

# SENTIMENT ANALYSIS OF TELEGRAM APPLICATION USER SATISFACTION ON GOOGLE PLAY STORE USING NAÏVE BAYES, LOGISTIC REGRESSION AND SVM

## ANALISIS SENTIMEN KEPUASAN PENGGUNA APLIKASI TELEGRAM PADA GOOGLE PLAY STORE MENGGUNAKAN METODE NAÏVE BAYES, LOGISTIC REGRESSION DAN SVM

Adellia Septiani Putri<sup>1</sup>, Anief Fauzan Rozi<sup>2</sup>

<sup>1,2</sup> Universitas Mercu Buana Yogyakarta, Daerah Istimewa Yogyakarta, Indonesia

*email:* padellia62@gmail.com<sup>1</sup>, anief@mercubuana-yogya.ac.id<sup>2</sup>

**Abstract** - Sentiment analysis is a technique for finding out how people feel about something and putting the polarity of text into groups of documents or words so that they can be labeled neutral, positive, or negative. We will use the Naïve Bayes algorithm, logistic regression, and SVM to conduct sentiment analysis on how happy Telegram app users are. The purpose of this study is to see what people who use the app think and group their thoughts into three groups: neutral, positive, and negative. The three methods' results will be compared to see which is most accurate for this study. The results of this sentiment analysis show that many users are dissatisfied with the verification code they need to register or log in to their accounts. This makes it difficult for new users to get the verification code because the app itself sends it. The SVM approach has an accuracy value of 89.73%, which means it is more accurate in this study. The Naïve Bayes approach is accurate by 75.61%, while the logistic regression method is accurate by 87.49%.

**Keywords** - Sentiment Analysis, Logistic Regression, Naïve Bayes, SVM, Telegram.

**Abstrak** - Analisis Sentimen adalah teknik untuk mengetahui perasaan orang terhadap sesuatu dan memasukkan polaritas teks ke dalam kelompok dokumen atau kata sehingga dapat diberi label netral, positif, atau negatif. Kami akan menggunakan algoritma Naïve Bayes, Regresi Logistik, dan SVM untuk melakukan analisis sentimen tentang seberapa bahagiannya pengguna aplikasi Telegram. Tujuan dari studi ini adalah untuk melihat apa yang dipikirkan orang-orang yang menggunakan aplikasi tersebut dan mengelompokkan pemikiran mereka ke dalam tiga kelompok: netral, positif, dan negatif. Setelah itu, temuan akurasi yang lebih baik dari ketiga pendekatan tersebut akan dibandingkan untuk mengetahui mana yang lebih akurat untuk studi ini. Hasil analisis sentimen ini menunjukkan bahwa banyak pengguna tidak puas dengan kode verifikasi yang mereka perlukan untuk mendaftar atau masuk ke akun mereka. Hal ini membuat pengguna baru kesulitan memperoleh kode verifikasi karena aplikasi itu sendiri yang mengirimkannya. Pendekatan SVM memiliki nilai akurasi sebesar 89,73%, yang berarti lebih akurat dalam studi ini. Pendekatan Naïve Bayes akurat sebesar 75,61%, sedangkan metode Regresi Logistik akurat sebesar 87,49%.

**Kata Kunci:** Analisis Sentimen, Logistic Regression, Naïve Bayes, SVM, Telegram.

## I. PENDAHULUAN

Teknologi saat ini berkembang sangat cepat. Dimana perusahaan-perusahaan teknologi terus-menerus memberi inovasi demi kenyamanan pengguna media social[1]. Perkembangan teknologi ini menjadi hal penting bagi perkembangan dan perubahan cara kita dalam berbagi informasi maupun berinteraksi dengan orang lain. Hingga saat ini, sebagian besar sangat bergantung pada teknologi seperti media sosial, baik itu untuk pekerjaan, berinteraksi dengan teman, dan melakukan aktivitas dengan banyak orang dalam sebuah grup[2]. Salah satu aplikasi yang sangat berkembang pesat hingga saat ini ialah Telegram. Aplikasi telegram mempunyai beberapa keunggulan dalam hal fleksibilitas, aksesibilitas, dan berbagai fitur yang terdapat pada aplikasi seperti grup, channel, pesan rahasia, polling, hingga *live streaming*[3].

