



Analisa Sentimen Pengguna Transportasi Jakarta Terhadap Transjakarta Menggunakan Metode Naives Bayes dan K-Nearest Neighbor

Ismia Iwandini, Agung Triayudi*, Gatot Soepriyono

Fakultas Teknologi Komunikasi dan Informatika, Sistem Informasi, Universitas Nasional, Jakarta

Jl. Sawo Manila No.61, RW.7, Pejaten Bar., Ps. Minggu, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta, Indonesia

Email: ¹ismiaiwandini2019@student.unas.ac.id, ^{2,*}agungtriayudi@civitas.unas.ac.id, ³gatot@civitas.unas.ac.id

Email Penulis Korespondensi: agungtriayudi@civitas.unas.ac.id

Submitted: 16/01/2023; Accepted: 28/01/2023; Published: 29/01/2023

Abstrak—Media sosial digunakan dalam berkomunikasi yang sangat populer di Indonesia. Salah satu media sosial yang paling populer yaitu Twitter. Twitter adalah situs media sosial di mana orang dapat berbagi informasi publik. Informasi ini dapat diproses untuk membuat analisis sentimen. Penelitian ini mencoba membuat suatu sistem yang dapat mendeteksi sentimen positif atau negatif dalam informasi publik. Metode yang dipakai untuk klasifikasi sentimen ini yaitu perbandingan metode Naive Bayes Classifier dan juga K-Nearest Neighbor Classifier dengan pembobotan TF-IDF. Masukan sistem ini berupa data tweet terhadap Transjakarta, dan hasil output sistem ini berupa visualisasi data sentimen positif dan negatif dengan menggunakan Streamlit yang merupakan library dari python. Berdasarkan pengujian didapatkan akurasi pendekatan Naive Bayes untuk analisis sentimen data Twitter terkait penggunaan transportasi Transjakarta sebesar 61,1%, dan akurasi metode K-Nearest Neighbor sebesar 75,7%. Untuk kedua metode yang dipakai dalam menentukan tingkat keakurasian dapat dilihat metode K-nearest-neighbor menghasilkan akurasi yang lebih tinggi.

Kata Kunci: Analisis Sentimen; Twitter, Streamlit; Naive Bayes; K-Nearest Neighbor.

Abstract—Social media used in communicating that is very popular in Indonesia. One of the most popular is Twitter. Twitter is a social media site where people can share information publicly. This information can be processed to make sentiment analysis. This research attempts to create a system that can detect positive or negative sentiments in public information. The method used for this sentiment classification is the comparison method of Naive Bayes Classifier and K-Nearest Neighbor Classifier using TF-IDF weighting. The input to this system is in the form of tweet data for Transjakarta, while the output of this system is in the form of visualization of positive and negative sentiment data using Streamlit which is a library from python. Based on testing the accuracy of the Naive Bayes approach for sentiment analysis of Twitter data related to the use of Transjakarta transportation is 61.1%, and the accuracy of the K-Nearest Neighbor method is 75.7%. For the two methods used in determining the level of accuracy, it can be concluded that the K-nearest-neighbor method produces better accuracy.

Keywords: Sentiment Analysis; Twitter, Streamlit; Naive Bayes; K-Nearest Neighbor.

1. PENDAHULUAN

Angkutan umum menjadi pilihan bagi warga Jakarta yang sangat banyak penduduknya. Apalagi dengan berkembangnya teknologi memaksa agar dapat mengatasi terjadi masalah karena kepadatan penduduk sehingga menyebabkan kemacetan di kota-kota. Jumlah kendaraan yang digunakan sehari-hari berbanding lurus dari jumlahnya penduduk. Sebagian pengguna jalan umum, maupun kendaraan pribadi dan angkutan umum di jalan kota, memilih angkutan umum dan tentunya mengurangi jumlah kendaraan yang dipakai. Transportasi umum meminimalkan kemacetan lalu lintas [1]. Dan Jakarta, ibu kota Indonesia, sebagai pusat pemerintahan dan segala proses kegiatan baik ekonomi maupun bisnis samapi saat ini. Tersedianya sarana dan prasarana yang sempurna untuk kegiatan bisnis membuat Jakarta menjadi tujuan para pendatang yang ingin mengadu nasib. Ibu kota yang setiap orang membutuhkan ataupun memiliki kendaraan pribadi menjadi menjadikan penyebab kemacetan. Oleh karena itu diperlukan transportasi umum yang aman, nyaman, memiliki akses jalan memadai [2].

