

Peningkatan Layanan Aplikasi Spotify Menggunakan Naïve Bayes

Era Wahyuni^{a1}, I Ketut Gede Suhartana^{a2}

^aProgram Studi Informatika, Universitas Udayana
Bukit Jimbaran, Bali, Indonesia
¹erawahyuni911@gmail.com
²ikg.suhartana@unud.ac.id

Abstract

Spotify is a platform that provides services for listening to music digitally that can be downloaded through the Google Play Store and The App Store can also be downloaded for PC. Currently, Spotify has 433 million users. With so many users, it is certainly inseparable from the reviews or ratings given to the Spotify application which will also have an impact on the platform. In order to improve the service to be better, these reviews need to be reviewed so that users are more comfortable in using the application. Through this research, it will be known the performance of the Naïve Bayes algorithm in conducting sentiment analysis on user reviews of the Spotify application on the Google Play Store. The results of this study show that the accuracy value of using Naïve Bayes is 85% from a ratio of 70:30 training data and testing data using a music dataset from 1 January 2022 to 9 July 2022, there were 61586 reviews from users taken from the Kaggle website.

Keywords: Spotify, Sentiment Analysis, Naïve Bayes

1. Pendahuluan

Dalam keseharian manusia musik telah masuk menjadi salah satu bagian penting yang selalu menemani, dimana manusia bisa mendengarkan musik ketika sedang belajar, bekerja, berolahraga, berbelanja di pusat perbelanjaan, pada transportasi umum, ruang publik dan masih banyak lagi. Setiap individu memiliki minat yang berbeda dalam pemilihan genre musik untuk didengarkan seperti pop, jazz, rock, rap dan sebagainya yang biasanya lebih mewakili emosi dari pendengarnya. Dari aspek psikologis musik sendiri berperan pada proses dalam mengatur emosi, berkomunikasi dan berinteraksi sosial. Area otak yang dilibatkan dalam musik sangat luas mulai dari *cerebellar, parietal, frontal, temporal, limbik/paralimbic* yang berkaitan dengan proses pada otak seperti perseptual kognitif, motorik dan emosional [3].

Cara mendengarkan musik pada saat ini sudah jauh berbeda jika dibandingkan dengan tahun 90-an dimana musik didengar melalui *tape deck, mini compo* atau pemutar musik portabel yang mana pada masa itu musik diputar menggunakan kaset atau compact disc (CD). Seiring dengan pesatnya perkembangan teknologi yang juga merambah hingga ke industri musik, pada saat ini musik sudah dapat didengar dan dinikmati keindahan alunan nadanya kapanpun dan dimanapun melalui ponsel pintar, komputer tablet, laptop dan diakses secara online menggunakan internet.

Semakin maraknya aplikasi yang menyediakan layanan *streaming* musik yang bergerak, membuat para pengembang harus saling berlomba untuk membuat aplikasi yang dapat memenuhi seluruh keinginan dan kebutuhan penggunanya. Untuk mencapai tujuan tersebut salah satu langkah yang harus ditempuh adalah dengan memikirkan setiap ulasan yang diberikan oleh pengguna baik itu merupakan ulasan yang berkonotasi negatif, positif, maupun berupa saran dan kritik yang membangun [4].

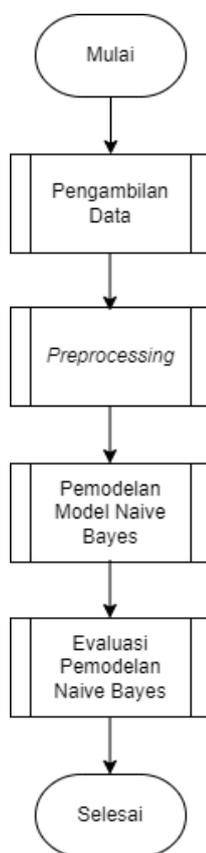
Pada penelitian terdahulu yang berjudul "Implementasi Algoritma Naïve Bayes Terhadap Analisis Sentimen Opini Masyarakat Indonesia Mengenai Drama Korea Pada *Twitter*" oleh Resti Amelia dkk. Dimana hasil menunjukkan bahwa perhitungan menggunakan Naïve Bayes memiliki performa yang termasuk sedang yaitu sebesar 69%. Demikian juga pada penelitian yang berjudul "Analisis Sentimen Ulasan Aplikasi *Buzzbreak* Menggunakan Metode Naïve Bayes *Classifier* pada Situs *Google Play Store*" oleh Dinda Putri Santoso dan Wahyu Wibowo diperoleh hasil Naïve Bayes untuk