Di balik keuntungan yang terdapat dari aplikasi ini pastinya mempunyai kekurangan menurut opini pengguna, hal tersebut dapat memberikan respon yang beragam dari pengguna aplikasi, seperti merasa kecewa dengan beberapa fitur yang terdapat di aplikasi. Namun, karena banyaknya ulasan dari pengguna yang diterima membuat pengembang aplikasi kesulitan dalam membaca semua ulasan dan juga akan banyak membuang waktu dan tenaga bila diharuskan membaca dan menganalisis semuanya secara manual. Cara ini, sangat tidak efisiensi dikarenakan tidak terlalu membantu dengan cepat. Untuk membantu pengembang aplikasi, diperlukan pendekatan analisis sentimen untuk memahami opini-opini pengguna aplikasi Telegram[4]. Analisis sentiment sendiri ialah proses implementasi teknologi pemrosesan bahasa alami dan analisis teks untuk mengenali pendapat subjektif di dalam sebuah dokumen[5]. Penelitian ini bertujuan untuk mengimplementasikan dan membandingkan hasil akurasi yang lebih tinggi dari tiga metode yang akan digunakan yaitu metode *Naïve Bayes*, *Logistic Regression*, dan SVM, juga dapat mengetahui berapa banyak ulasan-ulasan pengguna setelah dikelompokkan menjadi tiga bagian dan menganalisis kata yang sering muncul dalam ulasan. Penelitian ini diharapkan dapat mengidentifikasi algoritma mana yang paling efektif untuk menganalisis kepuasan pengguna aplikasi Telegram dan dapat menjadi dasar bagi pengembang aplikasi Telegram dalam mengambil keputusan dalam peningkatan fitur, perbaikan layanan, serta respon terhadap pengguna.

## II. SIGNIFIKANSI STUDI

### A. Penelitian Terdahulu

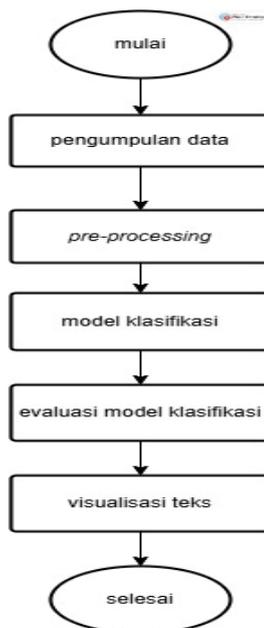
Friska et al. [6] melakukan penelitian pada tahun 2023 yang berjudul "Analisis Sentimen Aplikasi Tiktok Menggunakan Algoritma Naïve Bayes dan SVM." Terdapat 2000 ulasan dalam survei ini, dan 76,7% di antaranya positif dan 23,3% negatif. Pendekatan Naïve Bayes memiliki nilai akurasi sebesar 79%, sedangkan metode SVM mendapatkan nilai akurasi sebesar 84%, yang lebih tinggi dibandingkan metode *Naïve Bayes*. Pada tahun 2022, Andry Novantika dan kawan-kawan[7] melakukan penelitian berjudul "Analisis Sentimen Ulasan Pengguna Aplikasi Google Meet Video Conference Menggunakan Metode SVM dan Regresi Logistik". Penelitian ini menggunakan empat kernel SVM, yaitu *Linear*, *Radial Basic Function (RBF)*, *Sigmoid*, dan *Polinomial*. Angka akurasi untuk masing-masing kernel dalam penelitian ini adalah 87,02%, 84,59%, 86,63%, dan 71,12%, sedangkan untuk teknik Regresi Logistik adalah 85,17%. Kesimpulan yang diperoleh adalah algoritma terbaik dengan akurasi tinggi adalah algoritma SVM dengan kernel Linear, yaitu sebesar 87,02%. Fadhilah dan yang lainnya pada tahun 2023[8] melakukan penelitian berjudul "Perbandingan Algoritma *K-Nearest Neighbor* dan Regresi Logistik dalam Analisis Sentimen Vaksinasi Covid-19 di Media Sosial *Twitter* dengan Pelabelan Vader dan *Textblob*." Hasil penelitian tersebut mengungkapkan bahwa algoritma Regresi Logistik dengan pelabelan VaderSentiment lebih akurat dibandingkan teknik *K-Nearest Neighbor* dengan pelabelan TextBlob dan VaderSentiment[9].

