

Adaptive Systems and Artificial Intelligence

Dokumentation: Music Recommendation and
Emotion Prediction System

Team 3

Michelle Koops, 2651042

Avery Lönker, 2645331

Laura Siekierski, 2618388

Hochschule für Angewandte Wissenschaften Hamburg

14.02.2025

Inhaltsverzeichnis

1. Unsere Projektidee.....	3
1.1 Konzept.....	3
1.2 Inspiration durch andere Arbeiten.....	3
2. Materialien und Methoden.....	4
2.1 Herangehensweise.....	4
2.2 Feature Extraction.....	5
2.3 Umsetzung mit KNN.....	5
2.4 Realisierung in drei Modellen.....	6
2.5 Import in eine Huggingface Umgebung.....	7
3. Ergebnisse.....	8
4. Analyse.....	8
4.1 Gründe für die Ungenauigkeiten der Modelle.....	8
4.2 Evaluation des kleinen Modells.....	10
5. Projektmanagement & Versionierung.....	12
5.1 Vorgehensmodell Scrumban.....	12
5.2 Arbeitsweise im Team.....	12
6. Fazit.....	13
6.1 Endprodukt.....	13
6.2 Learnings.....	13
7. Quellen.....	15
8. Anhang.....	16

1. Unsere Projektidee

1.1 Konzept

Im Rahmen des Moduls Adaptive Systems and Artificial Intelligence haben wir uns dafür entschieden, ein Music Recommendation and Emotion Prediction System zu erstellen. Dieses soll in der Lage sein, einem User auf Grundlage eines eingegebenen Songs eine Playlist mit weiteren Liedern zu generieren, die musikalisch und emotional zum ursprünglichen Song passen.

Dabei soll es die Möglichkeit geben, selbst einen Song zur Verfügung zu stellen oder einen zufälligen Song aus dem von uns gewählten Datensatz zu verwenden. Der eingegebene Song wird daraufhin auf seine Features analysiert und mithilfe der K-Nearest-Neighbor-Classification (KNN) einem Label zugeordnet. Im Anschluss kann man sich eine Playlist mit den ähnlichsten Liedern ausgeben lassen.

1.2 Inspiration durch andere Arbeiten

Ähnliche Prinzipien findet man bei Plattformen wie iTunes oder Spotify. Auch sie haben eine Funktion, dem User eine Playlist basierend auf dem vorherigen Hörverhalten oder passend zu einem bestimmten Song, Genre oder Stimmung zu generieren.

Spotify nutzt dabei maschinelles Lernen für personalisierte Musikempfehlungen. Ein Empfehlungssystem analysiert das Hörverhalten der Nutzer, während kollaboratives Filtern durch den Vergleich von Hörmustern relevante Songs vorschlägt.

Der BART-Algorithmus optimiert Empfehlungen, indem er Songs, die von Usern über 30 Sekunden angehört werden, als positiv bewertet und den Startbildschirm täglich anpasst. Zudem setzt Spotify NLP ein, um Song-Metadaten und externe Texte zu analysieren, was kulturelle Trends besser erfasst und in die Empfehlungen einfließt.

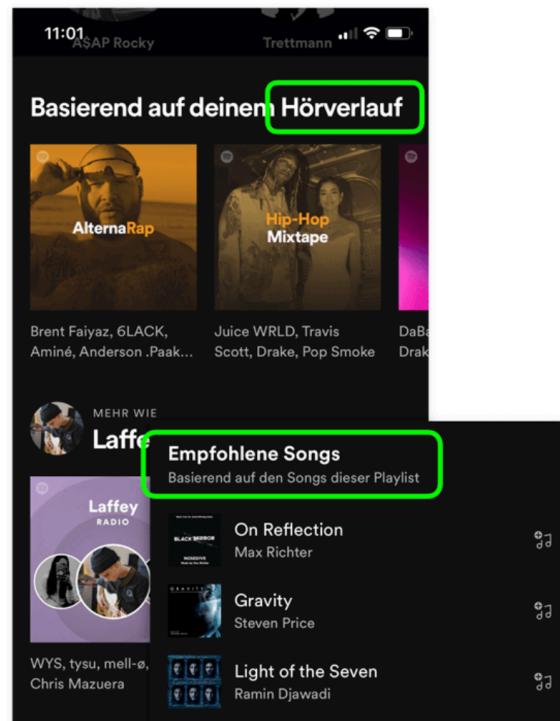


Abb. 1: Spotify trackt den Hörverlauf der eingeloggten User

Ein weiterer von Spotify verwendeter Algorithmus ist das sog. KI-Audio Modell (vgl. Katzlberger, M. 2021), welches das rohe Audiomaterial direkt analysiert, um Songs bestimmten Kategorien basierend auf ihren Audio-Features zuzuordnen. Dadurch können Songs unabhängig von ihrer Popularität empfohlen werden.

Es würde jedoch den Rahmen unserer Möglichkeiten sprengen zu versuchen, diese Vielfalt an intelligenten Systemen umzusetzen, vor allem da wir keinen Zugriff auf Verläufe der von Nutzern bereits gehörten Songs haben.

Aus diesem Grund nehmen wir uns für unser Projekt besonders das KI-Audio Modell als Inspiration, um unser Music Recommendation and Emotion Prediction System umzusetzen. Anders als Spotify steht uns dabei jedoch keine große Audiobibliothek zur Verfügung. Deswegen haben wir uns dafür entschieden, bereits gelabelte Datensätze zu benutzen.

2. Materialien und Methoden

Im folgenden Abschnitt erläutern wir die Materialien, die wir für unser Projekt herangezogen haben und die Methoden, die wir verwendet haben.

2.1 Herangehensweise

Unser Prozess bestand hier aus einem Brainstorming, bei dem wir zunächst verschiedene Themen und mögliche Modelle zusammengetragen haben, für die wir uns interessierten. Wir haben dann versucht, diese möglichst gut und sinnig zusammenzuführen und sind durch gemeinsame Interessen im Bezug auf Recommender Systeme und Emotion Recognition auf unser finales Projektthema gekommen.

Außerdem benötigten wir eine passende Entwicklungsumgebung, in der wir zusammen an unserem Projekt arbeiten können.

Da in unserer Gruppe bereits erste Erfahrungen mit [Google Colab](#) bestanden, einer kostenlosen, cloudbasierten Jupyter-Notebook-Umgebung, mit der man Modelle für maschinelles Lernen und Deep Learning auf CPUs, GPUs und TPUs trainieren kann (vgl. Ultralytics 2024), haben wir uns entschieden, dies mit der Kombination von Python's pandas, librosa und sklearn Bibliotheken zu benutzen.

Um unser Modell zu trainieren, benötigten wir einen passenden Datensatz, der sowohl Audiodateien als auch zugehörige Labels zur Stimmung des Songs beinhaltet. Wir haben schließlich das [Ludwig Music Dataset](#) ausgewählt, weil es der einzige von uns gefundene fehlerfreie Datensatz war, der unsere Anforderungen erfüllt.

Das Recommendation-System haben wir mittels K-Nearest-Neighbor umgesetzt, weil es in vielen Fällen bereits mit wenig Rechenleistung gute Ergebnisse erzielen kann und so in der Lage ist, effizient und schnell Playlists zu generieren.

2.2 Feature Extraction

In unserem Datensatz konnte zunächst jeder Song mittels der vorhandenen Emotions-Labels eines der folgenden Stimmungen zugeordnet werden: Happy, Sad, Aggressive, Party oder Relaxed. Damit sich die Ergebnisse des KNNs nicht verfälschen, haben wir von jeder Stimmung möglichst eine ähnliche Anzahl an Songs für unser KNN ausgewählt.