Dengan berkembangnya teknologi informasi, membawa banyak keuntungan ataupun manfaat bagi masyarakat. Teknologi informasi yang berkembang juga mengikuti dengan perkembangan transportasi. Pengguna transportasi dapat mengomentari transportasi yang digunakan. Aplikasi dan media sosial yang banyak dipakai untuk berkomentar salah satunya Twitter. Aplikasi Twitter berisi banyak informasi transportasi dan permintaan pengguna, yang sangat berguna untuk mengetahui kualitas lalu lintas pengujung. Dengan cara ini, pengguna lain dengan sangat mudah mendapatkan informasi yang mereka inginkan. Analisis sentimen yaitu jenis bahasa alami yang digunakan untuk menjangkau opini publik. Karena itu, dapat dilihat pendapat masyarakat yang menggunakan transportasi tentang bagaimana rasanya menggunakan transportasi, baik berupa saran maupun kritik terhadap permasalahan yang sering dialami, seperti: pelayanan yang buruk atau tidak memuaskan. Masalah-masalah ini dapat diselesaikan dengan menganalisis ulasan atau pendapat [3].

Analisis sentimen yaitu metode dalam menentukan sentimen untuk mengklasifikasikan polaritas teks untuk dimasukkan kedalam dokumen atau kalimat [4] untuk menentukan kategori seperti sentimen positif, negatif atau netral. Saat ini, analisis sentimen banyak digunakan oleh para peneliti sebagai bagian dari penelitian ilmu komputer. Pada jejaring sosial Twitter sering dipakai untuk menganalisa sentimen dalam menentukan audiens. Analisis sentimen adalah proses penggalian, pengolahan, dan pemahaman informasi secara otomatis dalam bentuk teks yang tidak teratur untuk mengambil data atau informasi berupa emosional yang terkandung dalam opini atau

kritik [5]. Analisis sentimen digunakan dalam memproses sentimen dan mengklasifikasikan polaritas teks pada dokumen dan suatu kalimat yang dapat ditentukan kategori sentimen baik positif, negatif dan netral [6].

Twitter dipakai sebagai sumber data dalam penelitian ini untuk menganalisis sebuah tweet dalam bahasa Indonesia [7] untuk menganalisa transportasi Transjakarta di Jakarta. Proses analisis sentimen Transjakarta di Twitter menunjukkan tanggapan pengguna Transjakarta, apakah secara umum positif, negatif, atau netral berdasarkan hasil tweet untuk kata kunci [8]. Naive Bayes dan juga K-Nearest Neighbor metode machine learning yang dipakai untuk klasifikasi sentiment dalam bentuk teks dan mendapatkan hasil sesuai. Pada penelitian ini kinerja kedua metode dalam analisis sentimen dibandingkan dengan metode lain menghasilkan nilai yang lebih baik pada kinerja metode klasifikasi [9].

Penelitian tentang klasifikasi sentimen sebelumnya pada jurnalnya, menganalisa sentimen pada transportasi KRL Commuter Line, mengekstrak secara acak sebanyak 127 data dengan mengolahnya dalam beberapa langkah, salah satunya dengan melakukan duplikasi untuk menghindari data. Memakai metode Naive Bayes, K-nearest Neighbor dan Decision Tree. Hasilnya menunjukkan bahwa terdapat perbedaan akurasi pada Naive Bayes dan juga K-nearest Neighbor [1].

Penelitian pada jurnal lainnya mencoba mencari tahu algoritma mana yang berfungsi paling baik saat menganalisis sentimen vaksin Covid-19 tertentu. Dengan memakai Naive Bayes dan juga K-Nearest Neighbor dengan validasi cross-validation melihat seberapa baik metode tersebut bekerja. Naive Bayes mendapatkan nilai 77,62% dalam akurasi sedangkan K-Nearest Neighbor mendapatkan 76,43. [10].

Pada jurnalnya peneliti melakukan analisa pendapat terhadap para penggunaan PeduliLindungi mengambil data dari komentar di play store. Dengan Naive Bayes dan juga K-Nearest Neighbor menggunakan Rapid Miner. Pengujian didapatkan hasil pada metode Naives Bayes sebesar 70,46% dan K-Nearest Neighbor 73,33% [11].

Penelitian analisis sentimen lainnya melakukan penelitian yang bermaksud untuk melihat hasil dari performa model Naive Bayes dan K-Nearest yang di ambil hasil sentimen terhadap Jasa Ekspedisi. Model Naive Bayes dan juga K-Nearest Neighbor berkembang secara signifikan setelah teknik oversampling diterapkan. Naive Bayes menghasilkan nilai terbesar terdapat JNT 82%, sedangkan K-Nearest Neighbor pada POS 71% [3].

