eISSN : 2621-4970
pISSN : 2621-4962

Volume 6 | No 2
Juli | **2023**

Analisis Sentimen Terhadap Pemutar Musik Online Spotify Dengan Algoritma Naive Bayes dan Support Vector Machine

Ginabila[1*], Ahmad Fauzi[2]
Fakultas Teknologi dan Informatika, Program Studi Sistem Informasi
Universitas Bina Sarana Informatika
gina.gnb@bsi.ac.id, ahmad.aau@bsi.ac.id

Abstrak

Manusia memiliki kebutuhan preferensi musik yang yang sangat beragam, oleh karena itu pemutar musik online menjadi salah satu solusi untuk memenuhi kebutuhan ini dengan menyediakan katalog musik yang luas. Analisis sentimen adalah proses untuk mengevaluasi dan mengklasifikasikan sentimen atau perasaan di balik teks atau data yang diberikan. Dalam konteks ini, analisis sentimen dilakukan pada pemutar musik online Spotify. Dua algoritma yang umum digunakan untuk analisis sentimen adalah Naive Bayes dan Support Vector Machine (SVM). Kedua algoritma ini dapat diterapkan dalam analisis sentimen pada pemutar musik online. Data teks seperti ulasan atau komentar pengguna dikumpulkan dan dilabeli dengan sentimen yang sesuai. Hasil dari penelitian menggunakan kedua algoritma ini menghasilkan nilai akurasi yang hampir sama baiknya. Algoritma Support Vector Machine menghasilkan tingkat akurasi sebesar 82,42%, sedangkan untuk Algoritma Naive Bayes mencapai 84,73%.
Kata kunci: Analisis Sentimen, Naive Bayes, Support Vector Machine

Abstract

Humans have diverse music preferences and online music players are a solution to meet these needs by providing a wide music catalog. Sentiment analysis is the process of evaluating and classifying sentiments or feelings behind given texts or data. In this context, sentiment analysis is performed on Spotify online music players. Two common algorithms used for sentiment analysis are Naive Bayes and Support Vector Machine (SVM). Both algorithms can be applied in sentiment analysis for online music players. Text data such as user reviews or comments are collected and labeled with corresponding sentiments. The results of the research using both algorithms yielded similar high accuracy. The Support Vector Machine algorithm achieved an accuracy rate of 82.42%, while the Naive Bayes algorithm reached 84.73%.
Keywords: ***Sentiment Analysis, Naive Bayes, Support Vector Machine***

PENDAHULUAN

Hiburan memiliki peran penting dalam kehidupan sehari-hari. Setelah melakukan segala aktivitas, hiburan dapat membantu kita menghilangkan penat pada pekerjaan dan segala hal yang telah dilakukan yang menguras energi dan fikiran. Kehidupan modern saat ini menjadikan segala hiburan mudah untuk didapatkan dalam aplikasi smartphone yang dimiliki. Dalam hal ini banyak orang menggunakan musik untuk merelaksasi fikiran dan memberikan hiburan setelah melakukan kegiatan bahkan bersamaan dengan melakukan kegiatan.

Musik merupakan seni yang melukiskan pemikiran dan perasaan

manusia lewat keindahan suara. (Iswandi, 2015). Pada zaman yang teknologi nya selalu berkembang saat ini, musik dapat diakses dengan sangat mudah. Sebelumnya kita hanya bisa mendengar musik dengan membeli kaset atau CD, lalu kita putar melalui alat pemutar. Zaman setelahnya, kita bisa mengunduh musik dari berbagai situs yang menyediakan *file* musik untuk diunduh. Dan pada saat ini kita bisa mendengar musik dari berbagai *platform* pemutar musik online. Kita hanya cukup mengunduh aplikasi pemutar musik, lalu berbagai musik tersedia dalam aplikasi tersebut.

Pemutar musik online adalah aplikasi atau *platform* digital yang memungkinkan pengguna untuk mendengarkan musik secara *streaming* melalui koneksi internet. Saat ini, pemutar musik online telah mengalami perkembangan yang signifikan dan menawarkan beragam fitur untuk memenuhi kebutuhan pendengar, salah satunya adalah Spotify. Spotify adalah layanan streaming musik yang populer, didirikan oleh Daniel Ek dan Martin Lorentzon pada tahun 2006 di Stockholm, Swedia. Layanan ini diluncurkan secara resmi pada tanggal 7 Oktober 2008. Spotify memungkinkan pengguna untuk mengakses katalog besar lagu, album, dan podcast dari berbagai genre musik dari seluruh dunia melalui internet.

Dengan munculnya sesuatu yang baru dalam kehidupan, banyak pendapat dari berbagai khalayak mengenai munculnya hal baru tersebut. Hal ini tidak luput dari munculnya pemutar musik online Spotify. Pendapat atau *review* dari sekelompok orang pada halaman playstore atau appstore dapat mempengaruhi orang lain terhadap kepopuleran aplikasi. Untuk itu pendapat atau review dari pengguna merupakan hal yang penting untuk menjadi masukan *developer* pemutar musik online. Pendapat atau *review* dari sekumpulan orang yang telah menggunakan aplikasi ini juga dapat menjadi indikator terhadap kualitas dari aplikasi tersebut.