Der zeitintensivste Schritt des Projekts lag darin, die Features jedes dieser Songs aus den Audiodateien zu extrahieren. Dabei haben wir verschiedene Features wie z.B. das Tempo, die MFCC-Werte oder die Zero-Crossing Rate betrachtet. Die Python Module Librosa und Pandas halfen uns dabei, die Features möglichst einfach und effizient zu laden und abspeichern. Trotzdem hat der Prozess je nach Anzahl der Features jedes Mal mehrere Stunden gedauert.

2.3 Umsetzung mit KNN

Nachdem für jeden Song die numerischen Features vorhanden waren, haben wir die Werte normalisiert, um keine fälschliche Gewichtung in Kauf zu nehmen. Alle Songs konnten nun auf Trainings und Testdaten aufgeteilt werden. Die Trainingsdaten wurden in das mittels des Python Moduls Scikit-Learn implementierte KNN-Modell geladen, in dem jeder Song nun durch den eigenen Feature-Vektor dargestellt wird.

Werden nun die Feature-Werte eines neuen Songs, zu dem eine passende Playlist generiert werden soll, eingegeben, wird dieser als neuer Vektor des KNNs betrachtet, von dem sich die Euklidische Distanz zu allen anderen geladenen Songs berechnen lässt. Die bei den K-nächsten Nachbarn am häufigsten vorkommende Stimmung wird nun als erwartete Emotion für den eingegebenen Song ausgegeben (Siehe Abbildung 2) und die Nachbar-Songs selbst bilden die erstellte Playlist.

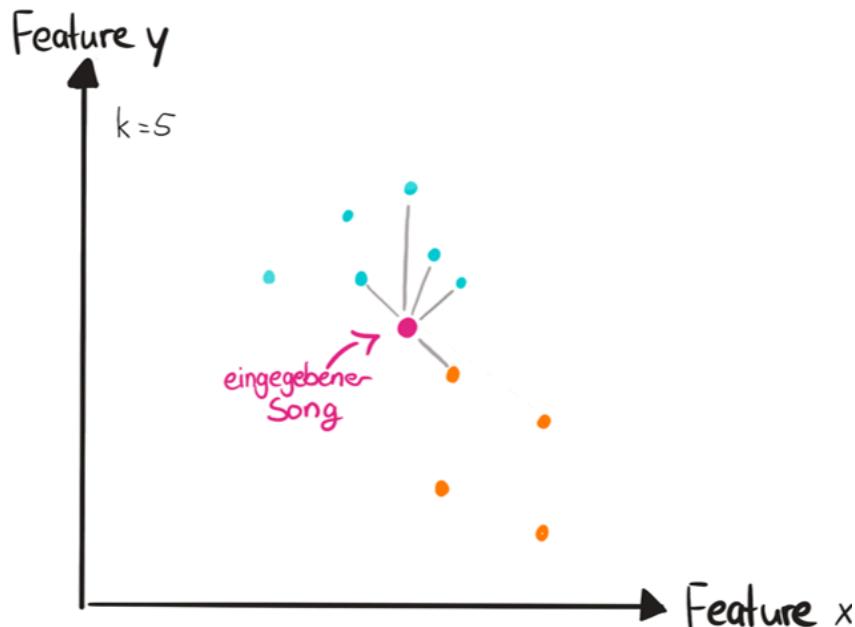


Abb. 2: Bestimmung des erwarteten Labels anhand der Labels der K-nächsten Nachbarn

Nachdem dieser Prozess für die gesamten Testdaten einmal durchgelaufen ist, kann die Accuracy, also der Anteil der Songs, bei denen das erwartete Label mit dem richtigen Label des Datensatzes übereinstimmt, ausgegeben werden.

2.4 Realisierung in drei Modellen

Durch eine Studie zur optimalen Feature Auswahl für Emotionserkennung in Musik haben wir herausgefunden, dass im Gegensatz zu einzelnen Features allein die richtige Kombination dieser Features zu guten Ergebnissen kommen kann (vgl. Le Cai et al., 2022). Dass einzelne Features nicht ausreichen würden, um die Emotion zu bestimmen, konnten wir an unserem Datensatz ebenfalls unschwer erkennen (Siehe Abbildung 3). Bestimmte Kombinationen sollten jedoch in der Lage sein, die Emotionen in einzelne räumliche hochdimensionale Cluster zu unterteilen, mit denen eine Stimmungserkennung durch ein KNN ermöglicht wird. Diesen Punkt visuell nachzuprüfen, ist aufgrund der höheren Dimensionen nicht möglich, aber ein Blick auf die Kombination durch zwei Features sah zumindest vielversprechender aus, als die Features einzeln (Siehe Abbildung 4).

Daher haben wir uns letztendlich dafür entschieden 3 verschiedene KNN Modelle der Songs zu erstellen, um die bestmöglichen Ergebnisse zu erzielen und den zweiten Teil der oben genannten Theorie zu testen: Ein kleines Modell mit 7 Audio-Features bzw. 19 numerischen Werten, ein mittleres Modell mit 8 Audio Features bzw. 73 numerischen Werten (dieses Modell hat unter anderem bestimmter Features auf mehrere Werte aufgeteilt und neben den Mittelwerten auch Standardabweichungen mit berechnet) und ein großes Modell mit 20 Audio Features bzw. 48 numerischen Werten, bei dem wir uns an die Auswahl von einer bereits existierenden Arbeit gehalten haben (vgl. IEEE Journals & Magazine, 2023).

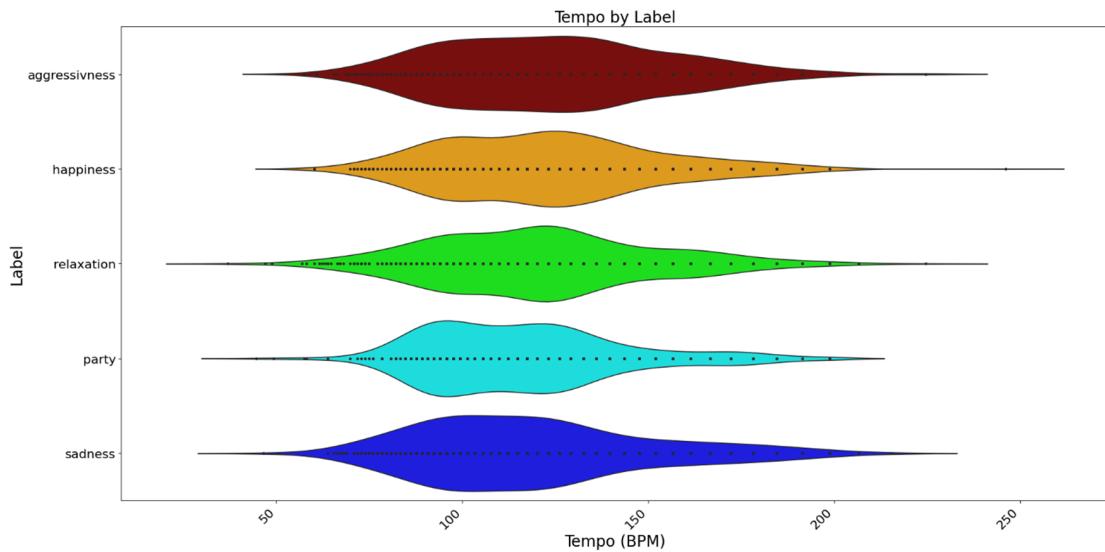


Abb. 3: Zeigt die Tempos der für unser KNN verwendeten Songs, aufgeteilt nach Label. Zu erkennen ist, dass das Bestimmen des Labels allein unter gegebenem Tempo keinen echten Erfolg haben kann.

