Analisis Sentimen Menggunakan Metode *Naive Bayes* dan *Support Vector Machine* pada Ulasan Aplikasi Spotify

1st Muhammad Rifqi Fauzi Ramdhani
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
muhammadrifqifr@student.telkomuniversity.ac.id

2nd Kemas Muslim Lhaksmana
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
kemasmuslim@telkomuniversity.ac.id

Abstrak—Pergeseran kebiasaan memutar lagu secara digital didukung oleh kemudahan akses yang tersedia di berbagai perangkat, membuat pengguna bisa mendegarkan lagu kapanpun dan dimanapun waktunya. Spotify merupakan platform nomor satu sebagai penyedia jasa musik dan audio gratis dengan hampir 422 juta pengguna aktif dan menguasai 31% pangsa pasar skala global. Dengan banyaknya unduhan yang sudah mencapai satu juta kali, Spotify mendapatkan nilai rating 4.4 dan ulasan oleh para penggunanya. Pengguna diberikan kebebasan untuk mengekspresikan hasil kepuasaan, kritik, dan saran terhadap aplikasi. Ulasan tersebut bisa digunakan sebagai umpan balik untuk perusaahan dalam meningkatkan layanan dan mengembangkan inovasi selanjutnya. Analisis sentimen diperlukan untuk mengolah ulasan menjadi informasi yang bermanfaat dengan melalui beberapa tahapan pembersihan data terlebih dulu. Pembobotan menggunakan TF-IDF dilakukan sebelum masuk kedalam proses klasifikasi menggunakan *Naive Bayes* dan *Support Vector Machine*. Nilai *F1-Score* terbaik didapatkan pada metode SVM kernel RBF dengan nilai C & gamma optimum menghasilkan nilai *F1-Score* tertinggi sebesar 81% pada dataset ulasan aplikasi Spotify di layanan GooglePlay Store.

Kata kunci—*naive bayes, support vector machine,* spotify, analisis sentimen, ulasan

Abstract—The shifting of playing songs digitally is supported by the ease of access on various devices, allowing a user to listen anytime and anywhere. Spotify is the number one platform of free music and audio services with nearly 422 million active users and has a 31% global market share of audio music platforms. With the number of downloads that have reached one million times, Spotify has received ratings and reviews by its users. Users are giving the freedom to express satisfaction, criticism, and suggestions for the application. These reviews can be more useful as a feedback for the company to improve services and develop for further innovations. Sentiment analysis is needed to process a review into helpful information with several stages of data cleaning. The weighting using TF-IDF has done before entering classification process using Naive Bayes and Support Vector Machine. The highest F1-score value using SVM kernel RBF with C and gamma optimum produce a F1-Score value 84% on the Spotify application review dataset in the GooglePlay Store.

Keywords— naive bayes. support vector machine, spotify, sentiment analysis, review

I. PENDAHULUAN

A. Latar Belakang

Pergeseran kebiasaan memutar lagu secara digital menjadikan banyaknya platform audio yang bermunculan menggantikan penjualan album fisik seperti CD dan piringan hitam. Kemudahan akses yang tersedia di berbagai perangkat seperti gawai dan komputer membuat pengguna bisa mendegarkan musik kapanpun dan dimanapun waktunya. Beragam fitur yang ditawarkan seperti pengguna premium yang terbebas dari iklan komersial bahkan konten yang bisa diunduh kedalam perangkat secara gratis. Setiap negara pun memiliki karakter demografi penggunanya masing-masing berdasarkan musisi asal dan lagu yang sedang populer [1]. Di Indonesia, berdasarkan survei penetrasi dan perilaku penggunaan internet tahun 2022 yang dilakukan oleh APJII (Asosiasi Penyedia Jasa Internet Indonesia), pengguna internet di dominasi oleh kelompok usia 19-34 tahun sebesar 98,64% dari total pengguna internet di Indonesia yaitu 210 juta orang. Mereka mengakses platform *streaming* musik sebesar 38,51% dengan menggunakan perangkat gawai sebesar 88,2% dibandingkan mengakses aplikasi tv berbasis internet sebesar 11,10% [3].

