



# Collective Intelligence

Performance verschiedener Lieder auf Spotify und YouTube

Sommersemester 2023



Hochschule **RheinMain**

Fachbereich Design Informatik Medien  
Studiengang Informatik (Master of Science)

Projektabgabe: 24. Juli 2023

Gruppe: Sebastian Braun und Sandra Kiefer

Dozentin: Dr. Biying Fu

[Link zum GitLab-Repository](#)

[Link zum verwendeten Datensatz](#)



# Inhaltsverzeichnis

<b>1 Explorative Datenanalyse</b>	<b>3</b>
1.1 Beschreibung des Datensatzes	3
1.2 Motivation für die Problemstellung	5
1.3 Beschreibung des Vorgehens	6
1.4 Visualisierung der Daten	7
<b>2 Lösungsweg</b>	<b>12</b>
2.1 Architektur	12
2.1.1 Aufbau 1 (Vorhersagen von Views, Likes, Kommentaren, Streams)	12
2.1.2 Herausforderungen	13
2.1.3 Aufbau 2 (Klassifizierung der Popularität)	14
2.2 Implementierung	14
2.3 Evaluationsergebnisse	14
2.3.1 Logistic Regression	15
2.3.2 K-Nearest-Neighbor	15
2.3.3 Support-Vector-Machine	16
2.3.4 Naive Bayes	16
2.3.5 Decision Tree	17
2.3.6 Random Forest	17
2.3.7 Gradient Boosting	18
<b>3 Diskussion</b>	<b>19</b>
3.1 State-Of-The-Art Lösungen	19
3.2 Abgrenzung und Unterschiede	20
3.3 Stärken und Schwächen des Modells	20
3.4 Interpretation der Eigenschaften in den Popularitätsklassen	20
<b>4 Fazit und Ausblick</b>	<b>25</b>
4.1 Zusammenfassung der Ergebnisse	25
4.2 Ausblick	26



# 1 Explorative Datenanalyse

Streaming gehört in der heutigen Zeit zur beliebtesten Art und Weise, Musik zu hören. Im Vergleich zu anderen Streaming-Diensten hat sich Spotify als einer der Pioniere und Marktführer etabliert. Mit seinem Start im Jahr 2008 revolutionierte das Unternehmen die Art und Weise, wie Musik online gehört und entdeckt wird. Doch auch YouTube spielt eine herausragende Rolle im Streaming-Bereich und hat die Art und Weise, wie Menschen digitale Inhalte konsumieren, maßgeblich beeinflusst. Als größte Video-Sharing-Plattform der Welt bietet YouTube seit 2005 eine schier endlose Vielfalt an Inhalten, die weit über Musik hinausgehen, darunter Vlogs, Tutorials, Unterhaltungsshow und vieles mehr. Insbesondere in der Musikindustrie hat YouTube eine enorme Bedeutung erlangt, da es Nutzern die Möglichkeit bietet, offizielle Musikvideos, Live-Konzerte, Cover-Versionen und Remixe zu entdecken und anzuhören. In dieser Arbeit wird die Performance verschiedener Top-Hits auf den beiden Plattformen miteinander verglichen sowie ein Modell erarbeitet, das die Performance einzelner Songs vorhersagen soll. Dafür werden Daten genutzt, die die musikalischen Eigenschaften der Songs genauer beschreiben.

## 1.1 Beschreibung des Datensatzes

Für unser Projekt haben wir uns für einen Datensatz von Kaggle entschieden, der diverse Informationen über Top-Songs verschiedener Künstler beinhaltet. Zu den Informationen gehören generelle Informationen über das Lied wie Track, Artist und Album sowie unterschiedliche Eigenschaften, welche den Song genauer beschreiben. Zu den Eigenschaften gehören Attribute wie Danceability, Energy, Key, Loudness, Speechiness, Acousticness, Instrumentalness, Liveness, Valence und Tempo. Der Großteil der Eigenschaften wird in Werten zwischen 0 und 1 angegeben, was vorteilhaft für einen einheitlichen Vergleich der Daten ist. Was die einzelnen Eigenschaften genau darstellen, wird nachfolgend im Detail erläutert. Weitere wichtige Informationen des Datensatzes sind die Anzahl an Streams auf Spotify und die Anzahl der Views, Likes und Kommentare auf YouTube, da anhand dieser Werte die Performance der einzelnen Lieder gemessen werden kann. Der Datensatz umfasst Einträge von 20717 Liedern.

Die Eigenschaften der Lieder werden folgend genauer erläutert:

1. Danceability: Beschreibt, wie geeignet ein Song zum Tanzen ist, basierend auf einer Kombination von musikalischen Elementen wie Tempo, Rhythmusstabilität,

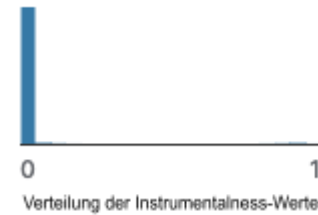


Beatstärke und allgemeiner Regelmäßigkeit. Der Wert reicht von 0,0 (am wenigsten tanzbar) bis 1,0 (am meisten tanzbar).

2. Energy: Ist ein Maß von 0,0 bis 1,0 und repräsentiert ein wahrnehmbares Maß an Intensität und Aktivität. Energetische Tracks fühlen sich typischerweise schnell und laut an. Zum Beispiel hat Death Metal eine hohe Energie, während ein klassisches Stück von Bach niedrig auf der Skala bewertet wird. Wahrnehmbare Merkmale, die zu dieser Eigenschaft beitragen, sind Dynamikumfang, wahrgenommene Lautstärke, Klangfarbe, Anfangsrate und allgemeine Entropie.
3. Key: Gibt die Tonart des Songs an. Ganze Zahlen werden mit der Standard Pitch Class Notation verknüpft. Zum Beispiel steht 0 für C, 1 für C♯/D ♭, 2 für D usw. Wenn keine Tonart erkannt wurde, ist der Wert -1.
4. Loudness: Die Gesamtlautstärke eines Tracks in Dezibel (dB). Die Lautstärkewerte werden über den gesamten Song gemittelt und sind nützlich, um die relative Lautstärke von Tracks zu vergleichen. Die Werte liegen typischerweise zwischen -60 und 0 dB.
5. Speechiness: Erkennt das Vorhandensein von gesprochenen Wörtern in einem Track. Je näher der Attributwert bei 1,0 liegt, desto mehr gleicht die Aufnahme gesprochener Sprache (z. B. Talkshow, Hörbuch, Poesie). Werte über 0,66 beschreiben Tracks, die wahrscheinlich ausschließlich aus gesprochenen Wörtern bestehen. Werte zwischen 0,33 und 0,66 beschreiben Tracks, die sowohl Musik als auch Sprache enthalten können, wie zum Beispiel Rap-Musik. Werte unter 0,33 repräsentieren höchstwahrscheinlich Lieder, in denen sehr wenig Text vorkommt.
6. Acousticness: Beschreibt die Wahrscheinlichkeit oder das Maß dafür, ob ein Song akustisch ist. Ein akustischer Song ist eine Aufnahme, die hauptsächlich oder ausschließlich mit akustischen Instrumenten erstellt wurde, wie z. B. akustische Gitarre, Klavier, Streichinstrumente oder Blasinstrumente. Im Gegensatz dazu werden nicht-akustische Songs oft durch elektronische Instrumente und Soundeffekte geprägt. Die Werte gehen von 0,0 bis 1,0 und geben die Wahrscheinlichkeit an, ob ein Track akustisch ist.



7. Instrumentalness: Prognostiziert, ob ein Track keine Gesangsstimmen enthält. "Ooh" und "aah" Klänge gelten in diesem Kontext als instrumental. Je näher der Wert an 1,0 liegt, desto wahrscheinlicher enthält der Track keinen Gesang. Da der Großteil der Werte jedoch sehr nah an der 0 liegt, wird diese Eigenschaft bei der genaueren Analyse der Daten vernachlässigt.



8. Liveness: Erkennt das Vorhandensein eines Publikums in der Aufnahme, zum Beispiel durch das Mitsingen und das Schreien der Zuhörer. Ein Wert über 0,8 deutet stark darauf hin, dass der Track live ist.
9. Valence: Ein Maß von 0,0 bis 1,0, das die musikalische Positivität eines Tracks beschreibt. Tracks mit hoher Valenz klingen positiver (z. B. glücklich, fröhlich, ekstatisch), während Tracks mit niedriger Valenz negativer klingen (z. B. traurig, deprimiert, wütend).
10. Tempo: Die geschätzte durchschnittliche Geschwindigkeit eines Tracks in Schlägen pro Minute (BPM). Das Tempo gibt die Geschwindigkeit eines Stücks an und leitet sich direkt von den durchschnittlichen Schlägen ab.

## 1.2 Motivation für die Problemstellung

Die Performance von Liedern anhand gewisser Eigenschaften eines Liedes vorhersagen zu können, ist in der Musikindustrie von entscheidender Bedeutung, da dies zahlreiche Vorteile und Möglichkeiten bietet. Zunächst ermöglicht eine solche Bewertung eine objektive Einschätzung der musikalischen Qualität und Popularität sowie eine Beurteilung der kommerziellen Erfolgschancen eines Liedes. Dies ist besonders wertvoll für Musikproduzenten und Plattenlabels, die ständig auf der Suche nach neuen Talenten und potenziellen Hits sind. Des Weiteren kann die Analyse von musikalischen Eigenschaften dazu beitragen, Zielgruppen besser zu verstehen und auf ihre musikalischen Vorlieben einzugehen. Dies wiederum erleichtert gezieltes Marketing und die Entwicklung maßgeschneiderter Kampagnen, um die Reichweite und den Erfolg eines Songs zu maximieren. Schließlich ermöglicht eine solche Bewertung auch, Schwachstellen in einem Lied zu identifizieren und gezielt an der Verbesserung von Liedern zu arbeiten.



Um die Performance vorhersagen zu können, muss ein Modell entwickelt werden, welches sich in die Themenfelder Supervised Learning und Klassifizierung einordnen lässt. Supervised Learning bezieht sich auf eine Art des maschinellen Lernens, bei dem ein Algorithmus mit einem gelabelten Trainingsdatensatz trainiert wird. Ein gelabelter Datensatz enthält sowohl Eingabevariablen (Features) als auch die zugehörigen Ausgabevariablen (Labels oder Zielwerte). Das Ziel besteht darin, eine Beziehung zwischen den Eingabevariablen und den Ausgabevariablen herzustellen, um Vorhersagen für neue, bisher ungesehene Daten zu treffen. Klassifizierung ist eine spezifische Aufgabe im Supervised Learning, bei der Daten in vordefinierte Klassen oder Kategorien eingeteilt werden sollen. Jede Klasse repräsentiert eine bestimmte Kategorie oder ein bestimmtes Label. Das Ziel besteht darin, ein Modell zu entwickeln, das die Eigenschaften der Lieder analysiert und diese den entsprechenden Klassen zuordnet. Der Prozess der Klassifizierung erfordert das Training eines Modells mit gelabelten Daten, bei dem das Modell lernt, die charakteristischen Merkmale jeder Klasse zu erkennen. Nach dem Training kann das Modell verwendet werden, um neue, nicht gelabelte Daten zu klassifizieren und ihnen die entsprechenden Klassen zuzuweisen. In unserem Fall sollen die Lieder anhand ihrer Eigenschaften in unterschiedliche Popularitätsklassen eingeteilt werden, um ihre Performance besser vorhersagen zu können.

