

Differenzierung zwischen musikalischer Klassik und Romantik mit Methoden der künstlichen Intelligenz

Im Bereich der diskriminativen KI-Modelle ist es möglich, Musikaufnahmen in Genres beziehungsweise Musikepochen oder -stile zu klassifizieren. Nachfolgend werden ein Modell des *Machine Learnings* sowie eines des *Deep Learnings* vorgestellt, die für die Klassifizierung von Klavieraufnahmen in klassischer und romantischer Musik erstellt wurden. Das Ziel dabei war es zu untersuchen, inwiefern sich diese musikalischen Stile mit Hilfe der KI klassifizieren lassen und welche *Features* für deren Unterscheidung am relevantesten sind.

Dafür wurde ein Datensatz erstellt, der aus 200 Klavierstücken der Komponisten Mozart, Beethoven, Brahms und Chopin besteht. Dabei wird davon ausgegangen, dass Mozart und Beethoven der Klassik, Brahms und Chopin der Romantik gehören. Aus den Aufnahmen wurden 43 *Features* extrahiert. Diese basieren auf Musikeigenschaften wie Klangfarbe, Harmonie und Tempo. Folgende *Features* wurden verwendet: 13 *Mel-Frequency Cepstral Coefficients (MFCCs)* sowie deren Differenzen, das *Chromagramm*, aufgeteilt in 12 Tonhöhenklassen, sowie dessen Entropie, die aus dem *Tempogramm* berechnete Entropie, ebenso wie Entropie, Standardabweichung und Varianz aus den Werten eines *Beat-Trackers*.

Da alle Aufnahmen unterschiedlich lang sind, wurden sie in gleich lange Segmente unterteilt, sodass aus jeder Aufnahme die gleiche Anzahl und Länge der Segmente für das Training eines Modells verwendet wurde. Der Trainingsprozess wurde mehrmals durchgeführt, wobei bei jedem Durchlauf sowohl die Anzahl als auch die Länge der Segmente variiert wurden. Damit kann herausgefunden werden, welche Parameterkombination die besten Ergebnisse liefert. Der Vorgang wurde insgesamt 167-mal wiederholt und dabei wurden elf unterschiedliche Parameterkombinationen verwendet. Diese werden in Tabelle 1 dargestellt. Die Aufnahmen wurden bei jedem Durchlauf nach einem Zufallsprinzip erneut segmentiert, um die Robustheit der Ergebnisse zu überprüfen. Anschließend wurden 80 % der Daten für den Trainings- und 20 % für den Testdatensatz verwendet. Welche Daten in den Trainings- und welche in den Testdatensatz kommen, wurde ebenfalls zufällig entschieden. Die Daten wurden vor dem Training der Modelle normalisiert. Der ausgewählte Klassifikator ist der *Random Forest*, der mit 30 *n_estimators* trainiert wurde. Das *Deep Learning* Modell besteht aus drei *Dense Layers* mit jeweils 128, 64 und 64 Neuronen, sowie jeweils ein *Dropout Layer* dazwischen. Das Modell wurde mit 10 *epochs* und eine *batch_size* von 32 trainiert.

Anzahl	Länge	Menge
5	5	15
5	10	15
5	20	15
5	30	15
10	5	15
10	10	16
10	20	15
20	5	15
20	10	15
20	5	15
15	15	16

Tabelle 1:
Parameterkombinationen

In Abbildung 1 sind die *Confusion Matrices* eines einzelnen Durchlaufs zu sehen, bei welchen die Verteilung in 10 Segmenten jeweils 10 Sekunden lang erfolgte. Die Klassifizierung mit dem *Random Forest* erzielte eine Genauigkeit (=Accuracy) von 75 %. Aus 40 Aufnahmen im Testdatensatz wurden vier Stücke falsch als klassisch und sechs falsch als romantisch eingestuft (siehe Abbildung 1a). Dabei hat sich das *MFCC-1* als relevantestes *Feature* für die Klassifizierung erwiesen (siehe Abbildung 2). Beim Vorgang mit dem *Deep Learning* Modell wurden die Ergebnisse verbessert: die Genauigkeit lag bei 82,5 % mit zwei falsch als klassisch und fünf falsch als romantisch eingestuften Stücken (Siehe Abbildung 1b).

