

# **Analisis Sentimen Ulasan Pengguna Aplikasi Netflix Pada Google Play Menggunakan Algoritma Naïve Bayes**

**Ananda Bagas Pranata, Allif Rizki Abdillah, Faldy Irwiensyah\***

Fakultas Teknologi Industri dan Informatika, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia

Email: <sup>1</sup>bagaspranta.bp@gmail.com, <sup>2</sup>allifrizki02@gmail.com, <sup>3</sup>\*faldy@uhamka.ac.id

Email Penulis Korespondensi: faldy@uhamka.ac.id

**Abstrak**—Perkembangan teknologi informasi sangat berkembang dengan pesat, termasuk perkembangan teknologi dalam kategori perfilman. Pada era yang canggih saat ini, menonton film tidak perlu lagi untuk datang ke bioskop, karena sudah ada aplikasi untuk menonton film dimanapun dan kapanpun. Salah satu aplikasi yang populer untuk menonton film bernama netflix, netflix merupakan platform streaming film atau serial yang banyak digunakan. Netflix juga merupakan platform streaming yang menempati peringkat ke-10 dalam hal akses di Indonesia. Penelitian ini berfokus untuk mengidentifikasi tingkat kepuasan pengguna terhadap aplikasi netflix berdasarkan ulasan di google play store. Adanya penelitian ini ditujukan untuk menganalisis sentimen ulasan pengguna suatu aplikasi yang ada di google play yaitu netflix. Ulasan tersebut akan peneliti gunakan untuk mengetahui seberapa puas pengguna dalam menggunakan aplikasi netflix. Peneliti mendapatkan ulasan tersebut dengan menggunakan web scrapper python dengan jumlah 1000 data yang belum peneliti olah. Setelah peneliti mengolah 1000 data tersebut dari duplikat dan simbol-simbol, peneliti memperoleh 893 data yang siap untuk dilakukan proses sentimen analisis menggunakan RapidMiner. Dari 893 data, peneliti melakukan pelabelan manual pada data sebanyak 635 data dan 258 data dilabeli dengan otomatis menggunakan machine learning yaitu naive bayes. Peneliti juga membuat confusion matrix untuk mengetahui tingkat keakuratan algoritma yang peneliti gunakan pada penelitian kali ini. Hasil keakuratan confusion matrix yang peneliti dapat pada penelitian ini sebesar 93,39%. Nilai class precision positif sebesar 85,52% menunjukkan bahwa sebagian besar ulasan positif teridentifikasi dengan baik, sementara nilai class precision negatif sebesar 100% menunjukkan kemampuan yang sangat baik dalam mengidentifikasi ulasan negatif. Kesimpulannya, aplikasi Netflix mendapatkan respon yang beragam dari pengguna, dan algoritma yang digunakan mampu mengidentifikasi ulasan dengan akurat.

**Kata Kunci:** Analisis Sentimen; Naive Bayes; Netflix; RapidMiner; Google Play Store

**Abstract**—The rapid development of information technology has advanced rapidly, including advancements in film technology. In this modern era, watching movies no longer requires going to the cinema, as there are applications available to watch movies anytime and anywhere. One popular application for watching movies is Netflix, a widely used streaming platform for films and series. Netflix also ranks 10th in terms of access in Indonesia. This study focuses on identifying user satisfaction levels with the Netflix application based on reviews on the Google Play Store. The research aims to analyze user review sentiment of an application available on Google Play, namely Netflix. These reviews will be used to gauge user satisfaction with the Netflix application. Researchers obtained these reviews using a Python web scraper with a total of 1000 unprocessed data points. After processing these 1000 data points by removing duplicates and symbols, researchers obtained 893 data points ready for sentiment analysis using RapidMiner. Out of the 893 data points, researchers manually labeled 635 data points, while 258 data points were labeled automatically using machine learning, namely Naive Bayes. Researchers also created a confusion matrix to determine the accuracy level of the algorithm used in this study. The accuracy result of the confusion matrix obtained by researchers in this study is 93.39%. The positive class precision value of 85.52% indicates that most positive reviews were identified accurately, while the negative class precision value of 100% demonstrates excellent capability in identifying negative reviews. In conclusion, the Netflix application receives diverse responses from users, and the algorithm used effectively identifies reviews accurately.

**Keywords:** Sentiment Analysis; Naïve Bayes; RapidMiner; Google Play Store

## **1. PENDAHULUAN**

Dalam beberapa dekade terakhir, perkembangan teknologi telah mengalami lonjakan pesat, terutama sejak kemunculan internet dan komputer pribadi pada tahun 1980-an. Kemajuan teknologi ini telah menyebabkan perubahan besar di berbagai sektor, termasuk komunikasi, transportasi, kesehatan, hiburan, dan pendidikan [1]. Peningkatan teknologi internet secara signifikan telah memperluas jangkauan distribusi informasi. Salah satu aspek yang mendukung peningkatan ini adalah media sosial, di mana pengguna tidak hanya berfungsi sebagai penerima informasi tetapi juga sebagai pembuat informasi [2]. Peningkatan jumlah pengguna internet di Indonesia disebabkan oleh berbagai kemudahan yang ditawarkan oleh media sosial dan internet. Melalui media sosial, orang bisa mengakses informasi dan berkomunikasi dengan sangat cepat [3].

