Spotify Daten - und Nutzerprofilanalyse

Projektarbeit im Fach: Artificial Intelligence

Genre Klassifikation mittels Machine Learning

Explorative Nutzer Analyse

…

…

…

Sebastian Buresch

Marco Lichtenberger

Thomas Elias Huber

31. Juli 2025

1. **Einleitung**

Musik ist ein kulturelles Phänomen mit hoher persönlicher Relevanz. Durch die Verfügbarkeit großer Musikkataloge auf Streaming-Plattformen wie Spotify entstehen umfangreiche Daten über das individuelle Hörverhalten. Ziel dieses Projekts war es, mithilfe der Spotify Web API sowie eines öffentlich verfügbaren Datensatzes Musikpräferenzen zu analysieren und mittels maschinellen Lernens eine automatische Genreklassifikation zu ermöglichen.

**2. Datengrundlage**

Für dieses Projekt wurden zwei Datenquellen verwendet: Einerseits die Spotify Web API zur Abfrage persönlicher Top-Tracks eines Nutzers inklusive zugehöriger Genres; andererseits ein Kaggle-Datensatz mit über 20.000 Songs und ihren akustischen Eigenschaften (z. B. Valence, Energy, Tempo, Danceability) sowie Genre-Tags. Die wichtigsten Merkmale für die Analyse sind:

- danceability, energy, loudness, speechiness

- acousticness, instrumentalness, liveness, valence, tempo

**3. Explorative Analyse**

Aus den Top-Tracks eines Spotify-Nutzers wurden mithilfe der API die häufigsten Genres extrahiert. Diese wurden mit einer Counter-Analyse gezählt und nach Häufigkeit sortiert. Anschließend wurden die Genres über ein Mapping in grobe Stilrichtungen (z. B. 'Rocker', 'Metalhead', 'Pop-Liebhaber') eingeordnet. Dies erlaubt eine vereinfachte Klassifizierung des Nutzerprofils.



**4. Genreklassifikation mit Machine Learning**

Zur Genreklassifikation wurde ein RandomForestClassifier mit 100 Bäumen auf dem Kaggle-Datensatz trainiert. Es wurden neun Audiofeatures verwendet. Um die Komplexität zu reduzieren und die Modellleistung zu verbessern, wurde die Klassifikation auf die 15 häufigsten Genres beschränkt. Die Klassifikation erreichte eine Genauigkeit von ca. 49 %.

Besonders gut erkennbar waren Genres wie 'minimal-techno', 'metalcore' oder 'party', während Pop und verwandte Subgenres (z. B. 'pop-film') schlechter differenzierbar waren.

Die wichtigsten Merkmale laut Modellanalyse waren:

- acousticness (13.2 %)

- danceability (12.2 %)

- instrumentalness (12.1 %)

Weitere wichtige Merkmale waren energy, valence und tempo.eingesetzter-Film.png

**5. Bewertung und Erkenntnisse**

Das Modell zeigt, dass bestimmte Genres klar erkennbare akustische Signaturen aufweisen. Die Audiofeatures ermöglichen eine solide Klassifikation in Fällen, in denen Genres technisch oder stilistisch eindeutig trennbar sind. Schwächen treten vor allem bei überlappenden Pop-Subgenres auf.

Die Feature-Wichtigkeit liefert zusätzlich interessante Erkenntnisse über die Struktur von Musikstilen und ihre Repräsentation in den Spotify-Daten.

**6. Fazit und Ausblick**

Dieses Projekt zeigt, wie mit einfachen Machine-Learning-Verfahren und öffentlich zugänglichen Daten ein Einblick in individuelle Musikpräferenzen sowie eine automatisierte Genreklassifikation möglich ist. Künftige Erweiterungen könnten Lyrics-Analyse, t-SNE-Clustering, oder Zeitverlaufsauswertungen umfassen.

Abbildung 1: Top 10 Genres basierend auf den persönlichen Spotify-Daten.

Abbildung 2: Wichtigste Audiofeatures laut Modell.

Abbildung 3: Konfusionsmatrix der Genreklassifikation.

**7. Literaturverzeichnis**

- Spotify Developer. (2023). Spotify Web API Documentation. https://developer.spotify.com/documentation/web-api/

- Kaggle. (2023). Spotify Dataset. https://www.kaggle.com/datasets

- Pedregosa, F. et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

- Python Docx. (2023). Create and Update Word Documents. https://python-docx.readthedocs.io/