Analisis sentimen adalah sebuah proses untuk menentukan sentimen atau opini dari seseorang yang diwujudkan dalam bentuk teks dan bisa dikategorikan sebagai sentimen posisif atau negatif. (Hadna, Winarno, & Santosa, 2016)

Beberapa penelitian terkait sebelumnya ditemukan yaitu analisis sentimen terhadap aplikasi PLN Mobile dengan menggunakan Algoritma Naive Bayes (NB) yang menghasilkan tingkat akurasi mencapai 70% (Asri, Nita Suliyanti, Kuswardani, & Fajri, 2022). Penelitian lainnya yang terkait dengan judul Perbandingan Akurasi Analisis Sentimen Tweet terhadap Pemerintah Provinsi DKI Jakarta di Masa Pandemi menghasilkan beberapa nilai akurasi dari berbagai algoritma yaitu *random forest classifier* dengan hasil akurasi sebesar 75,81%, algoritma naive bayes dengan hasil akurasi 75,22%, dan algoritma support vector machine 77,58% (Himawan & Eliyani, 2021).

Berdasakan uraian diatas, pada penelitian ini algoritma Naive Bayes dan Support Vector Machine akan digunakan dalam penelitian. Kedua algoritma tersebut digunakan untuk membandingkan metode yang menghasilkan tingkat akurasi yang lebih baik. Optimasi Particle Swarm Otimization (PSO) juga digunakan untuk mencoba meningkatkan nilai akurasi dari kedua algoritma yang digunakan.

Rumusan Masalah

1. Apakah Algoritma Naive Bayes dan Support Vector Machine dapat memprediksi dengan baik kelas pada sentimen analisis pemutar musik online?
2. Bagaimana Nilai Akurasi dari Algoritma Naive Bayes dan Support Vector Machine dalam memecahkan masalah analis sentimen pada pemutar musik online?
3. Manakah algoritma yang lebih baik digunakan dalam memecahkan masalah analis sentimen pada pemutar musik online?

Tujuan Penelitian

1. Untuk mengetahui Apakah Algoritma Naive Bayes dan Support Vector Machine dapat memprediksi dengan baik kelas pada sentimen analisis pemutar musik online
2. Untuk mengetahui bagaimana tingkat akurasi yang didapatkan dari algoritma Naive Bayes dan Support Vector Machine dalam menyelesaikan masalah pengklasifikasian sentimen pengguna pemutar musik online.
3. Untuk membandingkan algoritma yang paling baik digunakan dalam masalah pengklasifikasian sentimen pengguna pemutar musik online.

Manfaat Penelitian

1. Penelitian ini diharapkan dapat mempermudah para pengguna pemutar musik online dalam mencari aplikasi pemutar musik online terbaik dan dapat menjadi bahan evaluasi *developer* aplikasi untuk mengetahui apa yang dapat meningkatkan minat pengguna untuk mendownload aplikasinya.
2. Memberikan kontribusi keilmuan dalam bidang analisis sentimen dalam tema pemutar musik online.

Tinjauan Pustaka

1. Penelitian Terdahulu

Penelitian pertama dilakukan oleh Ferdiana dkk, mengenai Dataset Indonesia untuk Analisis Sentimen. Dataset ini mencakup data utama, yaitu 10.806 baris data berbahasa Indonesia yang diambil dari media sosial Twitter, yang telah dikategorikan ke dalam tiga label, yaitu positif, negatif, dan netral, beserta 454.559 baris data yang masih bersifat mentah.. Setelah dilakukan pengujian menggunakan model analisis sentimen sederhana yang menggunakan algoritme SVM, KNN, dan SGD. Penelitian ini menghasilkan tingkat akurasi dari masing-masing alroritma yaitu SVM dengan akurasi tertinggi 63%, KNN dengan akurasi paling tinggi 55% dan SGD dengan akurasi tertinggi 63% (Ferdiana, Jatmiko, Purwanti, Ayu, & Dicka, 2019).

Penelitian ke-dua dilakukan oleh Gunawan, Pratiwi dan Pratama. Penelitian ini menganalisis Ulasan Produk Menggunakan Metode Naive Bayes. Data review didapatkan dari hasil *review* yang ditinggalkan pada halaman web *female daily* dan berjumlah 1500 data latih. Algoritma yang digunakan dalam penelitian ini adalah Naive Bayes. Hasil dari penelitian ini menunjukkan nilai akurasi tertinggi yang didapatkan dengan menggunakan algoritma Naive Bayes adalah 77,78% (Gunawan, Pratiwi, & Pratama, 2018).

Penelitian ke-tiga dilakukan oleh Alita dan Rahman yang melakukan Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. Data *review* yang digunakan didapat

dengan metode *crawling* pada twitter tentang pelayanan publik dan dikumpulkan sebanyak 2027 Data. Pengujian dilakukan dengan menggunakan algoritma Random Forest. Tingkat akurasi yang didapatkan penelitian ini dengan algoritma Random Forest mencapai 77,79% (Alita & Rahman, 2020).

Penelitian ke-empat dilakukan Darwis, Siskawati dan Abidin, mengenai Penerapan Algoritma Naive Bayes untuk Analisis Sentimen *Review* Data Twitter BMKG Nasional. Dataset ini diambil dari twitter dengan tema pelayanan BMKG. Metode yang digunakan dalam penelitian ini adalah klasifikasi data adalah Naïve Bayes Classifier (NBC).. Hasil uji akurasi pada metode naive bayes untuk klasifikasi yaitu 69.97% (Darwis, Siskawati, & Abidin, 2021).