Google memiliki platform bernama *Play Store* yang menawarkan berbagai konten digital, termasuk permainan, aplikasi, film, buku, dan musik dalam berbagai kategori [4]. Netflix adalah *platform streaming* yang menempati peringkat ke-10 dalam hal akses di Indonesia. Ulasan pengguna terhadap layanan aplikasi Netflix dapat ditemukan di situs *Google Play Store* [5]. Netflix beroperasi seperti toko penyewaan film digital, menyediakan berbagai macam film. Serupa dengan layanan televisi berlangganan, Netflix tidak mengandalkan iklan, sehingga pengguna dapat menghindari jadwal penayangan televisi dan memilih konten sesuai keinginan mereka. Netflix menjadi pelopor dalam layanan sewa film online *global*. Berdiri sejak 29 Agustus 1997 di Scotts Valley, California, dan hadir di Indonesia pada tahun 2016, Netflix memiliki arsip film terlengkap di dunia dengan jangkauan operasional terluas [6]. Adanya netflix memudahkan pengguna untuk menonton berbagai film secara *streaming*. Pengambilan data netflix diambil melalui komentar pengguna tentang netflix di google playstore. Setelah memperoleh data tersebut, akan dilakukan *preprocessing* dan menerapkan metode algoritma Naive Bayes untuk melakukan evaluasi sentimen. Algoritma Naive Bayes adalah sebuah teknik klasifikasi yang

terbukti sangat efektif dan sering diterapkan dalam berbagai penelitian [7]. *Naïve Bayes* ialah sebuah algoritma klasifikasi yang memanfaatkan pendekatan statistik dan probabilitas secara sederhana, didasarkan pada teorema *Bayes*. *Naïve Bayes* ini mengoperasikan dengan asumsi bahwa setiap atribut bersifat independen. Dengan demikian, *Naïve Bayes* mengasumsikan keberadaan atau ketiadaan kesamaan teks dengan karakteristik kelas lainnya berdasarkan atribut tertentu [8].

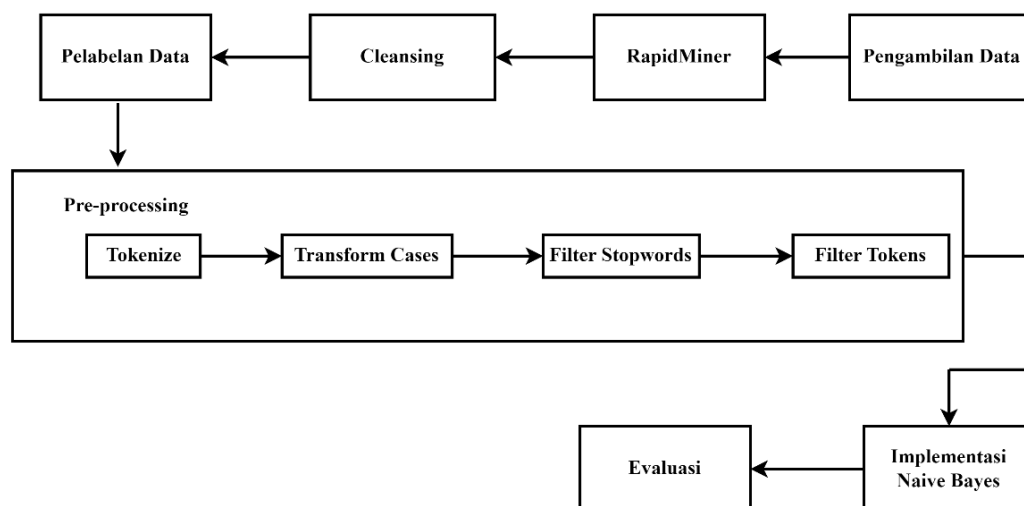
Analisis sentimen adalah teknologi baru yang sedang dikembangkan dan digunakan dalam Penelitian bertujuan untuk menganalisis topik serta pendapat yang diungkapkan. Penggunaan teknologi ini kerap dilakukan untuk mengevaluasi dan menganalisis sejauh mana pelanggan atau pengguna puas terhadap suatu produk atau kebijakan. Analisis sentimen merupakan langkah otomatis dalam mengekstrak, memahami, dan mengelola data teks dengan tujuan mendapatkan pemahaman atau informasi yang bermanfaat. Proses analisis melibatkan *tweet*, tulisan ulasan, forum, dan blog, sementara prapemrosesan data mencakup langkah-langkah seperti penghapusan *stopword*, *stemming*, identifikasi sentimen, tokenisasi, serta klasifikasi sentimen [9]. Tujuannya adalah mengidentifikasi bagaimana orang lain menilai suatu isu, apakah mereka cenderung memiliki pandangan positif atau negatif terhadapnya [10]. Pada beberapa penelitian, sentimen dapat dianggap sebagai data besar ketika jumlah data teksnya bertambah dan variasi konteksnya semakin kompleks [11]. *Text Mining* adalah disiplin baru yang sedang berkembang, yang bertujuan untuk mengekstrak makna dari teks alami [12]. Ini dapat dijelaskan sebagai proses menganalisis teks untuk mengambil informasi yang bermanfaat untuk tujuan khusus [13]. Tugas khusus dari *text mining* adalah melakukan penggolongan teks [14].

Peneliti terdahulu mengenai analisis sentimen yaitu prediksi calon presiden yang akan maju pada tahun 2024 di Indonesia. Penelitian ini membahas tentang sentimen netizen terhadap ke empat kandidat presiden yaitu Ganjar, Anies, Prabowo, dan Sandiaga Uno di *tweet* pada aplikasi *twitter*. Metode yang digunakan adalah *Naive Bayes* dan menggunakan perangkat lunak *RapidMiner*, sama seperti metode dan perangkat lunak yang peneliti gunakan. Perbedaannya terletak pada metode pengambilan data yang di mana penelitian terdahulu tersebut mengambil data langsung dari *rapid miner* tersebut dengan menggunakan API *twitter*, sedangkan penelitian yang dilakukan peneliti mengambil data menggunakan *web scrapping* dengan python di *Google Colab* dan mengatur pengambilan data tersebut ke ulasan *google play store netflix* [15]. Penelitian terdahulu lainnya membahas tentang analisis sentimen terhadap ulasan aplikasi ajaib pada *google play store*. Metode yang digunakan adalah *Naive Bayes* dan menggunakan perangkat lunak *RapidMiner*, sama seperti metode dan perangkat lunak yang peneliti gunakan. Metode pengambilan data yang digunakan peneliti terdahulu menggunakan *web scrapping* dengan python di *Google Colab*, sama seperti metode pengambilan data yang peneliti gunakan. Perbedaannya terletak pada jumlah pengambilan data ulasan yang dimana peneliti terdahulu mengambil dataset sebanyak 500 data, sedangkan penelitian yang dilakukan peneliti mengambil data sebanyak 1000 data yang setelah dilakukan pembersihan data, data menjadi 893 data [16]. Dalam penelitian ini, peneliti menggunakan metode *Naïve Bayes* dalam memprediksi apakah ulasan yang telah dikumpulkan sebelumnya bersifat positif atau negatif. Peneliti menggunakan perangkat lunak *RapidMiner*, di mana analisis sentimen juga dilakukan. Jadi, seluruh penelitian ini dilaksanakan menggunakan *RapidMiner*.

