PROSEMNASPROIT : Prosiding Seminar Nasional Ilmu Teknik
Volume. 2 Nomor. 2, Desember 2025
E-ISSN : 3063-5713; P-ISSN : 3063-4709; Hal. 208-225
DOI: https://doi.org/10.61132/prosemnasproit.v2i2.198
Tersedia: https://prosiding.aritekin.or.id/index.php/PROSEMNASPROIT

Analisis Sentimen Publik pada TikTok terhadap Rencana Penerapan Sistem Balik Nama Ponsel Bekas menggunakan Naive Bayes dan Support Vector Machine

Afif Lustyo Muji[1], Aziz Musthofa[2], Dihin Muriyatmoko [3]
[1.2.3] Universitas Darussalam Gontor, Indonesia

Alamat: Siman, Ponorogo, Jawa Timur
Korespondensi penulis: afifmuji42004G@mhs.unida.gontor.ac.id

Abstract. *Since the announcement of the policy plan for a name transfer system in the sale of used mobile phones, the issue has attracted widespread public attention and discussion. People have expressed their opinions on social media platforms, particularly TikTok. This study aims to classify the sentiment of TikTok users using Naive Bayes and Support Vector Machine (SVM) algorithms. The data were collected through a comment scraping technique on related content.The research stages include text preprocessing, sentiment labeling into positive, negative, and neutral categories, and feature extraction using TF-IDF. The classification process employs Naive Bayes and Support Vector Machine algorithms, which are then evaluated based on accuracy, precision, recall, and F1-score. The results of this study indicate that both methods are capable of classifying sentiment effectively. However, the Support Vector Machine method is superior to the Naive Bayes method with an accuracy rate of 99.57% compared to 94.30%. This study is expected to help the government understand public responses to the planned policy of the used mobile phone name transfer system.*

Keywords: *Sentiment, Naive Bayes, Support Vector Machine*

Abstrak. Sejak diumumkanya rencana kebijakan system balik nama pada jual beli ponsel bekas hal ini ramai di perbincangkan masyarakat. Merekapun menuangkan opini mereka di social media, salah satunya TikTok. Penelitian ini bertujuan untuk mengklasifikasikan sentiment pengguna TikTok menggunakan algoritma Naive Bayes dan Support Vector Machine. Data diambil dengan Teknik scraping komentar pada konten terkait. Tahapan penelitian terdiri dari preprocessing teks, pelebelan sentiment menjadi positif, negative, dan netral, dan ekstrasi fitur menggunakan TF-IDF. Proses Klasifikasi mengunkan algoritma Naive Bayes dan Support Vector Machine, kemudian dievaluasi berdasarkan akurasi, presisi, recall, dan F1-Score. Hasil dari penelitian ini menunjukan bahwasanya kedua metode dapat mengklasifikasikan sentiment dengan baik. Namun, metode Support Vector Machine lebih unggul dari metode Naive Bayes dengan perbandingan tingkat akurasi 99.57% banding 94.30%. Penelitian ini diharapkan dapat membantu pemerintah dalam memahami respon masyarakat terhadapa rencana kebijakan system balik nama ponsel bekas.

Kata kunci: Sentiment, Naive Bayes, Support Vector Machine

1. LATAR BELAKANG

Cepatnya perkembangan teknologi, terutama bidang komunikasi membuat hamper tidak mungkin untuk hidup tanpa memiliki smartphone. Dengan meningkatnya kebutuhan terhadap teknologi komunikasi, membuat pembelian dan penjualan smartphone bekas juga meningkat. Untuk mengurangi tingkat kejahatan pencuriat dan meningkatkan keamanan pada konsumen, pemerintah mengumumkan sebuah rencana kebijakan system balik nama pada jual beli smartphone bekas. Kebijakan ini bertujuan untuk mamastika kepemilikan perangkat dan menghindari penyalahgunaan perangkat yang berasal dari aktivitas illegal.

Naskah Masuk: 25 November 2025; Revisi: 30 November 2025; Diterima: 29 Desember 2025;
Terbit: 31 Desember 2025

Pengumuman rencana kebijakan tersebut mendapatkan banyak tanggapan dari masyarakat. Beberapa ada yang mendukung kebijakan ini karena dianggap menigkatkan keamanan dan transparasi dalam transaksi, sementara yang lain menggangap kebijakan ini akan mempersulit dan menambah biaya dalam transaksi. Tanggapan masyarakat ini banyak disampaikan melalui media sosial, yang mana menjadi salah satu cara masyarakat untuk mengespresikan pendapat mereka saat ini. Salah satu platform media sosial yang digunakan oleh masyarakat untuk menyampaikan opini mereka adalah TikTok, terutama melalui kolom komentar pada konten yang terkait.(Datau et al., 2025)

Dengan besarnya jumlah opini masyarakat yang di ungkapkan melalui media sosial, membuat hamper tidak memugkinkan untuk melakukan analisis manual dalam Waktu yang wajar(Anggraini et al., 2024). Oleh karena itu analisis sentiment otomatis dapat digunakan untuk mengkategorikan opini publik dalam komentar media sosial menjadi positif, negative, atau netral. Analisis sentimen, dalam hal ini, akan memberikan gambaran tentang bagaimana publik memandang situasi tersebut.(Husen et al., 2023)

2. KAJIAN TEORITIS

A. Analisis Sentiment

Analisis sentimen merupakan salah satu bidang data mining yang biasa digunakan untuk meganalisis data teks berupa opini yang terpolarisasi sehingga diperoleh informasi yang bernilai positif, negative, atau netral.(Simson Lende et al., 2024)