Penelitian ke-lima dilakukan oleh Fikri, Sabrila dan Azhar. Penelitian ini membandingkan metode Naive Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. Analisis dilakukan dengan mengklasifikasikan tweets yang berisi sentimen masyarakat mengenai UMM. Metode klasifikasi yang digunakan dalam penelitian ini adalah Naïve Bayes dan Support Vector Machine (SVM) dengan pembobotan menggunakan TF-IDF. Hasil komparasi kedua metode menunjukkan bahwa Naïve Bayes mendapatkan hasil akurasi yang lebih baik dari SVM dengan akurasi sebesar 73,65% sedangkan SVM hanya mencapai 70,20% (Fikri, Sabrila, & Azhar, 2020).

LANDASAN TEORI

1. Analisis Sentimen

Analisis sentimen merupakan analisa terhadap suatu peristiwa dari pendapat yang didasarkan pada sikap seseorang tentang suatu objek. Analisis sentimen biasanya dilakukan untuk mengumpulkan dan mengetahui opini masyarakat dalam postingan Blog, Twitter, Facebook, dan yang lainnya. Analisis sentimen dibutuhkan dengan tujuan untuk mengetahui opini publik terhadap suatu objek. Opini-opini tersebut bisa berupa opini negatif atau positif tergantung dari pandangan publik terhadap objek tersebut (Pintoko & L., 2018).

2. Algoritma Naive Bayes

Naïve Bayes merupakan machine learning yang menggunakan perhitungan probabilitas yang menggunakan konsep pendekatan Bayes. Penggunaan teorema Bayes pada algoritma Naïve Bayes adalah dengan menggabungkan *prior probability dan conditional probability* dalam suatu rumus yang dapat digunakan untuk menghitung probabilitas dari setiap kemungkinan klasifikasi (Yulita, Nugroho, & Algifari, 2021).

3. Support Vector Machine

Support Vector Machine merupakan sebuah metode yang membandingan suatu seleksi parameter *standart* nilai diskrit yang disebut kandidat set. Untuk mengklasifikasikan akurasi Support Vector Machinie (SVM) diperkenalkan oleh Vapnik, Boser dan Guyon pada tahun 1992. Metode Support Vector Machine memiliki 5 komponen yang berfungsi di antaranya (Riadi, Umar, & Aini, 2019):

a) *SVM Linear*
b) *SVM Polynomial*
c) *Kernel RBF* (Radio Basis Function)
d) Kernel MLP (*Multi Layer Perceptrom*)
e) *Tangent Hyperbolic* (sigmoid)

4. Particle Swarm Optimization

Particle Swarm Optimization merupakan teknik optimasi stokastik berbasis populasi yang diusulkan pada yahun 1995 oleh Kennedy dan Eberhart. Pengembangan PSO didasarkan pada metafora interaksi sosial dan komunikasi dari pergerakan kawanan burung atau ikan (*bird flocking atau fish schooling*). Pada algoritma Particle Swarm Optimization, pencarian solusi dilakukan oleh suatu populasi yang terdiri dari beberapa partikel. Ketika Populasi dibangkitkan secara random atau acak dengan batasan permasalahan yang dihadapi. Setiap partikel merepresentasikan partikel atau solusi dari permasalahan yang sedang dihadapi. Partikel tersebut mencari solusi yang optimal dengan melintasi ruang pencarian dengan cara partikel terkait melakukan penyesuaian terhadap posisi yang terbaik dari setiap partikel tersebut atau disebut *local best* dan posisi partikel terbaik dari seluruh kawanan atau disebut global best selama melintasi ruang pencarian. Jadi penyebaran informasi terjadi dalam partikel itu sendiri dan antara suatu partikel dengan partikel terbaik dari seluruh kawanan selama proses pencarian solusi. Setelah itu dilakukan proses pencarian untuk mencari posisi terbaik setiap partikel dalam jumlah iterasi tertentu sampai didapatkannya posisi relatif yang tetap (*steady*) atau mencapai batas iterasi yang telah ditetapkan. Pada setiap iterasi (t),setiap solusi yang direpresentasikan oleh partikel i, dievaluasi performanya dengan cara memasukkan solusi tersebut kedalam nilai *fitness function*. Setiap partikel memerlukan titik pada suatu dimensi ruang tertentu kemudian terdapat 2 faktor yang memberikan karakter terhadap status partikel pada ruang pencarian yaitu pada posisi partikel (X) dan pada kecepatan partikel (Y) (Mutiara, 2020).