Industri *streaming* musik global di Q2 tahun 2021 telah berkembang pesat sebesar 26% dibandingkan Q2 2020 berdasarkan penelitian yang dilakukan oleh (Midia,2021). Platform *streaming* musik seperti Spotify, Apple Music, dan Youtube Music mengalami peningkatan yang besar dari segi layanan dan sejumlah inovasi yang menarik. Pada Q2 2021 Spotify menguasai 31% pangsa pasar industri streaming musik skala global [2]. Spotify telah menjadi platform nomor satu sebagai penyedia jasa musik dan audio gratis dengan hampir 422 juta pengguna aktif setiap bulannya di seluruh dunia. Termasuk 182 juta penggunanya adalah pengguna

berlangganan akun premium. Dengan banyaknya unduhan yang sudah mencapai angka satu juta kali, Spotify mendapatkan penilaian 4.4 dari skala 5 dari 26 juta ulasan oleh para pengguna di GooglePlay. Tren memberikan ulasan terhadap salah satu produk adalah salah satu kebiasaan pengguna di era sosial media dan belanja daring sekarang. Pengguna diberikan kebebasan untuk mengekspresikan hasil kepuasaan, kritik, dan saran terhadap aplikasi yang digunakan. Ulasan bisa berbentuk sebuah kalimat, suka tidak suka, dan penilaian skala lima bintang. Data ulasan tersebut berguna untuk perusahaan sebagai evaluasi terhadap fitur produk yang telah diluncurkan[4][5]. Analisis sentimen diperlukan untuk mengolah data ulasan menjadi sebuah informasi umpan balik yang berguna.

Pada penelitian ini akan dilakukan analisis sentimen terhadap ulasan aplikasi Spotify. Data yang dikumpulkan adalah kalimat ulasan dan rating skala lima bintang aplikasi Spotify pada layanan GooglePlay Store di tahun 2022 . Kalimat ulasan di analisis melalui beberapa tahapan pembersihan data, seperti pelabelan, pembobotan, dan klasifikasi. Klasifikasi yang digunakan dalam penelitian ini adalah *Naive Bayes* (NB) & *Support Vector Machine* (SVM). Banyak penelitian tentang sentimen analisis yang menggunakan klasifikasi NB dikarenakan memiliki beberapa kelebihan seperti sederhana, cepat dan memiliki akurasi tinggi [4]. SVM diaplikasikan pada kasus klasifikasi teks dan efisien dalam menangani jumlah data yang besar [5]. Hasil akhir dari penelitian ini adalah melihat perbandingan nilai *F1-Score* terbaik untuk kasus analisis sentimen pada ulasan aplikasi Spotify.

B. Perumusan Masalah

Berdasarkan latar belakang diatas, rumusan masalah yang di angkat pada penelitian ini adalah untuk mendapatkan model klasifikasi menggunakan algoritma *Naive Bayes* dan *Support Vector Machine* pada ulasan aplikasi Spotify dan membandingkan hasil nilai *F1-Score*.

C. Tujuan

Merujuk pada perumusan masalah yang di angkat pada penelitian ini, maka tujuan pada tugas akhir ini adalah untuk menganalisis dan mengimplementasikan algoritma *Naive Bayes* dan *Support Vector Machine* pada dataset ulasan aplikasi Spotify dan melihat perbandingan hasil nilai *F1-Score*.

D. Batasan Masalah

Batasan masalah pada penelitian ini adalah :

1. Data yang dikumpulkan adalah kalimat ulasan dan rating bintang satu sampai bintang lima aplikasi Spotify pada layanan GooglePlay Store dari tanggal 1 Januari sampai dengan 9 Juli 2022.
2. Pelabelan data rating di asumsikan menjadi; rating 5,4,3 menjadi positif dan rating 2,1 menjadi negatif.
3. Kalimat ulasan pada data menggunakan Bahasa Inggris.
4. Nilai *F1-Score* digunakan sebagai metrik yang lebih baik dibandingkan nilai akurasi untuk data yang tidak seimbang.

II. KAJIAN TEORI

A. Studi Literatur

Beberapa penelitian yang sudah dilakukan mengenai analisis sentimen adalah penelitian [6] analisis sentimen terhadap opini masyarakat Jakarta mengenai salah satu penyedia aplikasi dompet digital. Pada penelitan tersebut metode yang digunakan adalah *Naive Bayes Classifier* dan *feature selection Particle Swarm Optimization* (PSO). Tahapan *preprocessing* yang dilakukannya dimulai dari tokenisasi, *stemming*, menghilangkan *stopword* dan *transform case*. Hasil akurasi klasifikasi yang didapatkan menggunakan *feature selction* (PSO) sebesar 83.60%, lebih baik dengan model yang tidak menggunakan PSO.

Penelitian [7] mengenai penerapan algortima *Suppot Vector Machine* (SVM) pada analisis sentimen data twitter berbahasa Indonesia yang mengklasifikasikan menjadi sentimen positif, netral, dan negatif. Pembobotan yang dipakai menggunakan (*Term Frequency – Inverse Document Frequency*) TF-IDF. Hasil dari klasifikasi menggunakan metode SVM. Dari hasil tersebut terdapat kesimpulan data cuitan twitter bersentimen negatif sebanyak 77% dengan akurasi sebesar 82%.