Sebagai subbidang yang sangat populer dalam ranah ilmu analisis data, terutama dalam konteks analisis teks, metode Analisis Sentimen dapat diaplikasikan untuk berbagai keperluan, mulai dari mencari sentimen di media sosial, mengevaluasi ulasan produk, hingga berbagai aspek lainnya.(Zhang et al., 2024)

B. Naive Bayes

Naïve bayes merupakan salah satu algoritma dari machine learning. Dalam perkembangannya Naïve Bayes termasuk kedalam golongan supervised learning yaitu salah satu machine learning yang membutuhkan sampel sebagai data latih yang memiliki label(Khoerunnisa et al., 2025). Supervised Learning dikelompokkan menjadi dua yaitu klasifikasi dan regresi. Klasifikasi merupakan salah satu proses pada yang bertujuan untuk

menemukan pola yang berharga dari data yang berukuran relatif besar hingga sangat besar.(Handhayani, 2025)

Dalam teori statistik dan probabilitas, teorema Bayes (juga dikenal sebagai aturan Bayes) adalah rumus matematika yang digunakan untuk menentukan probabilitas bersyarat suatu peristiwa.(Al Ahzuri, 2024) Intinya, teorema Bayes menggambarkan probabilitas suatu peristiwa berdasarkan pengetahuan sebelumnya tentang kondisi yang mungkin relevan dengan peristiwa tersebut.

C. Support Vector Machine (SVM)

Support Vector Machine merupakan salah satu metode klasifikasi dengan menggunakan metode machine learning (supervised learning) yang memprediksi kelas berdasarkan pola dari hasil proses training yang diciptakan oleh Vladimir Vapnik.(Huda Ovirianti et al., 2022)

Klasifikasi dilakukan dengan garis pembatas (hyperlane) yang memisahkan antara kelas opini positif dan opini negatif. Garis pembatas yang baik adalah yang memiliki jarak terbesar ke titik data pelatihan terdekat dari setiap kelas, karena pada umumnya semakin besar margin, semakin rendah error generalisasi dari pemilah. Margin adalah Jarak dari suatu titik vektor di suatu kelas terhadap hyperplane. Metode Support Vector Machine (SVM) digunakan untuk melakukan klasifikasi otomatis(Zakaria et al., 2023).

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

Metode Term Frequency-Inverse Document Frequency (TF-IDF) adalah salah satu teknik yang digunakan dalam pengolahan teks dan pemodelan bahasa alami. Tujuan utama dari metode TF-IDF adalah untuk mengevaluasi seberapa penting suatu kata (term) dalam sebuah dokumen dalam konteks koleksi dokumen yang lebih besar.(Septiani & Isabela, n.d.)

Metode TF-IDF memperhitungkan dua faktor penting yaitu Term Frequency (TF) dan Inverse Document Frequency (IDF).(Rofiqi et al., 2019) Term Frequency (TF) merupakan ukuran yang digunakan untuk mengetahui seberapa sering suatu kata muncul dalam sebuah dokumen. Sementara itu Inverse Document Frequency (IDF) merupakan ukuran yang digunakan untuk mengetahui seberapa penting suatu kata dalam konteks koleksi dokumen yang lebih besar.

E. Penelitian Sebelumnya

1) Dinda Franciska Mey Dina, Tining Haryanti, Muhamad Amirul Haq (2025)

Pada penelitian terdahulu yang berjudul "Analisis Sentimen Terhadap Komentar Pada Media Sosail Tiktok Yang Berpotensi Menyebabkan Depresi Menggunakan Metode Naive Bayes " analisis sentimen untuk mengidentifikasi kecenderungan opini pada suatu masalah, apakah opini tersebut cenderung positif atau negatif, analisis sentimen terhadap komentar dapat memberikan wawasan secara real-time dan relevan.

Pada tahap evaluasi di penelitian ini , terdapat perbedaan akurasi sebesar 4,48% antara data mentah dan data yang telah melalui pre-processing. Data yang telah diproses mencapai akurasi 71,55% dalam skenario pengujian dengan 50% data, dengan presisi untuk prediksi positif sebesar 67,77% dan prediksi negatif sebesar 77,31%. Nilai recall untuk data positif adalah 82,12%, sedangkan untuk data negatif adalah 60,95%. Selain itu, nilai F1-score tercatat sebesar 68,16%. Hasil klasifikasi menunjukkan 332 data dikategorikan sebagai sentimen positif (non-depresi) dan 216 data sebagai sentimen negatif.(Franciska Mey Dina et al., 2025)

2) Dianati Duei Putri, Gigih Forda Nama, Wahyu Eko Sulistiono (2022)

Pada penelitian terdahulu dengan judul "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier" Dalam penelitian ini akan dilakukan analisis sentiment masyarakat terhadap kinerja Dewan Perwakilan Rakyat (DPR) yang diungkapkan melalui media sosial twitter. Ada beberapa tahap untuk melakukan analisis sentimen , yaitu pengumpulan data (crawling), preporcessing data yang terdiri dari proses cleaning data, tokenization, stop remova dan case folding, splitting data dan klasifikasi data menggunakan metode Naive Bayes Classifier. Penelitian ini menggunakan sebanyak 1546 data tweet. Hasil dari penelitian ini didapatkan bahwa DPR mendapatkan 95 tweet positif dengan polaritas 0.75 atau 75% sentimen positif, 693 tweet netral dengan polaritas 0.79 atau 79% sentimen netral dan 758 tweet negatif dengan polaritas 0.82 atau 82 sentimen negatif dengan accuracy score 0.8 atau 80% berdasarkan data testing sebanyak 20%.4(Duei Putri et al., 2022)