5. SMOTE

Metode Synthetic Minority Over-sampling Technique (SMOTE) merupakan metode yang populer diterapkan dalam rangka menangani ketidak seimbangan kelas. Teknik ini mensintesis sampel baru dari kelas minoritas untuk menyeimbangkan dataset dengan cara sampling ulang sampel kelas minoritas (Siringoringo, 2018)

METODE PENELITIAN

1. Perencanaan Penelitian

Beberapa tahapan digunakan dalam metode penelitian ini, berikut ini adalah tahapan dalam penelitian yang akan dilakukan:

a. Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dari ulasan yang ditinggalkan oleh pengguna aplikasi pemutar musik online yaitu Spotify. Ulasan yang diambil untuk analisis sentimen adalah ulasan yang berbahasa Indonesia yang terkumpul sebanyak 121 data.

b. Pengolahan Data Awal

Pada tahap pengolahan data awal, 121 ulasan data yang terkumpul akan melaluai tahap normalisasi dan *preprocess*. Data tersebut akan diperbaiki kesalahan penulisan dan diperbaiki menjadi kata baku hingga sesuai dengan penulisan di KBBI (Kamus Besar Bahasa Indonesia).

c. Metode

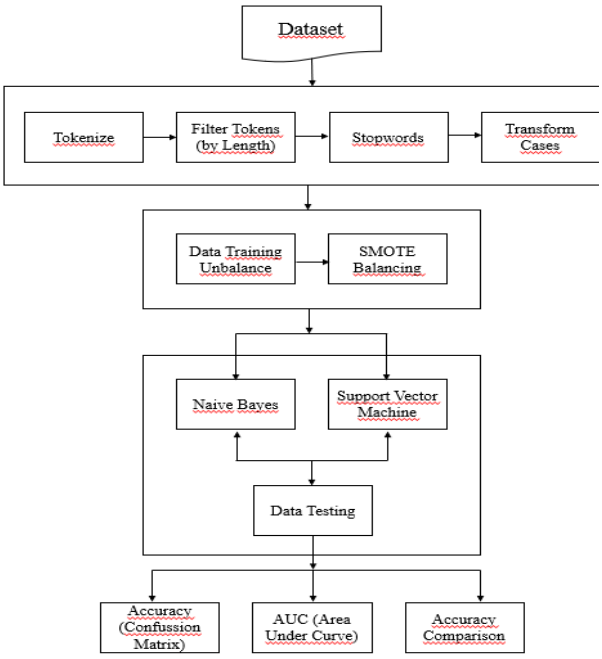
Metode yang diusulkan peneliti untuk melakukan eksperimen ini adalah dengan menggunakan dua algoritma klasifikasi yaitu Naive Bayes (NB) dan Support Vector Machine (SVM).

Sebelum pengujian dilakukan, metode SMOTE digunakan untuk menangani ketidakseimbangan pada kelas. Optimasi Particle Swarm Optimization disini juga digunakan untuk meningkatkan akurasi.

d. Evaluasi dan Validasi Hasil Evaluasi

Evaluasi pada tahap ini berfungsi untuk mengetahui tingkat akurasi dari algoritma naive bayes dan Support Vector Machine dalam penelitian ini. *Ten Fold Cross-Validation* dugunakan untik validasi dimana pada proses ini data dibagi secara acak ke dalam 10 bagian. Tahap pengujian dimulai dengan pembentukan model data pertama, kemudian model yang telah terbentuk akan diujikan pada 9 bagian data yang tersisa. Setelah tahap pengujian selesai *performance* dari algoritma klasifikasi *text mining* diukur menggunakan *Accuracy* (dihitung menggunakan *confusion matrix*), AUC dan kurva ROC.

Kerangka Pemikiran



Gambar 1. Kerangka Pemikiran

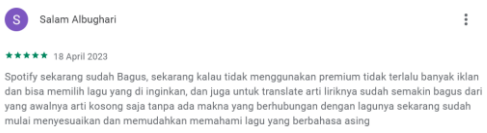
ANALISIS DAN PERANCANGAN

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan 121 data berbahasa Indonesia yang ditinggalkan oleh pengguna aplikasi pemutar musik online pada Playstore. Pemutar musik yang dipilih dalam penelitian ini adalah aplikasi "Spotify". Spotify merupakan aplikasi yang sangat banyak digunakan oleh pengguna android maupun iphone. Sampai saat ini aplikasi tersebut telah mencapai 26 juta ulasan dan merupakan aplikasi nomor satu populer di musik dan audio.

a. Contoh Ulasan Positif

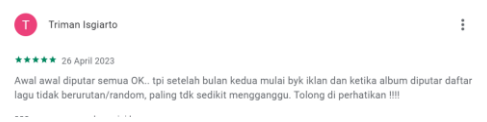
"Spotify sekarang sudah Bagus, sekarang kalau tidak menggunakan premium tidak terlalu banyak iklan dan bisa memilih lagu yang di inginkan, dan juga untuk translate arti liriknya sudah semakin bagus dari yang awalnya arti kosong saja tanpa ada makna yang berhubungan dengan lagunya sekarang sudah mulai menyesuaikan dan memudahkan memahami lagu yang berbahasa asing".



Gambar 2. Contoh Ulasan Positif

b. Contoh Ulasan Negatif

"Awal awal diputar semua OK.. tpi setelah bulan kedua mulai byk iklan dan ketika album diputar daftar lagu tidak berurutan/random, paling tdk sedikit mengganggu. Tolong di perhatikan !!!!"