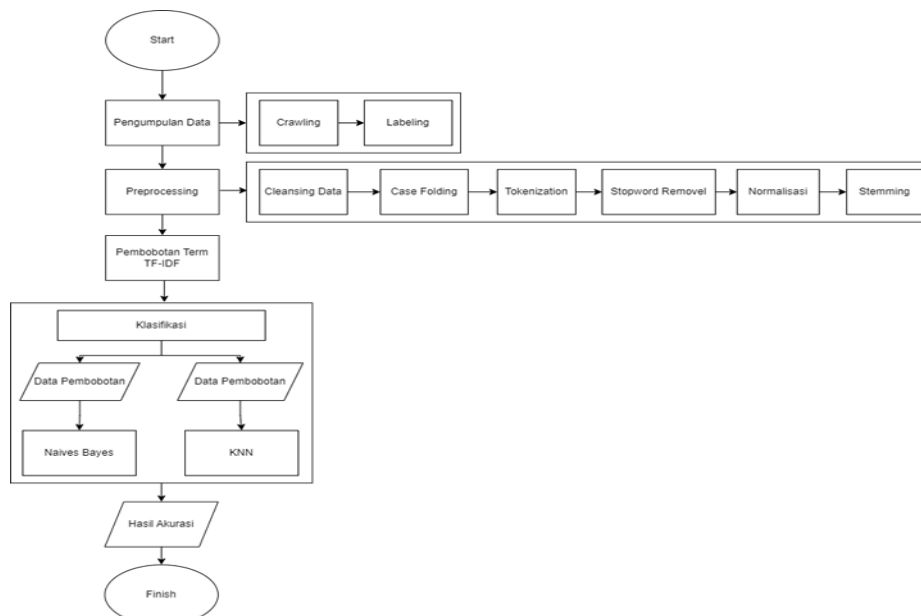
Lalu penelitian Analisis Sentimen Twitter MRT Jakarta menunjukkan bagaimana perasaan masyarakat umum tentang MRT Jakarta, menggunakan hasil dari tweet yang diperoleh dari Twitter tentang MRT Jakarta. Metode Naive Bayes Classifier teknik klasifikasi yang digunakan dalam studi sentimen ini. Akurasi sistem untuk analisis sentimen tweet pada akun Twitter terhadap MRT Jakarta adalah 95,88% [8].

Dari berbagai penelitian terkait sebelumnya dapat disimpulkan bahwa analisis sentimen dapat mengidentifikasi opini, opini, dan emosi dari sebuah teks dengan mengklasifikasikannya ke dalam kategori negatif, positif, dan netral. Penulisan ini akan fokus menganalisis sentimen Twitter terkait transportasi Transjakarta dengan Naive Bayes dan juga K-Nearest Neighbor.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Pada gambar 1 merupakan tahapan dalam penelitian.



Gambar 1. Proses Analisis Sentimen

Berikut penjelasan dari tahapan gambar 1

- Pengumpulan data dengan cara scraping dengan mengambil data dari twitter. Dalam pengambilan data ini kata pencarian yang di cari adalah “Transjakarta”, dimana data yang telah didapatkan di simpan dalam bentuk .csv sebanyak 4000 lebih data namun data tersebut masih data kotor dan akan diolah ke tahap selanjutnya untuk di jadikan data yang sudah bersih [12].
- Proses preprocessing yaitu pengurangan atribut yang sangat berpengaruh di proses klasifikasi dan pada proses ini masih data yang mentah sehingga diharapkan dapat mempermudah proses klasifikasi [13], terdiri dari cleansing, case folding, tokenizing, stopword removal, normalisasi lalu stemming.
- Langkah selanjutnya adalah menyederhanakan proses klasifikasi. Selama tahap pembuatan fitur, setiap kalimat atau dokumen mengalami pembobotan kata dengan Tf-Idf menggunakan tf-idf sehingga dataset dapat digunakan untuk pelatihan menggunakan perhitungan Naive Bayes dan juga K-Nearest Neighbor [3].
- Metode Naive Bayes dan juga K-Nearest Neighbor digunakan sebagai mengklasifikasikan data tweet untuk mendapatkan analisis opini. Untuk mengklasifikasikan sentimen. untuk memprediksi seberapa baik kinerjanya pada set pengujian baru. Kemudian menggunakan hasil prediksi ini untuk memperbaiki model [9]. Pada banyaknya jumlah data digunakan sebagai uji model yang diperoleh data training.
- Untuk mengetahui seberapa baik kinerja metode dilakukan evaluasi metode untuk mendapatkan nilai akurasi. Model dievaluasi dengan melihat akurasi pada metode dengan menggunakan Confusion matrix dan tabel akurasi masing-masing metode [12].