3. METODE PENELITIAN

Penelitian ini mengklasifikasikan keseluruhan data menjadi tiga kategori : positif, negatif, dan netral berdasarkan kelas yang telah di tentukan. Teknik analisis data pada penelitian ini menggunakan CRISP-DM, yang mana merupakan sebuah metodologi standar untuk menyusun dan mengerjakan proses data mineing dan data sience. Metode ini deiknal dengan sifatnya tang fleksibel dan umum.(Schröer et al., 2021) Tahapan utama dalam metodologi CRISP-DM dimulai dari business understanding, data understanding, data preparation, modeling, evaluation, dan deployment(Cholifah Sastya & Nugraha, n.d.). Gambar yang menjelaskan mengenai alur tahpan utapa pada CRISP-DM bisa di lihat pada Gambar 1.

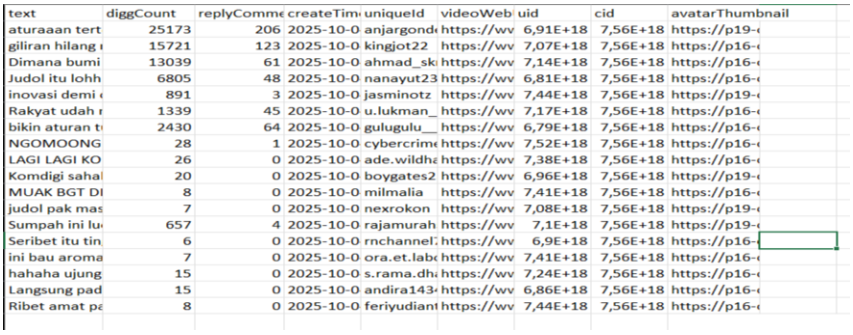
Gambar 1. Metode Penelitian CRIPS-DM

A. Business Understanding

Ini adalah tahap pertama dalam melakukan CRISP-DM. Tahapan ini dilaksanakan dengan penyatuan ide penulis atas permasalahan yang sedang populer di lingkungan masyarakat. Berdasarkan rumusan masalah, bahwa permasalahan yang diangkat dalam karya tulis ini adalah banyaknya opini masyarakat yang diungkapkan pada media sosial TikTok. Khususnya tanggapan terhadap rencana kebiakan penerapan system balik nama pada jual beli ponsel bekas. Ide ini di angkat berdasarkan permasalahan yang masih menjadi polemik di masyarakat. Sehingga di mungkinkan untuk melakukan analisis sentiment pengguna media sosial TikTok terhadap opini-opini masyarakat menaggapi renca kebiakan peneaan system balik nama pada jual beli ponsel bekas. Opini masyarakat tersebut kemudian di kelompokkan menjadi positif, negatif, dan netral.

B. Data Understanding

Pada tahap ini dimulailah proses pengambilan data dan pelabelan data yang dilakukan menggunakan metode pengambilan data scraping data didapatkanlah data acak sebanyak 2000 data yang diinginkan yang berupa komentar pengguna media sosial Tiktok terhadap topik penelitian.



text	diggCount	replyComme	createTim	uniqueId	videoWeb	uid	cid	avatarThumbnail
aturaaan tert	25173	206	2025-10-0	anjargond	https://wv	6,91E+18	7,56E+18	https://p19-
giliran hilang	15721	123	2025-10-0	kingjot22	https://wv	7,07E+18	7,56E+18	https://p16-
Dimana bumi	13039	61	2025-10-0	ahmad_sk	https://wv	7,14E+18	7,56E+18	https://p16-
Judol itu lohh	6805	48	2025-10-0	nanayut23	https://wv	6,81E+18	7,56E+18	https://p16-
inovasi demi	891	3	2025-10-0	jasminotz	https://wv	7,44E+18	7,56E+18	https://p19-
Rakyat udah	1339	45	2025-10-0	u.lukman_	https://wv	7,17E+18	7,56E+18	https://p16-
bikin aturan t	2430	64	2025-10-0	gulugulu__	https://wv	6,79E+18	7,56E+18	https://p16-
NGOMOONG	28	1	2025-10-0	cybercrim	https://wv	7,52E+18	7,56E+18	https://p16-
LAGI LAGI KO	26	0	2025-10-0	ade.wildh	https://wv	7,38E+18	7,56E+18	https://p16-
Komdigi sahal	20	0	2025-10-0	boygates2	https://wv	6,96E+18	7,56E+18	https://p19-
MUAK BGT DI	8	0	2025-10-0	milmalia	https://wv	7,41E+18	7,56E+18	https://p16-
judol pak mas	7	0	2025-10-0	nexrokon	https://wv	7,08E+18	7,56E+18	https://p19-
Sumpah ini lu	657	4	2025-10-0	rajamurah	https://wv	7,1E+18	7,56E+18	https://p19-
Seribet itu tin	6	0	2025-10-0	rnchannel	https://wv	6,9E+18	7,56E+18	https://p16-
ini bau aroma	7	0	2025-10-0	ora.et.lab	https://wv	7,41E+18	7,56E+18	https://p16-
hahaha ujung	15	0	2025-10-0	s.rama.dh	https://wv	7,24E+18	7,56E+18	https://p16-
Langsung pad	15	0	2025-10-0	andira143	https://wv	6,86E+18	7,56E+18	https://p16-
Ribet amat pa	8	0	2025-10-0	feriyudian	https://wv	7,44E+18	7,56E+18	https://p16-