analisis ketepatan klasifikasi adalah sebesar 76,52% dapat dikatakan bahwa masuk ke klasifikasi yang akurasinya cukup baik. Berbeda dengan kedua penelitian sebelumnya, pada penelitian yang berjudul “Analisis Sentimen Dengan Naïve Bayes Terhadap Komentar Aplikasi *Tokopedia*” oleh Rita Apriani dkk. Telah menunjukkan hasil performa yang baik pada penggunaan metode Naïve Bayes yaitu sebesar 97,13%.

Berdasarkan uraian diatas, peneliti mengusulkan penelitian yang berjudul “Peningkatan Layanan Aplikasi *Spotify* Menggunakan Naïve Bayes” yang menggunakan dataset sebanyak 61586 berisikan data dari tanggal 1 Januari 2022 hingga 9 Juli 2022 pada *Google Play Store* yang diunduh dari *website Kaggle* dan dipublikasikan oleh Hars Singh pada 19 September 2022. Dengan adanya penelitian ini diharapkan dapat membantu pengembang aplikasi untuk dapat meningkatkan aplikasi dan layanan yang diberikan agar lebih baik lagi kedepannya.

2. Metodologi Penelitian

Penelitian ini menggunakan data sebanyak 61586 ulasan mengenai ulasan aplikasi *Spotify* sejak tanggal 1 Januari 2022 hingga 9 Juli 2022 yang diperoleh dari *website Kaggle* yang diakses pada 26 September 2022 pukul 15:54:23 WITA. Diagram alir dari alur pelaksanaan penelitian ini dapat dilihat pada gambar 2.1.



Gambar 2.1 Diagram alir alur penelitian

Penelitian dimulai dengan (1) pengambilan data di website Kaggle dengan kata kunci *Spotify review* dataset, diperoleh hasil dataset sebanyak 61586 data ulasan pelanggan di *Google Play Store* data ini di publikasikan oleh Harsh Singh pada tanggal 19 September 2022 yang berisi data ulasan dari tanggal 1 Januari 2022 hingga 9 Juli 2022 (2) Selanjutnya data akan di *preprocessing* pada tahap ini akan dilakukan *cleaning* dimana tanda baca dihilangkan, kemudian *hashtag* dan *mention* juga dihilangkan, lalu tahap *case folding* untuk mengubah semua huruf pada teks dari yang tadinya terdapat huruf kapital diubah menjadi huruf kecil, lalu *filtering* penghapusan *stop word*, perubahan kata yang memiliki imbuhan menjadi kata dasar atau bisa juga disebut proses *stemming*, dan yang

terakhir adalah pengubahan kata dari yang bentuknya tidak baku menjadi kata dengan bentuk baku (3) Kemudian akan dilanjutkan dengan pembuatan model Naïve Bayes. Persamaan Naïve Bayes secara sederhana adalah sebagai berikut :

$$P(A|X) = \frac{P(X|A) * P(A)}{P(X)}$$

Dimana :

$P(A|X)$ = Probabilitas *posterior* dari A berdasarkan kondisi X (probabilitas *prior*)

$P(X|A)$ = Probabilitas *posterior* dari X berdasarkan kondisi A (*likelihood*)

$P(A)$ = Probabilitas *prior* dari A (probabilitas kelas *prior*)

$P(X)$ = Probabilitas *prior* dari X (prediktor kelas *prior*)

Prior secara sederhana dapat diartikan sebagai probabilitas dari kemungkinan dugaan suatu objek data pada *prior* atau saat sebelum dilakukan pengamatan. *Posterior* dapat dikatakan sebagai probabilitas bersyarat dari kemungkinan dugaan apa yang akan terjadi setelah dilakukan pengamatan. Proses penghitungan tingkat probabilitas data dimulai dengan pendefinisian *likelihood*. *Likelihood* merupakan proporsional terhadap probabilitas. Setelah *likelihood* didapatkan maka selanjutnya adalah mengalikan *likelihood* terhadap probabilitas dari setiap kelas. Hasil akhirnya akan digunakan sebagai patokan untuk melakukan klasifikasi pada data yang baru.