## 2. METODOLOGI PENELITIAN

### 2.1 Alur Penelitian

Dalam penelitian ini, peneliti menggunakan klasifikasi *naive bayes* untuk menentukan apakah sentimen dari ulasan pengguna aplikasi Netflix di *Google Play Store* bersifat positif atau negatif. Data diambil menggunakan platform *Google Colab* yang menggunakan bahasa pemrograman Python. [17]. Di bawah ini terdapat gambar 1 yang menunjukkan urutan langkah-langkah penelitian yang dilakukan oleh peneliti, dengan rincian sebagai berikut:



Gambar 1. Alur Penelitian

## 2.2 Pengambilan Data

*Scraping* merupakan proses pengambilan dokumen yang semi-terstruktur dari internet, sering kali berupa halaman web yang menggunakan bahasa *markup* seperti HTML atau XHTML. Tujuan dari aktivitas *scraping* ini adalah untuk mengekstraksi informasi dari halaman-halaman tersebut, baik secara keseluruhan maupun sebagian, dengan tujuan penggunaan yang berbeda-beda [18]. *Scraping* data ulasan dilakukan melalui proses dengan menggunakan *Google Collaboratory*. Dengan memanfaatkan metode *Scraping* data, dataset ulasan dan komentar aplikasi Netflix diambil dari *Google Play Store* menggunakan *Google Colab*. Sebanyak 1000 data berhasil diambil, kemudian data tersebut diubah ke format csv pada *Google Colab* dan disimpan.

## 2.3 RapidMiner

*RapidMiner* merupakan perangkat lunak sumber terbuka yang digunakan untuk melakukan penambangan data, penambangan teks, serta analisis prediktif. Selain itu, *RapidMiner* juga dapat terintegrasi dengan berbagai sumber data termasuk *database*, file teks, dan jenis data lainnya yang beragam [19]. Dengan memanfaatkan berbagai teknik deskriptif dan prediktif, *RapidMiner* bertujuan untuk memberikan wawasan kepada pengguna dan mendukung mereka dalam membuat keputusan yang lebih baik. Dalam mencapai tujuan ini, *RapidMiner* menyediakan berbagai operator *data mining*, termasuk operator untuk visualisasi, *preprocessing data*, *output* dan *input*, yang dapat digunakan dengan efisien dan mudah untuk menjalankan operasi data mining. Setelah mendapatkan dataset ulasan penggunaan aplikasi Netflix dari *Google Collab*, peneliti melanjutkan tahap penelitian selanjutnya menggunakan *RapidMiner* untuk dilakukan pembersihan dataset.

## 2.4 Cleansing

Cleansing adalah proses menghilangkan komponen yang tidak relevan dari dokumen, seperti karakter atau simbol, angka, emotikon, dan tautan URL [20]. Tahap ini sangat penting untuk menunjang proses penelitian khususnya pada saat pelabelan otomatis menggunakan *naive bayes*. Tahap *cleansing* dilakukan di dalam *tools RapidMiner* untuk membersihkan data dari simbol simbol yang tidak diperlukan dalam proses analisis sentimen, simbol yang dimaksud seperti tanda titik, tanda seru, tanda koma, sampai simbol emotikon. Bukan hanya itu, peneliti membersihkan dataset dari kalimat yang duplikat. Dari 1000 data yang diperoleh peneliti, setelah dilakukan *cleansing* data menggunakan *tools RapidMiner*, peneliti memperoleh 893 data yang sudah dibersihkan.

## 2.5 Pelabelan Data

Pelabelan data merupakan proses menetapkan atau memberi label kepada setiap instance atau contoh data dalam dataset dengan kategori positif dan negatif [21]. Setelah data *cleansing* total jumlah data akan dipisah menjadi 2, setengah untuk data *testing*, dan untuk data *training* dipakai untuk tahap implementasi *naïve bayes*. Proses pelabelan data dilakukan secara manual. Dalam penelitian ini, peneliti melabelkan secara manual dataset yang sudah dibersihkan tadi menggunakan *microsoft excel* sejumlah 635 data. Lalu sisa datanya yaitu 258 data akan dijadikan sebagai *data testing* yang akan diproses menggunakan *tools RapidMiner* dengan menggunakan algoritma *naïve bayes classifier*.

## 2.6 Pre-processing

*Pre-processing* yaitu langkah untuk mengolah *text mining* agar data menjadi lebih bermakna [22]. *Pre-processing* bertujuan untuk menyusun ringkasan teks, dan proses ini melibatkan serangkaian tahapan seperti memecah kalimat, mengurangi kapitalisasi huruf, melakukan tokenisasi, dan menghapus kata-kata penghenti (*stopword*). Pada *preprocessing* ada empat tahap yaitu *Tokenizing*, *Tranform Cases*, *Filter Stopwords*, *Filter Token by Length* yang akan peneliti lakukan di *RapidMiner*.

## 2.7 Implementasi Naïve Bayes

Dalam penelitian yang telah peneliti lakukan, peneliti menggunakan metode algoritma *naïve bayes* dalam melakukan analisis sentimen ulasan pengguna Netflix yang diambil dari *Google Play Store*. Klasifikasi *Naïve Bayes* memiliki kemampuan untuk memproyeksikan probabilitas keanggotaan suatu kelas dengan dasar pada asumsi independensi. Oleh karena itu, potensi untuk mengidentifikasi kemungkinan masa depan didasarkan pada pengalaman sebelumnya. Kelebihan dari algoritma *Naïve Bayes* terletak pada efisiensinya, yang memungkinkan proses analisis sentimen menjadi lebih singkat. Di samping itu, meskipun memiliki jumlah data latih atau data *training* yang terbatas, algoritma yang memiliki tingkat akurasi yang tinggi yaitu dengan metode *naive bayes* [23].

