

Proyecto Inteligencia Artificial: Clasificación de Canciones

Integrantes:

- Samuel Soto
- George Trujillo

Tipo de problema:

El problema que enfrentamos es un problema de clasificación binaria. Queremos clasificar automáticamente las canciones en dos géneros: 'Hip-Hop' y 'Rock' en función de sus atributos musicales sin necesidad de escucharlas.

Metodología:

1. Exploración de datos:

- Analizar y visualizar la distribución de los atributos musicales, así como la proporción de canciones de cada género en el conjunto de datos.

2. Preprocesamiento de datos:

- Normalizar los atributos numéricos si es necesario y codificar las variables categóricas.

3. División del conjunto de datos:

- Separar el conjunto de datos en conjuntos de entrenamiento y prueba para evaluar el rendimiento del modelo.

4. Selección de modelo:

- Probar varios algoritmos de clasificación como Regresión Logística, Árboles de Decisión, Bosques Aleatorios, SVM, Redes Neuronales, etc.

5. Entrenamiento del modelo:

- Ajustar los modelos seleccionados utilizando el conjunto de entrenamiento.

6. Evaluación del modelo:

- Evaluar el rendimiento de los modelos utilizando métricas adecuadas.

7. Ajuste de hiper parámetros:

- Optimizar los hiper parámetros del modelo para mejorar el rendimiento.

8. Validación cruzada:

- Realizar validación cruzada para garantizar la generalización del modelo.

9. Predicciones:

- Utilizar el modelo finalmente ajustado para clasificar nuevas canciones en 'Hip-Hop' o 'Rock'.

Métricas para medir el progreso:

- **Exactitud (Accuracy):** Proporción de predicciones correctas sobre el total de predicciones.

- **Precisión (Precision):** Proporción de verdaderos positivos sobre el total de predicciones positivas.
- **Recuperación (Recall):** Proporción de verdaderos positivos sobre el total de casos positivos reales.
- **F1-score:** Media armónica de precisión y recuperación.

Datos recolectados:

- **Conjunto de datos:** Compuesto por una cantidad de canciones con varias columnas que representan los atributos de cada pista, incluyendo el género musical ('Hip-Hop' o 'Rock').

Análisis Exploratorio de Datos (EDA):

1. Exploración de la distribución de la variable objetivo (género):

- Verificar la proporción de canciones de cada género.

2. Análisis de las características numéricas:

- Calcular estadísticas descriptivas (media, desviación estándar, mínimo, máximo, etc.) de los atributos musicales.
- Visualizar la distribución de cada atributo mediante histogramas o gráficos de densidad.

3. Análisis de las características categóricas:

- Explorar la diversidad de artistas y nombres de canciones.

4. Correlación entre atributos:

- Calcular la matriz de correlación para entender las relaciones entre los atributos musicales y la variable objetivo.
- Visualizar la matriz de correlación con un mapa de calor.

5. Visualización de datos:

- Utilizar gráficos de dispersión o gráficos de barras para explorar posibles relaciones entre los atributos musicales y la variable objetivo.

Siguientes pasos en el proyecto:

1. Preprocesamiento de datos:

- Manejo de datos faltantes si los hay.
- Codificación de variables categóricas si es necesario.
- Normalización de atributos numéricos si es necesario.

2. División del conjunto de datos:

- Separar el conjunto de datos en conjuntos de entrenamiento y prueba.

3. Selección de modelo:

- Probar varios algoritmos de clasificación como Regresión Logística, Árboles de Decisión, Bosques Aleatorios, SVM, Redes Neuronales, etc.

4. Entrenamiento del modelo:

- Ajustar los modelos seleccionados utilizando el conjunto de entrenamiento.

5. Evaluación del modelo:

- Evaluar el rendimiento de los modelos utilizando métricas como la exactitud, precisión, recuperación, F1-score.

6. Ajuste de hiper parámetros:

- Optimizar los hiper parámetros del modelo para mejorar el rendimiento.

7. Validación cruzada:

- Realizar validación cruzada para garantizar la generalización del modelo.

8. Predicciones:

- Utilizar el modelo finalmente ajustado para clasificar nuevas canciones.

Estrategias para Conseguir más Datos:

1. Explorar más playlists de plataformas de streaming:

- Utilizar una variedad de playlists de diferentes géneros musicales y décadas para aumentar la diversidad del conjunto de datos.
- Buscar playlists específicas de temas o estados de ánimo que contengan canciones de 'Hip-Hop' y 'Rock'.

2. Ampliar la búsqueda en plataformas de streaming adicionales:

- Explorar otras plataformas de streaming de música además de Spotify, como Apple Music, YouTube Music, Deezer, etc.
- Utilizar las APIs de estas plataformas para acceder a datos de canciones y características de audio.

Análisis de Aspectos Éticos:

1. Privacidad y protección de datos:

- Es importante proteger la privacidad de los usuarios al recopilar y utilizar datos sobre sus preferencias musicales. Esto implica obtener el consentimiento informado de los usuarios antes de recopilar sus datos y garantizar que los datos se almacenen y utilicen de manera segura y confidencial.

2. Sesgo algorítmico y equidad:

- Los algoritmos de IA pueden estar sujetos a sesgos inherentes en los datos utilizados para entrenarlos, lo que puede llevar a resultados sesgados o discriminatorios. Es importante realizar una evaluación continua del sesgo algorítmico y tomar medidas para mitigarlo, como la recopilación de datos equilibrados y la implementación de técnicas de corrección de sesgo.

3. Transparencia y explicabilidad:

- Los usuarios deben poder comprender cómo se toman las decisiones de IA, especialmente en el contexto de predecir géneros musicales. Es fundamental proporcionar transparencia y explicabilidad en los procesos de toma de decisiones algorítmicas para fomentar la confianza y la comprensión por parte de los usuarios.