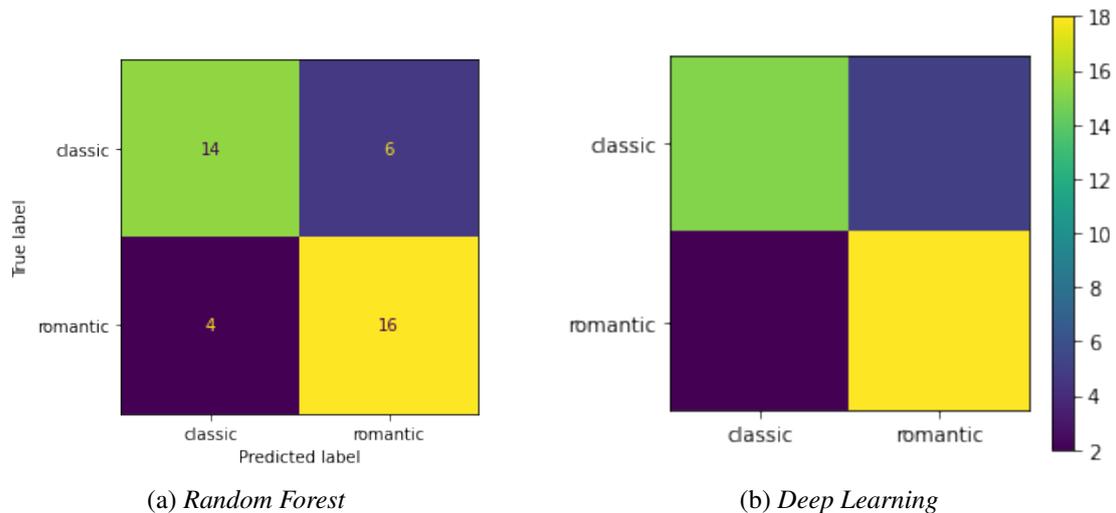


Abbildung 1: Confusion Matrices eines Durchlaufs

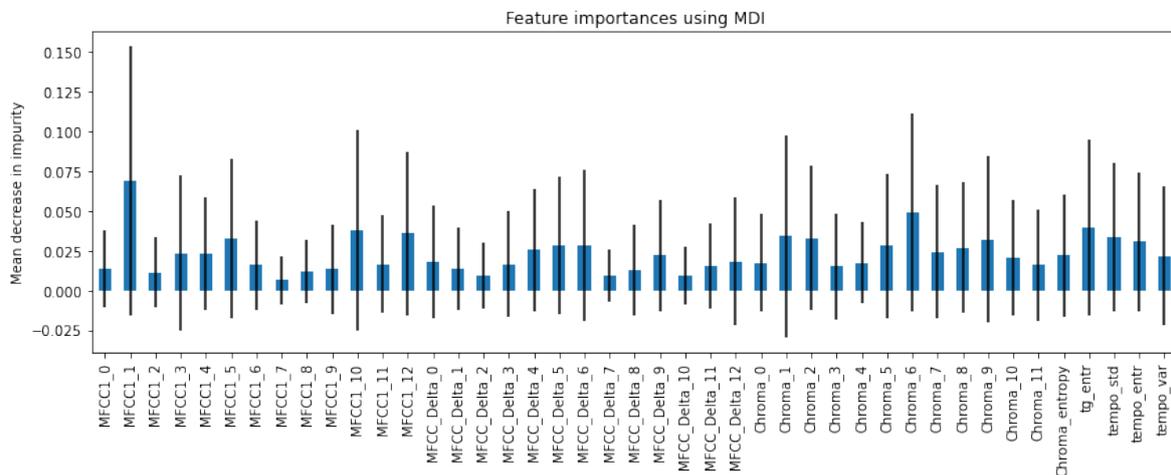


Abbildung 2: Feature importances

Nach der Durchführung der 167 Durchläufe lag die Genauigkeit beim *Random Forest* zwischen 50 % und 95 %. Dabei schnitten die Durchläufe, bei denen die Aufnahmen in fünf Segmente geteilt wurden, am schlechtesten ab. Die besten Ergebnisse wurden mit der Parameterkombination aus 15 15-sekündigen Segmenten mit einer durchschnittlichen Genauigkeit von 80 % erzielt, wie es auf Abbildung 3 zu sehen ist. Beim *Deep Learning* Modell waren die Ergebnisse einiger Durchläufe besser, bei anderen jedoch schlechter als beim *Random Forest*, sodass keine signifikante Steigerung der durchschnittlichen Genauigkeit – von 76,4 % auf 76,8 % – zu sehen ist, sondern nur eine größere Verteilung. Die Einstellungen beim *Random Forest* 15–15; 20–10 und 10–20 sowie beim *Deep Learning* Modell 15–15 und 20–10 erzielten die besten Genauigkeiten.