### 1.3 Beschreibung des Vorgehens

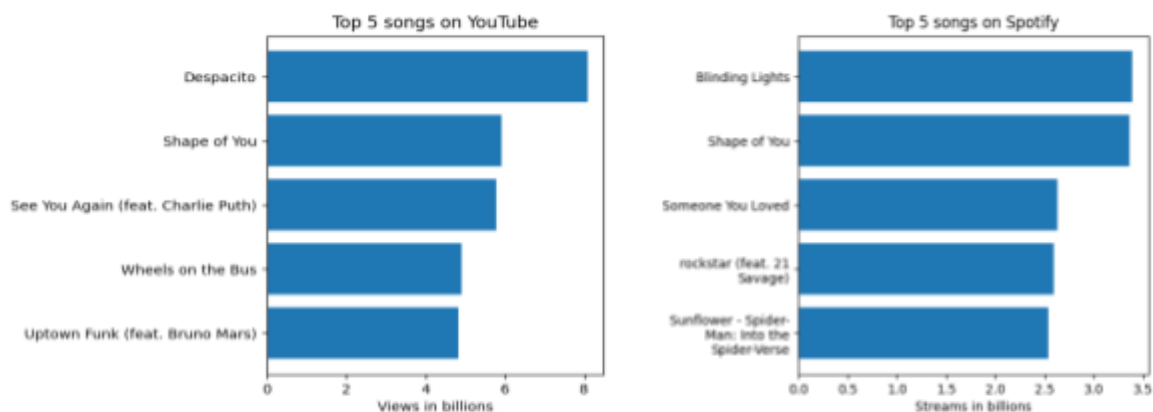
Um die Performance der Lieder erfolgreich vorhersagen zu können, müssen wir uns zuerst ein einheitliches Bild der vorhandenen Daten verschaffen und diese verstehen. Dazu erfolgt die explorative Datenanalyse, bei der erste Zusammenhänge und Vorhersagen getroffen werden können. Anschließend wird ein erstes Modell (eine Baseline) entwickelt, auf der nachfolgende Modelle aufbauen können. Zu Beginn wird das Modell mit den gegebenen Parametern trainiert und anschließend die Performance des Modells überprüft. Sollten die Ergebnisse unzufriedenstellend sein, wird der Datensatz um weitere Attribute ergänzt, die sich aus mehreren Attributen zusammensetzen und eine Popularitätsklasse ergeben. Mithilfe dieser Werte wird das Modell erneut trainiert, um die Baseline zu verbessern. Es gibt ebenfalls verschiedene Algorithmen, die für die Klassifizierung verwendet werden können, wie zum Beispiel k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), Decision Trees, Random Forests und andere. Schließlich wird dann überprüft, welches Modell die besten Vorhersagen treffen kann und wie stark sich dieses von der Baseline unterscheidet.



## 1.4 Visualisierung der Daten

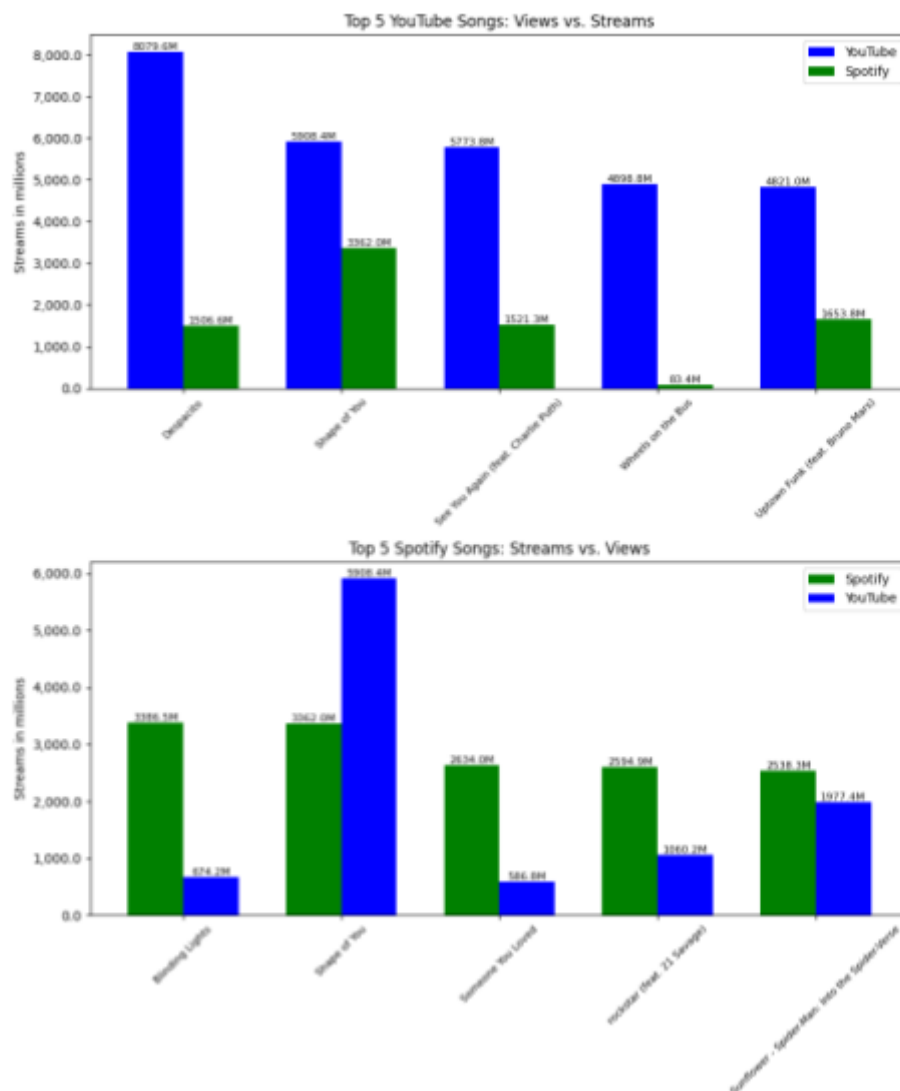
Die explorative Datenanalyse ermöglicht uns, die Daten zu verstehen und erste Einblicke in mögliche Zusammenhänge und Muster zu gewinnen. Auf Grundlage dieser Erkenntnisse können wir angemessene Modelle entwickeln, um die Performance von Songs auf Spotify und YouTube vorherzusagen.

Die nachfolgenden Visualisierungen wurden in Python mit der Bibliothek Matplotlib erstellt und sollen zum besseren Verständnis des Datensatzes beitragen. Zu Beginn wird erst einmal betrachtet, in welcher Größenordnung sich die Streaming-Zahlen der beiden Plattformen befinden. Dafür werden die fünf populärsten Songs der beiden Plattformen miteinander verglichen.



Übersicht der populärsten Lieder auf YouTube und Spotify mit den Views/Streams in Billionen

Der Vergleich macht deutlich, dass Lieder auf YouTube noch wesentlich häufiger aufgerufen werden als auf Spotify. Despacito hat auf YouTube knapp acht Milliarden Aufrufe erreicht. Das ist im Vergleich zu Platz 1 auf Spotify mit knapp 3,5 Milliarden Streams mehr als doppelt so viel. Interessant ist hier auch Platz zwei beider Plattformen. Shape of You von Ed Sheeran ist auf YouTube und Spotify der zweite meistgestreamte Song weltweit. Hier lassen sich die Aufrufe sehr gut miteinander vergleichen. Mit knapp sechs Milliarden Aufrufen auf YouTube ist die Performance fast wieder doppelt so hoch wie auf Spotify, mit ungefähr 3,4 Milliarden Aufrufen. Naheliegender wäre jetzt eine erste Schlussfolgerung, dass die Top 5 Songs auf beiden Plattformen ähnlich gut abschneiden. Dass dies jedoch nicht der Fall ist, zeigt die nachfolgende Abbildung relativ deutlich.



Genauer Vergleich der Streaming-Zahlen von Spotify mit den Aufrufzahlen von YouTube

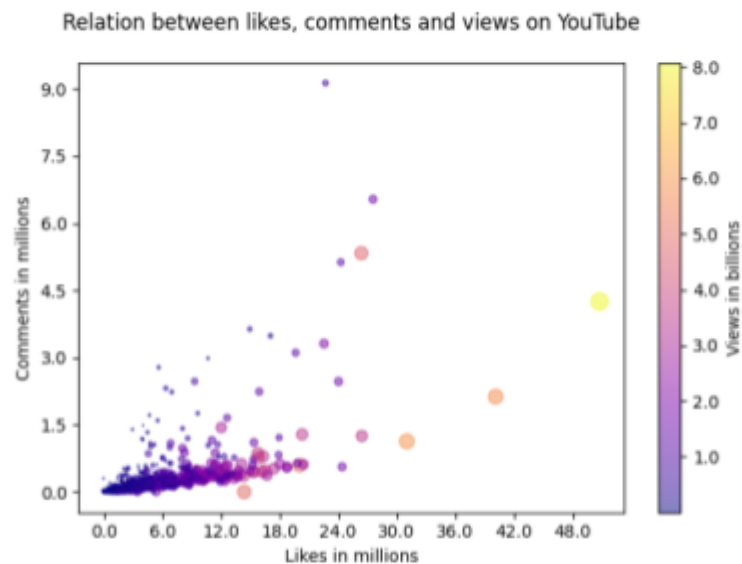
Die beiden Diagramme vergleichen die Anzahl der Streams der Top-Hits beider Plattformen. Es ist gut ersichtlich, dass sich auf den beiden Plattformen keine ähnliche Verteilung der Aufrufe abbildet. Im Gegenteil, die Verteilung erscheint eher willkürlich. So lässt sich als erste Schlussfolgerung ableiten, dass hohe Aufrufzahlen auf YouTube nicht auch für hohe Aufrufzahlen auf Spotify stehen und umgekehrt. Also müssen wir nach weiteren Indikatoren für eine hohe Performance eines Liedes suchen.

Ein weiterer Ansatz ist der Einfluss von Likes und Kommentaren auf die Anzahl der Aufrufe auf YouTube. Da Spotify diese Features nicht hat, bezieht sich dieser Ansatz nur auf die YouTube Performance. Die Idee ist, dass mithilfe von vielen Kommentaren und Likes, der YouTube Algorithmus beeinflusst werden soll. Dies würde bedeuten, dass eine hohe Anzahl an Likes und Kommentaren zu einer besseren Sichtbarkeit auf YouTube führen soll und dass





Lied so mehreren Personen vorgeschlagen werden würde. Um diesen Ansatz genauer zu untersuchen, betrachten wir den nachfolgenden Scatter-Plot.



