotify y un algoritmo de Random Forest. Se

espera que esta investigación aporte una perspectiva valiosa para mejorar la experiencia de escucha de música y abrir nuevas posibilidades en la exploración y descubrimiento musical basados en emociones.

2. Objetivos

El objetivo general para el presente trabajo es :

1. Desarrollar un modelo capaz de predecir si una canción es Alegre o Triste de acuerdo a los audio features de la misma

Los objetivos específicos son :

1. Definir un mecanismo para capturar información de las canciones por medio de la API de Spotify
2. Generar diagramas y visualizaciones de datos que permitan detallar la distribución de las canciones respecto a sus audio features
3. Identificar las relaciones que pueden existir entre las diversas variables de los audiofeatures en las canciones para poder clasificarlas como Alegres o Tristes
4. Desarrollar un modelo de K-means que etiquete de manera manual las canciones obtenidas como Triste o Alegres

3. Marco Teórico

4. Metodología

En esta sección se describen los métodos utilizados.

5. Desarrollo e implementación

6. Discusión de resultados

7. Resultados

Aquí se presentan y analizan los resultados obtenidos.

8. Conclusiones

Finalmente, se presentan las conclusiones del estudio.

Referencias

- [1] Autor1, A., y Autor2, B. (2023). Título del artículo. *Revista, número de volumen*(número de edición), páginas.
- [2] Autor3, C. (2023). Título del libro. Editorial.