Analisis sentimen berbasis aspek pada ulasan *Female Daily* salah satu forum yang membahas produk kecantikan yang berbahasa *Multilangual* telah dilakukan oleh Clarisa [8] menggunakan pembobotan TF-IDF dan algortima *Naive Bayes*. Dengan menggunakan 4 aspek label dengan kelas positif, netral, dan negatif. Dari keempat aspek, tiga diantara distribusi datanya tidak seimbang sehingga performansi menjadi rendah dibandingkan satu aspek sisanya. Dari skenario yang dilakukan, performansi tertinggi didapat pada tahapan *preprocessing* data yang tidak menggunakan *stopword removal*, karena kata-kata yang dapat menentukan sentimen tidak dihapus dan data yang diterjemahkan ke dalam Bahasa Inggris lalu kemudian diterjemahkan kembali ke Bahasa Indonesia mendapatkan nilai *F1-Score* 62,81%.

Penilitian lain menggunakan metode TF-IDF dan Naive Bayes [9] pada produk *game* di *e-commerce* melalui tahapan *preprocessing* yaitu *Case folding, Tokenizing, Stopwords Removal,* dan *Stemming*. Jumlah data yang dikumpulkan sebanyak 1000 ulasan produk *game* dibagi menjadi tiga label yaitu positif, netral, dan negatif menghasilkan nilai akurasi sebesar 80,23%.

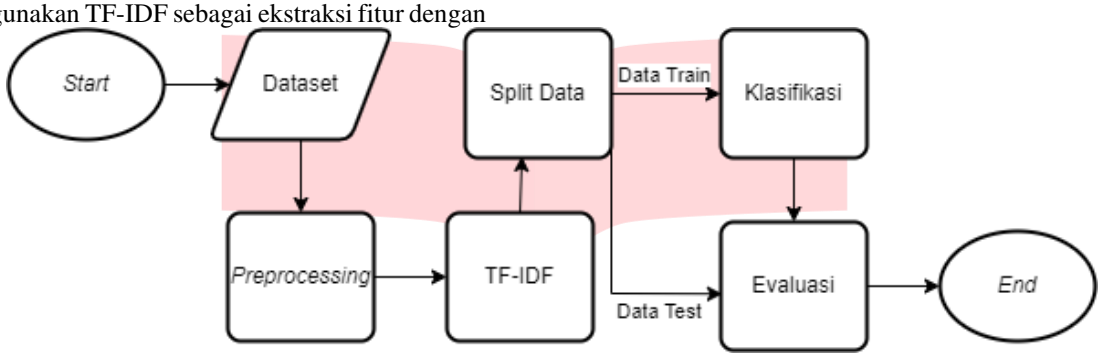
Susanti [10] melalukan analisis sentimen pada 8.925 ulasan *provider* by.U di layanan GooglePlay Store dengan label positif dan negatif menggunakan pembobotan TF-IDF. Tahapan dimulai dari mengambil data ulasan dengan cara *scraping*, pemberian label, *preprocessing* data termasuk *stopword removal* dan *negation handling*. Pengujian dilakukan dengan membagi dataset menjadi 80:20 menggunakan proses klasifikasi SVM dan 5-*Fold validation*. Hasil klasifikasi sentimen menggunakan ekstraksi fitur TF-IDF+SVM menghasilkan rata-rata nilai akurasi 84,7% dan nilai akurasi tertinggi pada nilai *fold* 2 sebesar 86,1%.

Pada penilitian tugas akhir ini, penulis menggunakan TF-IDF sebagai ekstraksi fitur dengan melalui beberapa tahapan *preprocessing* untuk persiapan data. Metode NB dan SVM digunakan sebagai metode klasifikasi sentimen pada dataset ulasan aplikasi Spotify. Hasil akurasi dari kedua metode tersebut akan dibandingkan untuk dilihat nilai akurasi terbaiknya.

III. METODE

A. Gambaran Umum Sistem

Pada bagian ini menjelaskan perancangan sistem yang akan digunakan dalam penelitian ini. Gambaran umum pada sistem sistem klasifikasi teks yang akan dibuat dari dataset ulasan aplikasi Spotify adalah sebagai berikut :