## 2.8 Evaluasi

Evaluasi menggunakan *Confusion Matrix* sebagai metode untuk melakukan perhitungan keakuratan sebuah algoritma *naive bayes*. *Confusion Matrix* adalah suatu teknik yang sering digunakan untuk menyajikan hasil akurasi dari model yang telah dibuat. Pada tahap ini, dilakukan penghitungan untuk mengevaluasi akurasi, presisi, dan *recall*. Dalam membuat *confussion matrix*, peneliti masih tetap menggunakan *tools RapidMiner* dalam melakukan perhitungan keakuratan data *training* yang telah peneliti buat, karena dengan menggunakan *RapidMiner* untuk menghitung keakuratan akan lebih efisien dan hasilnya pun juga tidak terlalu jauh berbeda dengan perhitungan manual.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Pengumpulan Data

Pada penelitian kali ini peneliti mengambil dataset ulasan pada aplikasi netflix sebanyak 1000 data dengan rentang waktu antara April 2024 sampai Juni 2024, tetapi pada saat dilakukan pembersihan data dari simbol-simbol yang tidak diperlukan dalam proses analisis sentimen dan kalimat duplikat, dataset menjadi 893 data saja yang siap untuk dilakukan ke tahap cleaning yang akan dijelaskan pada tahap selanjutnya. Gambar 2 di bawah merupakan ilustrasi pengambilan data.

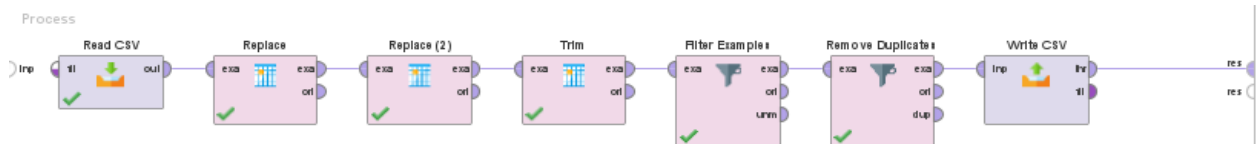


Gambar 2. Ilustrasi Pengumpulan Data

Peneliti memperoleh dataset tersebut menggunakan *Goggle Collab* dengan *python scripting*. Peneliti mengatur jumlah data ulasan yang diambil pada *python* sebanyak 1000 data terbaru. 1000 data ini akan peneliti lakukan pembersihan menggunakan *RapidMiner*.

#### 3.2 Tahap Cleaning

Pada saat data sudah diperoleh peneliti, tahap pertama yang perlu dilakukan peneliti yaitu membersihkan dataset dari simbol-simbol yang tidak diperlukan dan kalimat yang duplikat. Berikut gambar 3 adalah proses pembersihan dataset yang peneliti lakukan di *RapidMiner*.



Gambar 3. Ilustrasi Pembersihan Dataset

Gambar 3 di atas merupakan proses pembersihan dataset di *RapidMiner*, operator *Read CSV* adalah operator yang berfungsi untuk mengambil atau mengimpor file dataset yang ingin digunakan. Dalam hal ini, peneliti telah mengimpor dataset yang belum dibersihkan tadi ke dalam *tools RapidMiner* menggunakan operator *Read CSV* tersebut. Operator selanjutnya yaitu operator *replace* yang berguna untuk menghilangkan simbol-simbol yang tidak diperlukan, berikut merupakan simbol-simbol yang peneliti tulis ke dalam operator tersebut seperti `[.,?!@#%&^`~£δÿ~£δÿ~£‘(E)~‘□¼∩...-¢™Ž"©ãï,à»^áŠ~“r”¥/]`. Operator *trim* berguna untuk menghapus *space* yang tidak diperlukan dalam sebuah sentimen, karena biasanya terdapat *space* yang masih tertinggal setelah dilakukan proses *Replace* sebelumnya. Operator *Filter Examples* berguna untuk memilih kolom yang ingin dibersihkan, kolom yang peneliti pilih untuk dibersihkan yaitu kolom *content* yang berisi ulasan dari pengguna aplikasi Netflix di *Google Play Store*. Operator terakhir adalah operator *Remove Duplicates* berguna untuk menghilangkan kalimat yang duplikat, operator ini juga penting dalam proses analisis sentimen, karena pada saat peneliti melakukan pembersihan sebuah sentimen, terdapat beberapa pengguna yang memberikan ulasan dengan kalimat yang sama seperti “aplikasi bagus”, “Good”, “Ok”, dan lain-lain. Setelah semua operator tersebut dijalankan, dataset yang tadinya berjumlah 1000 data, tersisa 893 data yang sudah bersih dan siap untuk dilakukan pelabelan. Berikut gambar 4 yang menunjukkan hasil dataset setelah dibersihkan.

Row No.	content
1	Nonton film di tv ngga full screen layar jadi nontonnya agak males Beberapa film doang yg support full layar ngga smuanya jdi KECEWA mana pengaturan nya cuma buat su...
2	Selamat sore Netflix ini pengguna netflix di smartphone maupun di tv Mau tanya apa netflix ada gangguan jaringan Lewat Hp maupun tv Trimakasih
3	netflix masih jd apk kecintaan buat nonton film
4	Aplikasi nya ga bisa masuk
5	buat nonton ini seruuuu
6	APLIKASI BUSUK SUDAH GAK DOWNLOAD MASIH NYEDOT OTOMATIS UANG GUE DI GOPAY
7	Ribet banget masuk akun juga
8	Film nya sampah ga update update tagihan mah lancar masih lengkapan di streaming google
9	Sangat seru cocok untuk pencinta drakoranime
10	Jelek: kali anjeng akun ku di bikin ilang
11	Kok netflix sekarang gak bisa di buka setelah update Mentok di logo netflix doang pas mau masuk Udah coba instal ulang lagi gak bisa Ini apa masalah dan solusinya
12	Gak bisaaa
13	Terbaik
14	Udah download tapi ga bisa masuk dengan akun google Padahal sudah benar PARAH
15	Dasar kocak bayar mahal buat nonton anime tapi disuruh nunggu 1 minggu setelah rilis Nyesel gua bayar

Gambar 4. Hasil Pembersihan Dataset

Gambar 4 diatas merupakan hasil dari rapidminer setelah dilakukan pembersihan dataset dari simbol dan kalimat yang duplikat.