2.2 Data Mining

Data mining ialah mencari sumber informasi sebuah data. Informasi ini dapat ditemukan dengan menggunakan teknik yang kompleks seperti kecerdasan buatan (AI), metode statistik, matematika, pembelajaran mesin, dan sebagainya. Teknik-teknik ini nantinya dapat melihat informasi yang berguna dari database yang besar. Penambangan data berasal dari pembelajaran mesin dan statistik, tetapi juga ada di bidang ilmu komputer dan disiplin ilmu lainnya [14].

Data Mining adalah teknik yang menggunakan pengetahuan, seperti statistik dan matematika, untuk menemukan informasi yang berguna dari sekumpulan data. Ini dapat bermanfaat bagi perusahaan, karena dapat membantu mereka mengetahui apa yang terjadi dengan bisnis mereka, dan membuat prediksi tentang kejadian di masa mendatang [15].

2.2 Metode Naives Bayes dan K-Nearest Neighbor

Pengklasifikasi Naive Bayes biasa digunakan untuk menemukan kategori yang paling mungkin di mana sekumpulan data uji dapat diklasifikasikan [16]. Dalam penelitian ini, data uji berupa tweet dokumen. Tahap pertama dari algoritma menggunakan kategori yang diketahui untuk melatih pengklasifikasi. Kemudian klasifikasi digunakan untuk mencari kategori yang paling mungkin untuk dokumen baru yang belum pernah dilihat sebelumnya.

K-Nearest Neighbor yaitu metode machine learning yang dipakai untuk klasifikasi suatu objek dari dokumen latih dengan jaraknya yang paling mendekati dengan suatu objek. Metode K-Nearest Neighbor adalah metode yang sangat sederhana, sangat mudah digunakan, dan tahan pada data memiliki data noise, dan juga efektif untuk memproses pengelompokan [17].



Gambar 2. Proses Analisis Sentimen menggunakan Naives Bayes dan K-Nearest Neighbor

Pada gambar 2 menunjukkan tahapan dalam menggunakan metode Naives Bayes yang digunakan pada penelitian ini. Dimana tahapan awal data di crawling dari twitter lalu masuk ke tahap preprocessing tahap ketiga yaitu ekstrasi fitur dengan Tf-idf dan terakhir tahap pengujian dan evaluasi menggunakan Naives Bayes.

2.2 Pengujian dan Evaluasi

Setelah melatih model, menggunakannya untuk memprediksi seberapa baik kinerjanya pada set pengujian baru. Kemudian menggunakan hasil prediksi ini untuk memperbaiki model. Untuk mengetahui seberapa baik kinerja model dilakukan evaluasi model. Model dievaluasi dengan mengevaluasi tingkat akurasi metode dengan menggunakan matriks konfusi dan tabel akurasi masing-masing model. Setelah menguji data pengujian dengan data pelatihan, lalu membuat daftar kelas yang disebut kumpulan prediksi kelas dari data pengujian. Kemudian, dengan menggunakan data pengujian yang sebelumnya tersembunyi, kelas yang diprediksi dikontraskan dengan kelas sebenarnya. Berikut perhitungan nilai accuracy, precision, recall, dan f1- score [12]. Berikut rumus masing-masing perhitungan:

Rumus accuracy (acc) : Accuracy sebagai perbandingan informasi yang berhasil ditanggapi oleh sistem dengan data informasi keseluruhan.

$$acc = \frac{TN+TP}{FN+FP+TN+TP} \tag{1}$$

Rumus precision (pre) : Precision adalah tingkat akurasi antara permintaan pengguna dan jawaban sistem.

$$pre = \frac{TP}{FP+TP} \tag{2}$$

Rumus recall (rec) : Recall sebagai ukuran ketelitian antara informasi pada sebelumnya dengan informasi yang sebelumnya.

$$rec = \frac{TP}{FN+TP} \tag{3}$$

Rumus F-Measure (F1) : F-Measure adalah harmonic mean hasil dari perhitungan precision dan juga perhitungan recall