(4) Setelah model berhasil dibuat akan dilakukan evaluasi model Naïve Bayes. Evaluasi model dihitung akurasi menggunakan *confusion matrix*, tabel ini akan merepresentasikan performa dari model dengan lebih spesifik .

		Prediksi	
		Positif	Negatif
Realita	Positif	True Positive (TP)	False Positive (FP)
	Negatif	False Negative (FN)	True Negative (TN)

Tabel 3.2.1 Tabel *confusion matrix*

Rumus yang digunakan untuk menghitung seberapa besar akurasi dari model yang telah dibuat dalam mengklasifikasikan data dengan benar adalah:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

3. Hasil dan Pembahasan

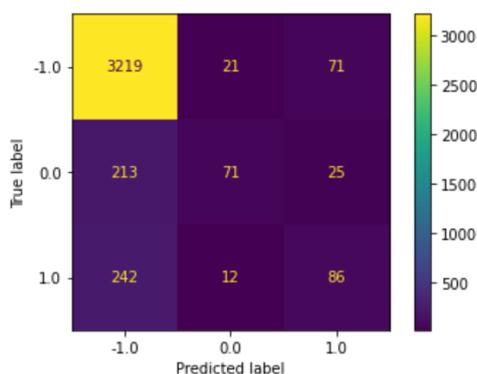
3.1 Preprocessing Data

Dalam penelitian ini sebanyak 61586 dataset ulasan mengenai aplikasi *Spotify* telah digunakan, dataset yang diperoleh dari *website Kaggle* ini perlu melewati suatu proses yang disebut dengan *preprocessing* agar data yang sebelumnya dalam bentuk mentah dapat berubah menjadi data dengan format yang lebih sesuai dan informasi yang terdapat didalamnya dapat ditafsirkan. Dimulai dengan proses *cleaning*, proses yang disebut *cleaning* ini bertujuan untuk membersihkan data ulasan yang mencakup *username*, *emoticons*, *hashtag* dan *mention* karena dianggap tidak perlukan. Selanjutnya tahap *case folding* pada tahap ini akan dilakukan pengubahan semua huruf pada teks dari yang tadinya terdapat huruf kapital diubah menjadi huruf kecil pada setiap kata atau disebut juga *lowercase*. Contohnya : "Universitas Udayana merupakan Universitas Terbaik Di Indonesia" berubah menjadi "universitas udayana merupakan universitas terbaik di indonesia". Kemudian proses *stemming* dalam proses ini dilakukan pengubahan kata yang memiliki imbuhan pada awalan, akhiran atau sisipan maupun kombinasi pada awalan dan akhiran menjadi kata dasar pada setiap

data ulasan. Contohnya kata “menguji” akan ditransformasikan menjadi kata “uji”. Dilanjutkan dengan tahap penghapusan *stop word*, proses ini akan membuang kata yang kurang ada pengaruhnya atau memiliki informasi yang rendah lalu akan memfokuskan pada kata-kata yang lebih penting. Contohnya dalam bahasa Indonesia adalah kata “di”, “yang”, “dari”, dan sebagainya. Kemudian masuk ke tahap *tokenizing*, pada tahapan ini sebelum dianalisa akan terjadi pemisahan teks menjadi potongan yang lebih kecil yang disebut sebagai token.