### 3.3 Tahap Pelabelan

Tahap selanjutnya setelah pembersihan dataset yaitu pelabelan dataset secara manual oleh peneliti, dataset yang peneliti peroleh pada awalnya berjumlah 1000 data dan setelah dilakukan pembersihan data, dataset menjadi 893 data. Dari 893 data ini, peneliti membuat kolom baru bernama kolom sentimen pada *microsoft excel* yang akan diisi negatif atau positif sebuah sentimen yang ada di sampingnya. Peneliti melabelkan data secara manual sebanyak 635 data dari 893 pada *microsoft excel*.

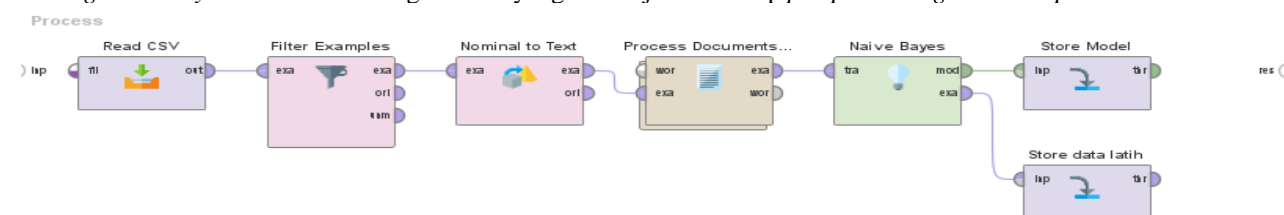
	A	B
18	sip	positif
19	Mantep Pisan	positif
20	please Netflix keluarin film baru dong jangan yg lama gue dr kapan hari nunggu Netflix keluarin kaya pengabdian setan 2 KKN baru sama lama pokok yg t	negatif
21	Very good service from Netflix members thank you very much don't forget to download Netflix guys	positif
22	Sampah	negatif
23	Bintang memberi jawaban	positif
24	Baru download aja suruh melakukan pembaruan terus menerus Tidak bisa login Susah benar mau masuk Netflix saja Uninstall sajalah	negatif
25	susah login nya	negatif
26	Apasih gajelas harus bayar segala dah cuma bikin akun	negatif
27	Sangat puas	positif
28	Film nya kurang banyak	negatif
29	Layanan ditangguhkan karena masalah pembayaran padahal udah bayar dan tagihan berikutnya tanggal 22 mei Sampah	negatif
30	bayar	positif
31	Lumayan	positif
32	Pembayaran udah berhasil tapi ga bisa nonton mana udah kepotong lagi aneh banget	negatif
33	Untuk berlangganan dipersulit sudah dibayar saldo sudah terpotong namun akun masih ditangguhkan Meminta menunggu 9Hari KECEWAAA	negatif
34	Nyesel gw langganan ni streaming Ternyata ada beberapa film populer yang udah lama dirilis tapi gak ada di sini Gimana sih Menurutku ini masih belu	negatif
35	Kenapa saya mo berlangganan kok sulit yah	negatif
36	Dengan adanya Netflix untuk menonton film layar lebar tidak perlu datang ke bioskop cukup di hptv android	positif
37	Platform nonton online yg paling lengkap	
38	Wihh semakin baik banyak film baru	

Gambar 5. Pelabelan Manual

Gambar 5 di atas merupakan gambar pelabelan secara manual yang dilakukan peneliti dalam proses analisis sentimen. Sekarang tersisa 258 data sentimen yang belum peneliti labelkan, sisa data ini akan dilabelkan secara otomatis menggunakan algoritma *naive bayes*. Tetapi sebelum masuk ke dalam tahap pelabelan otomatis menggunakan *naive bayes*, akan dilakukan proses *pre-processing* yang akan dijelaskan selanjutnya.

### 3.4 Tahap pre-Processing

Setelah pelabelan data secara manual telah selesai, tahap yang dilakukan adalah *pre-processing*. Pada tahap ini peneliti masih menggunakan *RapidMiner* dalam pengerjaannya, tahap *pre-processing* berguna untuk membuat model *data training naive bayes*. Berikut adalah gambar 6 yang menunjukkan tahap *pre-processing* dalam *RapidMiner*.



Gambar 6. pre-Processing pada RapidMiner

Operator yang peneliti gunakan yang pertama adalah *Read CSV* yaitu untuk mengimport dataset yang telah dilabeli secara manual sebelumnya, dilanjutkan dengan *filter examples* untuk mengambil kolom yang ingin difilter nanti, dalam hal ini peneliti memilih kolom *content* yang berisi ulasan pengguna Netflix dan memilih data yang memiliki nilai sentimen baik itu positif maupun negatif di barisnya. Seperti yang telah dijelaskan sebelumnya oleh peneliti, bahwa peneliti mengimport dataset yang sudah peneliti labelkan secara manual menggunakan *Read CSV*, karena dataset yang dilabelkan hanya 635 data, oleh karena itu terdapat 258 data yang belum diberikan sentimen secara manual oleh peneliti (258 didapatkan dari 893 dikurang dengan 635). Maka dari itu, digunakan *Filter Examples* hanya untuk memilih kolom *content* yang sudah memiliki label sentimen guna untuk membuat model *data training* nanti. Operator selanjutnya yaitu *nominal to text* yang berguna untuk mengubah *role* pada dataset yang pada awalnya polynominal menjadi *text* karena, pada operator selanjutnya yaitu *proses document from data* mengharuskan dataset tersebut bertipe *text* dan bukan *polynominal*. Dalam operator *process document from data*, seperti gambar 7 berikut.