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengujian penelitian ini didapatkan dari mengambil data di jejaring sosial Twitter, disebut tweet dengan menggunakan metode scraping yang memudahkan untuk menemukan pernyataan pada topik yang sama dan juga mengumpulkan tweet Transjakarta [18]. Jumlah data yang digunakan lebih dari 4.000 tweet untuk pernyataan tentang subjek tersebut. Pada saat proses pengumpulan data hanya di ambil text yang merupakan hasil scrapping. Hasil tweet pada penelitian ini adalah hanya tweet yang terdapat kata Transjakarta lalu di ambil acak oleh library yang dipakai yaitu scraping. Hasil teks dari scraping yaitu data yang masih sangat mentah maka di perlukan beberapa tahapan lagi lalu bisa diklasifikasikan. Pada gambar 3 berikut proses scraping di google colab dengan Bahasa pemrograman Python dan hasil output atau hasil yang sudah di scraping



Gambar 3. Hasil Scaraping

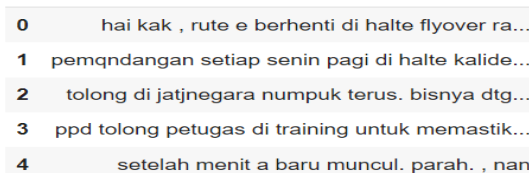
Setelah dataset di masukan ke dalam google colab selanjutnya dilakukan. Data hasil scaraping yang telah dikumpulkan akan diberi label positif atau negatif. Pada proses pelabelan dilakukan dengan menggunakan pelebelan secara otomatis dengan vader lexicon yang dimana hasilnya akan digunakan untuk ketahap selanjutnya.

3.2 Preprocessing

Dalam tahapan ini ada beberapa library yang harus di import terlebih dahulu untuk menjalan project, berikut beberapa library yang di import dalam project ini. Data yang dimasukkan pada proses ini masih data mentah, sehingga proses ini merupakan dokumen berkualitas tinggi yang diharapkan dapat mempermudah proses klasifikasi [13]. Proses pengolah preprocessing diantaranya:

a. Cleansing data dan Case Folding

Proses pembersihan untuk menghilangkan informasi yang tidak perlu, seperti tanda baca, angka, dan spasi dan mengubah kata menjadi huruf kecil untuk menyeragamkan huruf. Berikut hasil dari cleansing data dan case folding dari proses pemograman diatas



Gambar 3. Hasil Cleansing data dan Case Folding

b. Tokenization

Pada tahap ini beberapa library juga harus di import untuk menjalankan project. Tahap tokenization pada text processing merupakan proses pemisahan kata, simbol, frase, dan entitas penting lainnya dari sebuah teks.

Tokenizing Token Result :

```
0 [hai, kak, ,, rute, e, berhenti, di, halte, fl...
1 [pemandangan, setiap, senin, pagi, di, halte, ...
2 [tolong, di, jatjnegara, numpuk, terus, ., bis...
3 [ppd, tolong, petugas, di, training, untuk, me...
4 [setelah, menit, a, baru, muncul, ., parah, ,,...
Name: text_token, dtype: object
```

Gambar 4. Hasil Tokenization

Pada gambar 4 memperlihatkan hasil dari tokenization memisahkan kalimat menjadi perkata dan token token bagian tertentu untuk melanjutkan ke proses selanjutnya.

c. Stopword Removal

Digunakan untuk menghilangkan kata tidak dibutuhkan atau tidak sesuai dengan topik seperti kata “di”, “apa”, “untuk” dan lain lainnya. Karena kata yang tidak dibutuhkan dapat mempengaruhi akurasi dalam klasifikasi sentimen maka di butuhkan stopwords removal untuk mengurangi kata yang tidak digunakan.

```
0 [hai, kak, ,, rute, e, berhenti, halte, flyove...
1 [pemandangan, senin, pagi, halte, kalideres, gbk]
2 [tolong, jatjnegara, numpuk, ., bisnya, dtgnya...
3 [ppd, tolong, petugas, training, area, laki, w...
4 [menit, a, muncul, ., parah, ., ,]
Name: text_tokens_WSW, dtype: object
```