Scatter-Plot zur Betrachtung der Relationen zwischen den Views, Likes und Kommentaren auf YouTube

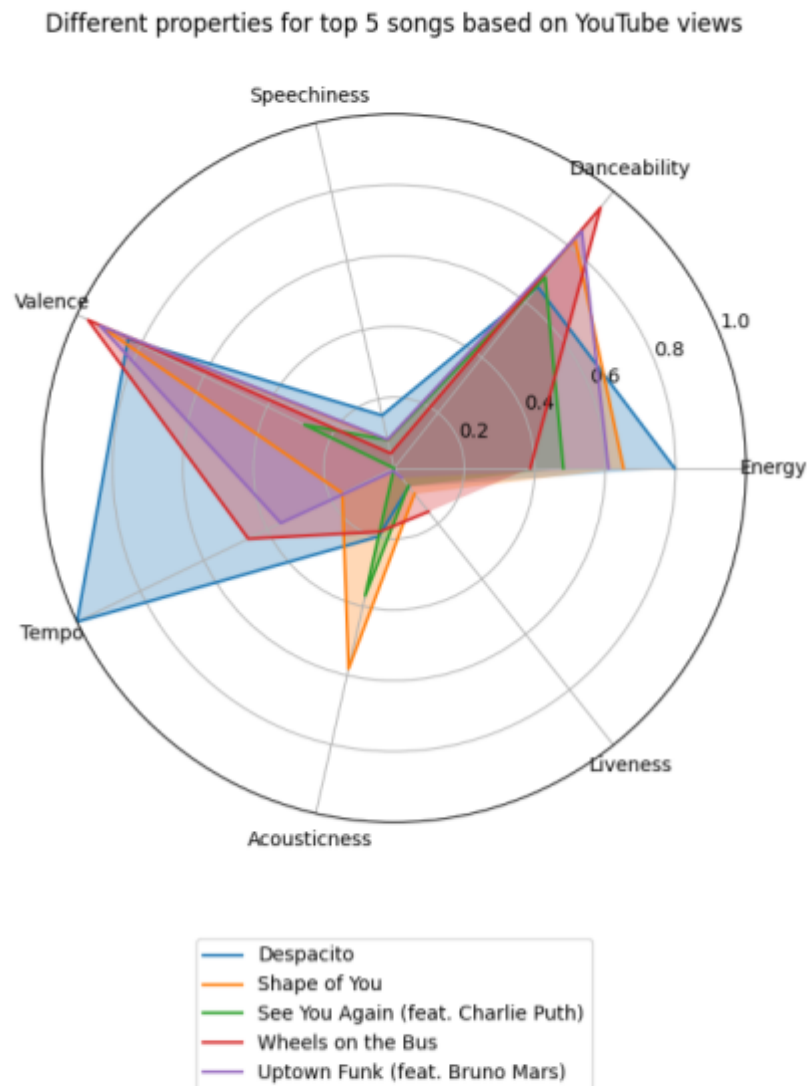
Auf der x-Achse ist die Anzahl an Likes in Millionen aufgeführt und auf der y-Achse die Anzahl an Kommentaren in Millionen. Die einzelnen Lieder werden als Kreise im Diagramm dargestellt. Die Größe und Farbe der Kreise steht für die Anzahl der Aufrufe der Lieder. Je größer und heller die Kreise sind, desto mehr Aufrufe haben sie.

Auffällig sind die größeren orangen und gelben Kreise im Diagramm. Diese sind in etwa wie eine Exponentialkurve angeordnet. Daraus lässt sich ablesen, dass Lieder mit hohen Aufrufzahlen tendenziell auch mehr Likes und Kommentare haben. Das Diagramm zeigt aber auch anhand der kleineren lila Kreise deutlich, dass eine hohe Anzahl an Kommentaren nicht automatisch zu hohen Aufrufzahlen führt. Diese könnten zum Beispiel auch entstehen, wenn ein Lied in den Kommentaren kontrovers diskutiert wird. Also lässt sich als zweite Schlussfolgerung ableiten, dass eine hohe Anzahl an Kommentaren und Likes nicht unbedingt ausschlaggebend für eine gute Performance sind, diese aber logischerweise höher ausfallen, je öfter ein Lied aufgerufen wurde.

Zuletzt untersuchen wir die Lieder genauer anhand ihrer Eigenschaften. Eventuell performen Lieder mit ähnlichen Eigenschaften gleich gut oder schlecht bzw. könnten die erfolgreichsten Lieder ähnliche Eigenschaften aufweisen. Dazu betrachten wir zunächst die Eigenschaften der Top 5 Songs von YouTube und Spotify mithilfe eines Radar-Charts. In dieser Art von Diagramm wurde nicht jede Eigenschaft verwendet, sondern nur jene, die auf einer Skala



von 0 bis 1 dargestellt werden können. Zusätzlich wurde der Wert Tempo in BPM auf eine Skala von 0 bis 1 umgerechnet, um die Geschwindigkeit der einzelnen Lieder miteinander vergleichen zu können.

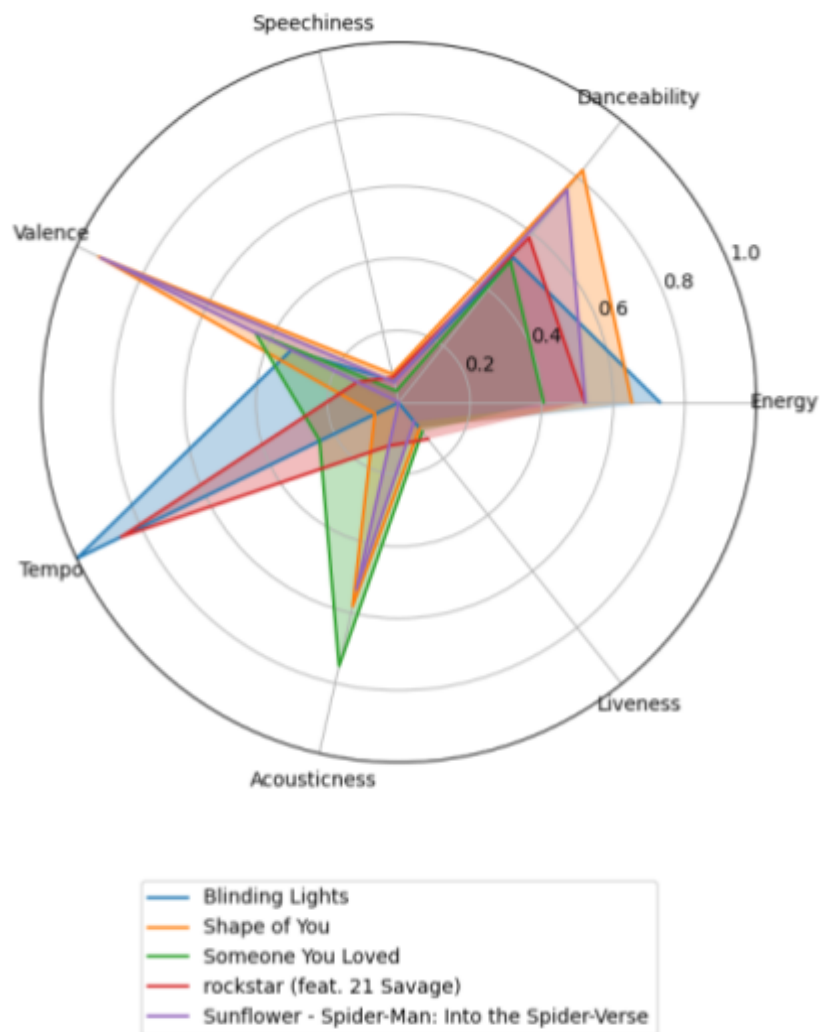


Visualisierung welche Eigenschaften besonders stark in den Top-Songs auf Youtube ausgeprägt sind

Auffallend sind einheitliche Ausschläge bei den Eigenschaften Valence, Danceability und Energy. Diese sind zwar unterschiedlich stark ausgeprägt, lassen aber grundsätzlich darauf schließen, dass fröhliche, energetische und tanzbare Lieder besser performen, als ruhigere, traurige Lieder. Auffallend ist auch, dass die Lieder einheitlich einen sehr geringen Speechiness und Liveness Wert aufweisen. Das deutet darauf hin, dass populäre Lieder einen hohen Anteil an Melodien haben und nicht live aufgezeichnet wurden.



Different properties for top 5 songs based on Spotify streams



Visualisierung welche Eigenschaften besonders stark in den Top-Songs auf Spotify ausgeprägt sind

Bei den Top 5 Songs von Spotify lässt sich ein ähnliches Muster erkennen. Einheitliche Ausschläge bei Valence, Danceability und Energy. Zusätzlich kommen hier noch größere Ausschläge bei Acousticness hinzu, was deutlich macht, dass sowohl elektronische als auch akustische Lieder gut performen. Speechiness und Liveness gehen wie zuvor gegen null. Der Tempo Wert ist bei beiden Diagrammen sehr durchmisch, was zeigt, dass ein bestimmter BPM-Bereich kein Kriterium für einen erfolgreichen Song ist.



## 2 Lösungsweg

Ziel des Projektes ist es, mithilfe der verschiedenen Eigenschaften eines Liedes die Performance vorherzusagen. Unter der Performance versteht man die Anzahl an Streams auf Spotify und die Anzahl der Views, Likes und Kommentaren vorherzusagen. Dazu werden verschiedene Modelle miteinander verglichen, um ein bestmögliches Ergebnis und Vorhersage zu schaffen. Mit den Ergebnissen können beispielsweise Aussagen überprüft werden, wie zum Beispiel, dass die meisten berühmten und gut performenden Lieder alle sehr ähnlich klingen. Demzufolge müssten alle Lieder mit bestimmten Eigenschaften wie Tonart, Tempo und Lautstärke ähnlich gut oder schlecht performen.

### 2.1 Architektur

Zur Bestimmung, ob ein Song gut oder schlecht auf Spotify und YouTube abschneidet, wurden verschiedene Methoden des überwachten Lernens angewandt. Im ersten Aufbau des Modells wurde versucht, durch Regression die Anzahl an Views, Likes, Kommentaren und Streams vorherzusagen. Jedoch wurde dabei schnell deutlich, dass keine aussagekräftigen Ergebnisse herausgekommen sind. Durch die genauere Betrachtung der Korrelationen der einzelnen Features wurde deutlich, warum dies der Fall ist. Aus diesem Grund wurde im zweiten Aufbau ein Wert für die Popularität jedes Liedes berechnet und darauf verschiedene Modelle zur Klassifizierung angewandt. Die Ergebnisse waren wesentlich aussagekräftiger und die Modelle wurden optimiert.

#### 2.1.1 Aufbau 1 (Vorhersagen von Views, Likes, Kommentaren, Streams)

Im ersten Aufbau des Modells wurde versucht, jeweils die Views, Likes, Kommentare und Streams eines Liedes mit linearer Regression und komplexer Regression (polynomiell) vorherzusagen. Jedoch waren die Ergebnisse nicht aussagekräftig und für eine Vorhersage der Popularität eines Liedes unbrauchbar. Deutlich wird dies beim Blick auf den MSE, der zwischen dem Milliarden- und Trilliarden-Bereich liegt. Das ist vor allem auf den großen Wertebereich der zu schätzenden Zahlen zurückzuführen, da beispielsweise die berühmtesten Lieder auf YouTube Aufrufzahlen im mehrstelligen Milliardenbereich haben.