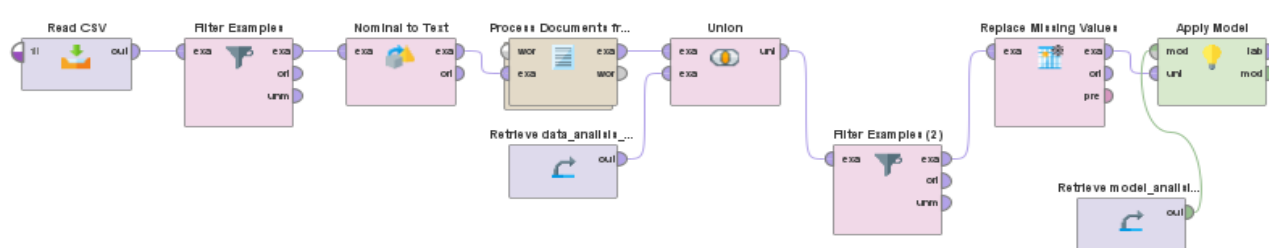
Gambar 7. Process Document from Data



Operator *tokenize* berguna untuk memisahkan kata-kata dalam suatu kalimat agar terpisah, operator *Transform Cases* digunakan untuk mengubah semua huruf yang pada awalnya ada huruf kapital menjadi huruf biasa. *Filter Stopwords* digunakan untuk menghilangkan sebuah kata yang tidak begitu berguna dalam proses analisis sentimen seperti di, yang, pada, ada, dll. *Stopwords* yang peneliti gunakan diambil dari situs kaggle.com karena *RapidMiner* belum menyediakan *stopwords* dalam bahasa Indonesia. Operator terakhir yaitu *Fiter Tokens by Length* yang berguna untuk menghapus sebuah kata dengan rentang jumlah tertentu, operator tersebut digunakan peneliti untuk menghapus kata-kata dengan jumlah kata kurang dari 3 dan lebih dari 25. Proses berikutnya yaitu *store* data latih, operator tersebut berfungsi untuk menyimpan dataset yang sudah diberikan label oleh peneliti secara manual sebelumnya, operator ini berguna untuk tahap selanjutnya yaitu tahap pelabelan secara otomatis. Operator terakhir yaitu operator *store model* berfungsi untuk menyimpan model algoritma naive bayes sebelumnya, operator ini digunakan untuk tahap selanjutnya dalam pelabelan otomatis suatu sentimen.

### 3.5 Implementasi Naive Bayes

Tahap terakhir yang peneliti lakukan yaitu membuat operator-operator yang berfungsi untuk proses pelabelan otomatis dataset peneliti, adapun operator yang peneliti gunakan seperti gambar 8 berikut.



Gambar 8. Tahap Pelabelan Otomatis RapidMiner

Pada gambar 8 di atas, operator yang peneliti gunakan sama seperti sebelumnya, hanya berbeda setelah *proses document from data* saja. Operator *union* berfungsi untuk menggabungkan dataset dari model *retrive* data analisis sentimen dan dataset dari operator *Read CSV*. Operator tersebut berfungsi untuk menggabungkan dataset dari *Read CSV* dengan dataset *retrive* data analisis sentimen pada tahap sebelumnya, tepatnya pada tahap *pre-processing* tadi. Karena dataset pada *retrive* data analisis sentimen hanya menyediakan data yang peneliti beri sentimen saja yaitu sejumlah 635 data, sedangkan dataset pada *Read CSV* adalah dataset keseluruhan yang lengkap sejumlah 893 data. Oleh karena itu, dibutuhkan operator *Union* untuk menggabungkan kedua dataset tersebut. Operator yang peneliti gunakan selanjutnya yaitu *replace missing value*, operator tersebut berfungsi untuk mengganti nilai yang kosong pada dataset setelah dilakukan penggabungan dengan operator *union* dengan angka nol. Karena apabila tidak menggunakan operator ini, dataset akan memiliki nilai kosong yang ditandai dengan tanda tanya (?). operator selanjutnya yaitu *retrive* model analisis sentimen, operator tersebut menyimpan algoritma naive bayes yang sudah dilatih dengan data *training* pada tahap sebelumnya, lalu diaplikasikan ke dalam operator *apply model* untuk dilihat kinerjanya apakah berhasil atau tidak. Berikut adalah gambar 9 dari hasil penerapan algoritma *naive bayes*.

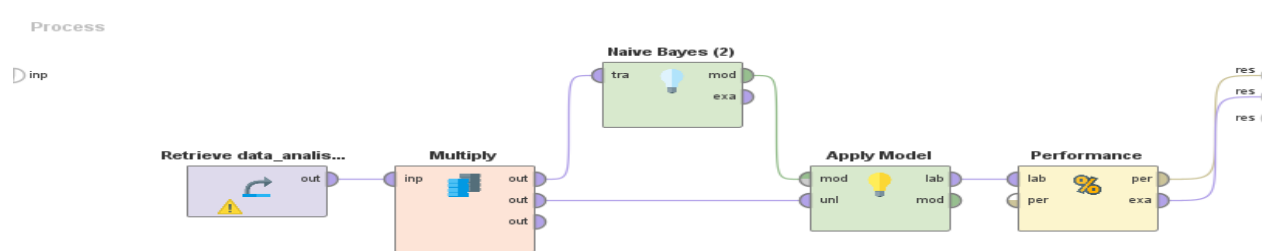
Ro...	prediction(sentimen)	text
1	positif	platform nonton online lengkap
2	positif	wihh film
3	negatif	aplikasi bagus filem tayangkan rekomendasi
4	negatif	ruwet daftar
5	positif	suka banget series netflix
6	positif	kualitas audio video nya bagus
7	negatif	filmnya seru seru dijamin gak bosan nontonnya
8	positif	akses masuknya sdh akun
9	positif	moga menghibur disaat santay
10	negatif	saldo berkurang berlangganan padahal gak menyetujui tolong duit abis gegara ludes maap hapus apk
11	negatif	dana kepotong pemberitahuan kocak sih
12	negatif	kualitas videonya bagus bgt subtitle nya lengkap premium penayangan beda jam korea admin nya gercep kesalahan teknis recommended sih apk over
13	positif	netflix bagus
14	positif	susah masuk

Gambar 9. Hasil Ananlisis Sentimen Menggunakan Naive Bayes

Gambar 9 di atas menunjukkan hasil dari penerapan metode algoritma *Naive Bayes* untuk mengevaluasi sentimen ulasan pengguna Netflix di *Google Play Store*. Bisa dilihat bahwa hasil tidak sempurna 100% karena memang ada beberapa kesalahan analisis yang salah diberikan oleh *naive bayes*, karena semakin banyak *naive bayes* diberi *data training*, maka hasil analisis yang dilakukan otomatis oleh *naive bayes* semakin akurat.