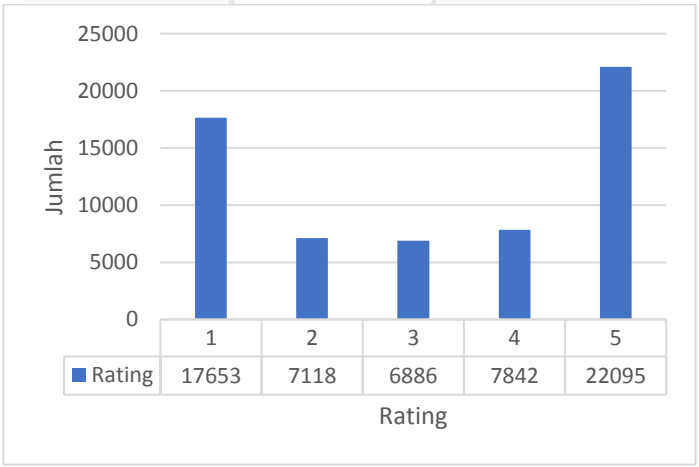


GAMBAR 3.1
DIAGRAM ALIR SISTEM YANG DIBANGUN

Setelah proses *preprocessing* dilakukan, data akan melalui tahap pembobotan atau *feature extraction* menggunakan TF-IDF. Lalu selanjutnya dataset dibagi menjadi 80% data *train* dan 20% data *test*. Model klasifikasi menggunakan metode *Naive Bayes dan Support Vector Machine.* Setelah didapatkan hasil klasifikasi dari metode NB dan SVM, selanjutnya kedua metode tersebut di evaluasi sehingga menghasilkan nilai *F1-Score* dari masing-masing metode.

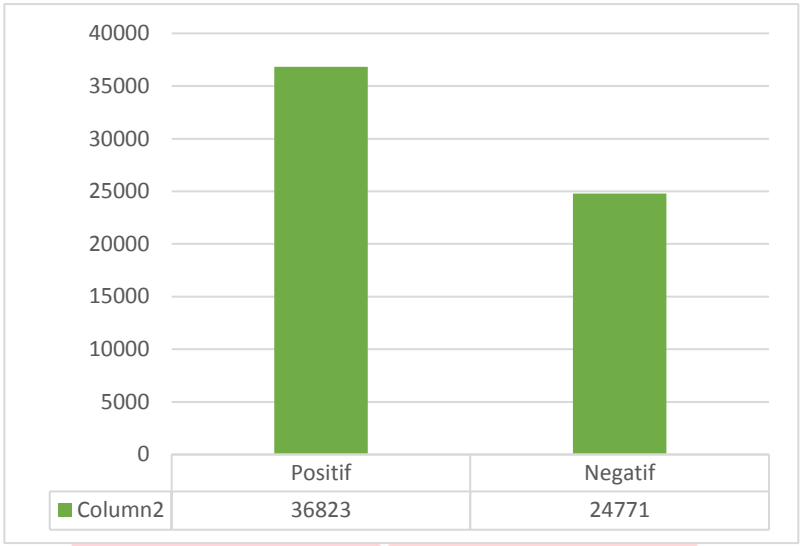
B. Dataset

Dataset yang digunakan berjumlah 61.356 baris dari tanggal 1 Januari 2022 sampai 9 Juli 2022. Dataset bersumber dari kaggle.com.



GAMBAR 3.2
GRAFIK PERBANDINGAN NILAI RATING PADA ULASAN APLIKASI SPOTIFY

Pelabelan data dibuat menjadi dua kelas yaitu kelas positif dan negatif yang berasal dari nilai *rating*. Dengan asumsi rating 5,4, dan 3 menjadi kelas positif, 2 dan 1 menjadi kelas negatif.



GAMBAR 3.3
JUMLAH PERBANDINGAN LABEL PADA ULASAN APLIKASI SPOTIFY

C. *Preprocessing*

Tahap data *preprocessing* dilakukan untuk mengurangi *noise* sebelum data diolah lebih lanjut untuk di ekstraksi sampai tahap klasifikasi. Tahap *preprocessing* yang dilakukan dalam penelitian ini adalah sebagi berikut:

TABEL 3.1
KALIMAT ULASAN SEBELUM DILAKUKAN PREPROCESSING

"The sound quality is really good and I can listen to my favo urite songs anytime and anywhere without any disturbance. That's really good experience 👍 I loved this app alot. Keep p roviding us the best service 👍👍"

1. *Case Folding* dan *Cleaning* data
   Tahap merubah semua huruf pada corpus menjadi huruf kecil. Hanya huruf a-z yang diterima, selain itu dianggap delimiter. Tahap *cleaning data* dilanjutkan dengan menghilangkan simbol khusus seperti emoji, tanda baca, dan angka.

TABEL 3.2
TAHAP CASE FOLDING DAN CLEANING DATA

"the sound quality is really good and i can listen to my favourite songs anytime and anywhere without a ny disturbance that s really good experience i loved t his app alot keep providing us the best service "

2. Tokenisasi
   Tahap ini merupakan proses pemisahan teks dalam dokumen menjadi potongan-potongan token untuk dianalisa. Kata, angka, simbol, tanda baca adalah termasuk token. Sintaks fungsi tokenisasi kata dalam bahasa pemrograman python adalah *word_tokenize()*.