Gambar 2. Hasil pengambilan data menggunakan metode Scraping

C. Data Preparation

Dalam tahapan ini data data akan dimasukan kedalam kategori - kategori yang telah ditentukan. Positif bagi komentar Tiktok yang mendukung dan setuju terhadap rencan kebijakan tersebut, Negatif bagi komentar yang menunjukan tanggapan buruk dan tidak setuju terhadap rencana kebijakan tersebut, dan netral bagi komentar yang tidak menujuru ke arah positif dan negative terhadap rencana kebijakan tersebut. Data preparation mmiliki dua tahapan yaitu Data Preprosessing dan Pembobotan TF-IDF.(Fatmawati, 2024)

1) Data Preprosessing

Data Preprocessing adalah tahap yang dilakukan guna menyelesaikan beberapa masalah seperti noisy data, data redundansi, nilai data yang hilang dan lain lain(Koukaras & Tjortjis, 2025). Langkah - langkah dalam data preprocessing adalah:

a) Case Folding dimana semua karakter dalam dokumen atau kalimat diubah menjadi huruf kecil

b) Tokenizing yang mana merupakan proses pemisahan suatu rangkain karakter berdasarkan karakter spasi, dan pada waktu yang bersamaan dilakukan juga penghapusan karakter tertentu

c) Filtering dimana kata - kata penting di ambil dari hasil term

d) Stemming yang mana digunakan untuk menyegerakan kata dan memperkecil daftar kata pada data latin(Hermawan et al., 2020)

2) Pembobotan TF-IDF

Pembobotan TF-IDF adalah metode yang digunakan untuk meneliti seberapa sering suatu kata muncul dalam suatu dokumen.ujuan dari TF-IDF ada dua yaitu:

a) TF (Term Frequency): menghitung seberapa sering suatu kata/token yang muncul dalam dokumen. Rumus TF sesuai dengan persamaan berikut ini:

$$F(tn,dn) = f(tn,dn)$$

Jadi $f(tn,dn)$ mendefinisikan jumlah kemunculan term-n pada sebuah dokumen-n

b) IDF (Invers Document Frequency): menilai seberapa penting suatu kata/token. Rumus dari IDF sesuai dengan persamaan berikut ini:

$$IDF(tn) = log\frac{D}{df(t)}$$

Jadi $log\frac{D}{df(t)}$ menjelaskan jumlah dokumen pada datased-D akan dibagi dengan jumlah dokument yang mengandung term-df(t)

Lalu menghitung keseluruhan nilai TF-IDF dengan mengkalikan nilai keduanya. Rumus sesuai dengan persamaan berikut:

$$W_{td} = TF(tn,dn) \times IDF(tn)$$

Keterangan :

W : Bobot kata-t terhadap dokument ke-d
t : Term atau kata ke-n dari kata yang di cari
d : Dokument ke-n
TF : Frekuensi/jumlah sebuah kata dalam satu dokumen
IDF : Inverse Document Frequency
D : Total Dokumen
Df : Jumlah dokumen yang mengandung Term/kata yang di cari

D. Modeling

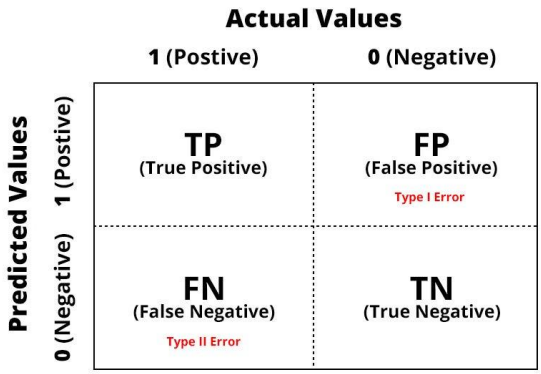
Pada tahap ini data dibagi menjadi data training dan data testing, data training merupakan data yang digunakan untuk mempelajari pola dan karakteristik data, sedangkan data testing merupakan data yang digunakan untuk menguji model yang telah mempelajari pola dan karakteristik data. Data training dan data testing ini

mendapatkan 70% data training dan juga 92% data testing. Metode yang digunakan untuk klasifikasi ini adalah metode klasifikasi Naive Bayes dan Support Vector Machine (SVM)

E. Evaluation

Pada tahap ini dilakukan pembelajaran data untuk mengenali pola dari data teks yang diolah dan diberi bobot. Setelah model mengenali pola di setiap kelas, model dapat mengklasifikasikan data baru. Kemudian setelah data diuji dan diolah maka akan dilakukan pengujian terhadap performa dan efektifitas dari sistem menggunakan Confusion Matrix.

Confusion Matrix adalah sebuah tabel statistik yang digunakan untuk menilai performa model klasifikasi. Tabel ini membandingkan hasil prediksi model (predicted) dengan keadaan sebenarnya (actual) dari data yang diuji.(Vanacore et al., 2024) Komponen utama dari Confusion Matrix dapat dilihat pada Gambar 3.



Gambar 3. Confusion Matrix

True Positives (TP): memprediksi data positif dengan benar.

True Negatives (TN): memprediksi data positif dengan benar.

False Positive (FP): Data negatif namun diprediksi sebagai data positif.

False Negative (FN): Data positif tetapi diprediksi sebagai data negatif.

F. Deployment

Dalam ini tahapan hasil dari sistem akan digunakan oleh masyarakat secara luas dan lembaga pemerintah yang berperan dalam perancangan rencana kebijakan terutama pada rencana penerapan system balik nama pada jual beli ponsel agar mempermudah dalam klasifikasi tanggapan masyarakatar terhadap rencana kebijakan kedepannya.