### 3.6 Tahap Evaluasi

Tahap selanjutnya peneliti akan membuat *confussion matrix* untuk melihat tingkat keakuratan *algoritma naive bayes* yang sudah peneliti latih menggunakan 635 data *training* dari 893 dataset. Berikut gambar 10 adalah operator yang digunakan untuk membuat *confussion matrix*.



Gambar 10. Tahapan Confussion Matrix RapidMiner

Operator *retrieve* di atas adalah operator yang menyimpan data *training* yang telah dibuat oleh peneliti dan disambung dengan operator *multiply* untuk membuat cabang lain karena akan disambung dengan dua operator yang lain yaitu operator *apply model* dan operator *naive bayes*. Terakhir yaitu operator *performance* yang digunakan untuk menghitung nilai akurasi pada data *training* yang telah peneliti sediakan. Berikut gambar 11 adalah hasil dari *confussion matrix* dari model di atas.

Table View Plot View

accuracy: 93.39%

	true negatif	true positif	class precision
pred. negatif	345	0	100.00%
pred. positif	42	248	85.52%
class recall	89.15%	100.00%	

Gambar 11. Hasil Confussion Matrix

Gambar 11 di atas merupakan hasil akurasi dari *confussion matrix* yang dilakukan dalam *RapidMiner*, dari akurasi di atas bisa dilihat bahwa *class precision* pada *pred. negatif* mencapai 100% sedangkan *class precision* untuk *pred. positif* mencapai 85,52% dengan nilai akhir akurasi dari data *training* mencapai 93,39%.

## 4. KESIMPULAN

Berdasarkan evaluasi sentimen terhadap data aplikasi Netflix menggunakan metode *Naive Bayes*, sebanyak 1000 data diambil dari April 2024 sampai dengan Juni 2024. Setelah proses pembersihan, tersisa 893 data yang kemudian dibagi menjadi 635 data untuk pelabelan secara manual agar dapat dijadikan data *training* dan sisanya yaitu 258 data akan peneliti jadikan data testing untuk menguji algoritma naive bayes dalam menganalisis sentimen secara otomatis. Hasil dari algoritma *naive bayes* dalam memberikan sentimen secara otomatis menunjukkan bahwa ada lebih banyak sentimen positif dibandingkan dengan sentimen negatif, sentimen positif yang didapatkan dari dataset di atas berjumlah 157 data yang mendapatkan sentimen positif, sedangkan sentimen negatif berjumlah 101 data yang mendapatkan sentimen negatif. Rata-rata ulasan pengguna yang memberikan ulasan negatif mengindikasikan ketidakpuasan pengguna terhadap fitur-fitur Netflix saat ini. Meskipun demikian, selisih antara sentimen positif dan negatif tidak terlalu besar, menunjukkan bahwa banyak pengguna yang masih tetap merasa puas dengan layanan ataupun fitur yang diberikan oleh Netfilix kepada pengguna. Hasil dari akurasi *confussion matrix* yang telah peneliti lakukan juga cukup tinggi untuk melakukan pelabelan secara otomatis yaitu sebesar 93,39% nilai akurasi. Kekurangan dari penelitian ini yaitu ada beberapa sentimen yang seharusnya sentimen itu bernilai positif tetapi naive bayes malah mengidentifikasikan sentimen tersebut sebagai sentimen negatif. Hal ini tentu saja bisa terjadi karena pada saat pelabelan manual dataset oleh peneliti, ada kata yang banyak masuk ke dalam kategori sentimen positif lalu pada saat pelabelan otomatis dengan *naive bayes*, kata tersebut malah masuk ke dalam kalimat yang berindikasi negatif. Saran untuk penelitian kedepannya adalah melabelkan dataset secara manual dengan lebih banyak lagi agar keakuratan naive bayes dalam memberikan sentimen dapat lebih tepat dan akurat.

## REFERENCES

- [1] A. Fricticarani, A. Hayati, R. R, I. Hoirunisa, and G. M. Rosdalina, "Strategi Pendidikan Untuk Sukses Di Era Teknologi 5.0," *J. Inov. Pendidik. dan Teknol. Inf.*, vol. 4, no. 1, pp. 56–68, 2023, doi: 10.52060/pti.v4i1.1173.
- [2] R. D. Himawan and E. Eliyani, "Perbandingan Akurasi Analisis Sentimen Tweet terhadap Pemerintah Provinsi DKI Jakarta di Masa Pandemi," *J. Edukasi dan Penelit. Inform.*, vol. 7, no. 1, p. 58, 2021, doi: 10.26418/jp.v7i1.41728.