Tabel 3.3 Tahap Tokenisasi

['the','sound', 'quality','is','really', 'good', 'and', 'i', ' can', 'listen', 'to', 'my', 'favourite', 'songs', 'anytime', 'and', 'anywhere', 'without', 'any', 'disturbance', 'tha t', 's','really', 'good', 'experience', 'i', 'loved', 'this', ' app', 'alot', 'keep', 'providing', 'us', 'the', 'best', 'serv ice']

3. *Stopword Removal*
   Tahap mengambil kata-kata penting dari hasil token dengan menghapus kata dengan frekuensi tinggi namun tidak memiliki arti menggunakan *stopword english dictionary* seperti yang disajikan pada Tabel 3.4.

TABEL 3.4
CONTOH KATA STOPWORD DALAM BAHASA INGGRIS

'i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", 'your', 'y ours', 'yourself', 'yourselves', 'he', 'him', 'his', 'himse lf', 'she', "she's", 'her', 'hers', 'herself', 'it', "it's", 'its ', 'itself', 'they', 'them', 'their', 'theirs', 'themselves', ' what', 'which', 'who', 'whom', 'this', 'that', "that'll", ' these', 'those', 'am', 'is', 'are', 'was', 'were', 'be', 'bee n', 'being', 'have', 'has', 'had', 'having', 'do', 'does', ' did', 'doing'.

TABEL 3.5
KALIMAT ULASAN SETELAH PROSES STOPWORD REMOVAL

"sound quality really good listen favourite songs a nytime anywhere without disturbance really good e xperience loved app alot keep providing us best ser vice"

4. *Stemming*
   Tahap menghilangkan infleksi kata ke bentuk dasarnya, namun bentuk dasar tersebut tidak berarti sama dengan akar kata. Contoh *stemming* dalam Bahasa Inggris adalah *"listening", "listened"* -> *"listen"*.

Proses *stemming* antara satu bahasa dengan bahasa lainnya berbeda. Jika dalam teks Bahasa Inggris prosesnya hanya menghilangkan sufiks [11]. Sufiks di akhir kata bisa mengubah arti dari kata tersebut. Contoh sufiks *-ness* makna imbuhannya berarti suatu kondisi yang sedang dialami, sehingga kata yang jadinya menjadi *happiness, sadness, sickness*.

TABEL 3.6
KALIMAT ULASAN SETELAH PROSES STEMMING

'sound qualiti realli good listen favourit song anyti m anywher without disturb realli good experi love a pp alot keep provid best servic '

D. TF-IDF (*Term Frequency – Inverse Document Frequency*)

Pembobotan kata adalah proses menambahkan nilai pada seluruh kata dalam kalimat ulasan yang telah melalui tahap *preprocessing*. Permbobotan menggunakan metode TF-IDF dituliskan dalam persamaan (1) dan (2) lalu hasilnya akan dilanjutkan pada tahap klasifikasi [9][12]. Langkah menghitung pembobotan kata dimulai dari menghitung frekuensi kemunculan kata keluar dalam kalimat (TF). Lalu menghitung angka frekuensi kata (DF) pada dokumen dan menghitung nilai invers (IDF).

Setelah mendapatkan nilai TF, tahap selanjutnya menghitung nilai invers. Persamaan menghitung nilai IDF adalah sebagai berikut :

$$IDF(w) = \log\left(\frac{N}{DF(w)}\right) \quad (1)$$

Tahap selanjutnya dalam TF-IDF adalah dengan mengkalikan nilai TF dengan IDF seperti persamaan (2).

$$W_{ij} = \mathrm{tf}_{ij} \log\left(\frac{\mathrm{D}}{\mathrm{df_j}}\right) \quad (2)$$

tf_{ij} = frekuensi sebuah kata i dalam dokumen j.

D = jumlah dokumen dalam dataset ulasan.

$\mathrm{df_j}$= jumlah dokumen dalam dataset ulasan yang mengandung nilai i.