4. HASIL DAN PEMBAHASAN

Pada tahapan ini berisi pembahasan dan hasil dari proses implementasi yang selesai dilakukan. Penelitian ini memiliki tujuan untuk memberi klasifikasi sentiment masyarakat pada media sosial TikTok mengenai rencana kebijakan penerpan system balik nama pada jual beli ponsel bekas menggunakan metode algoritma Naive Bayes dan Support Vector Machine (SVM). Data dikumpulkan menggunakan teknik scrapping dengan menggunakan aplikasi web Apify dengan data yang bersumber dari media sosial Tiktok. Data yang terkumpul dibagi menjadi 3 kategori, yaitu positif, negatif dan netral. Acuan yang digunakan dalam mengukur tingkat keberhasilan model adalah dari nilai akurasi yang dihasilkan dengan evaluasi yang dilakukan menggunakan Confusion Matrix dalam menentukan nilai akurasi.

A. Pengumpulan Data

Pada penelitian ini pengumpulan data dilakukan menggunakan aplikasi apify. Data yang digunakan diambil dari media sosial TikTok, dengan cara scrapping komentar yang terdapat pada konten yang berkaitan atau membahas tentang topik yang terkait sebanyak 2000 data. Contoh data yang berhasil dikumpulkan melalui scrapping dapat dilihat pada Tabel 1.

Tabel 1. Contoh Hasil Scrapping

No	Komentar	CreateTime	UsernameID
1	bayar apa gratis????	2025-10-04T05:01:27.000Z	elevatemind99
2	adaaaa aja cara malakin rakyat	2025-10-04T01:52:20.000Z	bastianahmad01
3	ada aja lu ide nya nyari duit, heran gue	2025-10-06T12:43:58.000Z	dama.ard
4	semakin rumit dan kalau hilang surat lengkap pun tidak mau membantu, dan bau bau mau dipajak	2025-10-04T07:18:42.000Z	starboy_258
5	balik nama bolah, asalkan tanpa pajak	2025-10-04T07:53:12.000Z	terserahh609

B. Pelebelan Data

Pada tahapan ini sentiment data yang telah didapat dibagi menjadi 3 label yaitu positif, negative, dan netral. Peroses pelebelan ini berlangsung selama 1 bulan dengan cara pelebelan mandiri, dan kemudian diserahkan kepada guru ahli untuk di kroscek mengenai ketepatan dalam pelebelanya sebelumnya. Contoh hasil dari proses pelabelan data bisa dilihat pada Tabel 2.

Tabel 2. Contoh Hasil Pelabelan Data

No	Teks	Label
1	bayar apa gratis????	netral
2	adaaaa aja cara malakin rakyat	negatif
3	ada aja lu ide nya nyari duit, heran gue	negatif
4	semakin rumit dan kalau hilang surat lengkap pun tidak mau membantu, dan bau bau mau dipajak	negatif
5	balik nama boleh, asalkan tanpa pajak	positif

Dari peroses pelebelen ini diketahui bahwasanya dari 3 kategori yang paling mendominasi adalah sentimen netral dengan jumlah sebanyak 1562 komentar, kemudian sentiment negatif sebanyak 376 komentar, dan yang paling terakhir sentimen positif sebanyak 62 komentar. Untuk hasil dari perbandingan ketiga kategori dapat dilihat pada Tabel 3.

Tabel 3. Data Yang Terkumpul

NO	**Kelas**	**Jumlah**
1	Netral	1562
2	Negatif	376
3	Positif	62
Jumlah		2000

C. Preprosesing

Pada proses ini dilakukan pemrosesan pada data teks untuk dibersihkan dan diseragamkan sehingga dapat memaksimalkan klasifikasi program kepada hasil yang diinginkan. pada proses ini dilakukan beberapa pemrosesan pada data teks yang sebagai berikut :

1) Case Folding

Pada proses ini data teks diseragamkan menjadi huruf kecil dan dibersihkan dengan beberapa pengaturan perubahan yang dibutuhkan untuk memaksimalkan proses klasifikasi dari program. Perubahan yang dilakukan adalah mengubah isi data menjadi huruf kecil, menghapus hyperlink, menghapus tanda koma, menghapus angka, menghapus semua karakter yang bukan huruf dan spasi. Tabel 4 menunjukan hasil perubahan pada data mentah hingga perubahan setelah dilakukan Case Folding.

Tabel 4. Hasil Case Folding

Data Awal	Case Folding
Gabut amat para pejabat ini , emang gak ada kerjaan lain . Inget lu juga warga RI , kok bisanya apa apa dibikin ruwet.	gabut amat para pejabat ini , emang gak ada kerjaan lain . inget lu juga warga ri , kok bisanya apa apa dibikin ruwet.

2) Tokenizing

Pada tahapan tokenizing, setiap teks akan dibagi menjadi bagian yang lebih kecil, bagian – bagian tersebut disebut token. Tujuan dari tokenizing adalah untuk memecah teks menjadi beberapa bagian yang lebih mudah di peroses untuk analisis. Hasil dari Tokenizing dapat dilihat pada Tabel 5.

Tabel 5. Hasil Tokenizing

Data Awal	Tokenizing
kenapa Mentri Prabowo semuanya gak ada yg punya ide masuk akal sih	kenapa', 'mentri', 'prabowo', 'semuanya', 'gak', 'ada', 'yg', 'punya', 'ide', 'masuk', 'akal', 'sih'

3) Filtering

Pada tahapan filtering data teks disaring dari kata-kata umum (stopword) yang mana kurang berguna untuk analisi sentiment. Pengklasifikasian stopword menggunakan library stopword bahasa indonesia yang ada pada library NTLK (Natural Language Tool Kit). Hasil dari filtering pada Tabel 6.

