Perbandingan Analisis Sentimen Untuk Prediksi Kepuasan Pelanggan Kedai Kopi Di Kofind Menggunakan Algoritma SVM Dan Naive Bayes

Fardeen[1], Ricky Eka Putra[2]

[1,2] Program Studi S1 Teknik Informatika, Universitas Negeri Surabaya
[1]fardeen.19022@mhs.unesa.ac.id
[2]rickyeka@unesa.ac.id

***Abstrak*— Penelitian ini bertujuan untuk membandingkan performa algoritma Support Vector Machine (SVM) dan Naïve Bayes dalam melakukan analisis sentimen terhadap ulasan pelanggan kedai kopi Kofind. Implementasi kedua algoritma dilakukan dengan pendekatan pembelajaran mesin yang memanfaatkan Term Frequency-Inverse Document Frequency (TF-IDF) sebagai metode ekstraksi fitur. Proses analisis meliputi tahap preprocessing data (pembersihan teks, tokenisasi, dan penghapusan stopwords), ekstraksi fitur menggunakan TF-IDF, pelatihan model dengan algoritma SVM dan Naïve Bayes, serta evaluasi kinerja model berdasarkan data uji. Prediksi sentimen diklasifikasikan ke dalam tiga kategori utama, yaitu positif, netral, dan negatif. Hasil penelitian menunjukkan bahwa SVM memiliki performa lebih baik dibandingkan dengan Naïve Bayes dalam menganalisis sentimen pelanggan. SVM mencatat akurasi sebesar 99%, sementara Naïve Bayes hanya mencapai 89%. Selain itu, presisi, recall, dan F1-score pada SVM juga lebih tinggi dibandingkan Naïve Bayes, terutama dalam klasifikasi sentimen positif dan netral. Hal ini menunjukkan bahwa SVM lebih efektif dalam menangkap pola sentimen dalam data ulasan pelanggan, sehingga lebih akurat dalam memprediksi tingkat kepuasan pelanggan. Dengan hasil yang diperoleh, penelitian ini menegaskan bahwa pemilihan algoritma yang tepat sangat berpengaruh terhadap akurasi analisis sentimen dalam konteks bisnis kedai kopi. Model SVM dapat menjadi solusi yang lebih optimal dalam mengembangkan sistem analisis sentimen yang digunakan untuk memantau dan meningkatkan kepuasan pelanggan. Penelitian selanjutnya dapat mengeksplorasi penggunaan model hybrid atau teknik deep learning untuk meningkatkan akurasi prediksi sentimen dalam skala yang lebih luas.**

***Kata Kunci*— SVM,Naive Bayes, sentimen, kepuasan, machine learning.**

I. PENDAHULUAN

Industri kopi telah mengalami perkembangan pesat dalam beberapa tahun terakhir, menjadikannya salah satu sektor yang paling kompetitif dalam industri makanan dan minuman. Konsumsi kopi tidak hanya menjadi kebiasaan, tetapi juga bagian dari gaya hidup masyarakat modern. Kedai kopi telah menjadi tempat favorit bagi berbagai kalangan untuk bersantai, bekerja, atau berdiskusi bisnis. Dengan meningkatnya jumlah kedai kopi, persaingan di industri ini semakin ketat, sehingga kepuasan pelanggan menjadi faktor utama dalam mempertahankan loyalitas dan meningkatkan daya saing bisnis [1].

Kepuasan pelanggan dalam industri kedai kopi dipengaruhi oleh berbagai aspek, seperti kualitas produk, harga, pelayanan, serta kenyamanan tempat. Untuk memahami tingkat kepuasan pelanggan, banyak bisnis menggunakan survei dan ulasan pelanggan sebagai sumber data utama [2]. Namun, menganalisis ulasan dalam jumlah besar secara manual sangat tidak efisien. Oleh karena itu, pendekatan berbasis analisis sentimen menjadi solusi yang efektif dalam mengekstrak wawasan dari opini pelanggan yang tidak terstruktur

Analisis sentimen merupakan teknik berbasis kecerdasan buatan yang digunakan untuk mengidentifikasi dan mengkategorikan opini atau emosi dalam teks. Teknik ini dapat membantu bisnis dalam memahami umpan balik pelanggan secara lebih mendalam, memungkinkan pengambilan keputusan yang lebih tepat berdasarkan data yang ada. Dengan analisis sentimen, pemilik usaha dapat mengukur kepuasan pelanggan secara lebih objektif dan menentukan strategi yang lebih efektif untuk meningkatkan kualitas layanan dan produk mereka [3].

Kofind Coffee.Co adalah salah satu kedai kopi yang terletak di perbatasan Surabaya dan Sidoarjo. Kafe ini menawarkan berbagai biji kopi serta pengalaman manual brew dengan spesialisasi pada specialty bean. Dengan 70% penjualannya dilakukan secara daring, Kofind memiliki banyak ulasan pelanggan di platform e-commerce seperti Tokopedia. Namun, meskipun menghadapi persaingan ketat di sektor kopi, Kofind belum dapat mengoptimalkan analisis data dari ulasan pelanggan untuk meningkatkan layanan dan pengembangan produk. Selain itu, belum tersedia sistem visualisasi data yang dapat menjelaskan informasi mengenai kepuasan pelanggan secara sistematis.

Untuk mengatasi tantangan ini, analisis sentimen dapat menjadi alat yang efektif dalam memahami opini pelanggan secara lebih mendalam. Teknik ini dapat dengan cepat dan akurat menganalisis volume besar data yang tidak terstruktur, seperti ulasan online dan postingan media sosial, yang sulit untuk dianalisis secara manual [4]. Dengan analisis sentimen, pemilik Kofind dapat memperoleh wawasan yang lebih komprehensif mengenai kepuasan pelanggan, mengidentifikasi tren atau pola tertentu, serta merancang strategi perbaikan yang lebih terarah.

