

ABSTRAK

Kemajuan teknologi yang terus berkembang membawa banyak perubahan bagi kehidupan manusia termasuk akses terhadap hiburan seperti musik. Musik digital dapat diakses secara *online* melalui platform *streaming* musik, salah satunya adalah Spotify. Platform yang memiliki jutaan pengguna ini unggul karena dapat mengkategorisasikan musik ke cakupan genre yang sangat luas. Penggunaan klasifikasi menjadi salah satu hal yang penting dalam memetakan setiap lagu ke genrenya masing-masing. Penelitian ini akan mengimplementasikan metode regresi logistik dan *K-Nearest Neighbors* (KNN) untuk mengklasifikasikan apakah sebuah lagu bergenre pop atau non-pop. Fitur-fitur lagu yang digunakan sebagai variabel independen adalah *acousticness*, *danceability*, *duration*, *energy*, *instrumentalness*, *key*, *liveness*, *loudness*, *mode*, *speechiness*, *tempo*, dan *valence*. Model regresi logistik mampu menghasilkan nilai *F1-Score* sebesar 81.1% dengan semua variabel dan 78.5% dengan variabel bebas multikolinearitas, sedangkan model KNN hanya menghasilkan nilai *F1-Score* sebesar 76% dengan semua variabel dan 76.2% dengan variabel bebas multikolinearitas. Model terbaik yang ditemukan berdasarkan asumsi yang terpenuhi dan metrik yang dihasilkan adalah model regresi logistik dengan variabel bebas multikolinearitas yang mampu mencapai *F1-Score* sebesar 78.5%. Hasil menunjukkan bahwa regresi logistik lebih unggul dibandingkan dengan KNN dalam mengklasifikasikan genre musik pop.

Kata kunci: Klasifikasi, Spotify, Regresi Logistik, KNN



ABSTRACT

Continuous enhancements in technology bring changes into human lives, including entertainments such as music. Digital music can be accessed through online music streaming platforms, one of them being Spotify. Spotify, having one of the largest numbers of paid subscribers, is one of the best digital music providers with the ability to categorize music into a wide pool of genres. The use of classification is important in order to map each song to its respective genres. This research will implement classification methods such as logistic regression and KNN in order to classify whether a song belongs to the pop or non-pop genre. Audio features used as independent variables include acousticness, danceability, duration, energy, instrumentalness, key, liveness, loudness, mode, speechiness, tempo, and valence. Logistic regression model obtained an F1-Score of 81.1% using all variables and 78.5% using variables free from multicollinearity, while KNN model obtained an F1-Score of 76% using all variables and 76.2% using variables free from multicollinearity. The best model determined by fulfilled assumptions and highest metrics was produced by logistic regression model using variables free from multicollinearity, obtaining an F1-Score of 78.5%. Results show that logistic regression has a better performance in comparison to KNN in the classification of pop music.

Keywords: Classification, Spotify, Logistic Regression, KNN