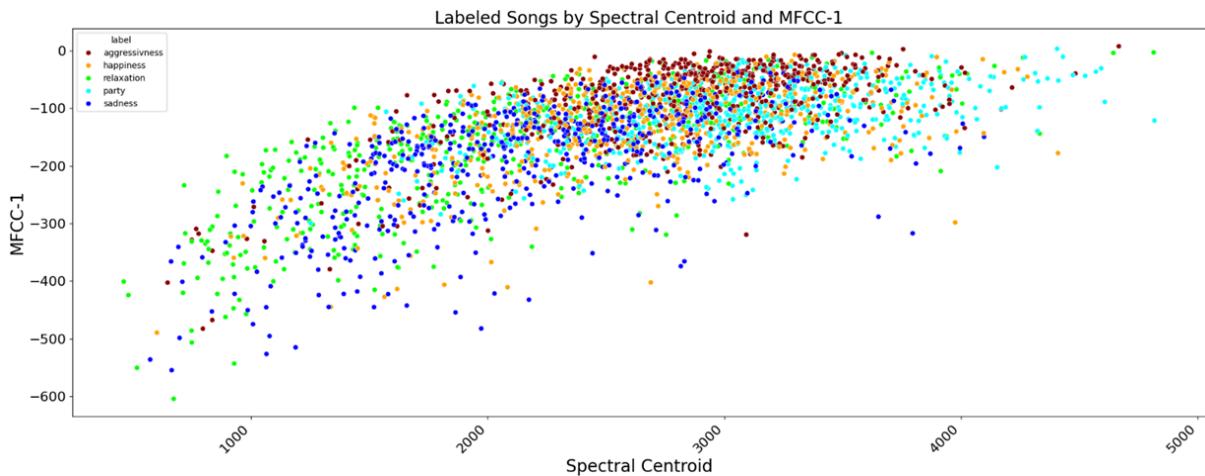


Abb. 4: Zeigt die Kombination aus Spectral Centroid und dem ersten MFCC-Wert der für unser KNN verwendeten Songs, aufgeteilt nach Label. Zu erkennen ist, dass das Bilden von Label Clustern noch nicht ausreichend ausgeprägt, aber zumindest angedeutet erkennbar ist.

2.5 Import in eine Huggingface Umgebung

Schließlich haben wir eine “Huggingface Space” eröffnet, in der das kleine Modell ohne zusätzlich benötigte Software ausgetestet werden kann. Wir haben uns für dieses Modell entschieden, da es mit einer ähnlichen Genauigkeit wie die anderen Modelle, aber weniger Features arbeitet, wodurch die Effizienz verbessert wird. Mithilfe der Weboberfläche Gradio musste dafür nur eine zusätzliche Funktion zum Ausgeben der erwarteten Emotion und der neuen Playlist programmiert werden.

3. Ergebnisse

Die angegebenen Genauigkeiten des kleinen, mittleren und großen Modells waren in der Reihenfolge 48,46%, 48,63% und 48,46%. Für eine Zuordnung von fünf möglichen Werten ist dies kein guter Wert, aber zumindest deutlich besser, als hätte das Modell zufällig geraten. Sehr interessant war, dass nicht nur die Accuracy, sondern auch die Confusion Matrizen der drei Modelle sehr große Ähnlichkeiten aufzeigen, ohne genau gleich zu sein. Dabei fiel es den Modellen erheblich leichter, die Aggressivität oder Party Emotion in einem Song zu bestimmen, während unter anderem entspannende Songs kaum erkennbar waren (siehe Abbildung 5).

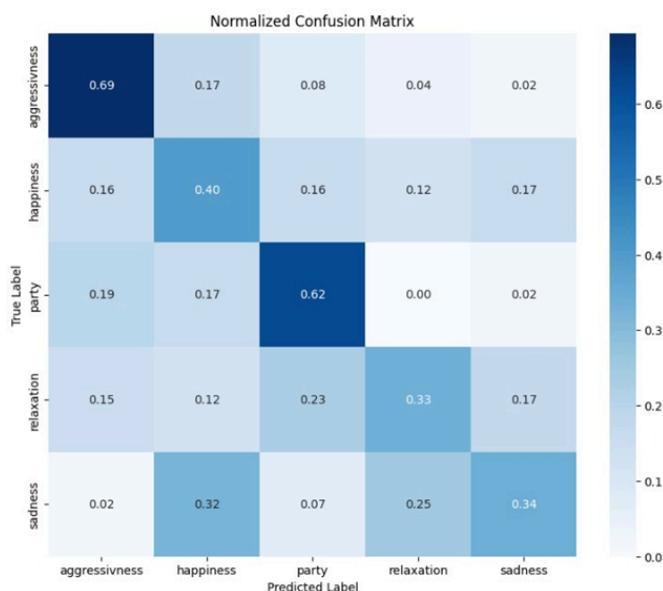


Abb. 5: Normalisierte Confusion Matrix des kleinen Modells (Die beiden anderen Modelle sind im Anhang zu finden, bieten aber keine großen Unterschiede.)

4. Analyse

Im folgenden Abschnitt gehen wir auf die Ergebnisse des Projekts ein und wie wir diese analysiert haben.

4.1 Gründe für die Ungenauigkeiten der Modelle

Die kaum unterscheidbaren Ergebnisse der drei Modelle implizieren entgegen unserer ursprünglichen Annahme, dass die Kombination verschiedener Features doch nicht der ausschlaggebende Punkt für den Erfolg eines KNN Modells zur Emotionserkennung ist. Unwahrscheinlich ist also, dass die Features für die schlechte Genauigkeit unserer Modelle verantwortlich sind. Dies bringt jedoch die Frage auf, was der stattdessen der Grund für die ungenauen Ergebnisse sein könnte.

Um einen besseren Überblick über das mögliche Problem zu bekommen, haben wir einen Teil der durch unsere Modelle fehlerhaft gelabelten Songs erneut per Hand einer der fünf Emotionen zugewiesen. Dafür haben zwei Personen, denen die erwarteten oder tatsächlichen Labels der Songs unbekannt waren, unabhängig voneinander 60 der fehlerhaft erkannten Songs nochmal gelabelt (siehe Tabelle 1).

Dabei heraus kam, dass beide Personen das Label nicht viel besser erkannt haben, als wenn sie dieses zufällig zugewiesen hätten. Generell lässt sich keine signifikante Übereinstimmung zwischen den Zuweisungen einer Person und den vom Datensatz oder Modell zugewiesenen Labels der Songs feststellen. Die beiden haben jedoch fast die Hälfte der Songs dem gleichen Label zugewiesen, eine, wenn auch nicht besonders große, deutlich auffälligere Übereinstimmung.

	Person 1	Person 2
Vergebenes Label stimmt mit dem Label des Datensatzes (True Label) überein	13	17
Vergebenes Label stimmt mit dem Label des Modells (Predicted Label) überein	13	10
Vergebenes Label stimmt nicht mit dem Label des Datensatzes oder Modells überein	34	33
Häufigkeit der Übereinstimmung zwischen den labelnden Personen	29	

Tabelle 1: Auswertung des manuellen Zuweisen der Label. Eine detaillierte Übersicht befindet sich im Anhang

Diese Ergebnisse lassen ein paar Vermutungen zu den möglichen Problemen unseres Modells zu.

Eine davon suggeriert, dass der von uns gewählte Datensatz nicht optimal gelabelt ist. Durch mehrere Stichproben konnten wir Lieder finden, deren Emotion wir bspw. klar als traurig oder entspannend einstufen würden, jedoch mit dem Label aggressiv im Datensatz gespeichert war.