Gambar 5. Stopword Removal

d. Normalisasi

Pada tahap ini, kata tidak baku diubah menjadi kata baku menggunakan kamus baku yang ada untuk menghilangkan duplikasi kata dengan arti yang sama. Pada kamus standar diperoleh dari web atau Internet oleh penulis dengan berbagai tambahan [3].

```
0 [hai, kak, ,, rute, e, berhenti, di, halte, fl...
1 [pemandangan, setiap, senin, pagi, di, halte, ...
2 [tolong, di, jatjnegara, numpuk, terus, ., bis...
3 [ppd, tolong, petugas, di, training, untuk, me...
4 [setelah, menit, a, baru, muncul, ., parah, ,,...
5 [tetapi, dia, datang, siang, jam, ., jadi, mau...
6 [hai, kak, ,, rute, l, e, puribeta, halimun, m...
7 [hai, kak, ,, koridor, cbd, ciledug, tendean, ...
8 [bebs, ini, yang, saya, maksud, kemaren, wkwkw...
9 [emang, situ, naik, tj, sehari, ?, , , nan]
Name: normalisasi, dtype: object
```

Gambar 6. Normalisasi

e. Stemming

Dilakukan suatu proses dimana kata-kata di dalam dokumen diubah menjadi kata dasar. Proses pembuatan bahasa Indonesia dilakukan dengan menggunakan Python Sastrawi, karya pengembangan dari proyek PHP Sastra. Python Sastrawi adalah library ringan yang dapat mengonversi sufiks bahasa Indonesia ke bentuk dasarnya [10].

```
Streaming output truncated to the last 5000 lines.
6221
-----
tiketnya : tiket
6221
-----
didapatkan : dapat
6221
-----
vendor : vendor
6221
-----
jaclinko : jaclinko
6221
-----
pool : pool
6221
-----
```

Gambar 7. Stemming

Hasil stemming pada gambar 7 melihatkan salah satu stemming bekerja yaitu menghilangkan kata imbuhan pada setiap kata, contohnya tiketnya lalu dihilangkan imbuhan nya menjadi tiket.

3.3 Klasifikasi Naive Bayes dan K-Nearest Neighbor

Langkah pertama proses ekstraksi fitur adalah mengubah dataset menjadi representasi vektor dengan bantuan CountVectorizer library Python. Frekuensi munculnya istilah dalam dokumen yang dimaksud disebut sebagai TF sedangkan IDF adalah perhitungan sejauh mana istilah tersebut didistribusikan ke seluruh kumpulan dokumen yang dipermasalahkan. Library Python TfidfVectorizer digunakan selama proses pembuatan bobot kata dalam penelitian ini.

```
#melakukan ekstraksi fitur menggunakan tf-idf
from sklearn.feature_extraction.text import TfidfVectorizer
vectorizer= TfidfVectorizer()
X = vectorizer.fit_transform(data1['text'])
y = data1['Sentiment']
```

Gambar 8. Ekstrasi Fitur TF-IDF

Langkah selanjutnya adalah membuat class yang didalamnya terdapat 3 siklus berturut-turut yaitu mengubah dataset menjadi representasi vektor menggunakan library CountVectorizer, diikuti pembagian atau split data, dan yang ketiga adalah klasifikasi. Pada penulisan ini menggunakan 40 persen data uji yang dipilih secara acak dalam proses klasifikasi. Maka untuk mendapatkan hasil prediksi data dimasukkan lalu setelah itu proses klasifikasi memakai perhitungan antar kalimat untuk setiap kelas telah selesai. Kinerja algoritma dapat dihitung setelah semua proses selesai.

```
X = X.toarray()
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.4, random_state=0)

print("Train: ",X_train.shape,y_train.shape,"Test: ",(X_test.shape,y_test.shape))

Train: (1901, 4985) (1901,) Test: ((1268, 4985), (1268,))
```

Gambar 9. Splitting Data

Untuk melihat hasil dari Algoritma yang digunakan, selanjutnya proses melakukan pengujian kepada model tersebut. Hasil pembobotan tf-idf pada proses sebelumnya lalu selanjutnya membuat pembagian data latih lalu data uji untuk memprediksi. Tujuan dalam membagi data latih dan data uji dalam preprocessing menjadi perbandingan. Dibagi dengan presentase 60% data latih lalu 40% data uji.

	precision	recall	f1-score	support
Negative	0.51	0.75	0.61	381
Positive	0.75	0.52	0.61	561
accuracy			0.61	942
macro avg	0.63	0.63	0.61	942
weighted avg	0.66	0.61	0.61	942

AKURASI NB: 0.611

Gambar 10. Akurasi Naive Bayes

	precision	recall	f1-score	support
Negative	0.73	0.63	0.68	381
Positive	0.77	0.84	0.80	561
accuracy			0.76	942
macro avg	0.75	0.74	0.74	942
weighted avg	0.75	0.76	0.75	942

AKURASI KNN: 0.757

Gambar 11. Akurasi K-Nearest Neighbor

Pada gambar 10 dan 11, dari hasil perbandingan diatas dapat dilihat hasil dari klasifikasi model K-Nearest Neighbor nilai akurasi lebih unggul jika dibandingkan model Naive Bayes, dimana nilai accuracy, yang sejatinya

nilai pada tingkat kedekatan setiap nilai prediksi dengan nilai sebenarnya, hasil yang didapatkan dari pengujian pada metode K-Nearest Neighbor 75.7% sedangkan Naive Bayes 61.1%.