Tabel 6. Hasil Filtering

Data Awal	**Filtering**
semakin rumit dan kalau hilang surat lengkap pun tidak mau membantu, dan bau bau mau dipajak	'rumit', 'hilang', 'surat', 'lengkap', 'membantu', 'bau', 'bau', 'dipajak'

4) Stemming

Pada tahapan steaming kata – kata dalam teks di ubah ke kata dasarnya dengan menghilangkan akhiran atau awalan tertentu. Tujuan dari proses steaming sendiri adalah untuk mengurangi variasi dalam kata atau entitas yang sama. Hasil dari peroses Stemming bisa di lihat pada tabel 7.

Tabel 7. Hasil Stemming

Data Awal	Stemming
dunia lain sedang berperang dalam dunia keuangan digital, disini masih ngurus balik nama hp bekas.	dunia lain sedang perang dalam dunia uang digital sini masih urus balik nama hp bekas

D. Pembobotan TF-IDF

Data yang telah melewati proses preprosessing diberi bobot menggunakan teknik pembobotan TF-IDF. Teknik ini digunakan dalam pemrosesan teks untuk menilai seberapa penting sebuah kata dalam suatu dataset. Teknik ini merupakan teknik yang paling umum untuk pemrosesan kata pada penambangan data. Manfaat dalam penggunaan pembobotan ini dapat membantu mengatasi masalah ketidakseimbangan kata kemudian pemberian bobot kata yang lebih akurat memiliki dampak yang besar pada analisis sentimen. Hasil dari pembobotan TF-IDF dapat di lihat pada Gambar.4

Top 15 Kata Berdasarkan Bobot TF-IDF:

	Kata/Token	Bobot TF-IDF (Mean)
0	yang	0.024563
1	ada	0.020880
2	tidak	0.020517
3	di	0.019952
4	kebijakan	0.019597
5	pajak	0.018654
6	hp	0.017867
7	ini	0.014311
8	aja	0.013249
9	komdigi	0.012747
10	ribet	0.012353
11	kalau	0.012191
12	makin	0.011727
13	nya	0.011454
14	rakyat	0.011099

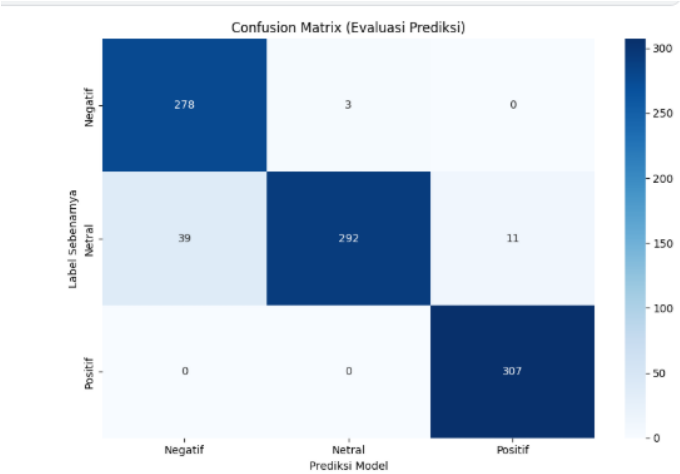
Gambar 4. Hasil Pembobotan TF-IDF

E. Evaluation

Setelah tahapan preprocessing dan pembobotan menggunakan TF-IDF pemodelan dilanjutkan menggunakan algoritma Naive Bayes dan Support Vector Machine dengan data training sebanyak 2000 data dan data testing sebanyak 4647 data . Kemudian data yang sudah melewadi tahap klasifikasi selanjutnya akan di evaluasi menggunakan Confusion Matrix

1) Hasis Evaluasi Confusion Matrix Naive Bayes

Data yang melalui proses klasifikasi menggunakan metode Naive Bayes selanjutnya akan dievaluasi menggunakan Confussion Matrix



Gambar 5. Hasil Confussion Matrix Naive Bayes

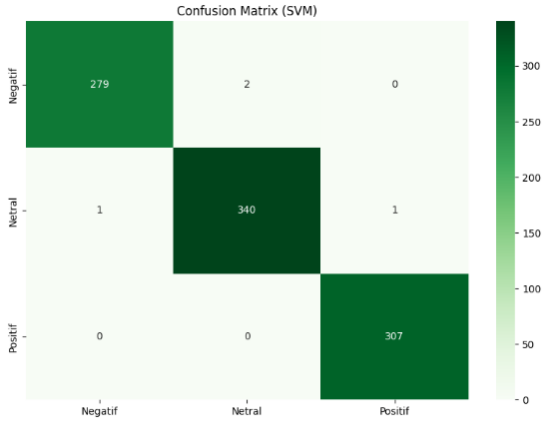
Pada Gambar 5 menunjukan bahwasanya Naive Bayes berhasil megklasifikasin data dengan nilai True Negatif sebesar 278, True Neutral sebesar 292, dan True Positif sebesar 307. Dan juga False Negatif sebesar 3, False Neutral 50, dan False Positif sebesar 0. Dengan memiliki akurasi sebesar 94.30% dengan rincian pada Gambar 6.

```
Akurasi Model: 94.30%

Laporan Klasifikasi:
              precision    recall  f1-score   support

     Negatif       0.88      0.99      0.93       281
      Netral       0.99      0.85      0.92       342
     Positif       0.97      1.00      0.98       307

    accuracy                           0.94       930
   macro avg       0.94      0.95      0.94       930
weighted avg       0.95      0.94      0.94       930
```

