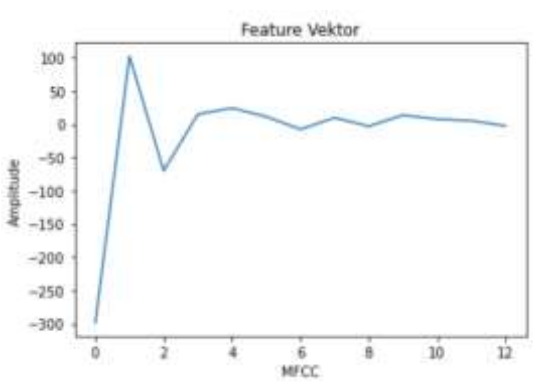
Automatische Erkennung von Vogelgesang

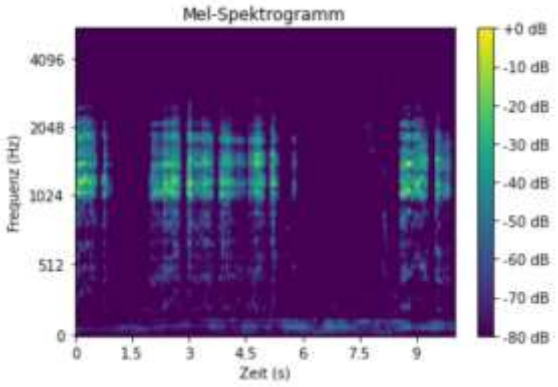
Fragestellung

Die Erkennung von Vogelstimmen in Tonaufnahmen ist häufig ein erster Schritt vor der weiteren Analyse, z. B. vor der Klassifizierung verschiedener Vogelarten für Vogelerkennungsapps oder deren Zählung in freier Wildbahn für den Artenschutz. Sie könnte auch für die automatische Erstellung von Untertiteln für Hörgeschädigte bei Filmen genutzt werden.
Algorithmen zur Erkennung von Vogelstimmen ermöglichen die Arbeit mit großen Datensätzen (z. B. bei einem kontinuierlichen 24-Stunden-Monitoring), indem aus den Daten bestimmte Merkmale herausgefiltert werden.
Ziel des vorliegenden Programmierprojekts ist die Erkennung von Vogelstimmen in zehnsekündigen Audiodateien. Um die künstliche Intelligenz (KI) möglichst realitätstauglich zu trainieren, wurde sie mit verschiedensten Hintergrundgeräuschen (Stimmen, andere Tiere, Straßenlärm etc.) konfrontiert, sowohl in den Dateien mit als auch in denen ohne Vogelgeräusch.