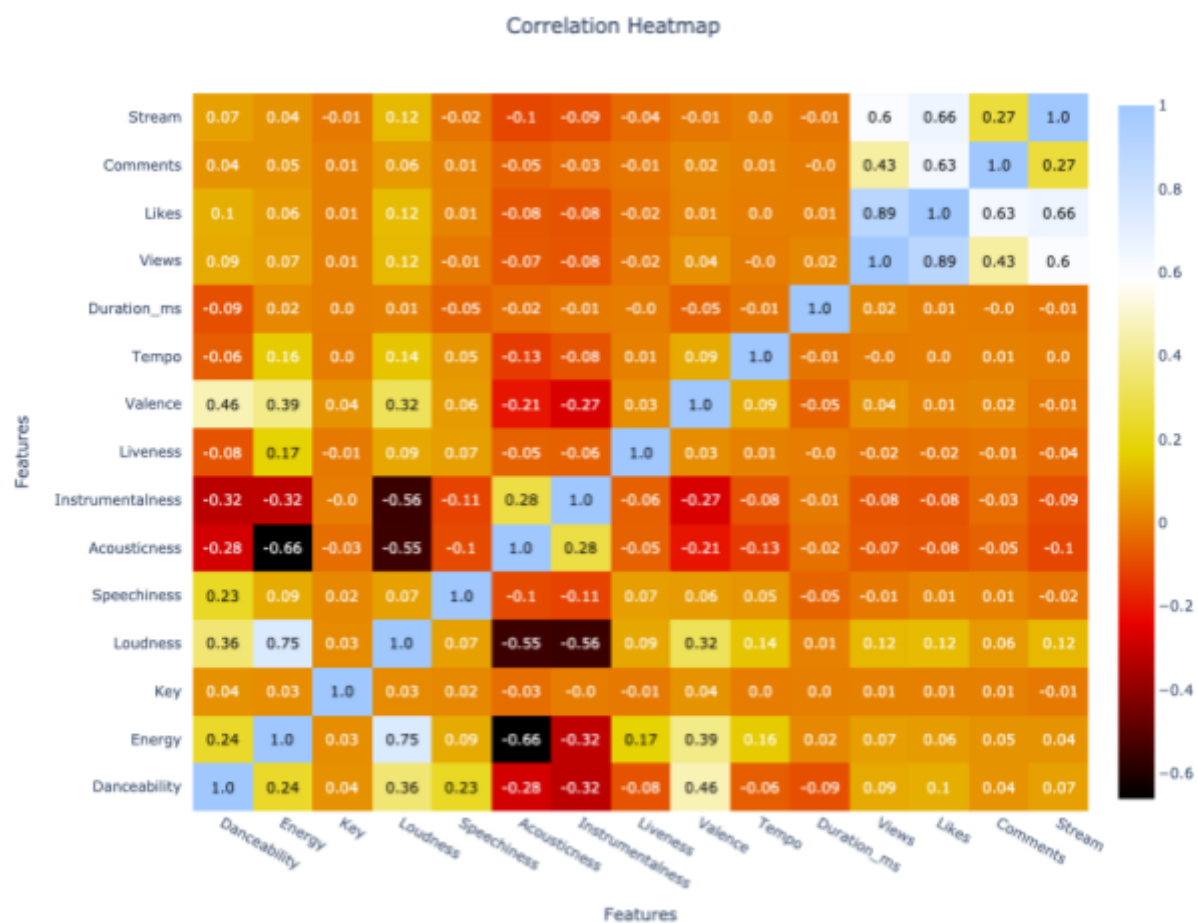
Vorhersage	Linear Regression (MSE)	Complex Regression (MSE)
Views	13 845 475 698 058 648	3 186 447 608 515 025 408
Likes	332 489 855 901	51 371 805 025 899
Kommentare	8 512 847 112	30 114 110 934
Streams	29 869 146 780 298 188	167 281 257 616 288 192

Tabelle zur Übersicht des Mean Squared Errors (MSE) bei Linearer und Komplexer Regression



### 2.1.2 Herausforderungen

Durch den ersten Ansatz wird deutlich, dass die genaue Vorhersage der View und Stream Zahlen sehr schwierig ist, da der einzuschätzende Wertebereich sehr groß ist. Bei Betrachtung der Korrelationen zwischen den numerischen Werten des Datensatzes fällt auf, dass die meisten Features eine geringe Korrelation haben (d.h. Wert um 0 herum) und somit sich teilweise zufällig zueinander verändern. Jedoch stehen die Features Streams, Kommentare, Likes und Views in Relation zueinander, was zufolge hat, dass populäre Lieder sowohl auf YouTube als auch auf Spotify gut performen.



Korrelations-Heatmap zur näheren Beschreibung der Beziehungen (Korrelationen) zwischen den verschiedenen numerischen Eigenschaften eines Liedes

Damit trotzdem eine Aussage über die Popularität eines Liedes getroffen werden kann, muss ein alternativer Lösungsweg gefunden werden. Die genaue Vorhersage von Views, Likes, Kommentaren und Streams ist eventuell gar nicht von so großer Bedeutung, sondern die grobe Einschätzung, ob ein Lied aufgrund seiner musikalischen Eigenschaften gut oder schlecht ankommt. Aus diesem Grund wird ein Popularitätswert und eine Popularitätsklasse aus den vier Werten (Views, Likes, Kommentare, Streams) zur Messung der Performance



berechnet. Dazu werden die vier Werte normiert, ein Gewicht darauf angewandt (Views = 0.3, Streams = 0.3, Likes = 0.2, Kommentare = 0.2) und summiert. Die genaue Zuteilung der Wertebereiche und der Klassen kann aus der untenstehenden Tabelle entnommen werden.

Popularitätsklasse	Low Popularity	Moderate Popularity	Good Popularity	High Popularity	Very High Popularity
Popularitätswert (%)	[0-30]	[31-50]	[51-80]	[81-90]	[91-100]

Tabelle zur Berechnung und Einordnung des Popularitätswertes und der Populairitätsklasse

### 2.1.3 Aufbau 2 (Klassifizierung der Popularität)

Im zweiten Aufbau werden die zuvor berechneten Popularitäten jedes Liedes dazu genutzt, um verschiedene Klassifizierungsalgorithmen darauf anzuwenden. Folgende Algorithmen und Methoden zur Klassifizierung der Performance eines Liedes wurden auf den Datensatz angewandt: Logistic Regression, K-Nearest-Neighbor, Support-Vector-Machine und Naive-Bayes. Des Weiteren wurden folgende Algorithmen und Techniken genutzt, die nicht im Rahmen der Lehrveranstaltung vorgestellt wurden, aber aus anderen Modulen bekannt waren: Decision Tree, Random Forest und Gradient Boosting. Alle genannten Modelle wurden mithilfe von Grid Search auf die besten Hyperparameter, welche die höchste Genauigkeiten liefern, optimiert und anschließend angewandt. Die Daten wurden in Trainings-, Validierungs- und Test-Sets (70 %, 15 %, 15 %) unterteilt.

## 2.2 Implementierung

Der Datensatz wurde aus der CSV-Datei in ein Pandas Dataframe eingelesen und von fehlenden Daten bereinigt. Für die Implementierung der Modelle wurden die Bibliotheken von Scikit-Learn verwendet. Zum Visualisieren der Ergebnisse in geeigneten Grafiken wurde die Bibliothek von Matplotlib verwendet. Der genaue Aufbau des Codes kann aus dem Repository entnommen werden.

## 2.3 Evaluationsergebnisse

Mit dem zweiten Aufbau und der Klassifizierung der Popularität eines Liedes konnten gute Ergebnisse erzielt werden und eine zuverlässige Aussage über die Performance eines Liedes auf Grundlage der musikalischen Eigenschaften getroffen werden. Die untenstehende Tabelle liefert einen Überblick der verwendeten Modelle und die erzielte Genauigkeiten der Klassifizierung auf den Test-Daten. Nachfolgend werde ich auf die einzelnen Modelle genauer eingehen. Für jedes Modell wurde eine Confusion-Matrix, ROC-Kurven und eine Tabelle mit den Precision und Recall Werten für jede Popularitätsklasse erstellt. Des Weiteren werden die bestmöglich gewählten Hyperparameter festgehalten.

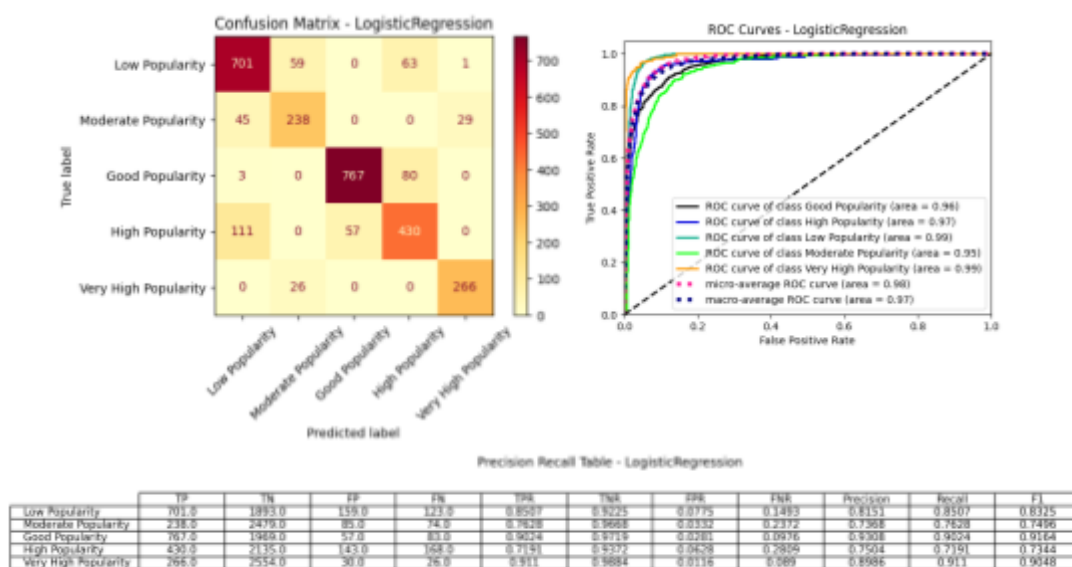


Modell	Accuracy	Precision	Recall
Gradient Boosting	97.71 %	97.35 %	97.35 %
KNN	95.41 %	95.02 %	94.53 %
SVM	95.27 %	94.73 %	94.56 %
Random Forest	94.85 %	95.09 %	92.88 %
Naive-Bayes	88.46 %	87.07 %	86.25 %
Logistic Regression	83.52 %	82.63 %	82.92 %
Decision Tree	79.55 %	79.88 %	75.61 %

Tabelle zum Vergleich der unterschiedlichen Genauigkeit der Modelle

### 2.3.1 Logistic Regression

Das Modell der logistischen Regression konnte eine Genauigkeit von 83,52 % erzielen und zählt damit zu den schlechter performenden Modellen. Der Hyperparameter C für die Regularisierungsstärke hat mit 0.01 die besten Ergebnisse erzielt.