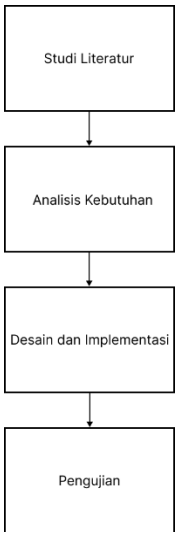
Beberapa penelitian sebelumnya telah membahas penerapan analisis sentimen dalam berbagai sektor. Penelitian yang membahas Analisis Sentimen Konsumen KFC Berdasarkan Pendekatan Naive Bayes dan AdaBoost Berbasis Data Twitter [5], sementara penelitian yang mengeksplorasi Metode SVM dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter [6], yang berfokus pada polaritas dan intensitas sentimen. Selain itu, studi yang melakukan Studi Komparasi Metode SVM dan Naive Bayes pada Data Banjir di Indonesia, meneliti fitur ekstraksi dan model deep learning dalam analisis sentimen [7]. Penelitian lainnya mencakup Sentimen Analisis Publik terhadap Joko Widodo terkait Wabah Covid-19 Menggunakan Metode Machine Learning [8], serta Analisis Algoritma SVM dan Naive Bayes dalam Klasifikasi Pinjaman pada Koperasi Simpan Pinjam [9].

Dalam penelitian ini, pendekatan konvensional menggunakan Support Vector Machine (SVM) dan Naive Bayes akan digunakan untuk melakukan analisis sentimen terhadap ulasan pelanggan Kofind. Algoritma SVM dikenal memiliki performa yang baik dalam menangani data berdimensi tinggi dan dataset kecil, sementara Naive Bayes menawarkan pendekatan probabilistik yang sederhana namun efektif dalam tugas klasifikasi teks. Dengan membandingkan kedua algoritma ini, penelitian ini bertujuan untuk mengidentifikasi metode yang lebih optimal dalam menganalisis sentimen pelanggan kedai kopi.

Penelitian ini memiliki beberapa kebaruan dibandingkan penelitian sebelumnya. Meskipun banyak studi telah membahas analisis sentimen dalam berbagai domain, masih sedikit yang secara khusus membandingkan algoritma SVM dan Naive Bayes dalam konteks kepuasan pelanggan kedai kopi. Dengan demikian, penelitian ini diharapkan dapat memberikan wawasan lebih lanjut mengenai efektivitas kedua algoritma tersebut dalam analisis sentimen pelanggan, serta memberikan rekomendasi bagi pemilik Kofind untuk meningkatkan kualitas layanan dan strategi bisnis mereka.

II. METODE PENELITIAN

Penelitian ini dilakukan melalui serangkaian tahapan sistematis, dimulai dari studi literatur untuk memahami konsep analisis sentimen, algoritma Support Vector Machine (SVM) dan Naïve Bayes, serta implementasi metode pada industri kopi. Data ulasan pelanggan dikumpulkan dari platform e-commerce Tokopedia dan melalui tahap preprocessing untuk memastikan kualitas data sebelum dianalisis. Selanjutnya, dataset dibagi menjadi data latih dan data uji, yang kemudian digunakan untuk membangun serta mengevaluasi model berbasis SVM dan Naïve Bayes. Perbandingan kinerja kedua model dilakukan berdasarkan metrik evaluasi seperti akurasi, precision, recall, dan F1-score, guna menentukan algoritma yang lebih optimal dalam memprediksi kepuasan pelanggan.



Gbr. 1 Diagram Alur Tahapan Penelitian

Berdasarkan gambar di atas, dapat dilihat tahapan-tahapan yang akan diterapkan dalam penelitian ini, yaitu:

A. *Studi Literatur*

Tahap ini merupakan langkah krusial di mana peneliti melakukan eksplorasi serta mengumpulkan berbagai informasi dari penelitian terdahulu yang relevan dengan topik ini. Proses ini bertujuan untuk memperoleh pemahaman mendalam mengenai perkembangan terkini dalam penelitian, mengidentifikasi celah dalam literatur, serta menentukan metodologi yang paling sesuai untuk diterapkan. Dalam tahap ini, peneliti akan merujuk pada berbagai sumber, seperti buku, artikel, dan jurnal ilmiah yang membahas analisis sentimen serta penerapan algoritma Support Vector Machine (SVM) dan Naïve Bayes.

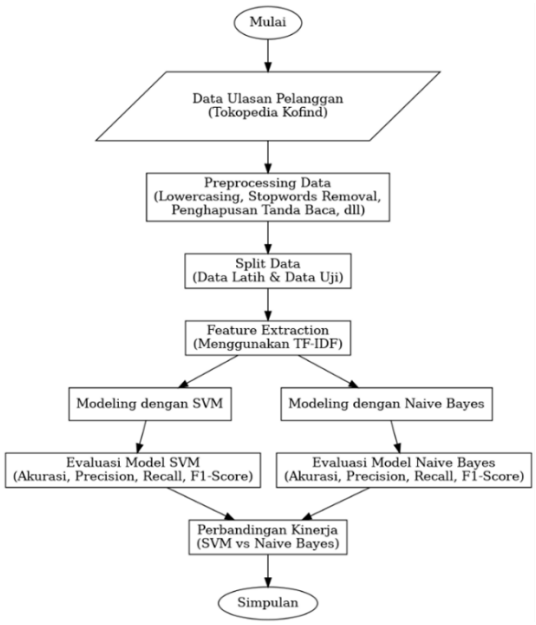
B. *Analisis Kebutuhan*

Hardware dan software yang digunakan untuk membantu penelitian dimasukkan dalam analisis kebutuhan. Berikut ini akan dijelaskan software dan hardware yang dapat menunjang