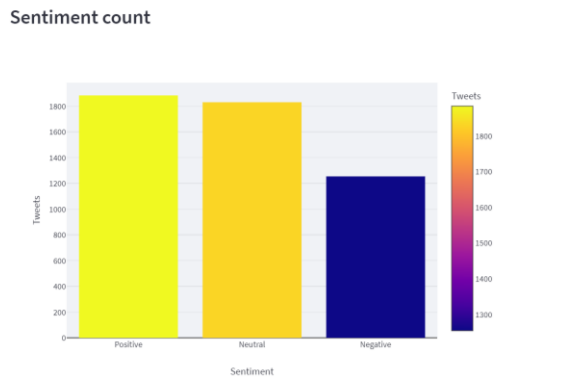
3.4 Implementasi

Dalam penelitian dilakukan implementasi hasil perancangan sistem yang dibuat. Implementasi ini adalah interface dari Aplikasi Sentimen yang di buat dengan framework streamlit. Dengan bantuan visualisasi data, pengguna dapat melihat-lihat dan mendapatkan informasi yang mereka butuhkan. Visualisasi analisis sentimen untuk Transjakarta dapat dilihat di bawah ini.



Gambar 12. Tampilan Dashboard Data set

Pada gambar 12 memperlihatkan hasil dari data set yang sudah di lakukan tahap preprocessing menggunakan google colab dan di visualisasikan kedalam streamlit



Gambar 13. Histogram label

Pada gambar 13 memperlihatkan Histogram yang dimana hasil dari scaraping data dilakukan pelebalan untuk melihat seberapa banyak analisis sentimen pada setiap kategorinya.

Accuracy Results

	precision	recall	f1-score	support
Negative	0.51	0.75	0.61	381
Positive	0.75	0.52	0.61	561
accuracy			0.61	942
macro avg	0.63	0.63	0.61	942
weighted avg	0.66	0.61	0.61	942

AKURASI NB: 0.611

Accuracy Results Naives Bayes

Berikut adalah hasil dari Metode Klasifikasi Naives Bayes dimana hasil akurasi nya sebesar 61.1%

Gambar 14. Akurasi Naive Bayes

Accuracy Results

	precision	recall	f1-score	support
Negative	0.73	0.63	0.68	381
Positive	0.77	0.84	0.80	561
accuracy			0.76	942
macro avg	0.75	0.74	0.74	942
weighted avg	0.75	0.76	0.75	942

AKURASI KNN: 0.757

Accuracy Results K-nearest Neighbor

Berikut adalah hasil dari Metode Klasifikasi K-nearest Neighbor dimana hasil akurasi nya sebesar 75.7%

Gambar 15. Akurasi k-nearest neighbor



Pada gambar gambar 14 dan 15 adalah hasil evaluasi metode yang sebelumnya di proses di google colab dengan berbagai tahapan yang digunakan sebelumnya, di tahap akhir memperlihatkan hasil akurasi pada masing masing metode dan di ditampilkan di streamlit.

4. KESIMPULAN

Kesimpulan berikut dapat diambil dari temuan pengujian dan analisis yang dilakukan untuk pembahasan sebelumnya. Akurasi pendekatan Naive Bayes untuk analisis sentimen data Twitter terkait penggunaan transportasi Transjakarta sebesar 61,1%, dan akurasi metode K-Nearest Neighbor sebesar 75,7%. Hasil analisis sentimen yang mengungkapkan bahwa sikap positif mendominasi, semakin menunjukkan bahwa masyarakat umum peduli dengan adanya teknologi. Dalam penyelidikan ini, ditentukan bahwa pendekatan K-Nearest Neighbor lebih baik daripada Naives Bayes.