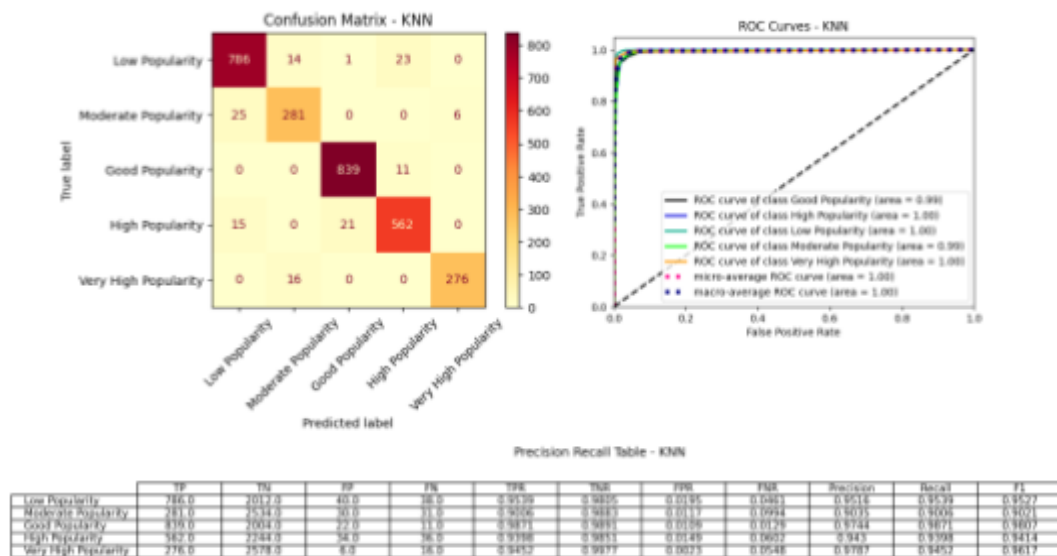


Evaluationsergebnisse (Konfusionsmatrix, ROC-Kurven, Precision-Recall-Tabelle) für Logistische Regression

### 2.3.2 K-Nearest-Neighbor

Das Modell des K-Nächsten-Nachbarn konnte eine Genauigkeit von 95,41 % erzielen und zählt damit zu den sehr gut performenden Modellen. Mit einer Anzahl von 13 nächsten Nachbarn und der gleich Gewichtung dieser konnte das beste Ergebnis erzielt werden.

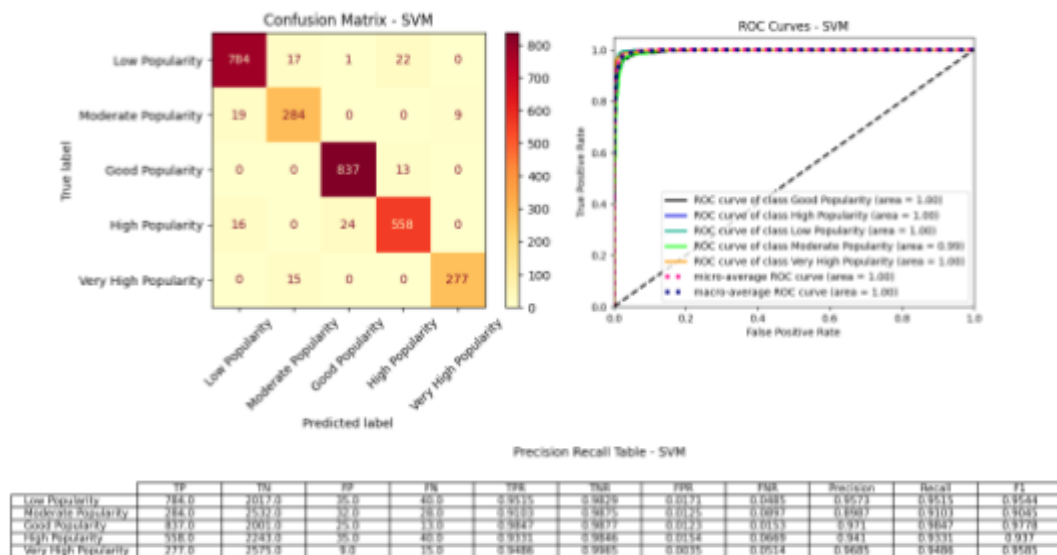




Evaluationsergebnisse (Konfusionsmatrix, ROC-Kurven, Precision-Recall-Tabelle) für K-Nearest-Neighbor

### 2.3.3 Support-Vector-Machine

Das Modell der Support-Vector-Machine konnte eine Genauigkeit von 95,27 % erzielen und zählt damit zu den sehr gut performenden Modellen. Mit dem Regularisierungsparameter C mit 5 und dem Kernel Algorithmus rbf konnten die besten Ergebnisse erzielt werden.

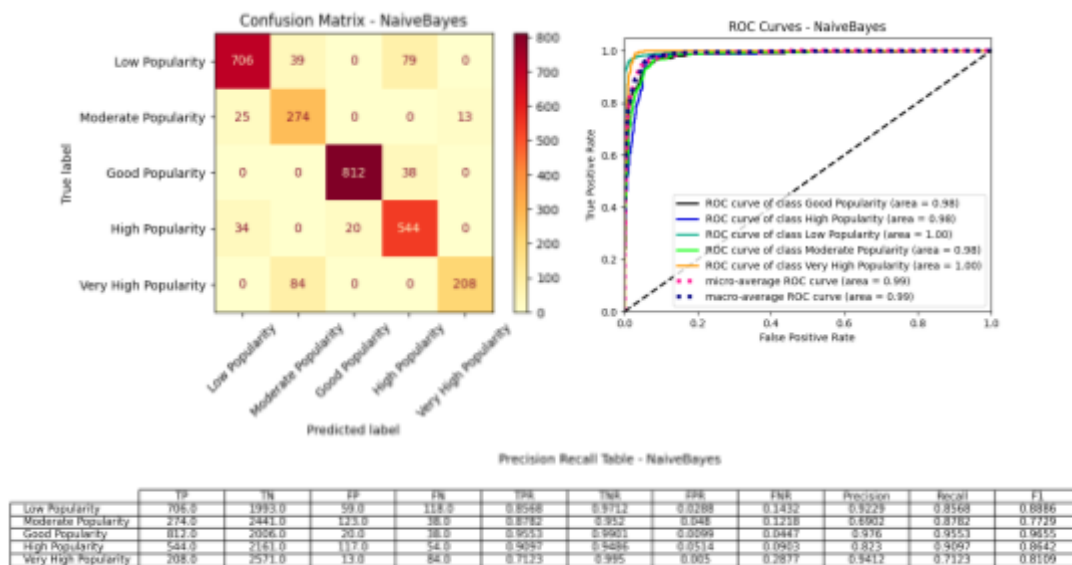


Evaluationsergebnisse (Konfusionsmatrix, ROC-Kurven, Precision-Recall-Tabelle) für Support-Vector-Machine

### 2.3.4 Naive Bayes

Das Modell des Naive-Bayes konnte eine Genauigkeit von 88,46 % erzielen und zählt damit zu den schlechter performenden Modellen. Für die Berechnung zur Stabilität der Varianzen hat der Wert 8e-06 für die besten Ergebnisse gesorgt.

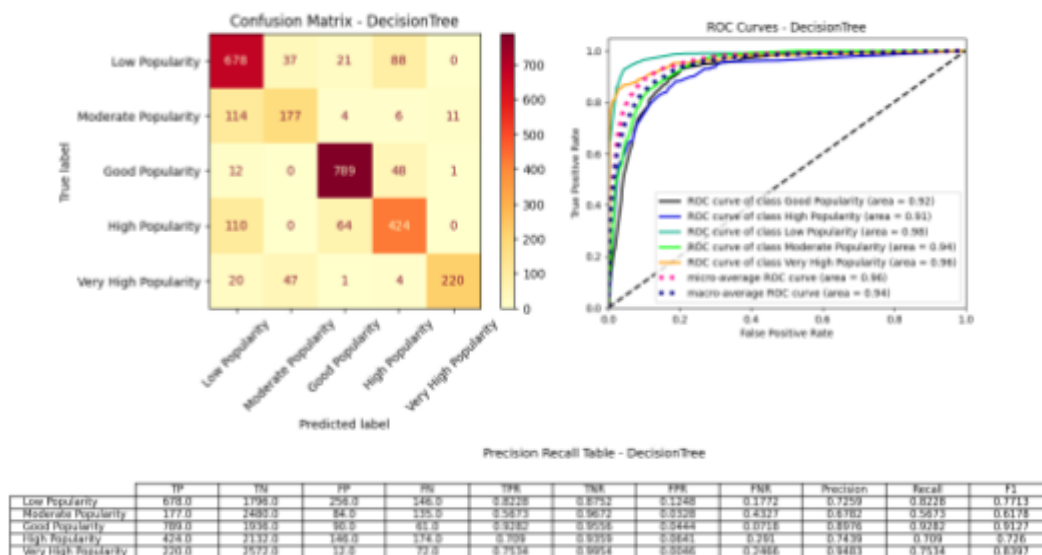




Evaluationsergebnisse (Konfusionsmatrix, ROC-Kurven, Precision-Recall-Tabelle) für Naive-Bayes

### 2.3.5 Decision Tree

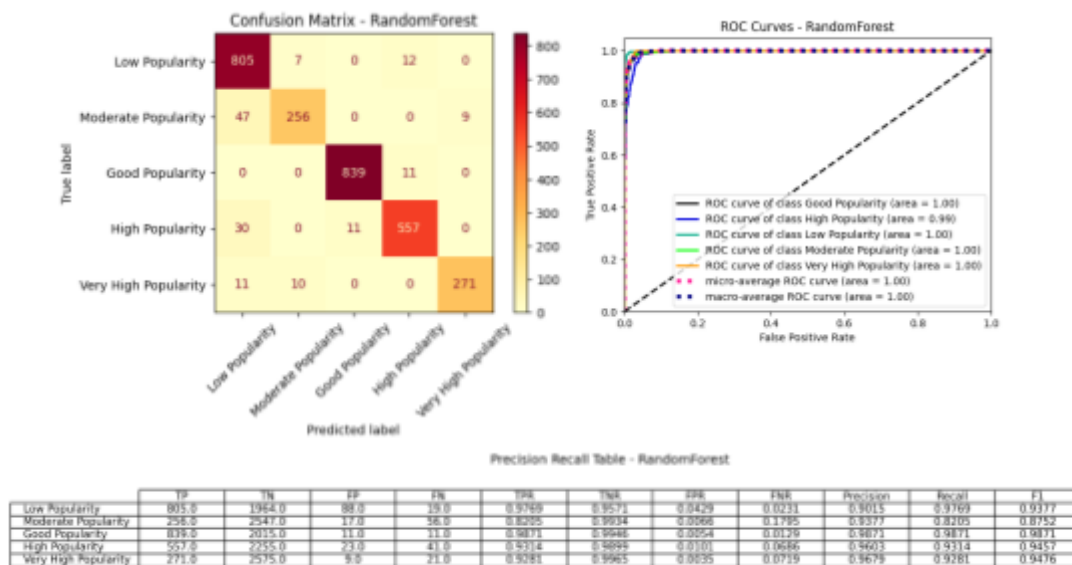
Das Modell des Decision Trees konnte eine Genauigkeit von 79,55 % erzielen und ist damit das am schlechtesten performende Modell. Mit einer maximalen Tiefe von 8 konnten die besten Ergebnisse erzielt werden.