E. Klasifikasi

1. *Naive Bayes*

Metode *Naive Bayes* merupakan metode *supervised learning* yang ditemukan oleh Thomas Bayes pada abad ke-18 [13]. Klasifikasi Naive Bayes digunakan untuk memprediksi probabilitas keanggotan dalam satu kelas. Kelebihan algoritma Naive Bayes yaitu mudah diimplementasikan, efisiensi waktu, dan dapat menangani data yang besar [14]. Salah satu model Naive Bayes yang sering digunakan untuk klasifikasi teks adalah *Multinomial Naive Bayes*. Perumusan *Multinomial Naive Bayes* menggunakan pembobotan TF-IDF dituliskan pada persamaan (3) [15].

$$\hat{P}(t|c) = \frac{W_{ct}+1}{(\sum_{w\prime\in V} W_{ct\prime})+B\prime} \quad (3)$$

Keterangan :

$\hat{P}(t|c)$: Probabilitas bersyarat term t di dokumen pada kelas c

W_{ct} : Bobot TF-IDF term t di dokumen dengan kategori c

$\sum_{w\prime\in V} W_{ct\prime}$: Jumlah bobot TF-IDF seluruh term pada kelas c

B' : Jumlah IDF seluruh term pada dokumen

2. *Support Vector Machine*

Support Vector Machine (SVM) adalah salah satu algoritma *machine learning* yang biasa digunakan dalam proses klasifikasi. Algoritma SVM bertujuan untuk mencari *hyperlane* terbaik untuk memisahkan kedalam kelas positif dan negatif, dan memaksimalkan margin diantara dua kelas tersebut [16]. *Hyperlane* adalah sebuah fungsi yang digunakan untuk pemisah antar kelas. Proses klasifikasi pada penelitian ini menggunakan bahasa pemrograman python dan *library* scikit learn svm. Fungsi kernel yang umum digunakan pada SVM adalah kerner linear, polynomial, dan RBF.

F. Evaluasi

Tahap terakhir dalam proses analisis sentimen pada penelitian ini adalah pengujian menggunakan *confussion matrix* untuk mendapatkan nilai *F1-Score* dari hasil klasifikasi algoritma NB dan SVM.

TABEL 3.7
CONTOH CONFUSSION MATRIX

	Actual Value		
Predicted Value	Label	*Positive*	*Negative*
	Positive	*TP*	*FP*

	Negative	*FN*	*TN*

Keterangan :
True Positive (TP) : Jika klasifikasi diprediksi positif dan aktualnya benar positif
True Negative (TN) : Jika klasifikasi diprediksi negatif dan aktualnya benar negatif
False Positive (FP) : Jika klasifikasi diprediksi positif dan aktualnya salah
False Negative (FN) : Jika klasifikasi diprediksi negatif dan aktualnya salah
Actual Value : Nilai klasifikasi sebenarnya yang terdiri dari label positif dan negatif
Predicted Value : Nilai prediksi hasil dari pemodelan *machine learning*

Dalam *confussion matrix* terdapat empat pengukuran yang digunakan yaitu Akurasi, Presisi, *Recall*, dan *F1-Score*. Akurasi menggambarkan akurasi model dalam mengklasifikasikan dengan benar. Presisi menggambarkan tingkat akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model. Presisi merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. *Recall* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar. *Nilai F1-Score* merupakan perbandingan rata-rata presisi dan recall yang dibobotkan. Penulisan rumus akurasi, presisi, recall dan *F1-Score* dapat disajikan pada persamaan (4),(5),(6),(7).

$$Akurasi = \frac{(TP+TN)}{(TP+FP+FN+TN)} \quad (4)$$

$$Presisi = \frac{TP}{(TP+FP)} \quad (5)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (6)$$

$$F1\text{-}Score = \frac{2*(Recall*Presisi)}{(Recall+Presisi)} \quad (7)$$

IV. HASIL DAN PEMBAHASAN

A. Hasil Pengujian

Pengujian dilakukan dengan membagi dataset menjadi 80% data *train* & 20% data *test* yang telah melalui tahapan pembobotan dan data *preprocessing*. Skenario pengujian pertama mendapatkan nilai *F1-Score* metode NB dan tidak menambahkan nilai parameter pada kernel linear, polynomial dan RBF pada metode SVM. Skenario pengujian kedua menambahkan nilai parameter C, *degree*, dan *gamma* pada kernel linear, polynomial, dan rbf (*gaussian radial basis function*) pada metode SVM.

1. Skenario Pengujian Pertama
   Pada skenario pengujian pertama di metod e SVM dilakukan tanpa menggunakan param eter pada kernel linear, polynomial & RBF.

TABEL 7
CONFUSSION MATRIX SVM TANPA PARAMETER & NB

	Kernel		
F1-Score SVM	Linear	Polynomial	RBF
	Akurasi tanpa parameter		
	80%	77%	81%
F1-Score NB	Multinomial NB		
	76%		

Berdasarkan tabel 3.7 untuk *confussion matrix* nilai dimasukkan kedalam persamaan (4) untuk metode *multinomial naive bayes* sehingga mendapatkan nilai *F1-Score* 76%. Nilai *F1-Score* digunakan sebagai metrik untuk data yang tidak seimbang. Berdasarkan pengujian pertama, dapat dilihat nilai *F1-Score* pada metode SVM tanpa menggunakan paramater pada kernel tersaji pada tabel 7 untuk selanjutnya dibandingkan dengan hasil pengujian kedua yang menambahkan nilai parameter pada kernel SVM.