Bei den *Feature*-Wichtigkeiten des *Random Forest* Klassifikators, deren Verteilung auf Abbildung 4 zu sehen ist, scheint die des *MFCC-1* mit einem durchschnittlichen Wert von 9 % am relevantesten zu sein. Darunter lagen *MFCC-5*, *10*, *11* und *12*; *Chroma 1*, *8* und *11* sowie *Tempo-Standarddeviation* etwas relevant mit einem Durchschnitt von zwischen 3,3 % und 4,6 %. Die Werten des *MFCC-1* variieren jedoch zwischen 3 % und 17 %, die der anderen relevanteren *Features* zwischen 0,78 % und 9,87 %.

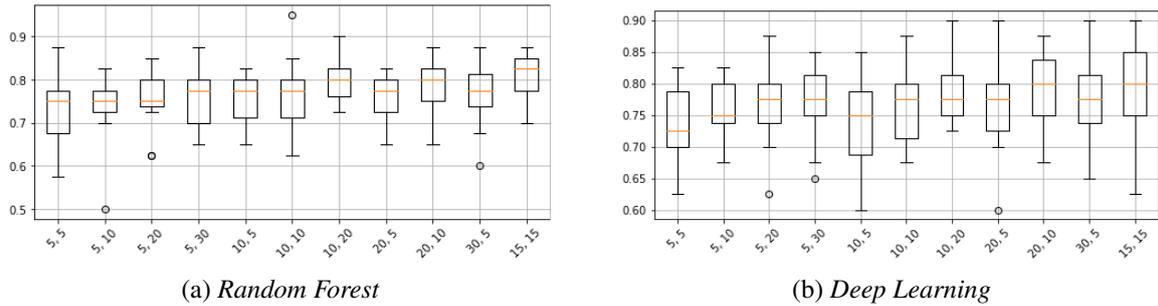


Abbildung 3: Vergleich der Accuracies

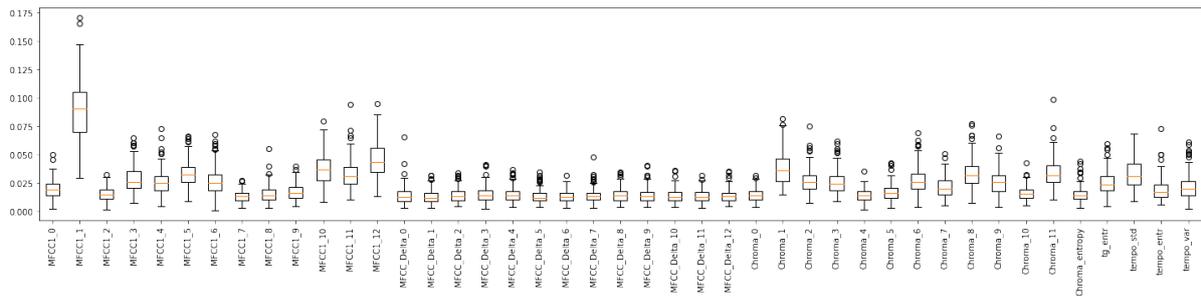


Abbildung 4: Vergleich der Wichtigkeiten nach mehreren Durchläufen

Mit einer durchschnittlichen Genauigkeit von etwa 76 % bei beiden Modellen kann die Behauptung aufgestellt werden, dass es möglich ist, musikalische Klassik und Romantik anhand der KI voneinander zu unterscheiden. Durch die Begrenzung auf vier Komponisten besteht allerdings die Gefahr, dass die Modelle lediglich den Stil der Komponisten zu unterscheiden vermögen. Das *MFCC-1* zeigte sich durchschnittlich als das relevanteste *Feature* bei der Unterscheidung. Die *Chromas 1, 8 und 11*, welche die Noten Cis, Gis und H entsprechen, waren unter den *Chromas* am relevantesten, was suggeriert, dass bestimmte Tonarten bei der Erkennung der Musikstile eine Rolle spielen können.