Gambar 6. Hasil Akurasi Confussion Matrix Naive Bayes

2) Hasis Evaluasi Confusion Matrix Support Vector Machine

Data yang melalui proses klasifikasi menggunakan metode Support Vector Machine (SVM) selanjutnya akan dievaluasi menggunakan Confussion Matrix



Gambar 7. Hasil Confussion Matrix Support Vector Machine

Pada Gambar 7 menunjukan bahwasanya Support Vector Machine (SVM) berhasil megklasifikasin data dengan nilai True Negatif sebesar 279, True Neutral sebesar 340, dan True Positif sebesar 307. Dan juga False Negatif sebesar 2, False Neutral 2, dan False Positif sebesar 0. Dengan memiliki akurasi sebesar 99.57% dengan rincian pada Gambar 8.

```
Akurasi Model SVM: 99.57%

Laporan Klasifikasi SVM:
              precision    recall  f1-score   support

     Negatif       1.00      0.99      0.99       281
      Netral       0.99      0.99      0.99       342
     Positif       1.00      1.00      1.00       307

    accuracy                           1.00       930
   macro avg       1.00      1.00      1.00       930
weighted avg       1.00      1.00      1.00       930
```

Gambar 8. Hasil Akurasi Confussion Matrix Support Vector Machine

Deploymen

1. Visualisai WordCloud

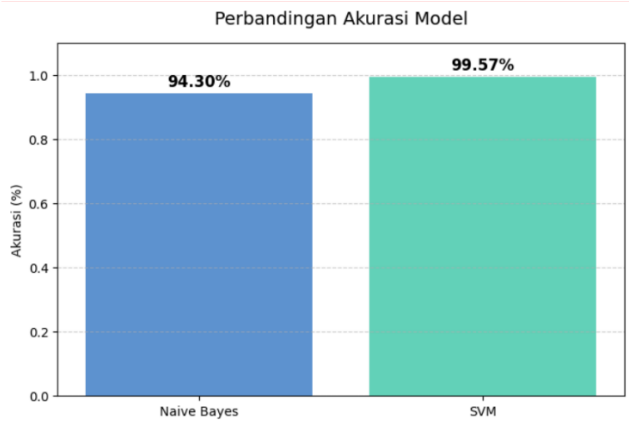
Data yang telah melalui proses preprocessing hingga evaluasi di presentasikan dengan menyoroti seberap penting kata – kata itu dalam data. Kata – kata yang sering muncul akan terlihat lebih pekat warnanya daripada kata yang lain. Hasil dari representasi Wordcloud dapat dilihat pada Gambar 9, Dan Gambar 10.

Gambar 9. Visualisasi WordCloud Sentimen Negatif

Gambar 10. Visualisasi WordCloud Sentimen Positif.

2. Hasil Perbandingan Klasifikasi Naive Bayes dan Support Vector Machine (SVM)

Setelah model klasifikasi Naive Bayes dan Suppor Vector Machine(SVM) di evaluasi menggunakan Confussion Matrix, Model Klasifikai Support Vector Machine (SVM) memiliki akurasi yang lebih tinggi dari model Klasifiaki Naive Bayes. Perbandingan Akurasi dapat dilihat pada Gambar 11 dan Gambar 12



Gambar 11. Perbandingan Akurasi Model

	Precision		Recall		F1-Score	
Model	**Naive Bayes**	**SVM**	**Naive Bayes**	**SVM**	**Naive Bayes**	**SVM**
Label						
Negatif	87.70%	99.64%	98.93%	99.29%	92.98%	99.47%
Netral	98.98%	99.42%	85.38%	99.42%	91.68%	99.42%
Positif	96.54%	99.68%	100.00%	100.00%	98.24%	99.84%

Gambar 12. Perbandingan Ditail Klasifikasi

5. KESIMPULAN DAN SARAN

Kesimpulan

Pada penelitian ini dapat disimpulkan :

a. Model yang dibuat dapat mengklasifikasina sentimen masyarakat mengenai Rencana Penerapan Kebijakan Sistem Balik Nama Pada Jual Beli Ponsel Bekas. Dengan data yang didapat melalui komentar konten terkait di media sosial TikTok
b. Klasifikai menggunakan metode Naive Bayes memiliki akurasi sebesar 94.30%
c. Klasifikai menggunakan metode Support Vector Machine memiliki akurasi sebesar 99.57%
d. Model Klasifikai Support Vector Machine (SVM) memiliki akurasi yang lebih tinggi dari model Klasifiaki Naive Bayes

Saran

Dengan hasil yang didapat dari penelitian ini, peneliti menyadari bahwa masih banyaknya kekurangan dari penelitian ini, dengan begitu perlunya pengembangan lebih lanjut untuk mendapatkan hasil yang lebih maksimal, berikut beberapa saran yang dapat di ambil :

a. Agar menambah varian data yang lebih banyak dari berbagai media sosial yang berbeda.
b. Memodifikais pada tahapan Preprocessing supaya nantinya dapat memberi peningkatan akurasi dan klasifikasi
c. Mencoba untuk menambah lebih banyak model guna menghasilkan nailai akurasi yang berbeda, kemudian membandingkan nilai dari setiap model.

