

BAB 6

KESIMPULAN DAN SARAN

Pada bab ini akan dibahas kesimpulan yang didapatkan dari penelitian yang dilakukan dan saran yang dapat diberikan untuk pengembangan penelitian ini lebih lanjut.

6.1 Kesimpulan

Pada bagian ini akan dijelaskan kesimpulan yang didapatkan dari penelitian yang dilakukan dengan menggunakan data lagu *Top 200* Spotify tahun 2020-2021. Berikut ini adalah kesimpulan-kesimpulan yang didapatkan:

1. Telah dilakukan proses pembersihan data dan penyiapan data dengan melakukan pembuangan *missing value* terhadap dataset utama serta melakukan pemilihan baris data dengan nilai atribut *region* adalah Indonesia dan nilai atribut *chart* adalah *Top 200* terhadap dataset tambahan.
2. Telah dilakukan eksplorasi data univariat dan bivariat terhadap dataset yang digunakan dalam penelitian. Didapati hasil bahwa:
 - Sebagian besar dari *Top 10* artis dan *Top 10* genre dengan jumlah lagu terbanyak pada tangga lagu *Top 200* Spotify tahun 2020-2021 didominasi dengan lagu-lagu yang berhasil menempati peringkat tertinggi antara peringkat 1 hingga 50. Maka artis dan genre memiliki hubungan yang kuat dengan peringkat tertinggi yang dapat dicapai sebuah lagu dalam tangga lagu. Beberapa artis yang memiliki hubungan kuat dengan peringkat yang tinggi dari sebuah lagu adalah Bad Bunny, Taylor Swift, Justin Bieber, BTS, dan The Weeknd. Sedangkan beberapa genre yang memiliki hubungan kuat dengan peringkat tinggi adalah pop, rap, postteen pop, hip hop, dan latin.
 - Artis dan genre tidak memiliki hubungan yang kuat dengan jumlah kemunculan sebuah lagu dalam tangga lagu. Namun didapati sebagian besar lagu-lagu dari *Top 10* artis dan *Top 10* genre dengan jumlah lagu terbanyak pada tangga lagu *Top 200* Spotify tahun 2020-2021 berhasil masuk ke dalam tangga lagu sebanyak 1 hingga 49 kali. Hanya didapati beberapa artis dan genre dengan lagu yang berhasil masuk ke dalam tangga lagu lebih dari 50 kali.

Telah dilakukan juga eksplorasi data bivariat untuk mencari atribut dengan nilai korelasi tinggi yang digunakan dalam pembuatan model klasifikasi dan model *clustering*. Kemudian telah dibuat model-model klasifikasi dengan menguji atribut-atribut dengan nilai korelasi tinggi seperti berikut ini:

- Model klasifikasi untuk prediksi *highest charting position* dengan menguji atribut *Number of Times Charted*, atribut *Popularity*, dan atribut *Streams* sebagai atribut prediktor menggunakan algoritma KNN.
- Model klasifikasi untuk prediksi *number of times charted* dengan menguji atribut *Highest Charting Position*, atribut *Popularity*, dan atribut *Streams* sebagai atribut prediktor menggunakan algoritma KNN.

Telah dibuat juga model-model *clustering* dengan menguji berbagai atribut sebagai berikut:

- Model *clustering* dengan menggunakan empat atribut yaitu atribut *Highest Charting Position*, atribut *Number of Times Charted*, atribut *Popularity*, dan atribut *Streams* menggunakan algoritma K-Means.
- Model *clustering* dengan menggunakan atribut-atribut yang merepresentasikan elemen-elemen dalam musik yaitu atribut *Danceability*, atribut *Energy*, atribut *Loudness*, atribut *Speechiness*, atribut *Acousticness*, atribut *Tempo* dan atribut *Duration* menggunakan algoritma K-Means.

Lalu telah dibuat juga analisis untuk data baru dengan melakukan perbandingan data lagu *Top 200 Spotify* tahun 2020-2021 Global dengan data lagu *Top 200 Spotify* tahun 2020-2021 Indonesia. Didapati hasil bahwa:

- *Top 10* artis di tangga lagu Global dan tangga lagu Indonesia berbeda, hanya didapati beberapa artis yang sama seperti BTS dan Taylor Swift.
- Proporsi lagu yang berada di peringkat atas dari seorang artis pada tangga lagu Global dan Indonesia belum tentu sama.