1. *Hardware*
   a. Processor Intel core i5-13400
   b. Random Access Memory 12 GB
   c. Kapasitas SSD 512GB dan 1 TB

2. *Software*
   a. Sistem Operasi Windows 11
   b. Google Collab
   c. Visual Studio Code
   d. Bahasa Pemrograman Python

C. *Desain dan Implementasi*



Gbr. 2 Flowchart Desain dan Implementasi

Gambar 2 diatas adalah ilustrasi flowchart dari metode penelitian yang akan dilakukan penulis. Dataset dalam penelitian ini diperoleh dari ulasan pelanggan kedai kopi di platform Tokopedia. Pengambilan data dilakukan melalui web scraping dengan bantuan ekstensi Chrome Scraping.io serta script Python untuk mengotomatisasi proses pengumpulan. Scraping.io digunakan untuk mengumpulkan data dalam jumlah besar dengan cepat, sementara Python digunakan untuk membersihkan dan memformat data agar sesuai untuk analisis sentimen. Data yang telah dibersihkan kemudian diberi label menjadi tiga kategori: positif, netral, dan negatif. Dataset berjumlah 486 review yang diambil dari kolom review tokopedia pada bulan Februari–Juni 2023. Pada tahap ini ada beberapa proses:

a) Preprocessing dataset

1. Lowercasing

   Semua teks dikonversi menjadi huruf kecil untuk menghindari perbedaan makna akibat perbedaan huruf kapital.

2. Stopwords Removal dan Data Cleaning

   Stopwords adalah kata-kata umum yang tidak memiliki makna signifikan dalam analisis sentimen, seperti *"dan", "di", "ke", "yang"*. Penghapusan stopwords bertujuan untuk mengurangi dimensi teks tanpa kehilangan makna penting, sednagkan data cleaning yaitu kalimat dalam kumpulan data dibersihkan dari elemen apa pun yang dapat memengaruhi temuan analisis, termasuk tautan, simbol, kata dengan dua atau lebih pengulangan, dan tanda baca yang berlebihan. Proses manual ini memungkinkan penulis untuk memastikan bahwa setiap pembersihan dilakukan dengan cermat tanpa mengorbankan makna atau konteks ulasan yang dianalisis.

3. Penghapusan Tanda Baca dan Karakter Khusus

   Tanda baca seperti *.,!?;* serta karakter spesial *@#$%* dihapus karena tidak memberikan kontribusi dalam analisis sentimen.

4. Tokenisasi

   Kalimat dipecah menjadi kata-kata atau token untuk memudahkan pemrosesan.

5. Labelling

   Proses labelling data adalah langkah penting yang harus dilakukan dalam preprocessing. Setiap ulasan perlu dibagi secara manual ke dalam sejumlah kategori yang telah ditentukan sebelumnya, seperti netral, negatif, dan positif. Penulis menggunakan skala 3-5 untuk prosedur pemberian label, dengan peringkat 3 menunjukkan negatif, peringkat 4 menunjukkan netralitas, dan peringkat 5 menunjukkan positif.

b) TF-IDF

Pada penelitian ini, proses ekstraksi fitur dilakukan menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) untuk mengubah teks ulasan pelanggan menjadi representasi numerik. TF-IDF membantu mengukur pentingnya suatu kata dalam dokumen dibandingkan dengan keseluruhan kumpulan data. Dalam penelitian ini, jumlah fitur dibatasi hingga 1000 kata yang paling sering muncul dalam dataset. Model TF-IDF yang telah dilatih kemudian disimpan agar dapat digunakan kembali pada tahap pemodelan.

Selain itu, untuk mengatasi ketidakseimbangan kelas dalam dataset, digunakan teknik oversampling *Synthetic Minority Over-sampling Technique (SMOTE)*. SMOTE bertujuan untuk menyeimbangkan jumlah sampel di setiap kategori sentimen dengan menghasilkan data sintetis untuk kelas minoritas. Dengan demikian, model yang dibangun dapat memiliki performa yang lebih baik dalam mengklasifikasikan ulasan pelanggan. Dataset yang telah melalui proses TF-IDF dan SMOTE selanjutnya digunakan dalam tahap pemodelan menggunakan algoritma Support Vector Machine (SVM) dan Naive Bayes.

D. *Pengujian*

Tahap pengujian model bertujuan untuk mengevaluasi performa algoritma Support Vector Machine (SVM) dan Naïve Bayes dalam melakukan analisis sentimen terhadap ulasan pelanggan kedai kopi. Pengujian dilakukan dengan membandingkan hasil klasifikasi model terhadap data uji

yang telah dipisahkan sebelumnya dari dataset utama. Model yang digunakan dalam penelitian ini adalah SVM dan Naïve Bayes, yang keduanya umum digunakan dalam klasifikasi teks. SVM bekerja dengan mencari hyperplane terbaik untuk memisahkan data ke dalam kelas sentimen, sementara Naïve Bayes menggunakan pendekatan probabilistik berdasarkan distribusi kata dalam setiap kategori sentimen.

1. Pembagian Data Uji dan Data Latih

Sebelum pengujian dilakukan, dataset ulasan pelanggan yang telah melalui proses preprocessing dibagi menjadi dua bagian, yaitu Data latih (training set) digunakan untuk membangun model dengan memahami pola dalam data. Data uji (testing set) digunakan untuk mengukur sejauh mana model mampu mengklasifikasikan sentimen dengan benar pada data baru yang tidak dilihat sebelumnya. Pembagian dataset dilakukan dengan pendekatan hold-out validation, di mana data dialokasikan dengan rasio 80:20, yaitu 80% untuk pelatihan dan 20% untuk pengujian.