Pada penelitian yang dilakukan oleh Fakhri, dkk[10] tahun 2024 dengan judul “Pengaruh RFE Terhadap Logistic Regression dan SVM pada Analisis Sentimen Hotel Shangri-La Surabaya. Tujuan dari penelitian ini untuk mengetahui pengaruh dalam penerapan RFE terhadap kinerja kedua metode yang digunakan untuk menganalisis sentimen ulasan tamu di Hotel Shari-La Surabaya. Hasil dari penelitian ini menghasilkan bahwa pada metode Logistic Regression yang menggunakan RFE, terdapat peningkatan performa yang signifikan dalam presisi, sensitivitas, F1-Score, dan akurasi, dengan peningkatan rata-rata 9%. Sedangkan untuk metode SVM, mempunyai peningkatan yang lebih signifikan dengan peningkatan rata-rata 14%, dengan ini menunjukkan bahwa penerapan RFE secara efektif meningkatkan kualitas prediksi dua model yang digunakan dalam melakukan analisis sentiment ulasan hotel. Dewi et al.[11] menyelesaikan penelitian pada tahun 2023 yang berjudul “Analisis Sentimen Pengguna Media Sosial Twitter terhadap Perokok di Indonesia”. Penelitian tersebut menunjukkan bahwa 40,25% pengguna Twitter setuju bahwa ada perokok di Indonesia, sedangkan 59,74% tidak setuju. Saat menggunakan 60% data latih dan 40% data uji, Naïve Bayes memiliki skor akurasi tertinggi yaitu 62,1%. Pada penelitian yang dilakukan oleh Nandi, dkk[12] tahun 2024 dengan judul “Perbandingan Logistic Regression dengan Random Forest dalam Memprediksi Sentimen Pada IMDb Movie Review”. Hasil dari penelitian ini ialah kedua metode mempunyai hasil yang memuaskan, tetapi untuk metode Logistic Regression cenderung lebih baik dalam memprediksi sentimen dataset penelit. Pada penelitian yang dilakukan oleh Haris, dkk tahun 2024[13] dengan judul “Komparasi Logistic Regression dan Random Forest dalam Analisis Sentimen Ulasan Aplikasi DISDUKCAPIL”. Pada penelitian ini melakukan analisis sentiment dengan menggunakan dua metode yaitu Logistic Regression dan Random Forest, dimana nantinya setelah melakukan perbandingan kedua metode dengan data uji sebanyak 18810 data ulasan pengguna di Google Play Store. Hasil dari penelitian ini masing-masing metode memperoleh nilai sebagai berikut: akurasi 90%, presisi 91%, recall 89%, dan f1-score 90% untuk metode Logistic Regression, sedangkan hasil nilai dari metode Random Forest ialah 89% akurasi, 92% presisi, 86% recall, dan f1-score 89%. Pada penelitian yang dilakukan oleh Alisyia, dkk[14], tahun 2024 dengan judul “Analisis Sentimen Twitter Ibu Kota Negara Nusantara Menggunakan Algoritma Naïve Bayes, Logistic Regression, dan K-Nearest Neighbors”. Penelitian ini melibatkan pengolahan data teks dan word cloud, serta penerapan tiga metode, yaitu Naïve Bayes, Logistic Regression, dan K-Nearest Neighbors untuk menganalisis sentimen. Hasil dari penelitian ini Tweet bersifat positif lebih banyak, disusul oleh netral dan negatif. Akurasi dari metode Naïve Bayes lebih unggul dengan nilai akurasi 94.80%, dengan presisi 93.99%, recall 96.09%, dan f1-Score 95.03%.

Pada tahun 2024[15], Yuning dan yang lainnya melakukan studi yang disebut "Analisis Sentimen Proyek Strategis National Food Estate Menggunakan Algoritma Naïve Bayes, Regresi Logistik, dan SVM." Tujuan studi ini adalah untuk melihat bagaimana perasaan masyarakat terhadap Proyek Strategis National Food Estate dengan membandingkan tiga metode: Naïve Bayes, Regresi Logistik, dan SVM. Studi ini juga melakukan eksperimen dengan ekstraksi fitur TF-IDF dan Word2Vec. Hasilnya menunjukkan bahwa TF-IDF lebih baik dalam menemukan fitur relevan yang dapat meningkatkan kinerja klasifikasi daripada Word2Vec. Pendekatan Regresi Logistik + TF-IDF adalah yang terbaik dari ketiganya, dengan nilai akurasi 74%. Metode SVM + TF-IDF berada di urutan kedua, dengan nilai akurasi 73%, dan metode Naïve Bayes + TF-IDF berada di urutan ketiga, dengan nilai akurasi 72%[15]. Dari beberapa penelitian terdahulu, algoritma *Naïve Bayes*, *Logistic Regression*, dan SVM terbukti sering digunakan dalam penelitian analisis sentimen dan menunjukkan performa yang baik dalam penelitian tersebut, tergantung dengan karakteristik data dan konteks permasalahan. Oleh karena itu, dalam penelitian ini ketiga metode tersebut akan digunakan dalam penelitian analisis sentimen kepuasan pengguna terhadap aplikasi Telegram di *Google Play Store*. Hasil dari penelitian ini dapat mengetahui opini-opini pengguna yang nantinya akan membantu pengembang aplikasi untuk mengembangkan atau memperbaiki fitur-fitur yang ada di aplikasi Telegram.

### B. Metodologi Penelitian

Untuk mencapai tujuan penelitian ini, akan melalui beberapa proses yang harus dilakukan. Berikut adalah tahapan-tahapan yang akan dilakukan selama penelitian terdapat di Gambar 1.



Gambar 1. Tahapan Penelitian

- 1) *Pengumpulan Data*, Pada langkah awal ini, *Web Scrapping* akan menggunakan bahasa komputer *Python* untuk mendapatkan ulasan pengguna aplikasi Telegram dari *Google Play Store*. Ulasan baru yang diambil selama proses scrapping berbahasa Indonesia. *Web scraping* adalah proses mendapatkan data semi-terstruktur dari internet, yang biasanya dilakukan dengan HTML atau XHTML. Tujuannya adalah untuk mendapatkan informasi secara keseluruhan atau sebagian sesuai kebutuhan[16]
- 2) *Pre-Processing*, Selanjutnya, setelah melakukan *scrapping*, dataset akan dibersihkan terlebih dahulu sebelum dipakai untuk penelitian. *Pre-Processing* merupakan sebuah proses perubahan bentuk data terstruktur sembarang menjadi data yang terstruktur sesuai dengan kebutuhan dalam pemrosesan *text mining*.
- 3) *Model Klasifikasi*, Setelah dataset melalui tahapan *pre-processing*, selanjutnya ialah tahap pemodelan klasifikasi. Sebelum melakukan klasifikasi model, akan dilakukan pembobotan menggunakan TF-IDF. TF-IDF merupakan salah satu cara pemberian nilai bobot pada sebuah data yang telah dilakukan pengktrakan atau bisa disebut dengan pemberian nilai angka pada data yang tidak berbentuk angka. Di tahap ini, nilai hasil dari TF-IDF akan menentukan kepentingan atau makna suatu fitur atau kata dalam korpus. Berikut rumus dari TF-IDF[20]:

$$W dt = tf dt * ldf t (1)$$

Pada tahap ini, model klasifikasi yang dipakai ialah *Naïve Bayes*, *Logistic Regression*, dan SVM.