Außerdem fiel uns beim Zuweisen der Label auf, dass die Songs extrem divers waren. Dabei kamen unter anderem Audiodateien mit Musik aus den 1970er Jahren, Partyhits aus den 2000-2010er Jahren oder Symphonien von Johann Sebastian Bach vor. Um trotz dieser großen Diversität an Genres ein gutes KNN zu ermöglichen, wäre wahrscheinlich ein viel größerer Datensatz nötig.

Zusätzlich kamen wir auf die Idee, dass manche Emotionen allgemein schwer voneinander zu unterscheiden sind. Vor allem Fröhlichkeit und Entspannung oder Fröhlichkeit und Party klingen in der Theorie sehr ähnlich und hängen häufig mehr

von bereits gemachten Erfahrungen des Zuhörers ab. Genauso gut kann man dafür argumentieren, dass ein Song in beide dieser Kategorien fallen kann, was mögliche Kritik an der eindeutigen Zuordnung von Songs zu Labels erlaubt.

Das Hauptproblem ist aber vermutlich, dass die emotionale Wahrnehmung von Musik sehr subjektiv ist (vgl. Kessler et al., 2004). Menschen, die sich in unterschiedlichen Lebenssituationen, Kulturen oder Stimmungen befinden, werden einen Song häufig auf sehr unterschiedliche Weise wahrnehmen und daher auch eine unterschiedliche Stimmung zuweisen. Dass die beiden labelnden Personen mehr Übereinstimmungen untereinander als mit dem Datensatz haben, könnte auch nur darauf hindeuten, dass sie sich z.B. in einem ähnlichen Umfeld oder Alter befinden.

4.2 Evaluation des kleinen Modells

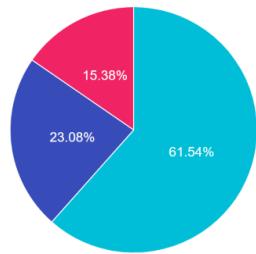
Die schlechte Genauigkeit unseres Modells veranlasste uns außerdem dazu, eine Evaluation für unsere Anwendung zu erstellen. Mithilfe dieser wollten wir herausfinden, ob unser kleines Modell in der Praxis besser ist, emotional ähnliche Songs zu finden, als die Label des Datensatzes angeben. Dabei können Anwender [einen Fragebogen](#) bezüglich der Genauigkeit unseres Modells anonym ausfüllen, nachdem sie es [hier](#) selbst getestet haben.

Ein Anwender hat die Möglichkeit, einen zufälligen Song aus dem Datensatz als Eingabe-Song zu benutzen oder selbst eine Datei hochzuladen. Nach dem Klicken auf "Submit" wird sowohl die analysierte Stimmung, als auch die auf dieser Stimmung basierende Playlist angezeigt. Die sich darin befindenden Songs können sich ebenfalls angehört werden.

Die Fragen der Umfrage gingen hauptsächlich um die von unserem Modell erwartete Stimmung des Eingabe-Songs sowie die Songs in der entsprechenden Playlist und ob sie mit dem persönlichen Empfinden der Testperson übereinstimmte.

Die Ergebnisse dieser Evaluation haben uns in der Annahme, dass das Modell besser sei, wenn es nicht von den Labels des Datensatzes abhängt, bestätigt. Bei jeder Frage, in der nach emotionaler bzw. musikalischer Ähnlichkeit zwischen Songs gefragt wurde, waren die Optionen "Applies to every song" und "Applies to most songs" die am häufigsten gewählten Antworten.

The generated playlist musically matched the provided song.
Beantwortet: 13 Übersprungen: 1



The songs in the generated playlist were emotionally coherent amongst each other.
Beantwortet: 13 Übersprungen: 1

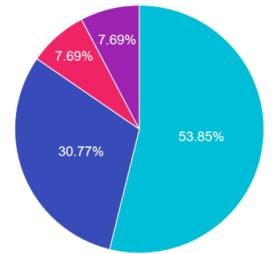


Abb. 6: Auszüge aus Auswertung der Evaluation

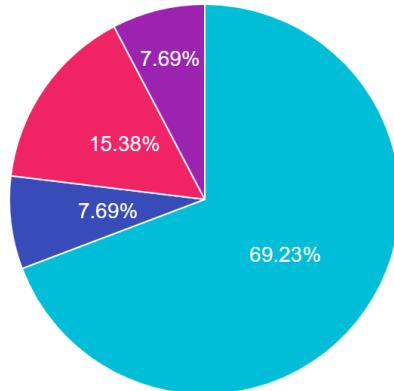
- Links: 84.62% empfinden die generierte Playlist als mindestens überwiegend musikalisch passend.
- Rechts: 84.62% empfinden die Songs der generierte Playlist als mindestens überwiegend kohärent untereinander.

Diese Fragen beziehen sich hier zunächst noch nicht auf die erwartete Emotion der Songs, sondern lediglich auf ihre Kohärenz untereinander. Demnach zeigt dieses Ergebnis, dass wir es mit unserem Vorgehen grundsätzlich geschafft haben, musikalische bzw. emotionale Ähnlichkeiten der Songs im KNN widerzuspiegeln.

Bei der konkreten Zuordnung des Gefühls zeigt die Frage nach der empfundenen Passung der Predicted-Emotion für die generierte Playlist nun aber eine leichte Abweichung.

The predicted emotion for the playlist was suitable for the provided song.

Beantwortet: 13 Übersprungen: 1



● Applicable. ● Rather applicable. ● Rather not applicable. ● Not applicable.

Abb. 7: Auszug aus Auswertung der Evaluation

- 76.92% empfinden die Predicted-Emotion als mindestens überwiegend passend.
- 23.07% empfinden die Predicted-Emotion als eher unpassend bis unpassend.

Hier zeigt sich nämlich, dass zwar ein Großteil diese Predicted-Emotion als zutreffend empfunden hat, allerdings ist die am zweithäufigsten gewählte Antwort hier, dass diese als eher nicht zutreffend empfunden wurde, was sich auch in der prozentualen Abweichung zu den anderen Fragen widerspiegelt.

Das bestätigt unsere Annahme, dass ein Song auch zu verschiedenen Emotionen passen kann. In dem Fall ist es subjektiv und abhängig von der jeweiligen Person, zu welcher Emotion der Song letztendlich zugeordnet wird.

Die vollständige Auswertung befindet sich im Anhang.

5. Projektmanagement & Versionierung

Das folgende Kapitel beleuchtet verschiedene Aspekte unserer Arbeitsweise im Team.

5.1 Vorgehensmodell Scrumban

Eine der ersten Entscheidungen, die wir im Team getroffen haben, war die Nutzung des Vorgehensmodells [Scrumban](#), jedoch angepasst auf unsere Bedürfnisse. Dabei handelt es sich um eine Vorgehensweise, die das agile Projektmanagement Framework Scrum mit dem ebenfalls agilen Projektmanagement Tool Kanban vereint.

Diese Entscheidung basierte auf vorheriger, positiver Erfahrung mit Scrumban als Vorgehensmodell. Außerdem passten die wöchentlichen Meetings mit dem kompletten Kurs ebenfalls gut zu diesem Modell, sodass wir unseren wöchentlichen Fortschritt gut im Blick behalten konnten.

5.2 Arbeitsweise im Team

Die einzelnen Sprint Termine erstrecken sich meist über mindestens zwei Stunden, da wir, anders als traditionell von Scrum vorgegeben, die Review, die Retrospektive und das Planning zusammengefasst haben. Die Dailys wurden komplett gestrichen.

Für kurze Absprachen zwischendurch wurde eine WhatsApp Gruppe verwendet, die Sprint Termine fanden über Discord statt.