3. Telah dilakukan evaluasi terhadap model klasifikasi dan model *clustering* dan didapati hasil seperti berikut ini:

- Model klasifikasi untuk prediksi *highest charting position* terbaik menggunakan atribut *Number of Times Charted* dan atribut *Popularity* sebagai atribut prediktor dengan nilai $k = 30$. Didapati akurasi model sebesar 46% yang artinya model klasifikasi masih kurang bagus, namun model dapat memprediksi kelas 1-50 cukup baik dengan *precision* (ketepatan prediksi) sebesar 59% dan *recall* (kelengkapan prediksi) sebesar 74%.
- Model klasifikasi untuk prediksi *number of times charted* terbaik menggunakan atribut *Popularity* dan atribut *Highest Charting Position* sebagai atribut prediktor dengan nilai $k = 33$. Didapati akurasi model sebesar 86% yang artinya model klasifikasi sudah bagus, namun model hanya dapat memprediksi kelas 1-20 dengan baik di mana didapati *precision* (ketepatan prediksi) sebesar 90% dan *recall* (kelengkapan prediksi) sebesar 97%.
- Model *clustering* dengan menggunakan atribut *Highest Charting Position*, atribut *Number of Times Charted*, atribut *Popularity*, dan atribut *Streams* dapat mengelompokkan dataset ke dalam 5 *cluster* berbeda di mana terdapat perbedaan yang cukup signifikan antar *cluster* berdasarkan atribut *Highest Charting Position*, atribut *Number of Times Charted*, dan atribut *Popularity*.
- Model *clustering* dengan menggunakan atribut-atribut yang merepresentasikan elemen-elemen dalam musik dapat mengelompokkan dataset ke dalam 5 *cluster* berbeda dengan perbedaan signifikan antar *cluster* berdasarkan atribut *Duration*. Namun pengelompokkan tersebut tidak dapat membagi data ke dalam 5 *cluster* berbeda yang menunjukkan perbedaan signifikan antar *cluster* berdasarkan atribut *Highest Charting Position*.

4. Telah dibuat perangkat lunak untuk menampilkan pola atau *trend* yang didapatkan berdasarkan hasil eksplorasi dan analisis data dalam bentuk *dashboard*, serta telah diluncurkan juga model klasifikasi untuk melakukan prediksi *number of times charted* dari sebuah lagu pada perangkat lunak yang dibangun.

6.2 Saran

Berikut ini adalah saran untuk memperbaiki dan mengembangkan penelitian ini lebih lanjut:

1. Mencari data pendukung untuk mendapatkan pola-pola lain seperti lagu-lagu dari negara manakah yang banyak mendominasi tangga lagu *Top 200* Spotify dan melengkapi dengan data dari web Spotify.
2. Melakukan analisis lebih dalam pada kelompok data lebih kecil seperti, data lagu-lagu yang berhasil masuk ke dalam tangga lagu lebih dari 100 kali, data lagu-lagu yang memiliki nilai popularity sangat tinggi, data lagu-lagu yang tidak dapat mencapai peringkat lebih tinggi dari 100 dan lainnya.

DAFTAR REFERENSI

- [1] Moertini, V. S. dan kawan kawan (2020) *Pengantar Data Science dan Aplikasinya bagi Pemula*, 1st edition. Unpar Press, Indonesia.
- [2] Han, J., Kamber, M., dan Pein, J. (2012) *Data Mining Concepts and Techniques*, 3rd edition. Morgan Kaufmann Publishers, USA.
- [3] Kirk, A. (2012) *Data Visualization: a successful design process*. Packt Publishing, United Kingdom.
- [4] Munzner, T. (2015) *Visualization Analysis & Design*. Taylor & Francis Group, USA.
- [5] III, W. M., Beaver, R. J., dan Beaver, B. M. (2013) *Introduction to Probability and Statistics*, 14th edition. Cengage Learning, USA.
- [6] Sarrazin, N. (2016) *Music and the Child*. Open SUNY Textbooks, USA.