- a. *Naïve Bayes*, merupakan cara termudah untuk mengklasifikasikan sesuatu dibandingkan dengan pendekatan lainnya. Di sisi lain, pendekatan ini cukup akurat dan bekerja dengan baik, terutama pada basis data besar [4]. Ada dua langkah dalam metode ini: langkah pelatihan dan langkah pengujian. Langkah pertama adalah pelatihan, dan langkah kedua adalah pengujian, yang dilakukan dengan melihat probabilitas kumpulan data pelatihan[21]. Teorema Bayes juga mencakup strategi ini, yang mengandaikan bahwa data tidak terhubung secara statistik. Ini adalah rumus untuk metode *Naïve Bayes*[22]:

$$P(c | x) = P(x | h) * P(c) / P(x) \quad (2)$$

Dimana:

$P(c | x)$  : Probabilitas *posterior* pada kelas c saat diberikan atribut x

$P(c)$  : Probabilitas prior pada kelas c

$P(x | c)$  : Probabilitas pada atribut x saat diberikan kelas c

$P(x)$  : Probabilitas prior pada atribut x

- b. *Logistic Regression*, ialah metode analisis yang digunakan dalam mendeskripsi hubungan antar variabel terikat yang mempunyai dua kategori atau lebih dengan atau variabel bebas, dengan skala kategori atau kontinu[23]. Metode ini memakai fungsi dari logistik untuk mengubah nilai output menjadi probabilitas
- c. *Support Vector Machine (SVM)*, ialah sekumpulan metode *supervised learning* untuk membuat *hyperlane* pada proses klasifikasi, regresi, dan *outlier detection*. Salah satu penggunaannya ialah melakukan pengelompokan *text* dan *hypertext*. SVM mempunyai beberapa kelebihan, yaitu[24]:
- Efektif dalam *high dimensional space*,
  - Efektif dalam menyelesaikan masalah dalam jumlah dimensi lebih banyak dibandingkan dengan jumlah sampel.
  - Menggunakan subset titik pelatihan supaya lebih efisien.

SVM merupakan teknik pembelajaran dengan banyak kualitas yang diharapkan oleh metode ini. Metode ini juga mempunyai dasar teoritis yang kuat dan melakukan klasifikasi yang lebih akurat dibandingkan dengan metode-metode algoritma lainnya. Sudah banyak penelitian-penelitian yang menggunakan metode ini dan melaporkan bahwa metode SVM adalah metode paling akurat dalam melakukan sebuah klasifikasi teks dan juga sudah banyak digunakan untuk melakukan sebuah penelitian analisis sentimen. Rumus umum dari metode SVM dapat ditulis sebagai berikut[25]:

$$F(x) = \text{sign}(w \cdot x + b) \quad (3)$$

Dimana,  $f(x)$  merupakan fungsi prediksi,  $w$  merupakan vector normal *hyperplane*,  $x$  merupakan vector input, dan  $b$  merupakan bias atau *intercept*.

- d. *Evaluasi Model Klasifikasi*, Tahap selanjutnya ialah mengevaluasi model klasifikasi yang menggunakan *confusion matrix*. Tujuan dari tahapan ini ialah untuk mengetahui keefektifan model klasifikasi. Pengukuran kinerja model pada penelitian ini menggunakan hasil akurasi, *recall*, *precision*, dan *f1-score*. Berikut rumus untuk menghitung nilai akurasi, presisi, *recall*, dan *f1-score*:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (4)$$

$$\text{Precision} = \frac{TP}{TP+FP} * 100\% \quad (5)$$

$$\text{Recall} = \frac{TP}{TP+FN} * 100\% \quad (6)$$

$$\text{F1-Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} * 100\% \quad (7)$$

- e. *Visualisasi Teks*, Tahapan terakhir dalam penelitian ini ialah visualisasi teks yang nantinya akan menggunakan metode *word cloud* yang mempunyai fungsi dalam mengumpulkan informasi tentang kata-kata apa saja yang sering disebut oleh pengguna dalam mengeluarkan opininya.