Zur Organisation und Visualisierung wurde ein Trello-Board eingesetzt. Das Trello-Board verwenden wir zur Verwaltung der Tickets und damit auch zur

Visualisierung des Arbeitsflusses. Wir wollen in der Lage sein, Tickets neu zu priorisieren, zuzuweisen und zu aktualisieren. Durch diese Kombination aus Scrum und Kanban erreichen wir eine für uns und unser Projekt geeignete Arbeitsweise.

Für unser Kanban-Board verwenden wir die Software [Trello](#).

Da wir uns für das gemeinsame Arbeiten in Google Colab entschieden haben, machte es für uns Sinn, all unsere Dateien in einem gemeinsamen Google Drive Ordner zu speichern. So hatten wir alle zu jeder Zeit vollen Zugriff auf unser Projekt.

6. Fazit

Im folgenden Abschnitt wollen wir unser Endergebnis sowie unsere Zusammenarbeit im Team kurz reflektieren und zusammenfassen.

6.1 Endprodukt

Mit der entstandenen Software sind wir insgesamt zufrieden. Wir haben uns zu Anfang eine bessere Genauigkeit erhofft und sind während des Projekts auf unerwartete Schwierigkeiten gestoßen. Wir haben aber auch gemerkt, dass die niedrigere Genauigkeit eventuell für die in der Analyse herausgearbeiteten Probleme gar nicht so schlecht ist und dass das Erstellen der Playlist relativ gut funktioniert. Die gewonnenen Erkenntnisse in Bezug auf Emotionen, Musik und das Arbeiten mit einem KNN im Zusammenhang mit unserem Projekt stellen für uns zusätzlich eine große Bereicherung dar.

[Link zur Anwendung](#)

6.2 Learnings

Im Folgenden wollen wir übersichtlich festhalten, was in diesem Projekt gut funktioniert hat und wo es für das nächste Projekt Verbesserungsbedarf gibt.

- Durch die Arbeit mit dem K-Nearest-Neighbor (KNN)-Algorithmus haben wir gelernt, dass einfache Modelle oft gute Ergebnisse z.B. bezüglich der Interpretierbarkeit liefern können, wie bei der Musik-Emotionerkennung, jedoch auch Schwächen mit sich bringen.
- Im Rückblick hätten wir früher mit der Erprobung verschiedener Algorithmen beginnen können, um unser Modell zu optimieren. Zeitlich haben wir es leider nicht mehr geschafft, weitere Algorithmen für einen Vergleich zu entwickeln.
- Wir haben ein besseres Verständnis für Feature-Engineering entwickelt, wie etwa die Extraktion relevanter Audio-Features (z.B. Tempo, Tonhöhe, Spektrale Merkmale).

- Wir konnten verschiedene Python Libraries für Maschinelles Lernen kennenlernen.
- Das Arbeiten mit verschiedenen Datensätzen hat uns gezeigt, wie wichtig Datenqualität ist und wie lange das Finden eines hochwertigen Datensatzes für einen bestimmten Use-Case dauern kann. Zudem wäre es sicher sinnvoll gewesen, noch vor der Projektwahl zu überprüfen, ob geeignete Datensätze zur Verfügung stehen.
- Durch das Arbeiten in Google Colab waren wir stark von der begrenzten GPU abhängig, die nach zu langer Laufzeit für eine Weile gesperrt wurde. Dadurch haben wir gemerkt, dass es wichtig ist, entweder Software-unabhängiger zu arbeiten oder zumindest eine alternative Umgebung einzurichten.
-
- Eine klare Aufgabenverteilung und regelmäßige Meetings waren entscheidend für den Projektfortschritt. Wir haben gelernt, wie wichtig es ist, bestimmte Entscheidungen gemeinsam zu reflektieren, um ein gemeinsames Verständnis zu entwickeln.
- In unserem Fall hat sich, statt einem starren Plan zu verfolgen, ein iterativer Ansatz als effektiver erwiesen. Wir haben das Modell basierend auf Zwischenergebnissen und Fehleranalysen kontinuierlich anpassen können, auch wenn es am Ende nicht perfekt war.

7. Quellen

Alle Quellen wurden zuletzt am 12.01.2025 abgerufen.

Amaai-Lab. (o. D.). A curated list of Datasets, Models and Papers for Music Emotion Recognition (MER). GitHub. <https://github.com/AMAAI-Lab/awesome-MER>

Audio Features for Music Emotion Recognition: A Survey. (1 Jan.-März 2023). IEEE Journals & Magazine | IEEE Xplore. <https://ieeexplore.ieee.org/document/9229494>

Cai, Le & Ferguson, Sam & Lu, Haiyan & Fang, Gengfa. (2022). Feature Selection Approaches for Optimising Music Emotion Recognition Methods..
https://www.researchgate.net/publication/366656906_Feature_Selection_Approaches_for_Optimising_Music_Emotion_Recognition_Methods

Forschung, A. M. (1982). Emotionspsychologie und Musik. peDOCS.
<https://doi.org/10.25656/01:24763>

Katzlberger, M. (2021, 25. Juni). Wie und wo Spotify Künstliche Intelligenz nutzt. Artificial Creativity.

<https://katzlberger.ai/2021/03/29/wie-und-wo-spotify-kuenstliche-intelligenz-nutzt/>

Kessler, A., & Puhl, K. (2004). Subjectivity, Emotion, and Meaning in Music Perception.

http://www.parncltt.org/CIM04/CIM04_paper_pdf/KesslerPuhlproceedings.pdf

Ludwig Music Dataset (Moods and Subgenres). (2022, 16. April). Kaggle.

<https://www.kaggle.com/datasets/jorgeruizdev/ludwig-music-dataset-moods-and-subgenres/data>

Ultralytics. (2024, 1. Oktober). Google Colab. Ultralytics YOLO Docs.