Gambar 3. Contoh Ulasan Negatif

2. Pengolahan Data Awal (*Preprocess*)

Pada tahap pengolahan data awal, dilakukan pelabelan data secara manual oleh peneliti. Pada tahap ini data ulasan juga diperbaiki karna data tersebut masih terdapat banyak kesalahan penulisan dan terdapat banyak emoticon. Data ulasan diolah agar tidak menurunkan tingkat akurasi karena teks yang diolah memiliki karakteristik dimensi yang tinggi, terdapat noise dan stuktur yang tidak baik. Untuk itu dalam pengolahan data awal *text mining* harus melaluai beberapa tahapan *preprocessing*. Tahapan daya yang dilakukan adalah sebagai berikut (Utami, 2017):

a) *Tokenize*

Tokenize merupakan proses untuk memisah-misahkan kata. Proses memotong setiap kata dalam teks dan mengubah huruf dalam dokumen menjadi huruf kecil. Hanya huruf yang diterima, sedangkan karakter khusus atau tanda baca akan dihilangkan.

b) *Filter Tokens* (By Length)

*Filter token*s (by length) merupakan proses mengambil kata-kata penting dari hasil token. Dalam proses ini, kata-kata yang memiliki panjang tertentu akan dihapus.

c) *Stopwords Removal*

Filter stopwords removal adalah proses menghilangkan kata-kata yang sering muncul namun tidak memiliki pengaruh apapun dalam ekstraksi sentimen suatu *review*. Kata yang termasuk seperti kata penunjuk waktu, kata tanya dan lain-lain.

d) *Transform Cases*

Transform Cases akan mengubah seluruh huruf kapital menjadi huruf kecil atau sebaliknya.

3. Metode yang Diusulkan

Metode yang diusulkan dalam penelitian ini adalah dengan menggunakan Algoritma Naive Bayes dan Support Vector Machine. Untuk mengatasi ketidakseimbangan data juga digunakan metode SMOTE. Metode *Synthetic Minority Over-sampling Technique* (SMOTE) merupakan metode yang populer diterapkan dalam rangka menangani ketidak seimbangan kelas. Teknik ini mensintesis sampel baru dari kelas minoritas untuk menyeimbangkan dataset dengan cara sampling ulang sampel kelas minoritas (Siringoringo, 2018). Optimasi Particle Swarm Optimization digunakan untuk meningkatkan tingkat akurasi dari kedua metode. Metode Naive Bayes merupakan salah satu metode klasifikasi yang digunakan dalam anaisis sentimen. etode ini didasarkan pada teorema Bayes yang mengasumsikan bahwa setiap fitur (kata atau frasa) dalam teks independen secara kondisional terhadap kelas sentimen yang diberikan. Sedangkan Support Vector Machine merupakan algoritma klasifikasi ensemble yang juga dapat digunakan dalam analisis sentimen. Ensemble berarti bahwa Support Vector Machine terdiri dari beberapa pohon keputusan yang digabungkan untuk menghasilkan prediksi akhir.

4. Eksperimen

Dalam melakukan eksperimen dan pengujian pada penelitian ini digunakan Aplikasi Rapid Miner Studio Versi 10.1 sebagai alat untuk menghitung tingkat akurasi. Dataset yang digunakan dalam penelitian ini merupakan dataset yang dikumpulkan oleh peneliti pada ulasan yang ditinggalkan pengguna spotify. Diambil sebanyak 121 data dan telah

dilakukan *preprocess* sebelum pengujian. Spesifikasi *Hardware* yang digunakan dalam penelitian ini dapat dilihat pada tabel berikut:

Tabel 1. Spesifikasi Sistem Komputer Minimum yang Digunakan

Processor	*Intel Core* 2duo 2.27 GHz
Memori	4 GB
Harddisk	500 GB
Sistem Operasi	*Microsoft Windows* 07
Aplikasi *Text Mining*	*Rapid Miner Studio* versi 10.1

5. Evaluasi dan Validasi Hasil

Validasi yang digunakan dalam penelitian ini merupakan validasi standar 10 *Fold Cross-Validation* dimana prosesnya data akan dibagi secara acak ke dalam 10 bagian. *Confusion matrix* disini juga digunakan untuk mengukur tingkar akurasi serta hasil evaluasi juga akan ditampilkan dalam bentuk kurva ROC.

IMPLEMENTASI DAN PEMBAHASAN

1. Pembahasan

A. Fase Pemahaman Bisnis (*Business Understanding Phase*)

Pada tahap ini, dilakukan pemahan terhadap objek penelitian dengan cara menggali informasi melalui aplikasi spotify pada playstore untuk mencari ulasan-ulasan dari para pengguna aplikasi. Ulasan yang ditinggalkan pengguna dapat digunakan untuk menjadi parameter penentu pemutar musik online yang digemari pengguna. Dalam hal ini developer juga dapat menjadikan ini sebagai bahan evaluasi aplikasi pemutar musik online yang mereka buat.

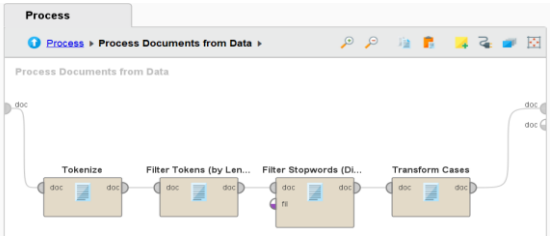
B. Fase Pemahaman Data (*Data Understanding Phase*)

Berikut adalah rincian dari pengambilan data yang dilakukan:

1) Data ulasan pengguna spotify dikumpulkan secara manual melalui halaman (https://play.google.com/ store/apps/ details?id=com.spotify. music&hl=id&gl=US) .
2) Data ulasan yang diambil merupakan data ulasan dari tanggal terbaru yang ditinggalkan oleh pengguna.
3) Setelah proses pengambilan data selesai dikumpulkan, data tersebut diberi label Positif atau Negatif.