2. Evaluasi Performa Model

Setelah model selesai dilatih, pengujian dilakukan dengan menggunakan data uji untuk mengukur performa model berdasarkan beberapa metrik evaluasi yang diperoleh dari confusion matrix, yaitu:

a) Akurasi (Accuracy): Mengukur persentase total prediksi yang benar terhadap keseluruhan data uji.
b) Presisi (Precision): Mengukur seberapa banyak prediksi positif yang benar dibandingkan dengan total prediksi positif yang dibuat oleh model.
c) Recall (Sensitivity): Menunjukkan seberapa baik model dalam menemukan semua contoh positif dalam data uji.
d) F1-Score: Rata-rata harmonik dari presisi dan recall yang memberikan keseimbangan antara kedua metrik tersebut.

Metrik-metrik ini dihitung untuk masing-masing algoritma, yaitu SVM dan Naïve Bayes, guna melihat keunggulan dan kelemahan masing-masing metode dalam menganalisis sentimen pelanggan.

3. Perbandingan Kinerja Model

Setelah evaluasi dilakukan, hasil perbandingan antara SVM dan Naïve Bayes dianalisis untuk menentukan algoritma yang lebih optimal dalam mengklasifikasikan sentimen pelanggan. Perbandingan dilakukan dengan melihat nilai akurasi serta metrik lainnya, seperti presisi, recall, dan F1-score.

Jika salah satu model memiliki kinerja yang lebih unggul secara konsisten dalam berbagai metrik, maka model tersebut direkomendasikan sebagai metode yang lebih efektif dalam analisis sentimen ulasan pelanggan di Kofind Coffee.co. Namun, jika perbedaan performa antara kedua model tidak signifikan, maka analisis lebih lanjut akan dilakukan untuk mengeksplorasi potensi peningkatan akurasi dengan metode optimasi lainnya, seperti penggunaan teknik feature extraction yang lebih kompleks atau penyesuaian parameter model.

Dengan adanya tahap pengujian ini, penelitian dapat memberikan rekomendasi berbasis data mengenai algoritma yang paling sesuai untuk mengolah dan memahami sentimen pelanggan secara otomatis, sehingga dapat membantu pemilik usaha dalam meningkatkan kualitas layanan mereka.

III. HASIL DAN PEMBAHASAN

Bab ini membahas mengenai implementasi sistem dari rancangan yang telah dibuat serta proses pengujian yang dilakukan sesuai dengan perencanaan sebelumnya.

A. Dataset

Jumlah Dataset yang digunakan adalah 486 review yang diambil dari kolom review tokopedia pada bulan Februari–Juni 2023. Data yang digunakan penulis sumbernya berasal dari web scrapping ulasan online shop Tokopedia. Teknik pengumpulan data(scrapping) dilakukan dengan python. Dataset akan berupa file csv dan json.

TABEL I
DATASET PENELITIAN

No	Ulasan
1	Oke dan mantap. Pengiriman oke, seller oke, kopi nikmat.
2	packing rapi, aman, dapat bonus bean
3	tolong diperbaiki lagi packingnya gan..kalau bisa lebih dari bubble wrap saja
4	Mantap fast respon fast delivery

Tabel 1 menampilkan beberapa contoh ulasan yang terdapat dalam dataset yang digunakan dalam penelitian ini. Sebelum digunakan, dataset telah melalui proses pengecekan untuk memastikan tidak ada nilai null, sehingga semua ulasan yang akan diproses pada tahap preprocessing data sudah lengkap dan valid.

B. Preprocessing Data

Di tahap preprocessing ini ada 6 bagian yaitu:

1) Lowercasing

TABEL II
HASIL LOWERCASE

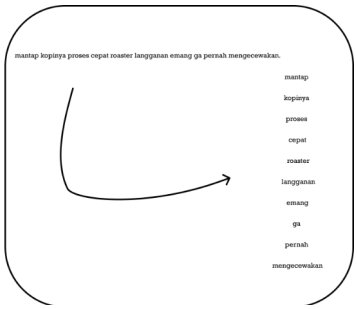
Sebelum	Sesudah
"Fast response! DAEBAK!"	fast response! daebak!
bold, ringan, aromatik, bikin penasaran tiap tegukan. JADI ASIK BANGET MINUMNYA 😆 SUKA!! Masuk ke list roastery favorit ku.	bold, ringan, aromatik, bikin penasaran tiap tegukan. Jadi asik banget minumnya suka!! masuk ke list roastery favorit ku.

2) Data Cleaning

TABEL III
HASIL DATA CLEANING

Ulasan	Hasil Data Cleaning
Robusta tp tidak terlalu pahit rasa dan aroma choco nya kurang pengiriman pk an**r *j* lmyn lama barang di lempar pdhal didlm rmh ada orangnya (cctv)	robusta tp tidak terlalu pahit rasa dan aroma choco nya kurang pengiriman pk lmyn lama barang di lempar pdhal didlm rmh ada orangnya cctv

3) Tokenisasi



Gbr. 3 Hasil Tokenisasi

4) Labelling

TABEL IV
HASIL LABELLING

No	Kategori	Jumlah	Total
1	Positif	413	486
2	Negatif	42	
3	Netral	31	

C. Split Data

Di sini, data training digunakan sebagai bentuk model. Data testing dimanfaatkan sebagai data uji yang berguna untuk menguji model dengan menggunakan data testing tersebut, sedangkan. Tujuannya adalah untuk mengetahui seberapa cocok model yang telah dibuat oleh penulis. Rasio 80% hingga 20% digunakan untuk memisahkan data ke dalam set pelatihan dan pengujian

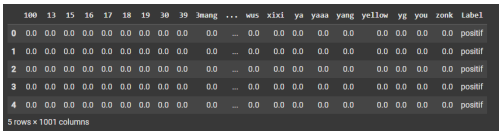
TABEL V
HASIL PEMBAGIAN DATA

No	Kategori	Tipe Data	Jumlah
1.	Positif	latih	330
		uji	83
2.	Netral	latih	25
		uji	6
3	Negatif	latih	33
		uji	8

Karena jumlah kategori positif, negatif dan netral yang sangat tidak seimbang maka akan dilakukan oversampling agar dataset menjadi lebih balance. Teknik ini akan melibatkan peningkatan jumlah contoh di kelas minoritas untuk menyamai jumlah contoh di kelas mayoritas (ulasan positif). Metode oversampling yang akan digunakan penulis adalah SMOTE. Untuk pengujian penulis menggunakan metode cross validation. Pengujian dilakukan sebanyak 10 kali (k=10) untuk mendapatkan hasil akurasi,precison,dan recall yang optimal.