Evaluationsergebnisse (Konfusionsmatrix, ROC-Kurven, Precision-Recall-Tabelle) für Entscheidungsbaum

### 2.3.6 Random Forest

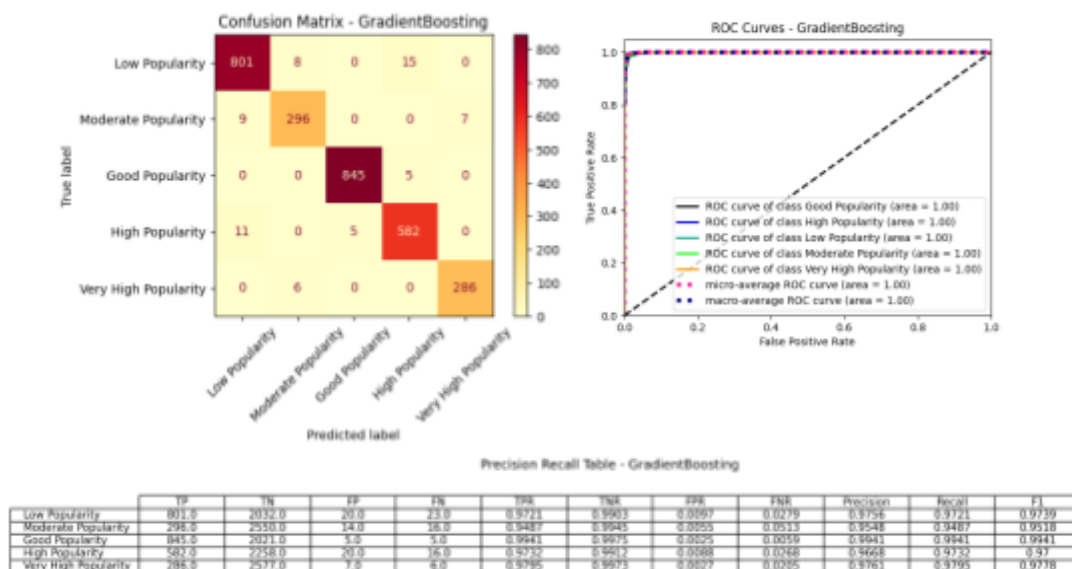
Das Modell des Random Forests konnte eine Genauigkeit von 94,85 % erzielen und zählt damit zu den sehr gut performenden Modellen. Mit einer maximalen Tiefe von 8 und maximal 200 Bäumen konnte das beste Ergebnis erzielt werden.



Evaluationsergebnisse (Konfusionsmatrix, ROC-Kurven, Precision-Recall-Tabelle) für Random Forest

### 2.3.7 Gradient Boosting

Das Modell des Gradient Boostings konnte eine Genauigkeit von 97,71 % erzielen und ist damit das am besten performende Modell. Mit einer maximalen Tiefe von 8 und einer Lernrate von 0.1 konnten die besten Ergebnisse erzielt werden.



Evaluationsergebnisse (Konfusionsmatrix, ROC-Kurven, Precision-Recall-Tabelle) für Gradient Boosting



## 3 Diskussion

Folgender Abschnitt befasst sich mit den aktuellen Lösungsansätzen aus der Literatur und dem Internet, um die Performance und Popularität eines Liedes auf Spotify und YouTube vorherzusagen. Der Fokus liegt dabei auf der Abgrenzung der bestehenden Versuche zu den hier beschriebenen Modellen und wo mögliche Stärken und Schwächen im Vergleich liegen. Zusätzlich erfolgt eine erste Einschätzung, welche Eigenschaften in welchen Popularitätsklassen am meisten vertreten sind.

### 3.1 State-Of-The-Art Lösungen

Die Performance eines Liedes auf Grundlage der musikalischen Eigenschaften vorhersagen zu können, ist ein mächtiges Werkzeug. Nicht nur Künstler können somit gezielter “Charts” komponieren, sondern auch die Musikindustrie kann speziellen Fokus auf gut performende Lieder legen und somit ihren Profit erhöhen.

In der Literatur gibt es eine Vielzahl an verschiedenen Versuchen, die Performance eines Liedes mithilfe von verschiedenen Machine Learning Algorithmen vorherzusagen. In der wissenschaftlichen Veröffentlichung “Predicting Music Popularity Using Spotify and YouTube Features”<sup>1</sup> werden die musikalischen Eigenschaften aus den Liedern extrahiert und damit verschiedene Modelle trainiert. Auch in diesem Beispiel werden verschiedene Popularitätsklassen definiert und versucht vorherzusagen. Das beste Modell war in diesem Fall Random Forest mit einer Genauigkeit von 79,6 %. Alleinstellungsmerkmal ist jedoch die Einbeziehung von Social-Media-Merkmalen in die Berechnung der Popularität eines Liedes. Der Artikel “Effect of Feature Selection on the Accuracy of Music Popularity Classification Using Machine Learning Algorithms”<sup>2</sup> konnte bessere Modelle trainieren, welche eine Genauigkeit von bis zu 95,15 % erreicht haben. Der Fokus lag dabei auf der Bereinigung des verwendeten Datensatzes, beispielsweise durch das Entfernen von Features mit geringer Korrelation und somit wenig Aussagekraft für die Popularität eines Liedes. Des Weiteren wurde in der Beurteilung der verschiedenen Algorithmen die Zeitkomplexität berücksichtigt und es wurde versucht, mit einer möglichst geringen Anzahl an Features ein bestmögliches und schnell zu berechnendes Ergebnis zu erzielen.

---

<sup>1</sup> [“Predicting Music Popularity Using Spotify and YouTube Features”](#) von Yap Kah Yee und Mafas Raheem

<sup>2</sup> [“Effect of Feature Selection on the Accuracy of Music Popularity Classification Using Machine Learning Algorithms”](#) von Faheem Khan, Ilhan Tarimer, Hatha Salama Alwageed, Buse Cennet Karadag, Muhammad Fayaz, Akmalbek Bobomirzaevich Abdusalomov und Young-Im Cho



Ein Großteil der Onlinequellen, welche denselben Datensatz von Kaggle benutzt haben, fokussieren sich rein auf die Visualisierung der Daten und explorative Datenanalyse. Wie im Code "Youtube/Spotify EDA and Simple Regression" <sup>3</sup> zu sehen ist, werden erste Ansätze zur Vorhersage der Performance eines Liedes mithilfe von linearer Regression versucht, jedoch wird nach den ersten schlechten Ergebnissen dieser Baseline aufgehört. Begründet wird dies meist durch die geringe Korrelation der Daten, jedoch werden keine alternativen Lösungswege gesucht.

### 3.2 Abgrenzung und Unterschiede

Zusammenfassend kann aus der Analyse der bestehenden Versuche in Abgrenzung zu dem in dieser Arbeit beschriebenen Ansatz festgehalten werden, dass auf den öffentlich zur Verfügung stehenden Datensatz verschiedene aus der aktuellen Literatur bekannte Ansätze angewandt wurden.

Die Views, Likes, Kommentare und Streams werden gebündelt in einem Wert, der Popularität, dargestellt. Diese Popularitätsklasse wird anschließend dazu genutzt, um diverse Machine Learning Modelle zur Klassifizierung zu trainieren. Die Ergebnisse der verschiedenen Modelle betragen im Durchschnitt eine Genauigkeit von über 90 %, was ein sehr gutes Ergebnis ist und damit mit einem Großteil der in der Literatur bestehenden Modelle sehr gut mithalten kann.

### 3.3 Stärken und Schwächen des Modells

Es ist wichtig zu berücksichtigen, dass der in dieser Arbeit beschriebene Ansatz keine genauen Vorhersagen treffen kann. Mithilfe der Modelle ist es lediglich möglich, eine erste grobe Einschätzung über die Performance eines Liedes auf Spotify und YouTube zu treffen. Eine Unterteilung zwischen der Performance auf den Plattformen findet dabei nicht statt. Demnach ist es nicht möglich, durch die Modelle vorhersagen zu können, ob ein Lied besser auf YouTube oder Spotify performt, beispielsweise durch ein besonders aufwändiges und gutes Musikvideo, da dies nicht in den Eigenschaften des Liedes berücksichtigt wird.

### 3.4 Interpretation der Eigenschaften in den Popularitätsklassen

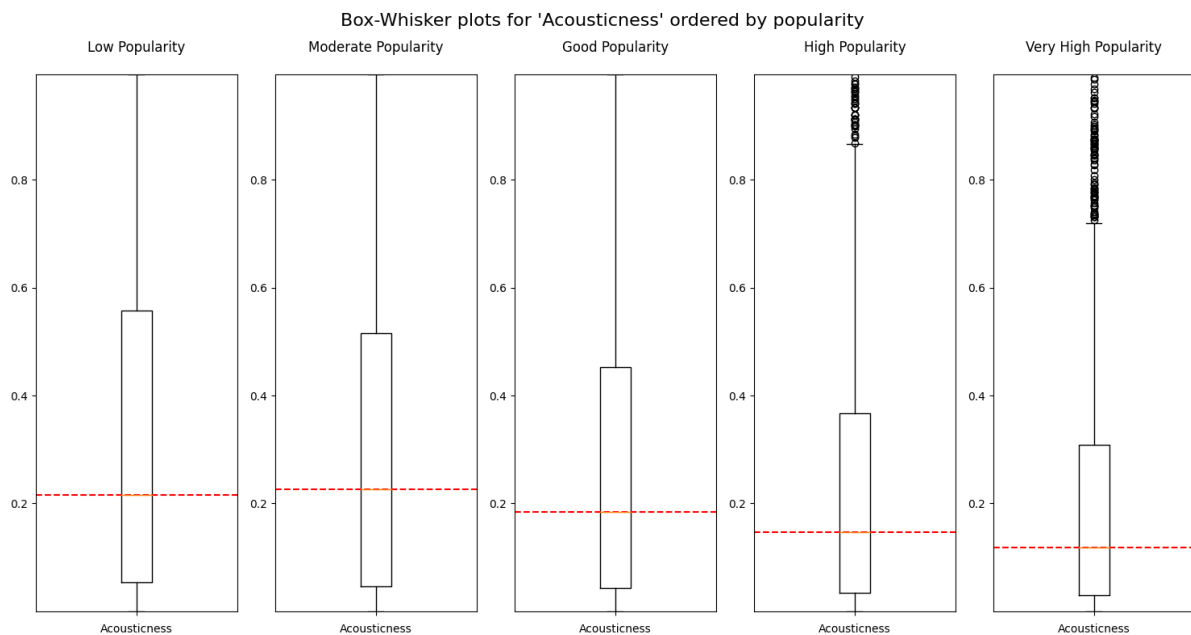
Um eine gute Übersicht der Eigenschaften innerhalb der einzelnen Popularitätsklassen zu erlangen, wurden für jede Eigenschaft Box-Whisker-Plots erstellt. Aus diesen lässt sich eine