C. Fase Pengolahan Data (*Data Preparation Phase*)

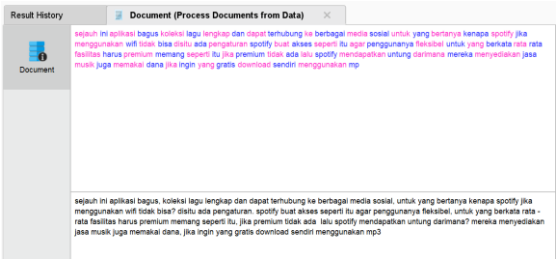
Tahap data *preparation* merupakan tahap proses perbaikan data yang berfungsi agar data bersih dan siap untuk digunakan untuk proses selanjutnya. Tahap *preprocess* data diaplikasikan pada RapidMiner Studio 10.1 dengan rangkaian sebagai berikut:



Gambar 4. Rangkaian *Preprocess* pada Rapid Miner

1) *Tokenize*

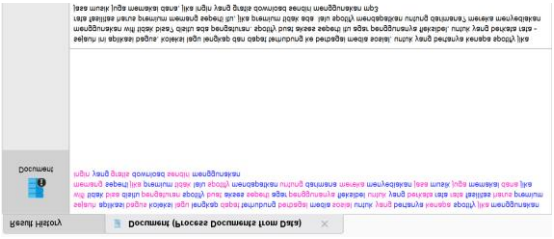
Pada tahap *tokenize* data yang telah melalui tahap normalisasi secara manual akan dihilangkan tnda baca, simbol dan karakter yang bukan merupakan huruf a-z. Berikut adalah hasil dari tahap *Tokenize* pada RapidMiner:



Gambar 5. Tahap *Tokenize* pada RapidMiner

2) *Filter Tokens* (*by Length*)

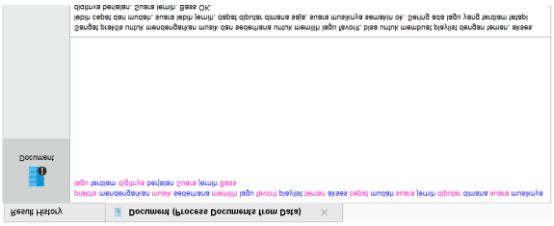
Setelah tahap *tokenize* selesai, tahap selanjutnya yang dijalankan pada RapidMiner adalah tahap *filter tokens (by length).* Pada tahap ini kata-kata yang berjumlah kurang dari 4 karakter dan lebih dari 25 karakter akan dibuang. Berikut adalah hasil dari proses *filter tokens (by length)* pada RapidMiner:



Gambar 6. Tahap *Filter Tokens (By Length)* pada RapidMiner

3) *Filter Stopwords* (*by Dictionary*)

Pada tahap *filter stopwords (by dictionary)* dilakukan, kata kata yang tidak memiliki makna penting atau informasi yang tidak dibutuhkan akan dihapus seperti kata sambung dan lainnya. Berikut adalah hasil data setelah melalui tahap *Stopwords* pada RapidMiner:



Gambar 7. Tahap *Filter Stopwords (By Dictinary)* pada RapidMiner

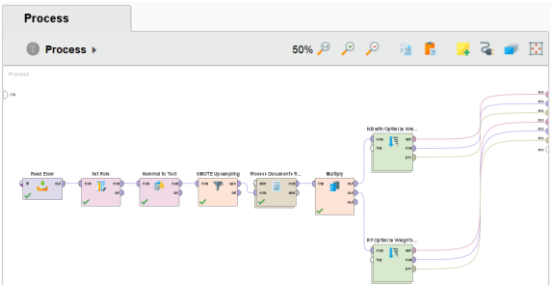
4) *Transform Cases*

Tahap selanjutnya setelah meakukan *stopwords*, data masuk ke dalam tahap *Transform Cases*. Pada tahap ini, semua huruf kapital (*Upper Case*) yang ada pada data kan di *Transform* atau diubah menjadi huruf kecil (*Lower Case*). Berikut adalah hasil dari tahap *Transform Cases* pada RapidMiner:

Gambar 8. Tahap *Transform Cases* pada RapidMiner

D. Fase Pemodelan (*Modelling Phase*)

Pada tahap ini, algoritma yang akan digunakan ditentukan oleh peneliti. Pada penelitian ini dua algoritma yaitu Naive Bayes dan Support Vector Machine digunakan untuk mencari algoritma yang paling tepat dan memiliki nilai akurasi tertinggi. SMOTE digunakan untuk menyeimbangkan antara kelas positif dan Negatif dan untuk meningkatkan nilai akurasi, optimasi juga dilakukan pada tahap ini dengan menggunakan Particle Swarm Optimization (PSO). Berikut adalah desain Fase Pemodelan pada Rapid Miner:



Gambar 9. Rangkaian Model dengan Algoritma Naive Bayes dan Support Vector Machine pada RapidMiner

E. Fase Evaluasi (*Evaluation Phase*)

Penentuan nilai kegunaan dari model yang telah dibuat akan dilakukan pada fase evaluasi. Model yang telah terbentuk akan diuji dengan *confusion matrix* untuk mengetahui tingkat akurasi.