D. Feature Extraction dengan TF-IDF

Setelah melewati tahap preprocessing, data kemudian diproses dalam tahap ekstraksi fitur. Pada tahap ini, digunakan metode TF-IDF untuk mengubah data teks menjadi representasi numerik. Dengan memanfaatkan TfidfVectorizer dari library scikit-learn, data hasil preprocessing dikonversi ke dalam bentuk TF-IDF.



	100	13	15	16	17	18	19	30	39	3mang	...	wus	xixi	ya	yaaa	yang	yellow	yg	you	zonk	Label
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif

5 rows × 1001 columns

Gbr. 4 Hasil TF IDF

Pada gambar 4 adalah sebagian dari hasil TF-IDF data 4 baris awal menggunakan TfidfVectorizer.

E. *Modelling SVM dan Naive Bayes*

Pada tahap ini, dilakukan pemodelan menggunakan dua algoritma yang dibandingkan, yaitu **Support Vector Machine (SVM)** dan **Naïve Bayes**. Model dibangun berdasarkan data ulasan pelanggan yang telah melalui tahapan preprocessing dan ekstraksi fitur menggunakan TF-IDF.

1) Training Model

```
# Training SVM
svm_model = SVC(kernel='linear') # You can experiment with different kernels
svm_model.fit(X_train, y_train)
svm_predictions = svm_model.predict(X_test)
print("SVM Accuracy:", accuracy_score(y_test, svm_predictions))
print(classification_report(y_test, svm_predictions))
```

Gbr. 5 Training SVM

Pada tahap awal modeling SVM dilakukan pelatihan model Support Vector Machine (SVM) untuk memprediksi sentimen pelanggan berdasarkan data yang telah melalui proses preprocessing dan ekstraksi fitur seperti yang dicantumkan pada *Gambar 5* diatas. Model dilatih menggunakan data latih (training set) yang telah dibagi sebelumnya, di mana SVM bertugas untuk menemukan hyperplane terbaik yang memisahkan kategori sentimen secara optimal. Proses pelatihan ini menggunakan kernel linear, yang umum digunakan dalam klasifikasi teks karena mampu menangani data berdimensi tinggi dengan baik. Setelah model selesai dilatih, dilakukan pengujian terhadap data uji (test set) untuk melihat sejauh mana model mampu mengklasifikasikan sentimen pelanggan dengan akurat. Evaluasi performa model dilakukan dengan menghitung *accuracy dan classification report*, yang menampilkan metrik seperti precision, recall, dan F1-score. Hasil evaluasi ini akan dibandingkan dengan model **SVM** untuk menentukan algoritma yang lebih optimal dalam analisis sentimen pelanggan di Kofind.

```
# Training Naive Bayes
nb_model = MultinomialNB()
nb_model.fit(X_train, y_train)
nb_predictions = nb_model.predict(X_test)
print("\nNaive Bayes Accuracy:", accuracy_score(y_test, nb_predictions))
print(classification_report(y_test, nb_predictions))
```

Gbr. 6 Training Naive Bayes

Untuk training model Naive Bayes masih sama seperti training SVM diatas. Proses pelatihan juga dilakukan dengan membiarkan model mempelajari pola dari data latih, sehingga dapat mengenali karakteristik teks dalam setiap kategori sentimen. Setelah itu, model digunakan untuk melakukan prediksi terhadap data uji guna mengklasifikasikan sentimen pelanggan.

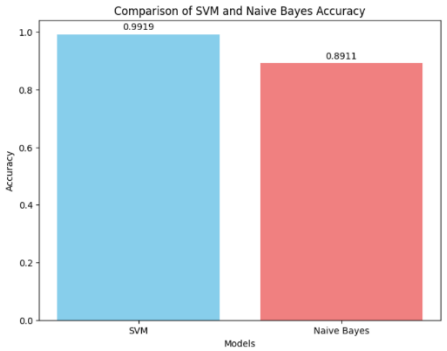
2) Hasil Prediksi Model

```
SVM Accuracy: 0.9919354838709677
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        88
           1       0.97      1.00      0.99        73
           2       1.00      0.98      0.99        87

    accuracy                           0.99       248
   macro avg       0.99      0.99      0.99       248
weighted avg       0.99      0.99      0.99       248


Naive Bayes Accuracy: 0.8911290322580645
              precision    recall  f1-score   support

           0       0.95      1.00      0.97        88
           1       0.77      1.00      0.87        73
           2       1.00      0.69      0.82        87

    accuracy                           0.89       248
   macro avg       0.90      0.90      0.89       248
weighted avg       0.91      0.89      0.89       248
```

Gbr. 7 Prediksi Model

Hasil yang didapatkan dari kedua model berbeda dimana Model SVM menghasilkan akurasi sebesar 99.19% (0.9919354838709677) dan Model Naive Bayes menghasilkan akurasi sebesar 89.11% (0.8911290322580645).