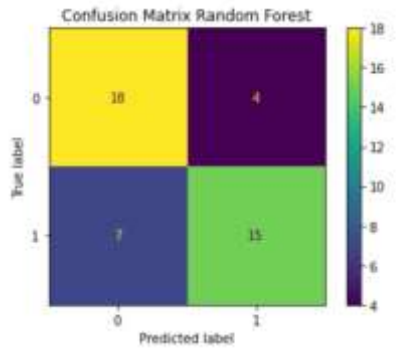
Vorgehensweise

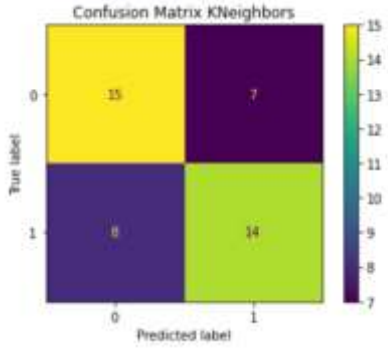
Zuerst wird ein geeigneter Datensatz ausgewählt. Auf der Webseite von DCase, auf der 2018 eine „Bird Audio Detection Challenge“ ausgeschrieben wurde, finden sich mehrere geeignete und annotierte Datensätze mit den Kategorien „Vogelstimme vorhanden“ und „Vogelstimme nicht vorhanden“.
Für einen ersten Durchlauf werden 200 Dateien ausgewählt, die Anzahl wird später noch auf 500 und 1000 erhöht. Der Anteil an Trainingsdaten beträgt jeweils 80%, der an Testdaten 20%.
Nachdem die Audiodateien mit Librosa geladen wurden, können aus ihnen MFC-Koeffizienten (Mel Frequency Cepstral Coefficients) erzeugt werden. Letztere werden häufig zur Spracherkennung verwendet, sind aber auch zur Analyse von Sounds und Musikstücken ein geeignetes Mittel. Häufig werden 13 MFC-Koeffizienten gewählt, so auch in meinem Beispiel. Die Erhöhung der Koeffizientenzahl auf z.B. 20 hat keinen signifikanten Effekt auf die Ergebnisse.
Der Feature Vektor bildet den Frequenzdurchschnitt über die Gesamtzeit der Audiodatei, führt also zu einer kompakten Darstellung des Frequenzspektrums. Das Mel-Spektrogramm visualisiert die Informationen nach Zeit, d.h. nicht gemittelt.



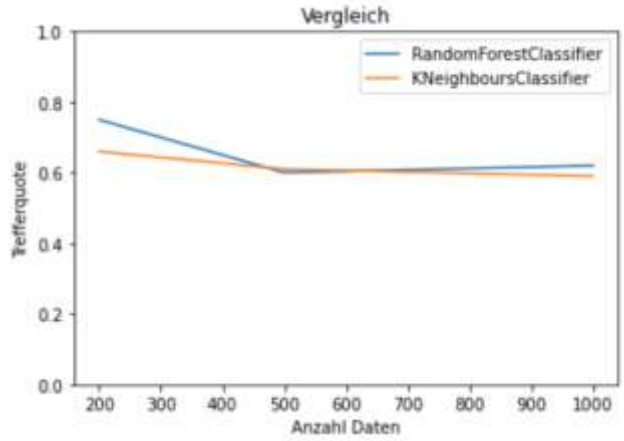


Mithilfe der MFC-Koeffizienten werden Daten für die Klassifikatoren generiert. Das Mel-Spektrogramm dient lediglich der Veranschaulichung. Im vorliegenden Beispiel werden die Klassifikatoren „RandomForest“ und „KNeighbors“ genutzt. Vor dem Einsatz im Klassifikator müssen die Daten normalisiert werden.
Bei 200 Dateien erzielt die KI eine Trefferquote von 75% mit dem RandomForestClassifier und eine von 66% mit dem KNeighborsClassifier.





In einem weiteren Schritt wird untersucht, ob die Implementierung einer größeren Datenmenge die Trefferquote der KI erhöhen würde. Daher wird der Algorithmus mit 500 und 1000 Dateien ausgeführt, jeweils wieder mit beiden Klassifikatoren.



Ergebnisse

Bei Verwendung unterschiedlich großer Datenmengen erzielt der Algorithmus eine Trefferquote von mindestens 59% und maximal 75%. Mit einer durchschnittlichen Trefferquote von rund 64% liegt er zumindest über der Zufallsbaseline (1/Anzahl der Klassen), die im Falle von zwei Klassen 50% beträgt.
Der Versuch eine größere Datenmenge zu verwenden, um die Trefferquote zu erhöhen, hat keinen Erfolg. Bei beiden Klassifikatoren sinkt die Trefferquote mit höherer Datenmenge.
Generell eignet sich der Klassifikator „Random Forest“ minimal besser für den Vogelerkennungsalgorithmus, da er im Durchschnitt eine höhere Trefferquote erzielt.
Mit einer Trefferquote von durchschnittlich rund 64% ist diese KI jedoch noch nicht ausgereift genug, um in der Realität erfolgreich angewendet zu werden.
Verbessert werden könnte der Algorithmus möglicherweise durch die Verwendung der zeitaufgelösten Information der Audiodateien, also der Mel-Spektrogramme, anstelle der MFC-Koeffizienten.