

Laboratorio

Clasificador de canciones de Spotify

Introducción



En este laboratorio, nos enfocaremos en un conjunto de datos repleto de atributos clave de distintas canciones. La meta es establecer a través de herramientas de aprendizaje automático, un clasificador que pueda predecir las inclinaciones musicales de un individuo y predecir si disfrutará o no de una determinada canción.

Conjunto de datos



Se proporcionará un archivo en formato CSV que alberga un conjunto de datos con 16 columnas. De estas, 13 representan atributos específicos de las canciones. Además, se incluyen columnas dedicadas al nombre de la canción y al artista respectivamente. La columna "destino" actúa como etiqueta, indicando las preferencias del usuario: un valor de "1" significa que la canción fue del agrado del usuario, mientras que un "0" indica lo contrario. Los atributos que caracterizan cada pista son:

- 1. Danceability (Bailabilidad):** Describe cuán adecuada es una pista para bailar basada en una combinación de elementos musicales, incluyendo el tempo, la estabilidad del ritmo y la fuerza del ritmo. Una puntuación de 0.0 es menos bailable y 1.0 es más bailable.
- 2. Energy (Energía):** Representa una medida perceptiva de intensidad y actividad. Las pistas con mucha energía suelen sentirse rápidas, ruidosas y rítmicas. Por ejemplo, el death metal tiene alta energía, mientras que un preludio de Bach tiene una puntuación baja en el espectro de energía.
- 3. Loudness (Volumen):** Es una medida general de la sonoridad de una pista en decibeles (dB). Las pistas sonoras tienen un valor promedio más alto.
- 4. Speechiness (Hablado):** Detecta la presencia de palabras habladas en una pista. Los valores más cercanos a 1.0 sugieren que la pista contiene más palabras habladas, valores por encima de 0.66 describen pistas que son probablemente hechas completamente de palabras habladas. Los valores entre 0.33 y 0.66 describen pistas que pueden contener música y palabras. Valores por debajo de 0.33 sugieren música o pistas no habladas.

5. **Acousticness (Acústica):** Una medida de cuán acústica es una pista. Una puntuación de 1.0 significa que es muy probable que la pista sea acústica.

6. **Instrumentalness (Instrumentalidad):** Predice si una pista no contiene voces. Los valores más cercanos a 1.0 significan que es muy probable que la pista sea instrumental, mientras que valores cercanos a 0.5 indican que la pista puede contener contenido vocal.

7. **Liveness (En vivo):** Detecta la presencia de una audiencia en la grabación. Valores más altos representan una mayor probabilidad de que la pista haya sido interpretada en vivo. Un valor por encima de 0.8 proporciona una gran probabilidad de que la pista esté en vivo.

8. **Valence (Valencia):** Describe la positividad musical transmitida por una pista. Las pistas con alta valencia suenan más positivas (por ejemplo, feliz, alegre, eufórico), mientras que las pistas con baja valencia suenan más negativas (por ejemplo, triste, deprimido, enojado).

9. **Tempo:** Es la velocidad general o ritmo de una pista y se deriva directamente de la duración promedio de la barra. Se mide en pulsos por minuto (BPM).

Informe



Se debe desarrollar un programa que categorice canciones según las preferencias del usuario. La tarea principal es crear un programa capaz de predecir si una canción podría ser del gusto del usuario o no. Este proyecto debe ser ejecutado dentro de un Notebook de Colab.

Dentro del Notebook, es importante que cada paso del proceso esté acompañado por bloques de texto que expliquen el razonamiento detrás de las decisiones tomadas. Se debe detallar por qué se eligieron ciertas técnicas, métodos o parámetros y cuál es su relevancia en el contexto del proyecto. Además, es necesario documentar los desafíos o inconvenientes que surgieron durante el desarrollo, proporcionando una visión clara y completa del proceso de diseño e implementación de la solución.

Flujo de Trabajo



Realizar varios modelos de Machine Learning:

- KNN (K-Nearest Neighbors)
- SVM (Support Vector Machines)
- Árbol de decisión
- Bayes (Naive Bayes)

Realizar algún tipo de validación:

- Validación Simple
- Validación Cruzada k-fold

Evaluación y análisis del rendimiento de cada modelo:

- Matriz de confusión
- Precisión, recall y F1-score

Ajustar para cada uno de los modelos los hiper parámetros:

- Grid Search
- Random Search

Realizar un ensamble de los modelos:

- Votación Mayoritaria

Evaluación y análisis del rendimiento:

- Matriz de confusión
- Precisión, recall y F1-score

Fecha de entrega

Miércoles, 8 de Noviembre, 23:00 hs