3) Perbandingan Kinerja Model



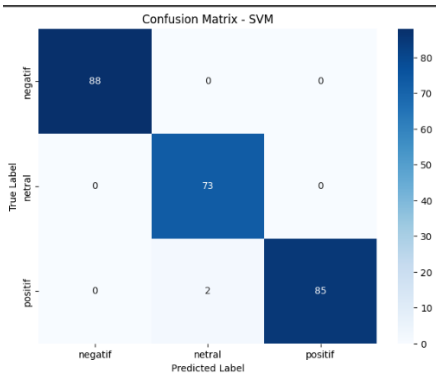
Gbr. 8 Grafik Perbandingan Model

Gambar 8 di atas menunjukkan perbandingan akurasi antara model Support Vector Machine (SVM) dan Naïve Bayes dalam melakukan klasifikasi sentimen ulasan pelanggan. Dari grafik tersebut, terlihat bahwa SVM memiliki akurasi sebesar 99.19%, sedangkan Naïve Bayes memiliki akurasi sebesar 89.11%. Perbedaan akurasi ini menunjukkan bahwa SVM lebih unggul dalam

mengklasifikasikan sentimen dibandingkan dengan Naïve Bayes. Nilai akurasi yang lebih tinggi pada SVM menunjukkan bahwa model ini mampu memisahkan kelas sentimen dengan lebih baik, serta lebih efektif dalam mengenali pola pada data yang digunakan.

Di sisi lain, meskipun Naïve Bayes masih mencapai akurasi yang cukup baik, hasilnya lebih rendah dibandingkan SVM. Hal ini dapat disebabkan oleh sifat asumsi independensi fitur pada Naïve Bayes, yang mungkin tidak sepenuhnya sesuai dengan karakteristik data teks yang memiliki ketergantungan antar kata.

4) Confusion Matrix SVM dan Naive Bayes



Gbr. 9 Confusion Matrix SVM

Confusion Matrix yang dihasilkan dari model Support Vector Machine (SVM) menunjukkan bahwa model ini memiliki kinerja yang sangat baik dalam mengklasifikasikan data. Berikut rincian dari setiap komponen yang ditampilkan:

- Klasifikasi label negatif: Dari 88 sampel data dengan label negatif, seluruhnya (88) diprediksi dengan benar sebagai negatif, tanpa ada kesalahan klasifikasi.
- Klasifikasi label netral: Dari 73 sampel data berlabel netral, seluruh prediksi juga tepat dengan nilai yang benar (73), menunjukkan tidak adanya kesalahan prediksi untuk kategori ini.
- Klasifikasi label positif: Dari 87 sampel data berlabel positif, 85 di antaranya diprediksi benar sebagai positif, sementara hanya 2 sampel yang diklasifikasikan salah sebagai netral.
    - Nilai akurasi, recall, dan skor f1 untuk pengujian SVM dapat kita amati dari temuan confussion matrix:

1. Akurasi

$$\frac{88 + 73 + 85}{248} = 0.99 \text{ atau } 99\%$$

2. Presisi

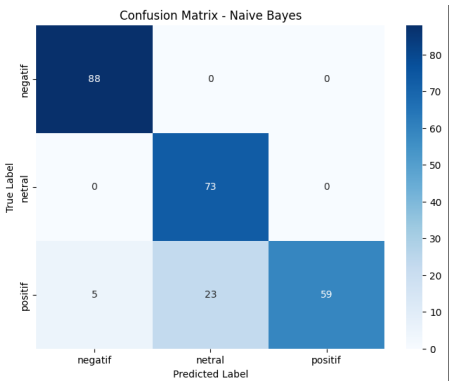
$$\frac{88}{88 + 0 + 0} = 1.00$$

3. Recall

$$\frac{85}{85 + 2} \approx 0.98$$

4. F1 Score

$$2 \times \frac{1.00 \times 0.98}{1.00 + 0.98} \approx 0.99$$



Gbr. 10 Confusion Matrix Naïve Bayes

Berbeda dengan model SVM, Confusion Matrix dari model Naive Bayes menunjukkan performa yang lebih rendah, terutama dalam klasifikasi label positif. Berikut rincian hasilnya:

- Klasifikasi label negatif: Sama dengan model SVM, dari 88 sampel data berlabel negatif, seluruh prediksi benar (88).
- Klasifikasi label netral: Sebanyak 73 sampel data berlabel netral diprediksi dengan benar sebagai netral, tanpa kesalahan.
- Klasifikasi label positif: Dari 87 sampel data positif, model ini hanya mampu memprediksi 59 sampel dengan benar sebagai positif. Terdapat 23 sampel yang salah diklasifikasikan sebagai netral dan 5

sampel yang salah diklasifikasikan sebagai negatif.65

- Nilai akurasi, recall, dan skor f1 untuk pengujian Naïve Bayes dapat kita amati dari temuan confussion matrix:

1. Akurasi

$$\frac{88 + 73 + 59}{248} = 0.89 \text{ atau } 89\%$$

2. Presisi

$$\frac{88}{88 + 0 + 5} = 0.95$$

3. Recall

$$\frac{59}{59 + 28} \approx 0.68$$

4. F1 Score

$$2 \times \frac{1.00 \times 0.98}{1.00 + 0.98} \approx 0.99$$

Dari perhitungan menggunakan rumus-rumus di atas, dapat dilihat bahwa model SVM memiliki metrik akurasi, presisi, recall, dan F1-Score yang lebih baik dibandingkan dengan Naive Bayes, terutama dalam memprediksi sentimen positif. Hal ini mendukung hasil evaluasi yang menunjukkan bahwa model SVM lebih unggul dalam klasifikasi sentimen yang digunakan dalam penelitian ini. Berikut tabel perbandingan antar hasil confusion matrix SVM dan naïve Bayes