DAFTAR REFERENSI

Al Ahzuri, Z. A. K. E. (2024). Teorema Bayes; Statistika Matematika Kecerdasan Buatan dan Pembelajaran Mesin. 1(2), 9–14. https://cendekia.co/index.php/cendekia/article/view/3/3

Anggraini, D., Rahmawati, S., & Kurniawan, R. (2024). Natural Language Processing For Automatic Sentiment Analysis In Social Media Data. International Journal of Information Engineering and Science, 1(1), 16–19. https://doi.org/10.62951/ijies.v1i2.54

Cholifah Sastya, N., & Nugraha, D. I. (n.d.). Penerapan Metode CRISP-DM dalam Menganalisis Data untuk Menentukan Customer Behavior di MeatSolution. http://ejournal.unis.ac.id/index.php/UNISTEK

Datau, R. R., Ichsanuddin Nur, D., Juliputra, F., Eka, A., Haryanto, P., & Fauzi, I. N. (2025). ANALISIS SENTIMEN TIKTOK TERHADAP COFFEE SHOP X DAN IMPLIKASINYA TERHADAP STRATEGI PEMASARAN DIGITAL. JAMBURA, 8(1). http://ejurnal.ung.ac.id/index.php/JIMB

Duei Putri, D., Nama, G. F., & Sulistiono, W. E. (2022). Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier. Jurnal Informatika Dan Teknik Elektro Terapan, 10(1). https://doi.org/10.23960/jitet.v10i1.2262

Fatmawati, D. (2024). Sentiment Classification of IT Service Feedback via TF-IDF. COGITO Smart Journal, 10(2). https://cogito.unklab.ac.id/index.php/cogito/article/download/701/358?utm_source=chatgpt.com

Franciska Mey Dina, D., Haryanti, T., & Amirul Haq, M. (2025). ANALISIS SENTIMEN TERHADAP KOMENTAR PADA MEDIA SOSIAL TIKTOK YANG BERPOTENSI MENYEBABKAN DEPRESI MENGGUNAKAN METODE NAIVE BAYES. In Jurnal Ilmiah Computing Insight (Vol. 7, Issue 1).

Handhayani, T. (2025). Perbandingan Kinerja Naïve Bayes dan Random Forest dalam Mendeteksi Berita Palsu. In Jurnal Informatika Sunan Kalijaga) (Vol. 10, Issue 2). MEI. https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset?select=True.csv.

Hermawan, L., Ismiati, M. B., Bangau, J., 60, N., & Charitas, M. (2020). Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval. TRANSFORMATIKA, 17(2), 188–199.

Huda Ovirianti, N., Zarlis, M., & Mawengkang, H. (2022). Support Vector Machine Using A Classification Algorithm. Jurnal Dan Penelitian Teknik Informatika, 6(3). https://doi.org/10.33395/sinkron.v7i3

Husen, R. A., Astuti, R., Marlia, L., Rahmaddeni, R., & Efrizoni, L. (2023). Analisis Sentimen Opini Publik pada Twitter Terhadap Bank BSI Menggunakan Algoritma Machine Learning. MALCOM: Indonesian Journal of Machine Learning and Computer Science, 3(2), 211–218. https://doi.org/10.57152/malcom.v3i2.901

Khoerunnisa, S., Shiddieq, D. F., & Nurhayati, D. (2025). Penerapan Algoritma Naive Bayes dengan Teknik TF-IDF dan Cross Validation untuk Analisis Sentimen Terhadap Starlink. MALCOM: Indonesian Journal of Machine Learning and Computer Science, 5(2), 566–577. https://doi.org/10.57152/malcom.v5i2.1852

Koukaras, P., & Tjortjis, C. (2025). Data Preprocessing and Feature Engineering for Data Mining: Techniques, Tools, and Best Practices. In AI (Switzerland) (Vol. 6, Issue 10). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/ai6100257

Rofiqi, M. A., Fauzan, Abd. C., Agustin, A. P., & Saputra, A. A. (2019). Implementasi Term-Frequency Inverse Document Frequency (TF-IDF) Untuk Mencari Relevansi Dokumen Berdasarkan Query. ILKOMNIKA: Journal of Computer Science and Applied Informatics, 1(2), 58–64. https://doi.org/10.28926/ilkomnika.v1i2.18

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. Procedia Computer Science, 181, 526–534. https://doi.org/10.1016/j.procs.2021.01.199

Septiani, D., & Isabela, I. (n.d.). SINTESIA: Jurnal Sistem dan Teknologi Informasi Indonesia ANALISIS TERM FREQUENCY INVERSE DOCUMENT FREQUENCY (TF-IDF) DALAM TEMU KEMBALI INFORMASI PADA DOKUMEN TEKS.

Simson Lende, A., Pati, G. K., & Adis, A. (2024). Analisis Sentimen Komentar Pengunjung Air Terjun Waikelo Sawa Menggunakan Metode Naive Baiyes Classifier. Jurnal Ilmu Komputer Dan Bisnis, 15(2), 42–50. https://doi.org/10.47927/jikb.v15i2.764

Vanacore, A., Pellegrino, M. S., & Ciardiello, A. (2024). Fair evaluation of classifier predictive performance based on binary confusion matrix. Computational Statistics, 39(1), 363–383. https://doi.org/10.1007/s00180-022-01301-9

Zakaria, Z., Kusrini, K., & Ariatmanto, D. (2023). Sentiment Analysis to Measure Public Trust in the Government Due to the Increase in Fuel Prices Using Naive Bayes and Support Vector Machine. International Journal of Artificial Intelligence & Robotics (IJAIR), 5(2), 54–62. https://doi.org/10.25139/ijair.v5i2.7167

Zhang, B., Xiao, J., Yan, Hao, Yang, Liziqiu, & Qu, P. (2024). Review of NLP Applications in the Field of Text Sentiment Analysis. Online) |, 2(3). https://doi.org/10.5281/zenodo.11403180