---

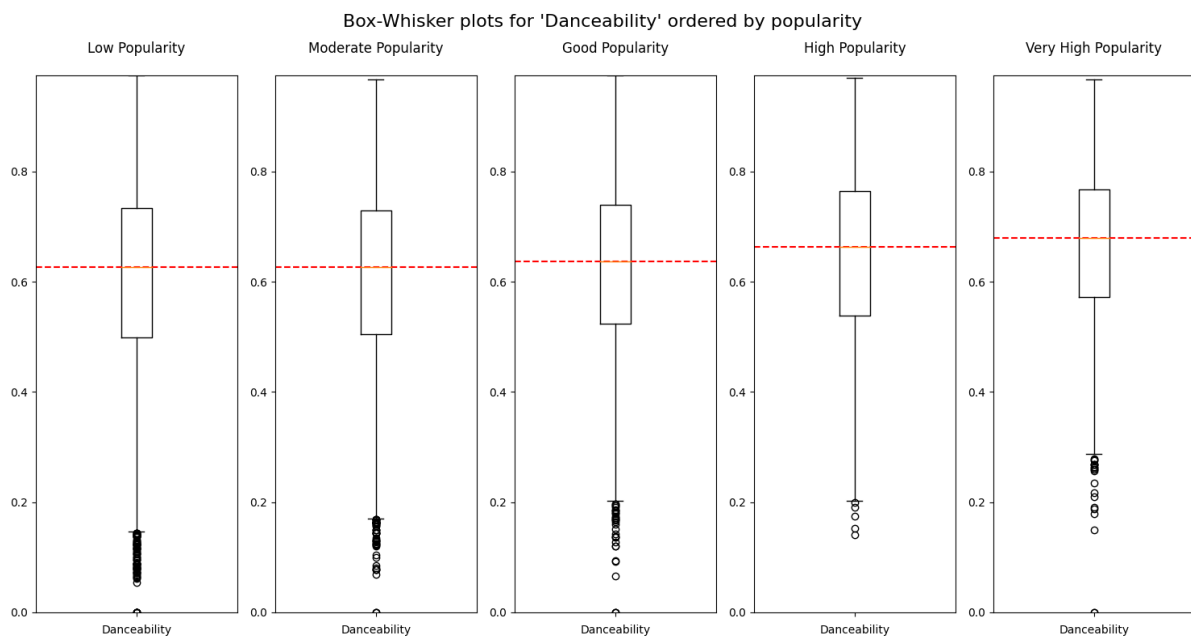
<sup>3</sup> "[Youtube/Spotify EDA and Simple Regression](#)" von Akipet98



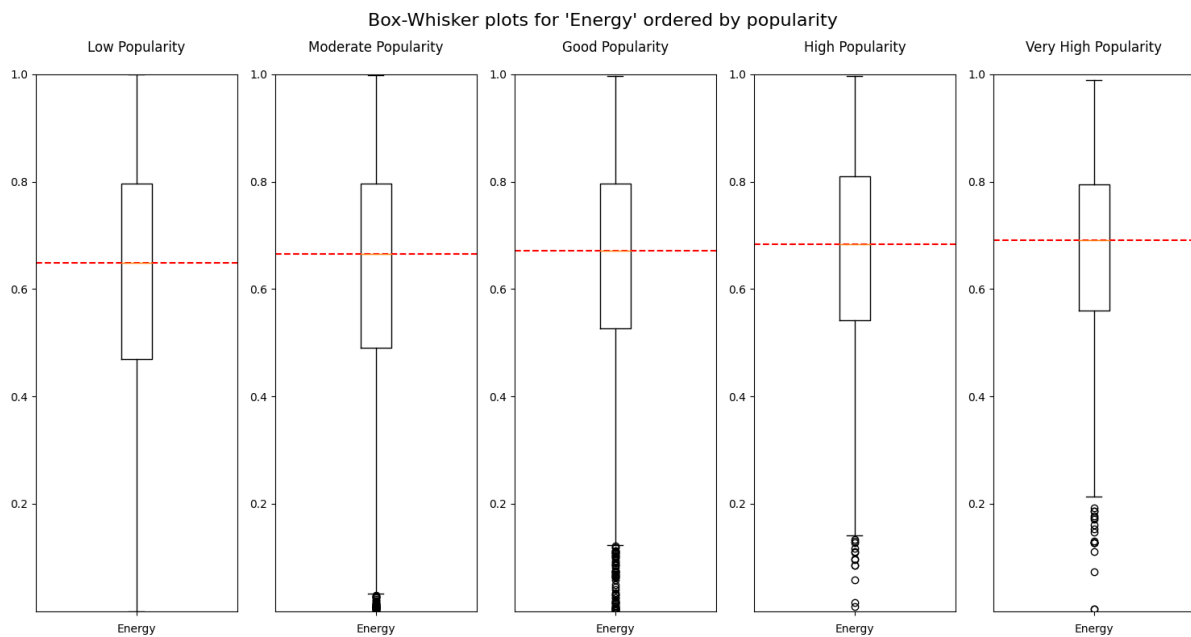
gute Tendenz ableiten, welche Ausprägung der Eigenschaften zu einer besseren Performance führen. Die rote Linie bildet den Median und ist ein guter Vergleichswert.



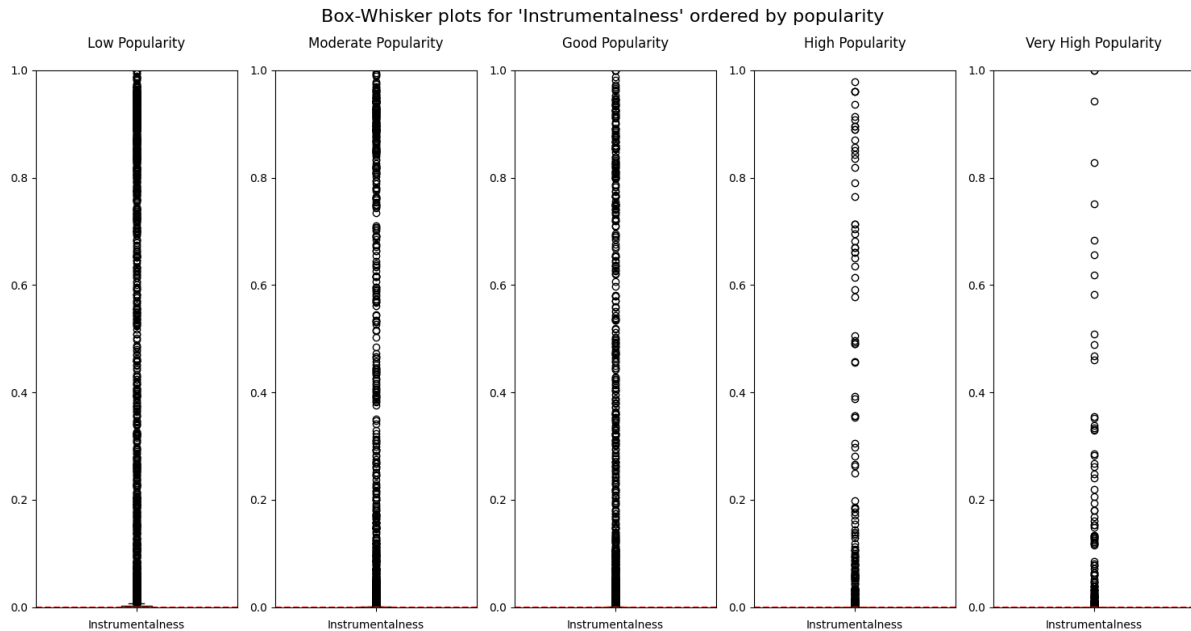
Für die Eigenschaft Acousticness lässt sich eine Tendenz zu einem geringeren Wert von ungefähr 0,6 ableiten. Dies zeigt, dass populäre Lieder weniger akustische Elemente enthalten.



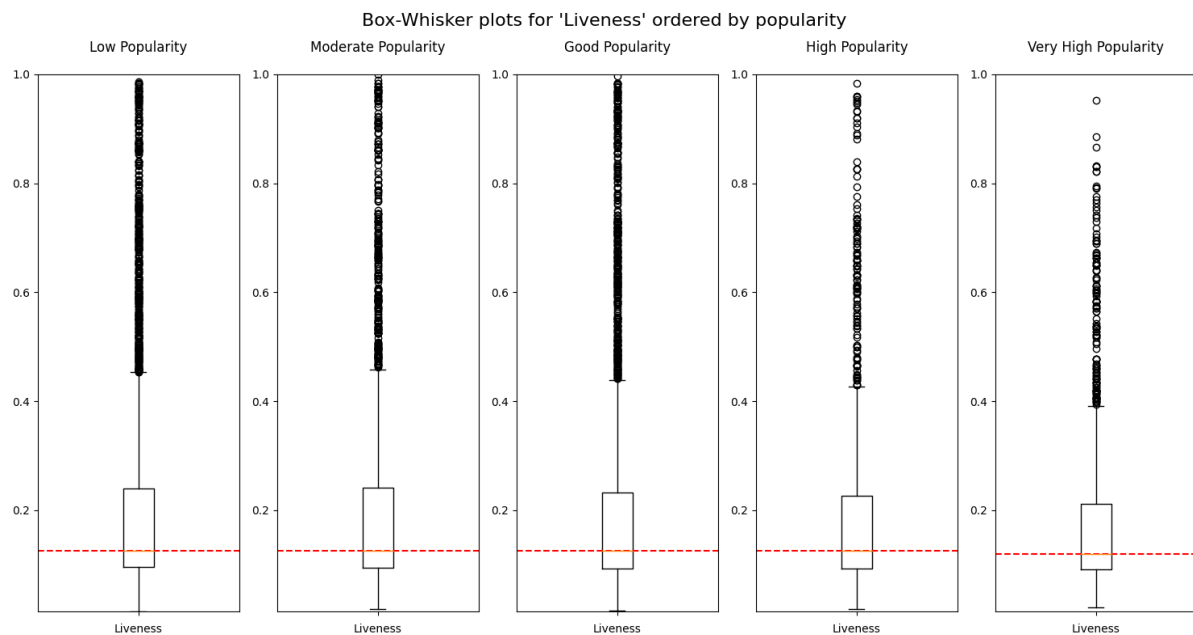
Die Danceability ist in allen Popularitätsklassen relativ hoch. Jedoch lässt sich auch hier eine positive Tendenz für eine höhere Popularität ableiten.



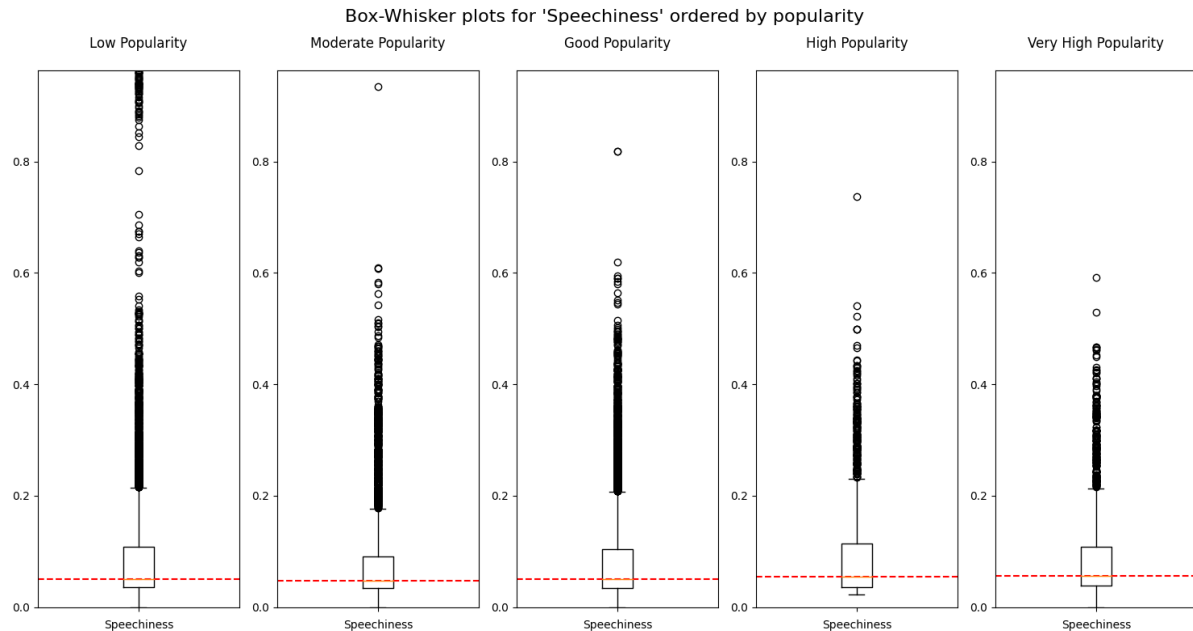
Auch die Energy ist in allen Klassen relativ hoch mit einem Wert über 0,6. Eine leichte positive Tendenz in Richtung 0,7 ist jedoch ersichtlich.