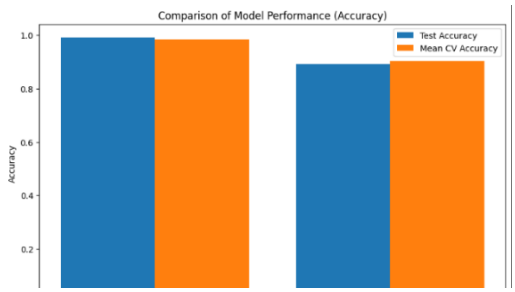
TABEL VI
PERBANDINGAN CONFUSION MATRIX

Metrik	SVM	Naive Bayes
Akurasi	99.19%	89.11%
Precision	98.99%	87.88%
Recall	99%	82.35%
F1 Score	98.99%	84.53%

5) Mean Cross Validation

Hasil dari rata-rata akurasi cross-validation (Mean CV) memainkan peran penting dalam mengevaluasi performa model dengan lebih menyeluruh. Cross-validation membantu membagi data latih menjadi beberapa subset atau lipatan (folds), di mana model dilatih pada beberapa lipatan dan diuji pada lipatan lainnya secara bergantian. Dengan demikian, Mean CV memberikan gambaran umum tentang bagaimana model berperilaku pada berbagai subset data, yang mengukur konsistensi dan kemampuan generalisasi model.

Cross-validation juga membantu mengurangi kemungkinan overfitting pada data latih. Tanpa cross-validation, model mungkin terlihat sangat akurat pada satu set data latih tetapi berkinerja buruk pada data baru. Mean CV membantu meminimalkan bias dalam pemilihan data latih dan uji sehingga hasil akurasi yang dilaporkan lebih stabil dan tidak terlalu dipengaruhi oleh keberuntungan dalam pemilihan data.



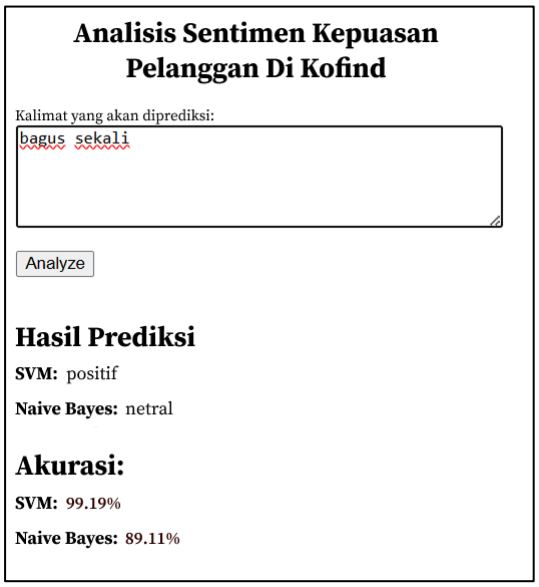
Gbr. 11 Grafik Perbandingan mean CV

Pada proses evaluasi kedua model diuji menggunakan metode cross-validation dan pengujian langsung pada data uji. Hasil validasi silang menunjukkan bahwa model SVM memiliki rata-rata akurasi cross-validation sebesar 98,33%, dengan skor yang sangat stabil pada setiap lipatan validasi (range antara 96,96% hingga 98,98%). Sebaliknya, model Naive Bayes mencatatkan rata-rata akurasi cross-validation sebesar 90,31%, dengan variasi skor yang lebih besar. Nilai rata-rata akurasi yang tinggi dan stabil pada SVM mengindikasikan bahwa model ini mampu generalisasi dengan baik pada data latih dan memiliki risiko overfitting yang rendah. Pada pengujian langsung menggunakan data uji, SVM kembali unggul dengan akurasi mencapai 99,19%, jauh lebih tinggi dibandingkan Naive Bayes yang hanya mencapai 89,11%. Performa yang konsisten pada data uji memperkuat kesimpulan bahwa SVM adalah model yang lebih efektif untuk kasus klasifikasi ini.

Dilihat dari laporan klasifikasi, model SVM menunjukkan presisi, recall, dan f1-score yang tinggi

untuk semua kelas, dengan nilai rata-rata tertimbang (weighted average) mendekati 1.00, mengindikasikan kemampuan prediksi yang hampir sempurna. Model ini secara khusus memprediksi kelas negatif dan positif dengan akurasi sempurna, meskipun terdapat sedikit ketidakakuratan pada kelas netral. Sebaliknya, model Naive Bayes mengalami kesulitan dalam memprediksi kelas netral, dengan f1-score hanya sebesar 0,74, akibat penurunan presisi dan recall. Ketidakseimbangan ini mempengaruhi performa keseluruhan model Naive Bayes, meskipun kelas negatif dan positif masih terprediksi dengan baik. Dengan demikian, berdasarkan analisis komprehensif dari kedua model, SVM dapat dikategorikan sebagai algoritma yang lebih optimal untuk memprediksi sentimen teks ulasan, berkat kestabilan dan akurasi yang lebih unggul pada data yang digunakan

6) Tampilan GUI Prediksi Sentimen



Gbr. 12 Tampilan GUI

Tampilan antarmuka grafis (GUI) yang dikembangkan dalam penelitian ini berfungsi sebagai alat untuk menganalisis sentimen kepuasan pelanggan di Kofind berdasarkan ulasan yang diberikan. Pengguna dapat memasukkan teks ulasan ke dalam kolom input, kemudian sistem akan memproses dan memprediksi sentimen menggunakan dua model pembelajaran mesin, yaitu Support Vector Machine (SVM) dan Naïve Bayes. Hasil prediksi ditampilkan dalam bentuk kategori sentimen (positif, netral, atau negatif) untuk masing-masing model, serta dilengkapi dengan informasi akurasi model yang menunjukkan tingkat kinerja masing-masing algoritma. Dengan adanya GUI ini, proses analisis sentimen menjadi lebih interaktif dan memudahkan pengguna dalam memahami hasil klasifikasi sentimen secara langsung.