### III. HASIL DAN PEMBAHASAN

#### A. Pengumpulan Data

Tahap pengumpulan data pada penelitian ini menggunakan bahasa python untuk melakukan sebuah web scrapping untuk mengambil data ulasan/opini pengguna aplikasi Telegram pada Google Play Store. Data yang diambil sebanyak 13.827 ulasan pengguna, dari tahun 2025 pada bulan Februari sampai bulan April. Pengambilan data ini menggunakan Google Collab. Berikut pada Gambar 2. Hasil dari web scrapping.

nama	tanggal	rating	komentar
Ope	19 April 2025	Diberi rating 1 bintang dari 5 bintang	mau daftar aja susah banget harus dapet kode d...
Ardil March	19 April 2025	Diberi rating 1 bintang dari 5 bintang	aneh kali telegram ini masa minta kode ga di d...
jevelline frience efendi wen	19 April 2025	Diberi rating 3 bintang dari 5 bintang	ga bisa nerima kode, dan tidak menggunakan VPN...
Saudah	19 April 2025	Diberi rating 5 bintang dari 5 bintang	good
Maulana Malik ibrahim	19 April 2025	Diberi rating 1 bintang dari 5 bintang	kode email kadaluarsa terus.. padahal baru beb...

Gambar 2. Hasil Pengumpulan Data

#### B. Pre-Processing

Tahap selanjutnya setelah mengumpulkan data-data yang akan digunakan untuk menganalisis sentimen ialah tahapan *pre-processing*. Tahapan *pre-processing* ialah tahapan paling penting dalam mengolah data sebelum digunakan untuk melakukan sebuah analisis sentimen. Tahapan ini digunakan untuk membersihkan data dan menyederhanakan data supaya dapat diproses lebih efektif oleh *machine learning*. Tahapan ini melewati beberapa proses, yaitu *case folding*, *cleansing*, *tokenization*, *filtering*, dan *stemming*. Dengan melewati tahapan ini, data yang nantinya akan digunakan ke dalam metode klasifikasi menjadi lebih terarah dan konsisten, sehingga dapat meningkatkan hasil akurasi dan performa metode. Berikut pada Gambar 3. Hasil dari tahapan *pre-processing*.

```

Teks asli:
mau daftar aja susah banget harus dapet kode d telegramnya sedangkan saya belum masuk si gimana

Setelah case folding:
mau daftar aja susah banget harus dapet kode d telegramnya sedangkan saya belum masuk si gimana

Setelah cleansing:
mau daftar aja susah banget harus dapet kode d telegramnya sedangkan saya belum masuk si gimana

Setelah tokenisasi:
['mau', 'daftar', 'aja', 'susah', 'banget', 'harus', 'dapet', 'kode', 'd', 'telegramnya', 'sedangkan', 'saya', 'belum', 'masuk', 'si', 'gimana']

Setelah normalisasi:
['mau', 'daftar', 'aja', 'susah', 'banget', 'harus', 'dapet', 'kode', 'd', 'telegramnya', 'sedangkan', 'saya', 'belum', 'masuk', 'si', 'gimana']

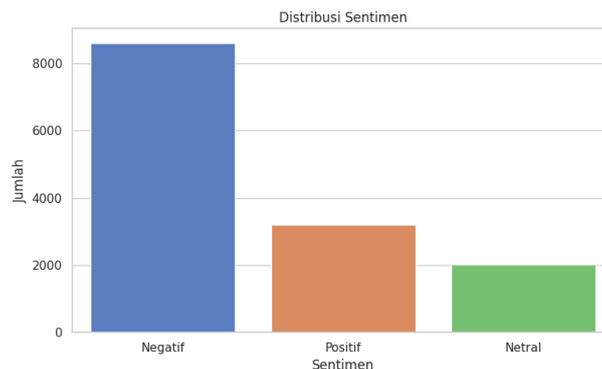
Setelah filtering stopwords:
['daftar', 'aja', 'susah', 'banget', 'dapet', 'kode', 'd', 'telegramnya', 'masuk', 'si', 'gimana']

Setelah stemming:
['daftar', 'aja', 'susah', 'banget', 'dapet', 'kode', 'd', 'telegram', 'masuk', 'si', 'gimana']

Hasil akhir preprocessing:
daftar aja susah banget dapet kode d telegram masuk si gimana
    
```

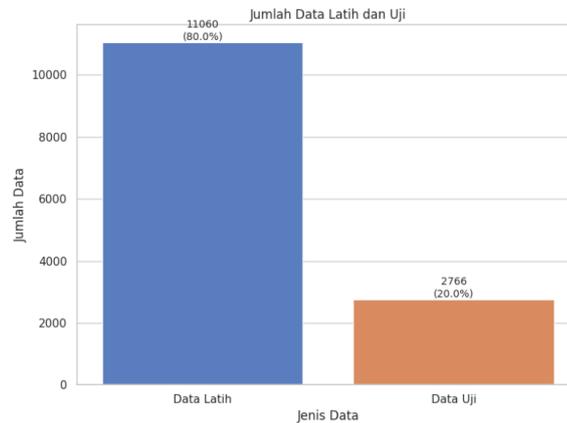
Gambar 3. Hasil *Pre-processing*

Setelah data dibersihkan, selanjutnya ialah tahapan *labelling*. Pelabelan data pada penelitian ini akan dibagi menjadi 3 kategori, yaitu kategori sentimen positif, sentimen negatif, dan sentimen netral. Pada Gambar 4. terlihat bahwa pengguna aplikasi Telegram lebih banyak memberi ulasan negatif dibandingkan dengan ulasan positif maupun netral.