<https://docs.ultralytics.com/de/integrations/google-colab/>

8. Anhang

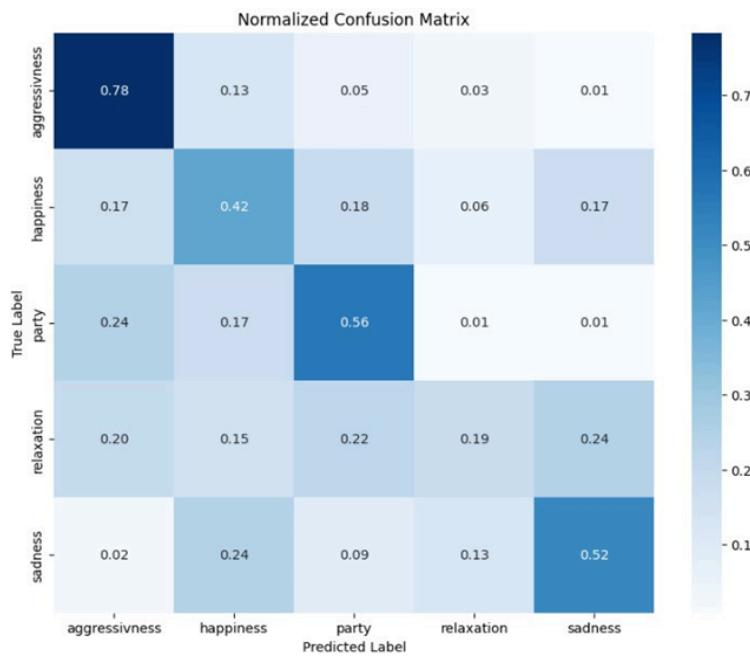


Abbildung: Normalisierte Confusion Matrix des mittleren Modells

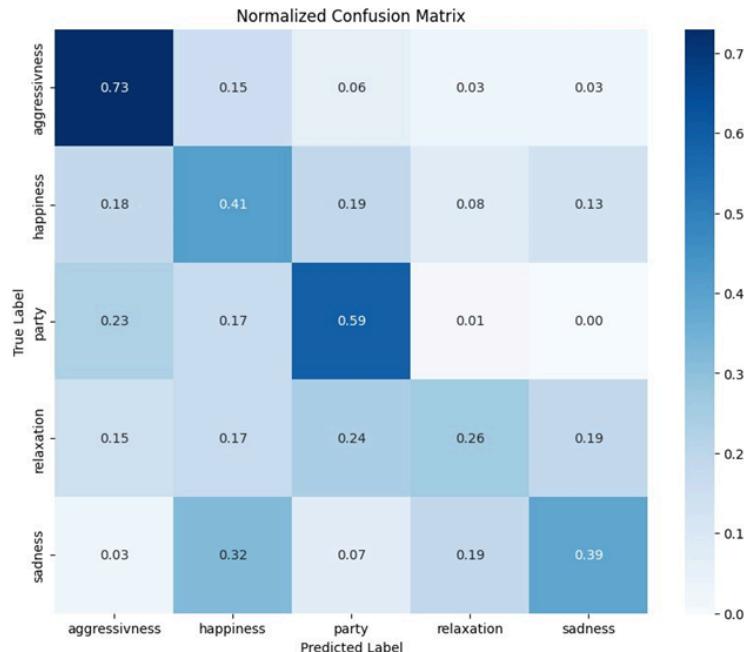


Abbildung: Normalisierte Confusion Matrix des großen Modells

ID	True Label	Predicted Label	Mimmie's Label	Laura's Label
2vcwbsUZS8Zarf5spZ02IQ	sadness	happiness	relaxation	happiness
5KHN4ERDcU8h2Vx6Xl8wKY	sadness	happiness	sadness	relaxation
7yTvcW3GVeFw3TkDUPvtLl	sadness	happiness	happiness	party
2ayfTxIwkeqWoEzmgJjKS	sadness	happiness	relaxation	sadness
0MNet5QmGhQ01ln7X8mVJ4	sadness	happiness	party	party
6zvGluRPhlmtPs8nPtgCR	sadness	happiness	happiness	relaxation
1IIydYQ6LTzPzgkLXYpNlNy	sadness	happiness	relaxation	aggressivness
72117vnmeQRAg8bptWypVG	sadness	happiness	sadness	relaxation
69jOKoPJuwpnbgWrfn7Yll	sadness	happiness	relaxation	happiness
0bETLttHjPjm7v94ce8x3D	sadness	happiness	happiness	happiness
1oFrzQ2srUSLeIPoamRmsu	sadness	happiness	happiness	relaxation
15UiSHRnDSrw5uAUtKOYSO	sadness	happiness	party	party
4K2fyKapgoneVla0cwQuwh	sadness	happiness	sadness	sadness
7d1bqTWYTYf2YrAAHVBYkL	sadness	happiness	party	party
3MdyGckdJyTuSylcK4Cn3e	sadness	happiness	aggressivness	happiness

ID	True Label	Predicted Label	Mimmie's Label	Laura's Label
71y7sPJm0DpCix0fHaIhi	party	aggressivness	aggressivness	happiness
3jaVs4si0Rt8woMSSH54yV	party	aggressivness	relaxation	aggressivness
1hDC618zxh7CTThfvIi0U2	party	aggressivness	aggressivness	party
5NOieEsnYibQkpRb8oaVsh	party	aggressivness	happiness	relaxation
6wVvi12xSRoDK2T7dmZbR	party	aggressivness	aggressivness	relaxation
07Jc7pNpyo5izZ6aLutNGL	party	aggressivness	party	party
4FzPIalyruWIiefI91qe4B	party	aggressivness	party	happiness
5YtcHwvM9O2BayUOsghqhO	party	aggressivness	happiness	party
18c7QvhQJqJwuUgs1k3H7E	party	aggressivness	relaxation	relaxation
2orX7XWXfT4eZNEFv90KFU	party	aggressivness	sadness	sadness
42qh86p7TLXyunxShn65kc	party	aggressivness	happiness	party
5ISPoYHI87FmHKV5mvS514	party	aggressivness	happiness	happiness
6vMNwB8iwOGEBk77NKGK77	party	aggressivness	party	party
70YvYr2hGl801kRlh01HN	party	aggressivness	happiness	happiness
6vkaGRe98A6x7s5dGmxsd30	party	aggressivness	happiness	relaxation

ID	True Label	Predicted Label	Mimmie's Label	Laura's Label
56uR6cc14YrbTW7QMhs9im	relaxation	sadness	sadness	relaxation
0IF7bHzCXCZeKnog5vBC4g	relaxation	sadness	sadness	sadness
118c2gbHcuYzzIVjsaNIpc	relaxation	sadness	relaxation	relaxation
6a08QZLi5a5vT6YmtxuAFU	relaxation	sadness	relaxation	relaxation
4axaxuvllF43svhCoadq1Jx	relaxation	sadness	aggressivness	party
5oHefuJW6udq0pphl5KXwN	relaxation	sadness	happiness	happiness
4i15m6DrqNMC8WP11Dn3h1	relaxation	sadness	happiness	relaxation
6d31c1mkWMxf5DPIW79f0	relaxation	sadness	happiness	party
0692fj6zoU8zkVfsY05p	relaxation	sadness	happiness	happiness
1rq1ODqNnghz78wCWg2obj	relaxation	sadness	relaxation	relaxation
3H3w2SDOp6ln2zBljFJEY	relaxation	sadness	happiness	happiness
3XFcsC3wl5SDZHPrH38co1	relaxation	sadness	happiness	happiness
7FQsdKvkCaNYnf8eboSgU	relaxation	sadness	relaxation	relaxation
7geFpUawLnGf11fnltiIR	relaxation	sadness	sadness	sadness
3W1NkQ3AzHuf00JeICQSER	relaxation	sadness	happiness	happiness

ID	True Label	Predicted Label	Mimmie's Label	Laura's Label
4tDa5P1so01pdvC5Yw16Or	relaxation	party	sadness	sadness
6yJxClgtgmGxIgKOS6ndnu	relaxation	party	happiness	sadness
0wrhLxy4aDAQ0z3FOa4ace	relaxation	party	party	aggressivness
31Lbh9SMQfaEuYubYRjYUp	relaxation	party	relaxation	relaxation
62WGawqtgtznIhLypZVlrK	relaxation	party	happiness	relaxation
5gnjaZqoAqE5GZRjby8RoT	relaxation	party	aggressivness	party
7w5By5oFAHsI71xAquFj9	relaxation	party	party	happiness
0Bof3JxdDYFxLjNer8n110	relaxation	party	happiness	happiness
2rgtCOVR2tpfae2VjURn84	relaxation	party	party	party
64Ocu3SrfYngj5V1sowMQG	relaxation	party	relaxation	party
7guWevC6eZXAw0w442BvOzZ	relaxation	party	relaxation	relaxation
0go0aA5ckbpAwuULJXqTNr	relaxation	party	aggressivness	aggressivness
4sbVo64EIP8kN44DcAyPlo	relaxation	party	sadness	sadness
1wSEVWRqzZdsPLB8zKqEoJ	relaxation	party	sadness	relaxation
6bj2r3CKqXhMnvb07VY4iR	relaxation	party	relaxation	sadness

Tabelle: Erneutes Labeln von von unserem Modell fehlerhaft erwarteten Songs durch Michelle und Laura, Übereinstimmungen mit dem True Label des Datensatzes in grün, Übereinstimmungen mit dem Predicted Label unseres Modells in blau.