IV. KESIMPULAN

Penelitian ini berfokus pada analisis sentimen ulasan pelanggan untuk memprediksi tingkat kepuasan pelanggan di kedai kopi Kofind dengan menerapkan metode machine learning. Dua algoritma yang digunakan dalam penelitian ini adalah Support Vector Machine (SVM) dan Naïve Bayes, yang memanfaatkan teknik Term Frequency-Inverse Document Frequency (TF-IDF) untuk mengekstraksi fitur dari teks ulasan pelanggan. Proses implementasi dilakukan melalui tahapan preprocessing data, ekstraksi fitur, pelatihan model, dan pengujian model guna mengklasifikasikan sentimen pelanggan ke dalam tiga kategori utama, yaitu positif, netral, dan negatif.

Berdasarkan perbandingan kinerja kedua model, hasil penelitian menunjukkan bahwa SVM memiliki performa lebih baik dibandingkan dengan Naïve Bayes. SVM mencapai akurasi 99%, sementara Naïve Bayes hanya mencapai 89%. Selain itu, presisi, recall, dan f1-score dari SVM lebih tinggi, terutama dalam klasifikasi sentimen positif dan netral, yang merupakan kategori penting dalam menentukan tingkat kepuasan pelanggan. Dengan demikian, dapat disimpulkan bahwa SVM lebih efektif dalam menganalisis sentimen berbasis ulasan pelanggan dan lebih direkomendasikan sebagai model untuk sistem prediksi kepuasan pelanggan dibandingkan dengan Naïve Bayes.

UCAPAN TERIMA KASIH

Puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat, taufik, serta hidayah-Nya sehingga penulis dapat menyelesaikan penyusunan jurnal ini dengan baik dan tepat waktu. Jurnal ini merupakan hasil dari proses penelitian yang telah dilakukan dengan berbagai tantangan, dan atas pertolongan-Nya, semua dapat terselesaikan dengan baik. Penulis juga ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan, baik secara akademik maupun moral. Ucapan terima kasih ditujukan kepada dosen pembimbing yang telah memberikan bimbingan, arahan, serta masukan yang sangat berarti dalam proses penelitian ini. Tak lupa, penulis juga mengucapkan terima kasih kepada keluarga dan teman-teman yang selalu memberikan dukungan, motivasi, serta semangat selama proses penyusunan jurnal ini.

Semoga segala bantuan, dukungan, dan ilmu yang telah diberikan mendapatkan balasan yang setimpal dari Allah SWT. Penulis menyadari bahwa jurnal ini masih jauh dari kesempurnaan, sehingga saran dan kritik yang membangun sangat diharapkan demi perbaikan di masa yang akan datang. Semoga jurnal ini dapat memberikan manfaat bagi pembaca serta menjadi kontribusi dalam pengembangan penelitian di bidang analisis sentimen dan pembelajaran mesin.

REFERENSI

[1] Wulandari, I. S. (2012). Analisis Pengaruh Kualitas Pelayanan Terhadap Kepuasan Konsumen pada Kedai Kopi Starbucks Cabang Alam Sutera. Jurnal ekonomi Manajemen Universitas Gunadarma

[2] Maris, E. R. (2015). Analisis Kepuasan Pelanggan Menggunakan Algoritma C4. 5. *Teknik Informatika, Fakultas Ilmu KomputerUniversitas Dian Nuswantoro, Semarang*.

[3] Wattimena, A. D. (2018). *Analisis Sentimen Teks Bahasa Indonesia Pada Media Sosial Menggunakan Algoritma Convolutional Neural Network (Studi Kasus: E-Commerce)* (Doctoral dissertation, Institut Teknologi Sepuluh Nopember).

[4] Hayatin, N., Marthasari, G. I., & Nuraini, L. (2020). Optimization of Sentiment Analysis for Indonesian Presidential Election using Naïve Bayes and Particle Swarm Optimization. Jurnal Online Informatika, 5(1), 81–88. https://doi.org/10.15575/join.v5i1.558

[5] Andhika, L. D., Cahyani, D. R., Saputra, D., Herawati, T., Khoiruddinsyah, M., & Saputra, D. D. (2023). Analisis Sentimen Kosumen KFC Berdasarkan Pendekatan Naive Bayes dan Ada Boost Berbasis Data Twitter. *Jurnal INSAN Journal of Information System Management Innovation*, *3*(1), 55-61.

[6] Khatib Sulaiman, J., Atmajaya, D., Febrianti, A., Darwis, H., Artikel Abstrak, I., & Kunci, K. (2023). Metode SVM dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter. Indonesian Journal of Computer Science Attribution, 12(4), 2173.

[7] Abdullah, R. K., & Utami, E. (2018). Studi Komparasi Metode SVM dan Naive Bayes pada Data Bencana Banjir di Indonesia. *JURNAL TECNOSCIENZA*, *3*(1), 103-122.

[8] Hikmawan, S., Pardamean, A., Nur Khasanah, S., Mandiri, N., Damai No, J., Jati Barat, W., & Selatan, J. (2020). Sentimen Analisis Publik Terhadap Joko Widodo Terhadap Wabah Covid-19 Menggunakan Metode Machine Learning (Vol. 20, Issue 2). http://ejurnal.ubharajaya.ac.id/index.php/JKI

[9] Riyadi, S., Siregar, M. M., Margolang, K. fadhli F., & Andriani, K. (2022). ANALYSIS OF SVM AND NAIVE BAYES ALGORITHM IN CLASSIFICATION OF NAD LOANS IN SAVE AND LOAN COOPERATIVES. JURTEKSI (Jurnal Teknologi Dan Sistem Informasi), 8(3), 261–270. https://doi.org/10.33330/jurteksi.v8i3.1483