Gambar 4. Diagram perbandingan jumlah sentimen

Setelah membagi ulasan pengguna menjadi tiga kategori, nantinya data tersebut akan dibagi menjadi dua, yaitu data latih dan data uji. Data latih ini akan diambil 80% data sedangkan untuk data uji ini akan diambil 20% data dapat dilihat lebih rinci pada Gambar 5. Data ini dibagi bertujuan untuk melatih model *machine learning* ataupun menguji data itu sendiri.



Gambar 5. Diagram Data Latih dan Data Uji

C. Klasifikasi Model dan Evaluasi Model Klasifikasi

Tahap selanjutnya ialah melakukan klasifikasi dan juga evaluasi model klasifikasi pada tiga metode yang dipakai dalam penelitian ini. Dalam evaluasi model klasifikasi ini akan memakai *confusion matrix*. Tetapi, sebelum melakukan proses klasifikasi model dan evaluasi klasifikasi model akan dilakukan proses pembobotan pada setiap kata terlebih dahulu menggunakan TF-IDF. Adanya proses pembobotan kata ini, supaya hasil dari klasifikasi model akan lebih akurat. Pada Gambar 6. Merupakan hasil dari sebuah pembobotan.

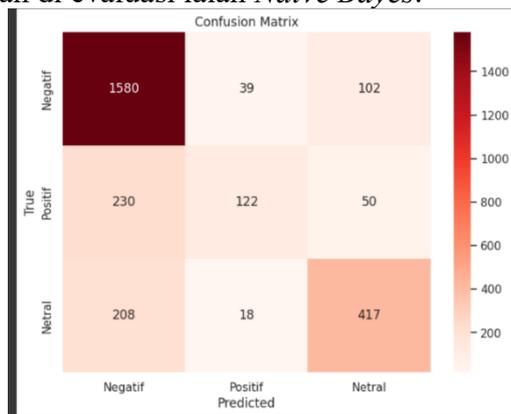
```

login nya susah bgt notif ny ga muncul muncul tolong baik login pakai akun google 0.458721
aja download udh coba masukin nomer nya udah kali download gw coba gw saranin download 0.482166
ok 0.134228
aplikasi jelek login login jelek telegram 0.242512
kode verifikasi 0.187742
terima kode 0.184335
bagus sih baguss cumann aneh akun hapus blok udah kali nya hapus ga ngelakuin apa maless cari nomer lainn 0.529572
aplikasi sampah download 0.225647
gabisa login ga dapat kode otp 0.318767
gabisa login kode telpon cmn kolom nya 0.328766
taik gak kirim kode 0.248444
oke 0.134228
please kak gw udh blokir blom buka sampe kak 0.344109
susah dapetin kode verifikasi kocak 0.276696
gk login daftar kocak 0.261927
ko gabisa masuk kmp ya 0.293525
udah ganti nomor login pas code verifikasi gak muncul tolong gak ngeribetin aja 0.432596
telegram nomor nya blokir min tolong jelas nya salah main blokir blokir aja 0.369298
kode verifikasi nya gak muncul muncul 0.260605
    
```

Gambar 6. Hasil Pembobotan Kata Menggunakan TF-IDF

1) Metode Naïve Bayes

Metode pertama yang akan di evaluasi ialah *Naïve Bayes*.



Gambar 7. Hasil Akurasi dan *Confusion Matrix* Metode *Naïve Bayes*

Metode *Naïve Bayes* menunjukkan performa yang cukup baik yang mempunyai akurasi sebesar 76,61% dengan performa terbaik pada kelas “Negatif”. Namun, performa menurun pada kelas “Netral” yang hanya mempunyai recall 0.30. Dalam *confusion matrix* metode ini dapat diketahui ada sebanyak 1.580 data diprediksi dengan benar negatif, 122 data diprediksi benar positif, dan sebanyak 417 data yang dapat diprediksi benar netral. Hal ini menunjukkan metode cenderung bias ke kelas mayoritas.

2) *Metode Logistic Regression*

Metode selanjutnya ialah metode *Logistic Regression*.