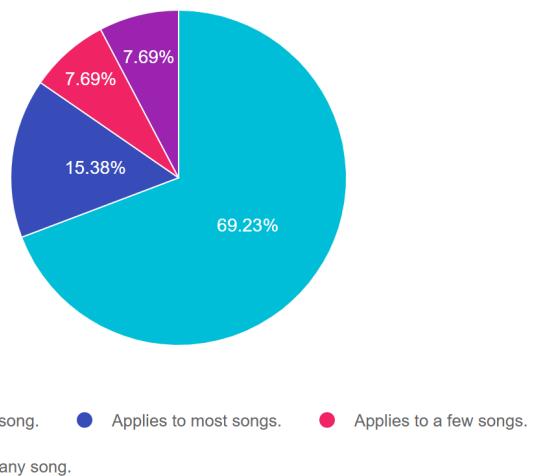
Vollständige Ergebnisse der Evaluation:

Response Statistics



The generated playlist emotionally matched the provided song.

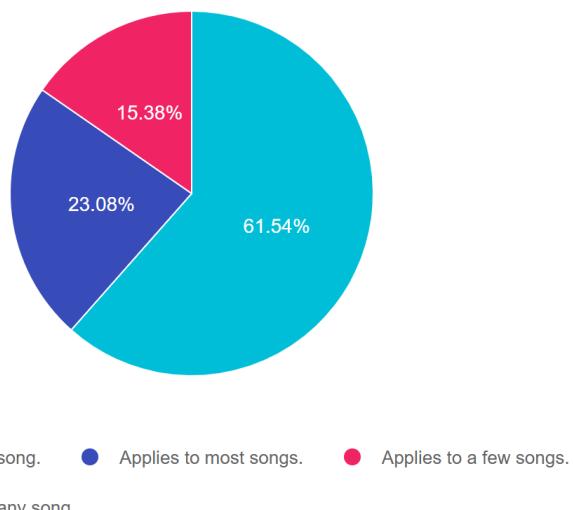
Beantwortet: 13 Übersprungen: 1



Auswahlen	Reaktionsprozentwert	Reaktionsanzahl
Applies to every song.	69.23%	9
Applies to most songs.	15.38%	2
Applies to a few songs.	7.69%	1
Doesn't apply to any song.	7.69%	1

The generated playlist musically matched the provided song.

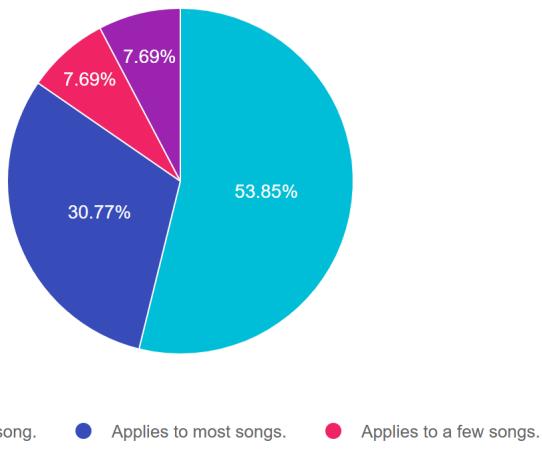
Beantwortet: 13 Übersprungen: 1



Auswahlen	Reaktionsprozentwert	Reaktionsanzahl
Applies to every song.	61.54%	8
Applies to most songs.	23.08%	3
Applies to a few songs.	15.38%	2
Doesn't apply to any song.	0.00%	0

The songs in the generated playlist were emotionally coherent amongst each other.

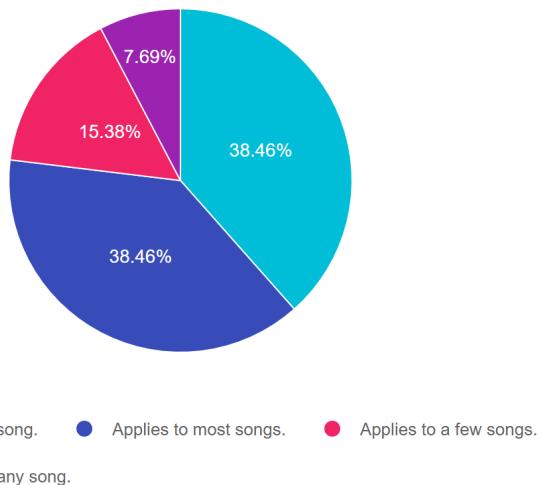
Beantwortet: 13 Übersprungen: 1



Auswählen	Reaktionsprozentwert	Reaktionsanzahl
Applies to every song.	53.85%	7
Applies to most songs.	30.77%	4
Applies to a few songs.	7.69%	1
Doesn't apply to any song.	7.69%	1

The songs in the generated playlist were musically coherent amongst each other.

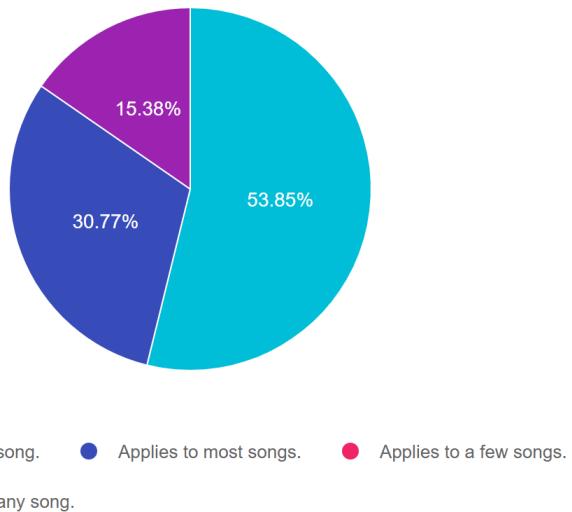
Beantwortet: 13 Übersprungen: 1



Auswahlen	Reaktionsprozentwert	Reaktionsanzahl
Applies to every song.	38.46%	5
Applies to most songs.	38.46%	5
Applies to a few songs.	15.38%	2
Doesn't apply to any song.	7.69%	1

The predicted emotion for the playlist was suitable for its songs.

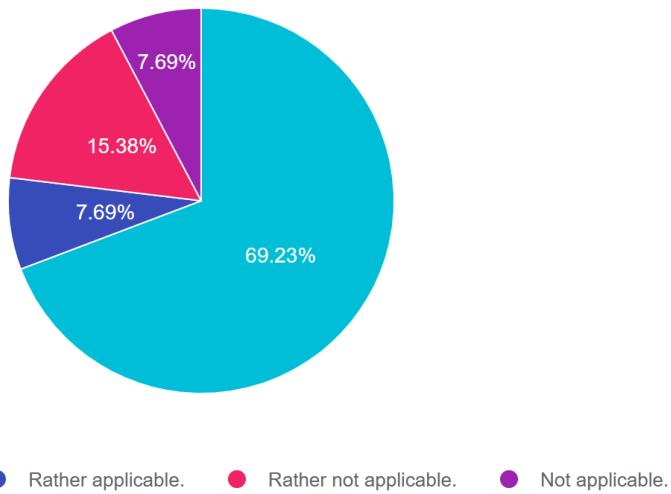
Beantwortet: 13 Übersprungen: 1



Auswahlen	Reaktionsprozentwert	Reaktionsanzahl
Applies to every song.	53.85%	7
Applies to most songs.	30.77%	4
Applies to a few songs.	0.00%	0
Doesn't apply to any song.	15.38%	2

The predicted emotion for the playlist was suitable for the provided song.

