

Vulgarität in Deutschrap

Studiengang: BSc in Informatik | Vertiefung: Data Engineering
Betreuer: Prof. Dr. Jürgen Vogel
Experte: Alain Joray

Deutschrap hat in den letzten Jahren stark an Popularität gewonnen mit steigenden Streaming- und Verkaufszahlen. Dennoch bleiben Vorurteile gegenüber Deutschrap bestehen, da er als vulgär, frauenfeindlich und unangemessen für Jugendliche betrachtet wird. In dieser Arbeit wird mittels Natural Language Processing analysiert, ob diese Vorurteile stimmen. Dazu wurden Songtexte von Anfang 2011 bis Ende 2021 gesammelt und weiterverarbeitet.

Einleitung

In dieser Bachelor Thesis wurde ein Deep Learning Modell entwickelt, das in der Lage ist, vulgäre Songtexte zu erkennen. Der verwendete Datensatz wurde bereits im letzten Semester im Project 2 erarbeitet. Dabei wurden alle Lieder von Alben gesammelt, die von Anfang 2011 bis Ende 2021 die Spitze der Charts erreichten oder eine Gold- oder Platin-Auszeichnung erhielten. Insgesamt wurden 2724 Songtexte erhoben, von denen 1000 Songtexte als Gold Standard verwendet wurden.

Methodik

Anhand der gesammelten Daten konnte eine Pipeline entwickelt werden. Im Preprocessing wurden die Daten bereinigt, wobei alle nicht relevanten Texte gelöscht wurden. In den Songtexten wurden häufig Umgangssprache und Abkürzungen verwendet. Zudem enthielten die Lieder oft lautmalende Wörter, die für

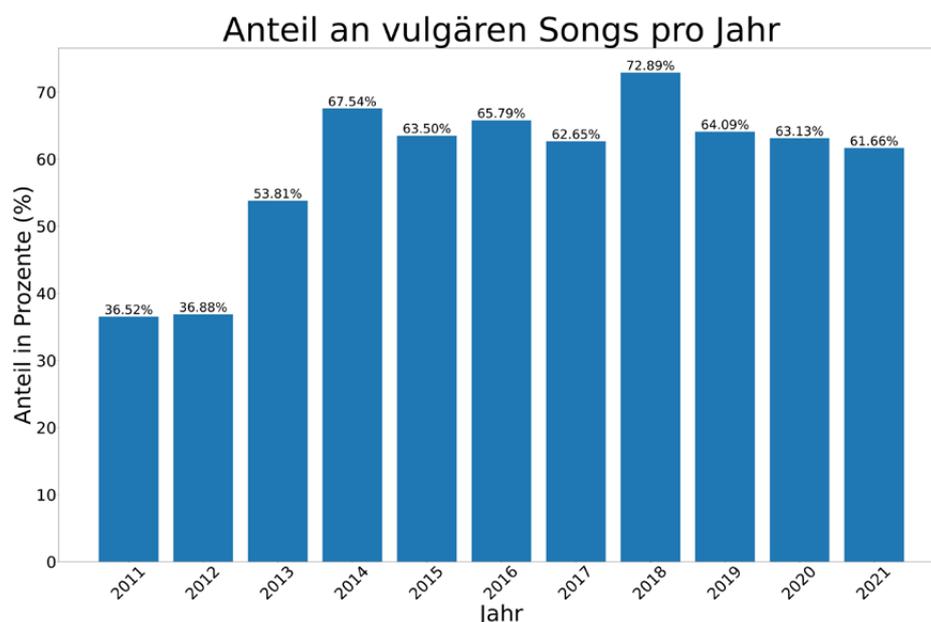
den Kontext keine Bedeutung hatten. Diese Aspekte wurden während des Preprocessings entfernt. Für das Deep Learning wurde ein deutsches Pretrained BERT Modell verwendet, das speziell für dieses Problem angepasst wurde.

Ergebnisse

Das entwickelte Modell hat eine Test Accuracy von 80%. Die Evaluation hat gezeigt, dass das Modell den Use Case verstanden hat und die Songtexte mit einer hohen Genauigkeit klassifizieren kann. Nach der Auswertung des Modells wurde deutlich, dass es mehr vulgäre Texte als nicht vulgäre Texte gibt. Interessanterweise waren die Lieder im Jahr 2011 überwiegend jugendfreundlicher im Vergleich zu den heutigen Liedern (siehe Abbildung).



Kirusihaan Sathiyapalan
kirusihaan.sathiyapalan@
hotmail.com



Anteil an vulgären Songs pro Jahr