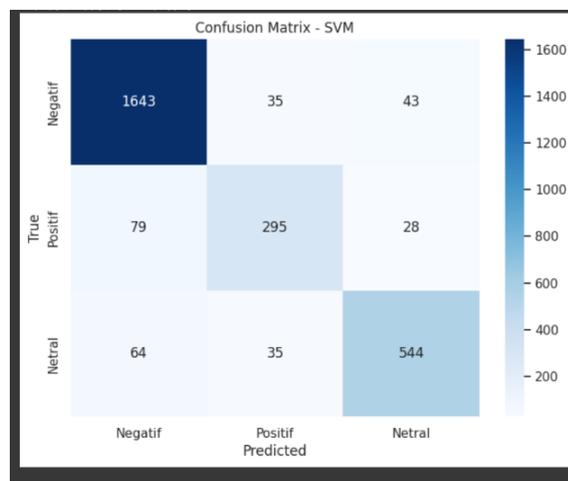


Gambar 8. Hasil Akurasi dan *Confusion Matrix* Metode *Logistic Regression*

Metode *Logistic Regression* mempunyai hasil akurasi lebih tinggi dibandingkan dengan akurasi metode sebelumnya sebesar 87,49%. Kinerja terbaik pada metode ini terlihat pada kelas “Negatif”, diikuti kelas “Positif”, lalu kelas “Netral”. Dalam *confusion matrix* pada metode ini dapat diketahui bahwa 1.650 data diprediksi benar negatif, 506 data diprediksi benar positif, dan sebanyak 264 data diprediksi benar netral. Secara keseluruhan hasil dari *confusion matrix*, metode ini cukup seimbang, walaupun masih ada kekeliruan dalam mengenali kelas netral.

3) *Metode Support Vector Machine (SVM)*

Metode terakhir dari penelitian ini ialah *Support Vector Machine (SVM)*.



Gambar 9. Hasil Akurasi dan *Confusion Matrix* Metode *Support Vector Machine*



Berdasarkan hasil dari visualisasi teks dapat diketahui kata yang sering muncul dalam sentimen pengguna. Pada sentimen netral kata yang sering muncul ialah “telegram”, “login”, dan “ga”. Untuk sentimen positif terdapat kata “telegram”, “premium”, dan “tele”. Lalu pada sentimen negatif terdapat kata “telegram”, “kode verifikasi”, dan “login”.

#### IV. KESIMPULAN

Penelitian ini membandingkan kinerja dari ketiga metode yang digunakan yaitu metode *Naïve Bayes*, *Logistic Regression*, dan SVM. Hasil dari penelitian ini menunjukkan bahwa metode SVM menjadi metode paling efektif digunakan dalam analisis sentimen pengguna aplikasi Telegram yang mempunyai nilai akurasi yang lebih baik dibandingkan dengan dua metode lainnya yaitu metode *Naïve Bayes* dan *Logistic Regression* dengan nilai akurasi 89,73%. Sedangkan untuk metode *Naïve Bayes* mempunyai nilai akurasi sebesar 75,61% dan *Logistic Regression* mempunyai nilai akurasi sebesar 87,49%. Perbedaan dari akurasi ini dikarenakan pengaruh oleh karakteristik masing-masing algoritma. SVM bisa lebih unggul dikarenakan metode ini mampu menangani data yang berdimensi tinggi dengan distribusi yang kompleks, sedangkan untuk metode *Logistic Regression* bekerja dengan baik pada data dengan relasi fitur yang linier dan relative seimbang dan untuk metode *Naïve Bayes* mempunyai keterbatasan dalam menangani depensasi antar kata yang umum terjadi pada teks ulasan. Dengan ini, dapat disimpulkan bahwa metode SVM ialah metode yang cocok untuk menganalisis sentimen dalam konteks data ulasan aplikasi yang bervariasi dan kompleks. Hasil analisis sentimen menunjukkan keluhan pengguna Telegram dalam mendaftar akun ataupun login dikarenakan kode verifikasi dikirimkan di dalam aplikasi tersebut. Padahal untuk yang mendaftar akun tentunya belum mempunyai akun Telegram. Tentunya hal ini dapat menjadi masukan untuk pengembang aplikasi supaya memperbaiki cara mendapatkan kode verifikasi, supaya calon pengguna dapat mudah mendaftar akun Telegram

#### REFERENSI

- [1] N. D. Pratidina and J. Mitha, “Dampak Penggunaan Media Sosial terhadap Interaksi Sosial Masyarakat: Studi Literature,” *J. Ilm. Univ. Batanghari Jambi*, vol. 23, no. 1, p. 810, 2023, doi: 10.33087/jiubj.v23i1.3083.
- [2] E. Fitriana and D. H. Siswanto, “KEEFEKTIFAN FITUR-FITUR PADA APLIKASI TELEGRAM SEBAGAI MEDIA PEMBELAJARAN MATEMATIKA SISWA SEKOLAH,” vol. 3, no. September, pp. 99–109, 2024.
- [3] I. N. Rois and C. E. Setyawan, “Penggunaan Aplikasi Telegram sebagai Media Pembelajaran pada Mata Pelajaran Bahasa Arab Peminatan di MAN 1 Kulon Progo,” *Proceeding Annu. Int. Symp. Arab. Lang. Cult. Lit.*, vol. 1, no. 1, pp. 59–68, 2021, [Online]. Available: <http://jurnal.uinbanten.ac.id/index.php/aisall/article/view/4348>
- [4] K. Kevin, M. Enjeli, and A. Wijaya, “Analisis Sentimen Penggunaan Aplikasi Kinemaster Menggunakan Metode Naive Bayes,” *J. Ilm. Comput. Sci.*, vol. 2, no. 2, pp. 89–98, 2024, doi: 10.58602/jics.v2i2.24.
- [5] K. S. Putri, I. R. Setiawan, and A. Pambudi, “Analisis Sentimen Terhadap Brand Skincare Lokal Menggunakan Naïve Bayes Classifier,” *Technol. J. Ilm.*, vol. 14, no. 3, p. 227, 2023, doi: 10.31602/tji.v14i3.11259.
- [6] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, “Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine,” *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 176–184, 2023, doi: 10.37373/tekno.v10i2.419.
- [7] A. Novantika, “Analisis sentimen ulasan pengguna aplikasi video conference google meet menggunakan metode svm dan logistic regression,” *Prism. Pros. Semin. Nas. Mat.*, vol. 5, pp. 808–813, 2022, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>