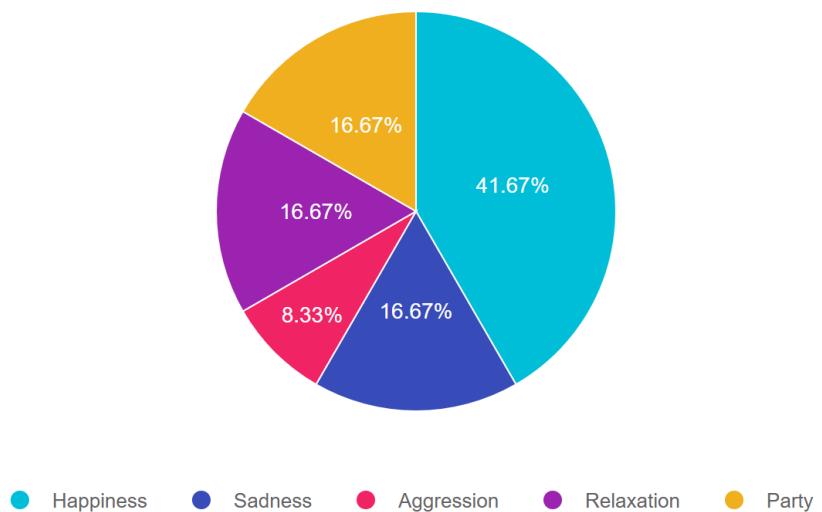
Beantwortet: 13 Übersprungen: 1



Auswahlen	Reaktionsprozentwert	Reaktionsanzahl
Applicable.	69.23%	9
Rather applicable.	7.69%	1
Rather not applicable.	15.38%	2
Not applicable.	7.69%	1

The predicted emotion for the playlist was:

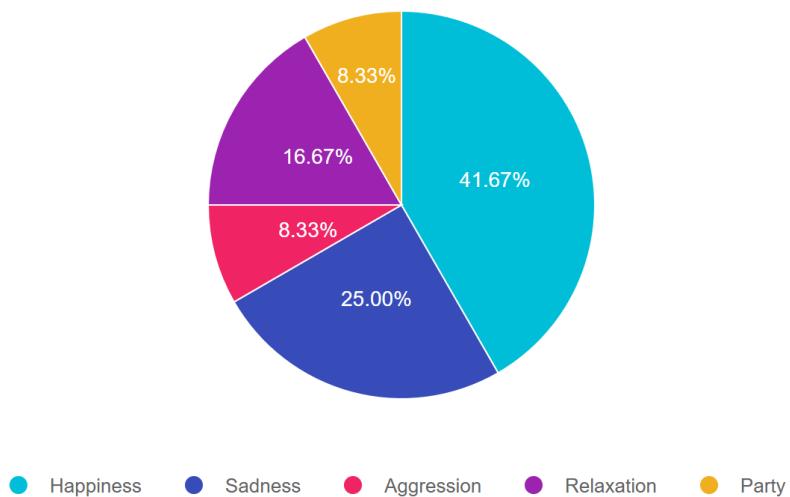
Beantwortet: 12 Übersprungen: 2



Auswählen	Reaktionsprozentwert	Reaktionsanzahl
Happiness	41.67%	5
Sadness	16.67%	2
Aggression	8.33%	1
Relaxation	16.67%	2
Party	16.67%	2

I would assign the following emotion to the provided song:

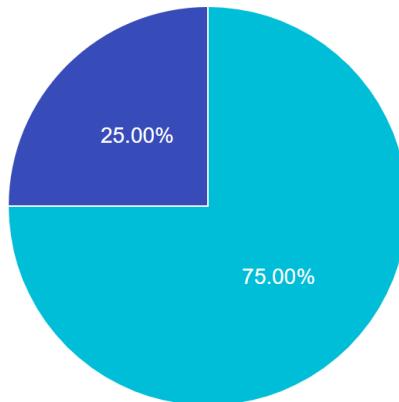
Beantwortet: 12 Übersprungen: 2



Auswahlen	Reaktionsprozentwert	Reaktionsanzahl
Happiness	41.67%	5
Sadness	25.00%	3
Aggression	8.33%	1
Relaxation	16.67%	2
Party	8.33%	1

Did you use a random song from the model?

Beantwortet: 12 Übersprungen: 2

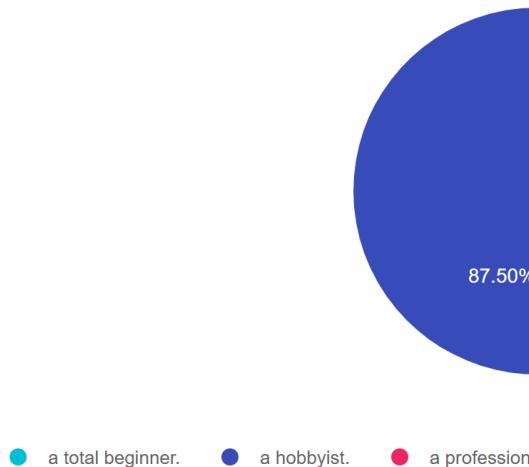


- Yes, I used a random song.
- No, I selected the song myself.

Auswählen	Reaktionsprozentwert	Reaktionsanzahl
Yes, I used a random song.	75.00%	9
No, I selected the song myself.	25.00%	3

When it comes to my knowledge about music, I consider myself...

Beantwortet: 8 Übersprungen: 6

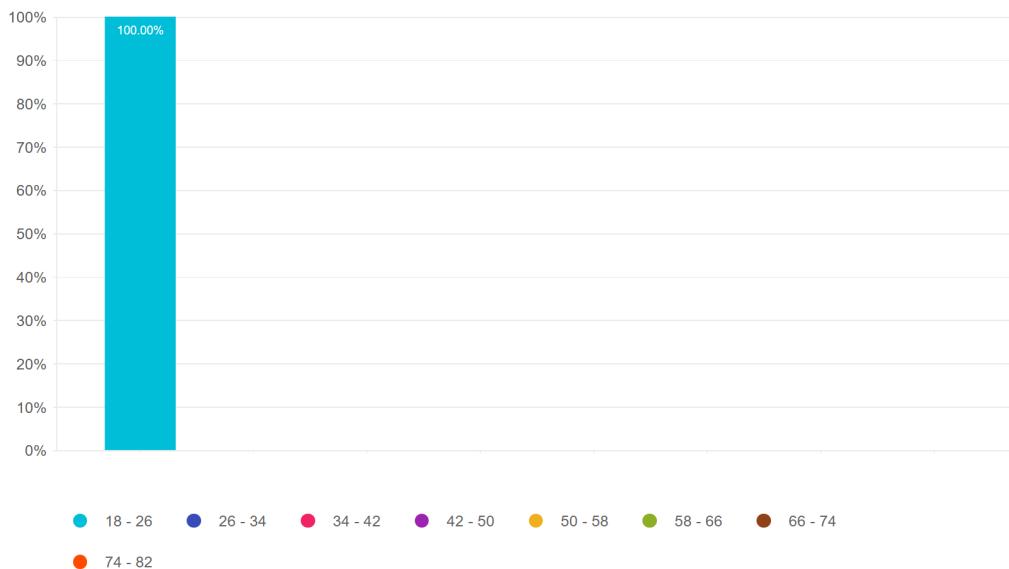


● a total beginner. ● a hobbyist. ● a professional.

Auswählen	Reaktionsprozentwert	Reaktionsanzahl
a total beginner.	12.50%	1
a hobbyist.	87.50%	7
a professional.	0.00%	0

My age:

Beantwortet: 8 Übersprungen: 6

**I identify as...**

Beantwortet: 8 Übersprungen: 6