- [3] N. Anas and S. Sapri, "Komunikasi Antara Kognitif dan Kemampuan Berbahasa," *EUNOIA (Jurnal Pendidik. Bhs. Indones.*, vol. 1, no. 1, p. 1, 2022, doi: 10.30821/eunoia.v1i1.997.
- [4] M. Fudhail Ferio Supeli and S. Setiaji, "Klasifikasi Sentimen Positif Dan Negatif Pada Aplikasi Vidio Dengan Algoritma K-Nearest Neighbor," *Indones. J. Comput. Sci.*, vol. 2, no. 1, pp. 7–15, 2023, doi: 10.31294/ijcs.v2i1.1874.
- [5] E. Lestari and O. R. C. Soesanto, "PREDICTING FACTORS THAT INFLUENCE ATTITUDE TO USE AND ITS IMPLICATIONS ON CONTINUANCE INTENTION TO USE SVOD: STUDY ON NETFLIX USERS OF INDONESIA," *DeReMa (Development Res. Manag. J. Manaj.*, vol. 15, no. 2, p. 183, 2020, doi: 10.19166/derema.v15i2.2541.
- [6] P. Adi Cakranegara, Y. Budiasih, H. Tannady, and Y. Totok Suyoto, "Analysis Of The Effect Of Perceived Value And Brand Image On Netflix Service Purchase Decisions Analisis Pengaruh Perceived Value Dan Brand Image Terhadap Purchase Decision Layanan Netflix," *Manag. Stud. Entrep. J.*, vol. 3, no. 5, pp. 3238–3247, 2022, [Online]. Available: <http://journal.yrpicpu.com/index.php/msej>.
- [7] A. Zy and Wahyu Hadikristanto, "Implementasi Algoritma Metode Naive Bayes dan Support Vector Machine Tentang Pembobolan dan Kebocoran Data di Twitter," *Bull. Inf. Technol.*, vol. 4, no. 1, pp. 49–56, Mar. 2023, doi: 10.47065/bit.v4i1.493.
- [8] H. Setiawan and I. Zufria, "Analisis Sentimen Pembatalan Indonesia Sebagai Tuan Rumah Piala Dunia FIFA U-20 Menggunakan Naïve Bayes," *J. Media Inform. Budidarma*, vol. 7, no. 3, pp. 1003–1012, 2023, doi: 10.30865/mib.v7i3.6144.
- [9] F. Sidik, I. Suhada, A. H. Anwar, and F. N. Hasan, "Analisis Sentimen Terhadap Pembelajaran Daring Dengan Algoritma Naive Bayes Classifier," *J. Linguist. Komputasional*, vol. 5, no. 1, p. 34, 2022, doi: 10.26418/jlk.v5i1.79.
- [10] A. Yahyadi and F. Latifah, "Analisis Sentimen Twitter Terhadap Kebijakan PpkM Di Tengah Pandemi Covid-19 Menggunakan Mode Lstm," *J. Inf. Syst. Applied, Manag. Account. Res.*, vol. 6, no. 2, pp. 464–470, 2022, doi: 10.52362/jisamar.v6i2.791.
- [11] M. A. Amrustian, W. Widayat, and A. M. Wirawan, "Analisis Sentimen Evaluasi Terhadap Pengajaran Dosen di Perguruan Tinggi Menggunakan Metode LSTM," *J. MEDIA Inform. BUDIDARMA*, vol. 6, no. 1, p. 535, Jan. 2022, doi: 10.30865/mib.v6i1.3527.
- [12] F. N. Hasan and M. Dwijayanti, "Analisis Sentimen Ulasan Pelanggan Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naïve Bayes Classifier," *J. Linguist. Komputasional*, vol. 4, no. 2, pp. 52–58, 2021, doi: <https://doi.org/10.26418/jlk.v4i2.61>.
- [13] A. N. Indraini and I. Ernawati, "Analisis Sentimen Terhadap Pembelajaran Daring Di Indonesia Menggunakan Support Vector Machine (SVM)," *J. Ilm. FIFO*, vol. 14, no. 1, p. 68, 2022, doi: 10.22441/fifo.2022.v14i1.007.
- [14] S. Y. Nursyiah, A. Erfina, and C. Warman, "Analisis Sentimen Pembelajaran Daring Pada Masa Pandemi Covid-19 Di Twitter Menggunakan Algoritma Naïve Bayes," *J. Media Inform. Budidarma*, pp. 117–123, 2021.
- [15] A. R. Abdillah and F. N. Hasan, "Analisis Sentimen Terhadap Kandidat Calon Presiden Berdasarkan Tweets Di Sosial Media Menggunakan Naive Bayes Classifier," *Smatika J.*, vol. 13, no. 01, pp. 117–130, 2023, doi: 10.32664/smatika.v13i01.750.
- [16] H. Z. Muflih, A. R. Abdillah, and F. N. Hasan, "Analisis Sentimen Ulasan Pengguna Aplikasi Ajaib Menggunakan Metode Naïve Bayes," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 3, pp. 1613–1621, 2023, doi: 10.30865/klik.v4i3.1303.
- [17] D. Wijaya, R. A. Saputra, and F. Irwiensyah, "Analisis Sentimen Ulasan Aplikasi Samsat Digital Nasional Pada Google Playstore Menggunakan Algoritma Naïve Bayes," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 4, pp. 2369–2380, 2024, doi: 10.30865/klik.v4i4.1738.
- [18] R. Wahyudi and G. Kusumawardana, "Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine," *J. Inform.*, vol. 8, no. 2, pp. 200–207, 2021, doi: 10.31294/ji.v8i2.9681.
- [19] A. Meiriza, E. Ali, Rahmiati, and Agustin, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Program BPJS Ketenagakerjaan," *Indones. J. Comput. Sci.*, vol. 12, no. 2, pp. 714–728, 2023, doi: 10.33022/ijcs.v12i2.3184.
- [20] M. W. A. Putra, Susanti, Erlin, and Herwin, "Analisis Sentimen Dompot Elektronik Pada Twitter Menggunakan Metode Naïve Bayes Classifier," *IT J. Res. Dev.*, vol. 5, no. 1, pp. 72–86, 2020, doi: 10.25299/itjrd.2020.vol5(1).5159.
- [21] D. Oktavia, Y. R. Ramadahan, and M. Minarto, "Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 407–417, 2023, doi: 10.30865/klik.v4i1.1040.
- [22] M. Arifiansyah Ayub, "Analisis Topik Ekonomi Dengan Algoritma K-Means Pada Media Online Era Pandemi Covid-19 Di Sulawesi Tenggara," *JIKO (Jurnal Inform. dan Komputer)*, vol. 4, no. 2, pp. 133–138, 2021, doi: 10.33387/jiko.v4i2.3235.
- [23] A. I. Tanggraeni and M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.