3.2 Pemodelan Naïve Bayes

Setelah dilakukan proses *preprocessing* pada data maka selanjutnya untuk analisis sentimen pada data ulasan aplikasi *Spotify* perlu dibuat pemodelan menggunakan Naïve Bayes. Data yang telah melewati tahap *preprocessing* adalah sebanyak 61578 data uji dan 162878 data latih. Untuk membuat model sentimen analisis ini data yang digunakan adalah data latih. Tahapan yang akan dilalui dalam pembuatan model ini diantaranya adalah dimulai dengan melakukan perubahan bentuk data teks menjadi bentuk angka agar data dapat terbaca oleh komputer dengan menggunakan *count vectorize*, pada pembagian data ini untuk membangun model analisis sentimen data yang digunakan adalah data latih sedangkan data uji digunakan untuk menghitung seberapa besar akurasi dari pemodelan yang dibuat. Tahapan berikutnya adalah pelatihan data yang dimulai dengan penentuan probabilitas *prior* pada data latih di setiap kategori data yang digunakan, kemudian dilakukan perhitungan probabilitas *likelihood*. Setelah probabilitas *likelihood* didapatkan selanjutnya adalah tahap prediksi kategori dengan mengalikan nilai probabilitas *prior* dengan probabilitas *likelihood*. Tahap terakhir adalah tahap evaluasi model. Evaluasi model dihitung akurasinya menggunakan *confusion matrix*. Hasil evaluasi dapat dilihat pada gambar 3.2.1 di bawah ini.



Gambar 3.2.1 Hasil evaluasi model menggunakan *confusion matrix*

Setelah dilakukan pemodelan menggunakan Naïve Bayes diperoleh hasil data train sebesar 90% dan data tes sebesar 83%. Kemudian dilanjutkan dengan penggunaan *hyperparameter tuning* untuk menganalisis parameter paling optimal sehingga diperoleh hasil yang lebih tinggi yakni sebesar 96% untuk data training dan 85% untuk data testing, serta diperoleh juga nilai alpha terbaik dengan nilai 0.01. Dari hasil evaluasi menggunakan *confusion matrix* pada gambar di atas diperoleh hasil bahwa akurasi dari model yang dibuat adalah sebesar 85%.

4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan untuk analisis sentimen pada data ulasan aplikasi Spotify di Google Play Store dengan menggunakan algoritma Naïve Bayes menghasilkan akurasi sebesar 85% dengan *hyperparameter tuning* yang dimana dinilai cukup tinggi, namun jika tidak menggunakan *hyperparameter tuning* hasil yang diperoleh lebih kecil. Sentimen yang diperoleh didominasi oleh sentiment negative sebanyak 96%, sentiment positif sebanyak 1.8% dan sentiment netral sebanyak 1.75%. Spotify perlu membenahi aplikasinya dan kualitas layanannya agar dapat menjawab permasalahan sebagian besar konsumen yang telah menyampaikan keluhannya pada Google Play Store sehingga konsumen tetap merasa nyaman menggunakan aplikasi Spotify.

References

- [1] Amelia, R., Prastiwi, N. S., & Purbaya, M. E. (2022). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Masyarakat Indonesia Mengenai Drama Korea Pada Twitter. *Jurnal Riset Komputer*, 9(2), 338–343. <https://doi.org/10.30865/jurikom.v9i2.3895>
- [2] Apriani, R., Gustian, D., Program, S., Sistem, I., Putra, U. N., Indonesia, S., Raya, J., Kaler, C., 21, N., & Sukabumi, K. (2019). Analisis Sentimen dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia. *Jurnal Rekayasa Teknologi Nusa Putra*, 6(1), 54–62. <https://rekayasa.nusaputra.ac.id/article/view/86>
- [3] Fiana, D. N., & Cahyani, A. (2019). Dampak Terapi Musik pada Fungsi Kognitif Pasien dengan Demensia. *JK Unila*, 3(1), 221–225. <http://repository.lppm.unila.ac.id/17010/1/2231-2951-1-PB.pdf>
- [4] Prasetyo, J., & Siahaan, D. O. (2017). Klasifikasi Ulasan Aplikasi Pada Toko Aplikasi Bergerak Dengan Memanfaatkan Issue Tracker Github. *JUTI: Jurnal Ilmiah Teknologi Informasi*, 15(2), 206. <https://doi.org/10.12962/j24068535.v15i2.a666>
- [5] Santoso, D. P., & Wibowo, W. (2022). Analisis Sentimen Ulasan Aplikasi Buzzbreak Menggunakan Metode Naïve Bayes Classifier pada Situs Google Play Store. *Jurnal Sains dan Seni ITS*, 11(2). <https://doi.org/10.12962/j23373520.v11i2.72534>

Halaman ini sengaja dibiarkan kosong