2. Skenario Pengujian Kedua
   a. Kernel Linear

TABEL 8
NILAI KERNEL LINEAR

Nilai C	Akurasi
0.1	80%
1	**80%**
10	78%

Berdasarkan tabel 8 kernel linear menggunakan nilai C sebagai pengoptimalan metode SVM untuk menghindari misklasifikasi di setiap sampel dalam data *training*. Nilai C yang tinggi, kemungkinan terjadinya kesalahan akan semakin kecil. Jika Nilai C rendah, semaking tinggi proporsi kesalahan yang terjadi pada penentuan solusi semakin besar[17]. Nilai C optimum pada kernel linear adalah C = 1 dengan nilai *F1-Score* 80%.

b. Kernel Polynomial

TABEL 9
NILAI KERNEL POLYNOMIAL

d	gamma		
	0.1	1	10

1	80%	**80%**	78%
2	2%	80%	76%
3	0%	71%	69%

Pada kernel polynomial memiliki parameter derajat (d) yang berfungsi untuk mencari nilai optimal pada setiap data. Semakin besar nilai d maka akurasi sistem menjadi tidak stabil. Semakin tinggi nilai parameter d maka semakin melengkung garis *hyperlane* yang dihasilkan. Nilai gamma menentukan seberapa jauh pengaruh dari satu sampel dataset. Jika gamma bernilai rendah maka titik yang jauh dipertimbangkan, dan jika gamma bernilai tinggi maka titik yang dekat akan dipertimbangkan untuk menentukan batas keputusan [17]. Berdasarkan tabel 9 parameter d = 1 dan gamma = 1 merupakan hasil yang optimum pada kernel polynomial dengan nilai *F1-Score* 80%.

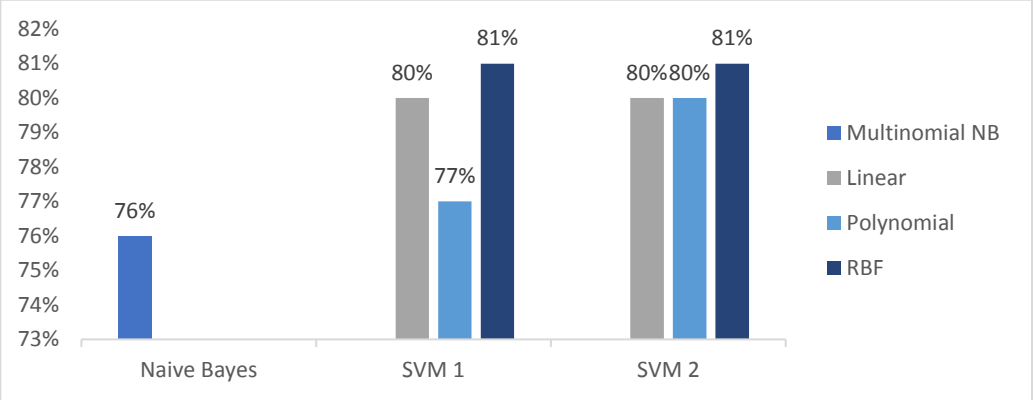
c. Kernel RBF

TABEL 10
NILAI KERNEL RBF

C	gamma		
	0.1	1	10
0.1	77%	79%	0%
1	80%	**81%**	1%
10	80%	79%	1%

Berdasarkan pengujian yang dilakukan pada kernel RBF terdapat nilai C dan gamma. Berdasarkan tabel 10 nilai *F1-Score* terbaik sebesar 81% dengan nilai C yang optimum = 1 dan gamma optimum = 1.

B. Analisis Hasil Pengujian



GAMBAR 4
HASIL F1-SCORE DARI PENGUJIAN

Berdasarkan hasil dua kali skenario pengujian nilai *F1-Score* dipilih karena pada gambar 3.3 pelabelan data tidak seimbang antara data positif & negatif. Nilai *F1-Score* menggunakan *multinomial* NB pada pengujian 1 sebesar 76%. Fungsi kernel linear terbaik pada metode SVM di pengujian kedua dengan nilai parameter C optimum = 1. Sedangkan pada kernel polynomial pengujian kedua penambahan nilai d optimum = 1 mempengaruhi nilai *F1-Score* dibandingkan dengan pada pengujian pertama tanpa nilai d. Pada fungsi kernel RBF nilai *F1-Score* paling tinggi didapatkan sebesar 81% menggunakan C optimum = 1 & gamma optimum = 1.