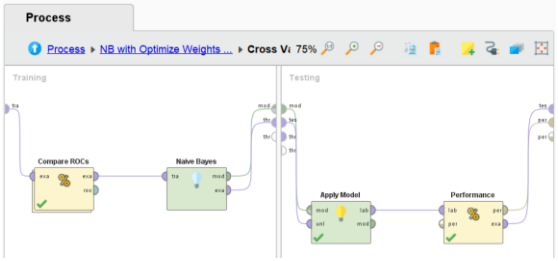
2. Hasil

A. Hasil Eksperimen Pengujian dengan Algoritma Naive Bayes

Berikut ini adalah hasil rancangan model klasifikasi, Hasil Eksperimen, Confusion Matrix dan Kurva ROC dengan Algoritma Naive Bayes:

1) Model Klasifikasi

Berikut adalah rangkaian model pengujian pada RapidMiner dengan Algoritma Naive Bayes:



Gambar 10. Rangkaian Model Pengujian dengan Algoritma Naive Bayes

2) Hasil Eksperimen

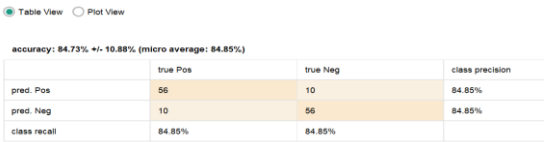
Setelah rangkaian model dijalankan dalam RapidMiner, hasil precision, recall dan akurasi dari pengujian akan didapatkan. Berikut adalah rincian hasil pengujian dengan algoritma Naive Bayes:

Tabel 2. Tabel Rincian Hasil Eksperimen dengan Algortima Naive Bayes

No.	Keterangan	Hasil
1	Precision	87,00%
2	Recall	85,48%
3	Akurasi	84,73%

3) *Confusion Matrix*

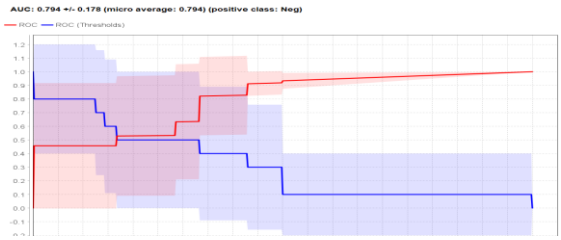
Berikut adalah *confusion matrix* yang dihasilkan dari pengujian data dengan algoritma Naive Bayes:



Gambar 11. *Confusion Matrix* Hasil Eksperimen dengan Algoritma Naive Bayes

4) Kurva ROC (*ROC Curves*)

Berikut adalah kurva ROC yang divisualisasikan berdasarkan perhitungan *confusion matrix*:



Gambar 12. Kurva ROC Pengujian Data dengan Algoritma Naive Bayes

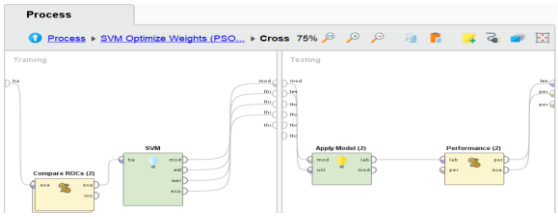
Berdasarkan gambar diatas dapat diketahui hasil dari nilai AUC pada Kurva ROC adalah sebesar 0,794 yang menyatakan bahwa kategori nilai yang didapatkan adalah *Fair Classification.*

B. Hasil Eksperimen Pengujian dengan Algoritma Support Vector Machine

Berikut ini adalah hasil rancangan model klasifikasi, Hasil Eksperimen, Confusion Matrix dan Kurva ROC dengan Algoritma Support Vector Machine:

1) Model Klasifikasi

Berikut adalah rangkaian model pengujian pada RapidMiner dengan Algoritma Support Vector Machine:



Gambar 13. Rangkaian Model dengan Algoritma Support Vector Machine pada RapidMiner

2) Hasil Eksperimen

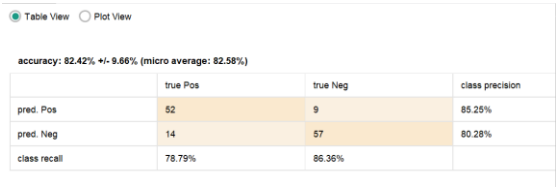
Setelah rangkaian model dijalankan dalam RapidMiner, hasil precision, recall dan akurasi dari pengujian akan didapatkan. Berikut adalah rincian hasil pengujian dengan algoritma Support Vector Machine:

Tabel 3. Tabel Rincian Hasil Eksperimen dengan Algortima Support Vector Machine

No.	Keterangan	Hasil
1	Precision	83,81%
2	Recall	85,95%
3	Akurasi	82,42%

3) *Confusion Matrix*

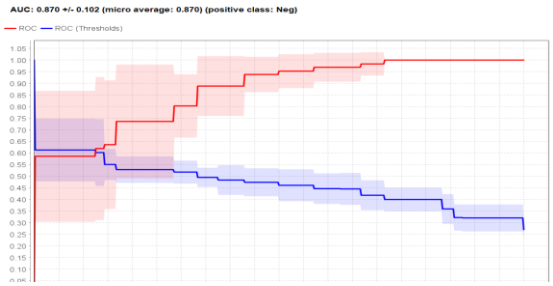
Berikut adalah *confusion matrix* yang dihasilkan dari pengujian data dengan algoritma Support Vector Machine:



Gambar 13. *Confusion Matrix* Hasil Eksperimen dengan Algoritma Support Vector Machine

4) Kurva ROC (*ROC Curves*)

Berikut adalah kurva ROC yang divisualisasikan berdasarkan perhitungan *confusion matrix*:



Gambar 14. Kurva ROC Pengujian Data dengan Algoritma Naive Bayes

Berdasarkan gambar diatas dapat diketahui hasil dari nilai AUC pada Kurva ROC adalah sebesar 0,870 yang menyatakan bahwa kategori nilai yang didapatkan adalah *Good Classification.* Dari kurva ROC dapat diketahui bahwa model klasifikasi yang digunakan dalam penelitian memiliki kinerja yang relatif baik dalam membedakan antara kelas positif dan kelas negatif. Untuk meningkatkan kinerja keseluruhan meningkatkan jumlah data training dengan teknik augmentasi data atau pengumpulan lebih banyak sampel data dapat dilakukan dengan menggabungkan beberapa ensemble seperti Voting, Bagging, atau Boosting.

KESIMPULAN

1. Algoritma Naive Bayes dan Support Vector Machine dapat memprediksi kelas dengan baik pada analisis pemutar musik online. Hal ini didukung dengan hasil nilai akurasi dari kedua metode yang didapatkan.
2. Nilai akurasi yang didapatkan dari penggunaan algoritma Naive Bayes adalah sebesar 84,73%. Sedangkan, hasil dari algoritma Support Vector Machine adalah sebesar 82,42%.
3. Dalam pemecahan kasus Analis Sentimen pada Pemutar Musik Online, dari kedua algoritma yang digunakan oleh peneliti dapat disimpulkan bahwa Algoritma Naive Bayes lebih unggul untuk mengklasifikasikan dataset hasil ulasan pengguna pemutar musik online yang dimiliki oleh penelti.

Saran

Untuk mendapatkan hasil terbaik pada penelitian kedepannya, dataset yang dikumpulkan juga harus diperbaharui berdasarkan tanggal terbaru. Hal itu dikarenakan akan semakin banyak kebutuhan-kebutuhan lain yang akan disampaikan pengguna melalui ulasan.

DAFTAR PUSTAKA

Alita, D., & Rahman, A. (2020). *Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest*

Classifier. *Jurnal Komputasi*, 50-58.

Asri, Y., Nita Suliyanti, W., Kuswardani, D., & Fajri, M. (2022). *Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam*. *PETIR:* Jurnal Pengkajian dan Penerapan Teknik Informatika, 275.

Darwis, D., Siskawati, N., & Abidin, Z. (2021). *Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional*. Jurnal TEKNO KOMPAK, 131-145.

Ferdiana, R., Jatmiko, F., Purwanti, D., Ayu, A., & Dicka, W. (2019). *Dataset Indonesia untuk Analisis Sentimen*. JNTETI, 334-339.

Fikri, M. I., Sabrila, T., & Azhar, Y. (2020). *Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter*. SMATIKA Jurnal, 71-76.

Gunawan, B., Pratiwi, H., & Pratama, E. (2018). *Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes.* JEPIN (Jurnal Edukasi dan Penelitian Informatika), 113-188.

Hadna, N. M., Winarno, W., & Santosa, P. (2016). *Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis.* Seminar Nasional Teknologi Informasi dan Komunikasi 2016 (SENTIKA 2016).

Himawan, R. D., & Eliyani. (2021). *Perbandingan Akurasi Analisis Sentimen Tweet terhadap Pemerintah Provinsi DKI Jakarta di Masa Pandemi.* JEPIN (Jurnal Edukasi dan Penelitian Informatika), 58-63.

Iswandi. (2015). *Refleksi Psikologi Musik Dalam Perilaku Masyarakat Sehari-Hari.* Humanus, 152-157.

Mutiara, E. (2020). *Algoritma Klasifikasi Naive Bayes Berbasis Particle Swarm Optimization Untuk Prediksi Penyakit Tuberculosis (TB)*. JURNAL SWABUMI, 46-58.

Pintoko, B. M., & L., K. (2018). *Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier*. e-Proceeding of Engineering , 8121-8130.

Riadi, I., Umar, R., & Aini, F. (2019). *Analisis Perbandingan Detection Traffic Anomaly Dengan Metode Naive Bayes Dan Support Vector Machine (Svm)*. ILKOM Jurnal Ilmiah, 17-24.

Siringoringo, R. (2018). *Klasifikasi Data Tidak Seimbang Menggunakan Algoritma Smote Dan K-Nearest Neighbor*. Jurnal ISD, 44-49.

Utami, L. A. (2017). *Analisis Sentimen Opini Publik Berita Kebakaran Hutan Melalui Komparasi Algoritma Support Vectormachinedan K-Nearest Neighborberbasisparticle Swarm Optimization*. Jurnal Pilar Nusa Mandiri, 103-112.

Yulita, W., Nugroho, E., & Algifari, M. (2021). *Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier*. JDMSI, 1-9.