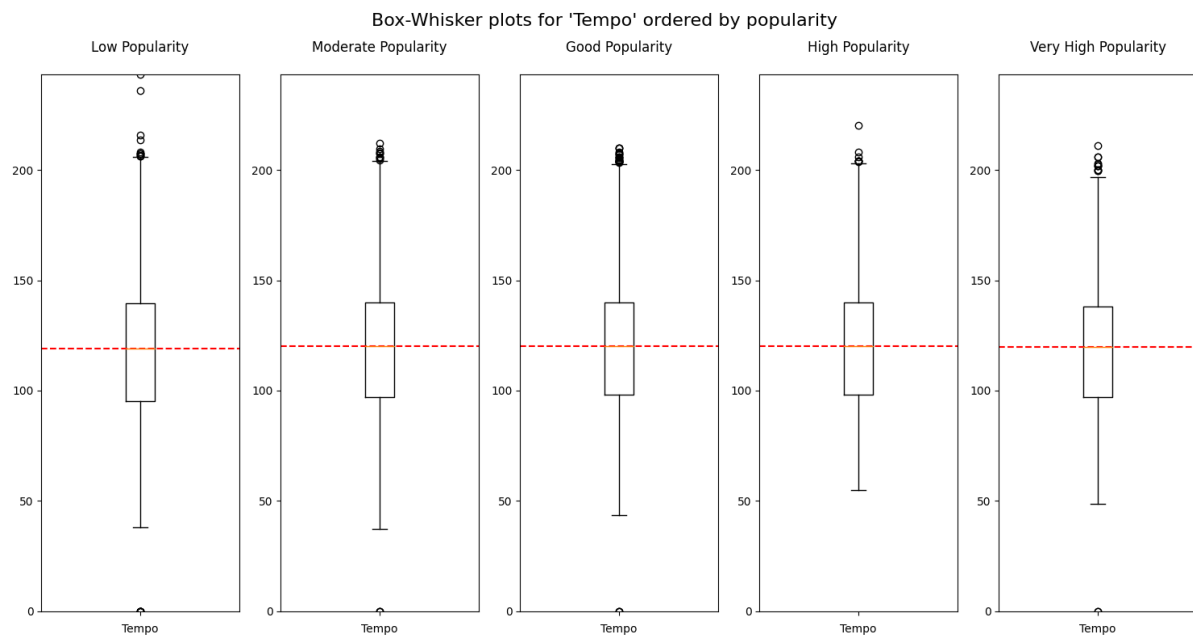
Wie zu Beginn erklärt, ergeben die Instrumentalness Werte keine zuverlässigen Daten. Es ist jedoch ersichtlich, dass bei geringer Popularität mehr Ausreißer über die ganze Skala verteilt sind. Dies nimmt bis hin zur höchsten Popularitätsklasse immer weiter ab. Daraus lässt sich ableiten, dass in beliebten Liedern Gesang vorhanden sein muss.



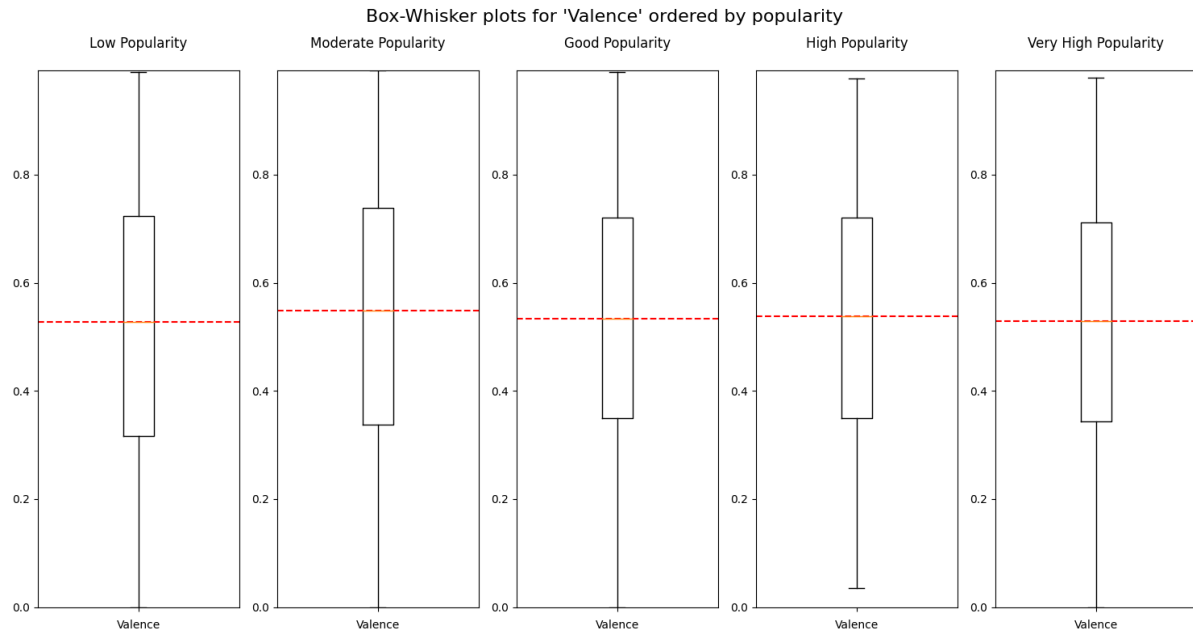
Die Liveness ist in allen Klassen konstant niedrig. Die Ausreißer nehmen jedoch auch bei erhöhter Popularität immer weiter ab. Dies zeigt, dass die meisten Lieder Studioaufnahmen sind.



Die Speechiness Werte sind in allen Klassen konstant niedrig. Das zeigt, dass in allen Klassen der Anteil der Melodie höher ist als der Anteil des Gesangs.



Das Tempo ist in allen Klassen relativ konstant auf ungefähr 120 BPM. Daraus lässt sich ableiten, dass die Geschwindigkeit des Liedes keinen großen Einfluss auf die Popularität hat bzw. dass die Lieder sich an den 120 BPM orientieren können.



Zuletzt betrachten wir den Wert der Valence. Die Werte bewegen sich in allen Klassen zwischen 0,5 und 0,6. Dies zeigt, dass grundsätzlich eher fröhliche Songs verbreitet sind. Daraus lässt sich jedoch kein Einfluss auf die Popularität eines Liedes ableiten.





## 4 Fazit und Ausblick

Dieses Kapitel fasst die Ergebnisse der Arbeit noch einmal zusammen und gibt einen Ausblick, was an dem Modell noch verbessert und welche zusätzliche Funktionalität noch implementiert werden könnte.

### 4.1 Zusammenfassung der Ergebnisse

Zusammenfassend kann man sagen, dass wir erfolgreich ein Modell entwickelt haben, das die Performance der Lieder anhand ihrer Eigenschaften vorhersagen kann. Zu Beginn wurde eine erste Baseline entwickelt, die die Anzahl an Views, Likes, Kommentaren und Streams anhand der Eigenschaften der Lieder vorhersagen sollte. Da der Wertebereich dieser Features jedoch sehr weit divergiert, ist eine zuverlässige Vorhersage nur schwer möglich gewesen, was man an einem extrem hohen Mean Squared Error gesehen hat. Daraufhin wurde der erste Ansatz überdacht, die Anzahl an Views, Likes, Kommentaren und Streams vorhersagen zu wollen. Stattdessen erfolgte die Einteilung in Popularitätsklassen, welche sich aus den Features Views, Likes, Kommentaren und Streams berechnen lassen. Anhand dieser Popularitätsklassen wurde eine zweite Baseline entwickelt. Die Klassifizierung in die einzelnen Popularitätsklassen führte zu besseren Ergebnissen und zeigte, dass der zweite Ansatz wesentlich besser funktioniert. Die zweite Baseline wurde durch den Einsatz verschiedener Algorithmen wie Logistic Regression, K-Nearest-Neighbor, Support-Vector-Machine, Naive-Bayes, Decision Tree, Random Forest und Gradient Boosting getestet und verbessert. Alle Algorithmen liefern Accuracy, Precision und Recall Werte über 79 %, was in der Tabelle in Kapitel 2.3. eingesehen werden kann. Das beste Ergebnis liefert jedoch der Gradient Boosting Algorithmus mit Accuracy, Precision und Recall von über 97 %. Dafür wurde in dem Modell eine maximale Tiefe von 8 und eine Lernrate von 0.1 verwendet. Mit diesem Ergebnis wurde sowohl die erste als auch die zweite Baseline weit übertroffen und die Performance des Modells optimiert.

Da die Vorhersage des Modells sich aktuell auf die Performance eines Liedes anhand seiner Eigenschaften beschränkt, kann eine Aussage über den Einfluss der Eigenschaften auf die Performance aktuell nur über die Box-Whisker-Plots abgeleitet werden. Daraus lässt sich eine Kernaussagen formulieren: Die Eigenschaften der populärsten Songs lassen darauf schließen, dass ein Lied eine hohe Erfolgschance hat, wenn es einen geringen akustischen Anteil hat, das Lied sehr tanzbar, energetisch und fröhlich ist, ein ausgewogenes Gleichgewicht zwischen Gesang und Melodie vorherrscht und sich die Geschwindigkeit des Liedes um 120 BPM befindet. Dabei erscheint es logisch, dass professionelle Studioaufnahmen besser performen als live aufgenommene Konzertausschnitte.



## 4.2 Ausblick

Es kann festgehalten werden, dass eine Klassifizierung der Popularität eines Liedes sehr gut mit dem Datensatz realisiert werden kann. Sowohl Gradient Boosting, die Support Vector Machine und K-Nearest-Neighbor liefern zuverlässige Aussagen mit einer Genauigkeit von über 95 %. Jedoch findet eine gewisse Abstraktion der Vorhersage der Popularität statt, da die Views, Likes, Kommentare und Streams auf nur fünf verschiedene Popularitätsklassen aufgeteilt werden. Um genauere Vorhersagen treffen zu können, wäre es denkbar, die Einteilung der zu bestimmenden Klassen noch kleinteiliger zu treffen. Denkbar wäre es ebenfalls, eigene Popularitätsklassen für jeweils die Views, Likes, Kommentare und Streams zu schaffen, um gezielte Aussagen über beispielsweise nur die Anzahl an Streams auf Spotify treffen zu können. Zuletzt wäre die Hinzunahme weiterer Features in den Datensatz denkbar, sodass noch mehr Informationen zu dem Lied vorliegen und dadurch eine genauere Klassifizierung stattfinden kann. Wichtig ist dabei die Betrachtung der Korrelationen.