V. KESIMPULAN

A. Kesimpulan

Berdasarkan hasil pengujian yang dilakukan nilai *F1-Score* metode SVM dengan menambahkan nilai parameter dan fungsi kernel memiliki nilai yang paling tinggi pada fungsi kernel RBF sebesar 81% dengan nilai C = 1 dan gamma =1 dibandingkan kernel lain. Metode SVM menghasilkan nilai *F1-Score* yang lebih baik dibandingkan NB sebesar 76% pada penelitan analisis sentimen pada ulasan aplikasi Spotify.

B. Saran

Pada penelitian selanjutnya pencarian nilai parameter C, d, dan gamma yang optimum akan mempengaruhi nilai *F1-Score* pada fungsi kernel SVM. Tahap evaluasi seperti *cross validation*¸ fitur ekstraksi dan metode klasifikasi lain bisa digunakan untuk mendapatkan nilai *F1-Score* yang lebih baik.

REFERENSI

[1] Way, S. F., Garcia-Gatright, J., & Cramer, H. 2020. Local Trends in Global Music Streaming. Proceedings of the International AAAI Conference on Web and Social Media, vol.14, pp. 705-714.

[2] APJII (Asosiasi Penyedia Jasa Internet Indonesia). 2022. Profil Internet Indonesia 2022. Survei.

[3] Stephen, Global Streaming music subscription market Q2 2021 MIDiA Research, [Online]. https://musicindustryblog.wordpress.com/2022/01/18/music-subscriber-market-shares-q2-2021/ [Diakases 25 April 2022].

[4] Putri, D. A. 2020. Comparison of Naive Bayes Algorithm and Support Vector Machine using PSO Feature Selection for Sentiment Analysis on E-Wallet Review. Journal of Physics: Conference Series, vol. 1641, No. 1, p.012085.
[5] Basari, A. S. H., Hussin, B., Ananta, I. G. P., & Zeniarja, J. 2013. Opinion Mining of Movie Review Using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. *Procedia Engineering*, *53*, 453-462.
[6] Saputra, S. A., Rosiyadi, D., Gata, W., & Husain, S. M. 2019. Analisis sentimen E-Wallet pada google play menggunakan algoritma naive bayes berbasis particle swarm optimization. *Jurnal RESTI*, *3*, 377-382.
[7] Darwis, D., Pratiwi, E. S., & Pasaribu, A. F. O. 2020. Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia. *Jurnal Ilmiah Edutic: Pendidikan dan Informatika*, *7*(1), 1-11.
[8] Yutika, C. H., Adiwijaya, A., & Al Faraby, S. 2021. Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *5*(2), 422-430.
[9] Kosasih, R., & Alberto, A. (2021). Sentiment analysis of game product on shopee using the TF-IDF method and naive bayes classifier. *LKOM Jurnal Ilmiah*, *13*(2), 101-109.
[10] Fransiska, S., Rianto, R., & Gufroni, A. I. (2020). Sentiment Analysis Provider by. U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method. *Scientific Journal of Informatics*, *7*(2), 203-212.
[11] Adriani, M., Asian, J., Nazief, B., Tahaghoghi, S. M. M., & Williams, H. E. (2007). Stemming Indonesian. ACM Transactions on Asian Language Information Processing, 6(4), 1–33.
[12] Ahuja, R., Chug, A., Kohli, S., Gupta, S., & Ahuja, P. (2019). The impact of features extraction on the sentiment analysis. *Procedia Computer Science*, *152*, 341-348.
[13] Suyanto. 2017. *Data mining* Untuk Klasifikasi Dan Klasterisasi Data.Bandung:Informatika Bandung.
[14] Wibawa, A. P., Kurniawan, A. C., Della Murbarani Prawidya Murti, Adiperkasa, R. P., Putra, S. M., Kurniawan, S. A., & Nugraha, Y. R. (2019). Naïve Bayes Classifier for Journal Quartile Classification. *Int. J. Recent Contributions Eng. Sci. IT*, *7*(2), 91-99.
[15] Sabrani, A., & Bimantoro, F. (2020). Multinomial Naïve Bayes untuk Klasifikasi Artikel Online tentang Gempa di Indonesia. *Jurnal Teknologi Informasi, Komputer, Dan Aplikasinya (JTIKA)*, *2*(1), 89-100.
[16] Pratama, A., Wihandika, R. C., & Ratnawati, D. E. (2018). Implementasi algoritme support vector machine (SVM) untuk prediksi ketepatan waktu kelulusan mahasiswa. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer E-ISSN*, *2548*, 1704-1708.
[17] Trivusi, Apa itu Kernel Trick? Pengertian dan Jenis-jenis Fungsi Kernel SVM, [Online]. https://www.trivusi.web.id/2022/04/fungsi-kernel-svm.html [Diakses 8 Agustus 2022]