- [8] F. Fazrin, O. N. Pratiwi, and R. Andreswari, "Perbandingan Algoritma K-Nearest Neighbor dan Logistic Regression pada Analisis Sentimen terhadap Vaksinasi Covid-19 pada Media Sosial Twitter dengan Pelabelan Vader dan Textblob," *J. e-Proceeding Eng.*, vol. 10, no. 2, pp. 1596–1604, 2023.
- [9] K. Perdana, J. Sitompul, A. R. Pratama, and K. A. Baihaqi, "Vector Machine , Dan Logistic Regression Pada Analisis Sentimen Pengguna Aplikasi," *Kumpul. J. Ilmu Komput.*, vol. 10, no. 01, pp. 27–38, 2023.
- [10] F. M. Herza and D. Pembimbing, "PENGARUH RFE TERHADAP LOGISTIC REGRESSION DAN SUPPORT VECTOR MACHINE PADA ANALISIS SENTIMEN HOTEL SHANGRI-LA SURABAYA," 2024.
- [11] Dewi Setiyawati and N. Cahyono, "Analisis Sentimen Pengguna Sosial Media Twitter Terhadap Perokok Di Indonesia," *Indones. J. Comput. Sci.*, vol. 12, no. 1, pp. 262–272, 2023, doi: 10.33022/ijcs.v12i1.3154.
- [12] N. A. Permana and H. Bunyamin, "Perbandingan Logistic Regression dengan Random Forest dalam Memprediksi Sentimen Pada IMDb Movie Review," vol. 6, no. November, pp. 391–399, 2024.
- [13] H. Junianto, R. E. Saputro, B. A. Kusuma, and D. I. S. Saputra, "COMPARISON OF LOGISTIC REGRESSION AND RANDOM FOREST IN SENTIMENT ANALYSIS OF DISDUKCAPIL APPLICATION REVIEWS.," *J. Tek. Inform.*, vol. 5, no. (6), pp. 1539–1547., 2024, [Online]. Available: <https://doi.org/10.52436/1.jutif.2024.5.6.1802>.
- [14] A. N. Sativa, A. Rizky, I. Putri, and J. A. Putri, "Analisis Sentimen Twitter Ibu Kota Negara Nusantara Menggunakan Algoritma Naive Bayes , Logistic Regression dan K-Nearest Neighbors," vol. 3, no. 2, pp. 34–40, 2024.
- [15] Y. R. Z. Mustopo and Afiyati, "Jurnal JTIC ( Jurnal Teknologi Informasi dan Komunikasi ) Analisis Sentimen Proyek Strategis Nasional Food Estate," vol. 9, no. June, pp. 485–494, 2025.
- [16] 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.
- [17] N. Hendrastuty, A. Rahman Isnain, and A. Yanti Rahmadhani, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, pp. 150–155, 2021.
- [18] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional," *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
- [19] N. Cahyono and Anggista Oktavia Praneswara, "Analisis Sentimen Ulasan Aplikasi Tiktok Shop Seller Center di Google Playstore Menggunakan Algoritma Naive Bayes," *Indones. J. Comput. Sci.*, vol. 12, no. 6, pp. 3925–3940, 2023, doi: 10.33022/ijcs.v12i6.3473.
- [20] Z. C. Dwinnie and R. Novita, "Penerapan Machine Learning Pada Analisis Sentimen Twitter Sebelum dan Sesudah Debat Calon Presiden dan Wakil Presiden Tahun 2024," *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 758, 2024, doi: 10.30865/mib.v8i2.7504.
- [21] A. Nurian, "Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naïve Bayes," *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3s1, pp. 829–835, 2023, doi: 10.23960/jitet.v11i3s1.3348.
- [22] R. A. Husen, R. Astuti, L. Marlia, R. Rahmadden, and L. Efrizoni, "Analisis Sentimen Opini Publik pada Twitter Terhadap Bank BSI Menggunakan Algoritma Machine Learning," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 211–218, 2023, doi: 10.57152/malcom.v3i2.901.
- [23] Z. B. Toyibah, Y. N. Putri, Z. M. Widodo, and A. Tsalitsatun, "Perbandingan Kinerja Algoritma Multinomial Naïve Bayes dan Logistic Regression pada Analisis Sentimen Movie Ratings IMDB," vol. 10, no. 2, pp. 181–189, 2024.
- [24] M. . Adrian, M. . Rafialdy, and N. . Rakhmawati, "Perbandingan Random Forest dan SVM Analisis Sentimen PSBB," *J. Inform. UPGRI*, vol. 7, no. 1, pp. 36–40, 2021.
- [25] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, "Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 153–160, 2023, doi: 10.57152/malcom.v3i2.897.