REFERENCES

- [1] N. Tri Romadloni, I. Santoso, and S. Budilaksono, "Perbandingan Metode Naive Bayes, Knn Dan Decision Tree Terhadap Analisis Sentimen Transportasi Krl Commuter Line," *J. IKRA-ITH Inform.*, vol. 3, no. 2, pp. 1–9, 2019.
- [2] C. Aksan, B. Pramono, and A. M. Sajiah, "Analisis Sentimen Pembangunan Mass Rapid Transit (Mrt) Jakarta Pada Twitter Menggunakan Metode Improved K-Nearest Neighbor," *semantik*, vol. 8, no. 1, p. 53, 2022, doi: 10.55679/semantik.v8i1.24653.
- [3] Z. P. Putra and A. Nugroho, "Pebandingan Performa Naive Bayes dan KNN pada Klasifikasi Teks Sentimen Jasa Ekspedisi," *JOINTECS (Journal Inf. Technol. Comput. Sci.)*, vol. 3, no. 1, p. 2022, 2018.
- [4] R. Safitri, N. Alfira, D. Tamitiadini, W. W. A. Dewi, and N. Febriani, *Analisis Sentimen : Metode Alternatif Penelitian Big Data*. Universitas Brawijaya Press, 2021.
- [5] B. Brahimi, M. Touahria, and A. Tari, "Improving sentiment analysis in Arabic: A combined approach," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 33, no. 10, pp. 1242–1250, 2021, doi: 10.1016/j.jksuci.2019.07.011.
- [6] D. Ramadhan and E. B. Setiawan, "Analisis Sentimen Program Acara di SCTV pada Twitter Menggunakan Metode Naive Bayes dan Support Vector Machine," *e-Proceeding Eng.*, vol. 6, no. 2, pp. 9736–9743, 2019, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/10708>
- [7] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.
- [8] D. Iksari, Y. Fajarwati, and Widiastuti, "Analisis Sentimen Dan Klasifikasi Tweets Berbahasa Indonesia Terhadap Transportasi Umum Mrt Jakarta Menggunakan Naive Bayes Classifier," *J. Ilm. Inform. Komput.*, vol. 25, no. 1, pp. 64–75, 2020, doi: 10.35760/ik.2020.v25i1.2427.
- [9] Azhar, S. U. Masruroh, L. K. Wardhani, and Okfalisa, "Perbandingan Kinerja Algoritma Naive Bayes Dan K-Nn Pendekatan Lexicon Pada Analisis Sentimen Di Media," *Pros. Semin. Nas. Fis. Univ. Riau IV*, no. September, pp. 978–979, 2019.
- [10] A. Y. Permana and H. Noviyani, "Komparasi Algoritma Naive Bayes Dan K-Nearest Neighbor Dalam Melihat Analisis Sentimen Terhadap Vaksinasi Covid-19," *Pros. SAINTEK*, vol. 1, no. 1, pp. 128–134, 2022.
- [11] A. P. Wibowo, W. Darmawan, and N. Amalia, "Komparasi Metode Naive Bayes Dan K-Nearest Neighbor Terhadap Analisis Sentimen Pengguna Aplikasi Pedulilindungi," *IC-Tech*, vol. 17, no. 1, pp. 18–23, 2022, doi: 10.47775/ictech.v17i1.234.
- [12] E. Dwianto and M. Sadikin, "Analisis Sentimen Transportasi Online pada Twitter Menggunakan Metode Klasifikasi Naive Bayes dan Support Vector Machine," *Format J. Ilm. Tek. Inform.*, vol. 10, no. 1, p. 94, 2021, doi: 10.22441/format.2021.v10.i1.009.
- [13] S. Rahayu, M. Z. Yumarlin, J. E. Bororing, and R. Hadiyat, "Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP," *Edumatic J. Pendidik. Inform.*, vol. 6, no. 1, pp. 98–106, 2022, doi: 10.29408/edumatic.v6i1.5433.
- [14] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, "Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner," *JBASE - J. Bus. Audit Inf. Syst.*, vol. 4, no. 1, pp. 13–21, 2021, doi: 10.30813/jbase.v4i1.2729.
- [15] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naive Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *J. Ris. Komput.*, vol. 8, no. 6, pp. 219–225, 2021, doi: 10.30865/jurikom.v8i6.3655.
- [16] D. A. Pratiwi, R. M. Awangga, and M. Y. H. Setyawan, *SELEKSI CALON KELULUSAN TEPAT WAKTU MAHASISWA TEKNIK INFORMATIKA MENGGUNAKAN METODE NAIVE BAYES*. Kreatif, 2020.
- [17] A. Salam, J. Zeniarja, and R. S. U. Khasanah, "Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia)," *Pros. SINTAK*, pp. 480–486, 2018.
- [18] M. Syarifuddin, "Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naive Bayes Dan Knn," *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, 2020, doi: 10.33480/inti.v15i1.1347.