ANALISIS SENTIMEN KOMENTAR BERITA DETIK.COM MENGGUNAKAN ALGORITMA SUPORT VEKTOR MACHINE (SVM)

Hendiana [1], Ade Irma Purnamasari [2], Irfan Ali [3]
[1,2] Teknik Informatika, STMIK IKMI Cirebon
[3] Rekayasa Perangkat Lunak, STMIK IKMI Cirebon
Jalan Perjuanagan No. 10 B Majasem, Kec.Kesambi, Kota Cirebon, Jawa Barat 45135, Indonesia
hendiana1933@gmail.com

ABSTRAK

Dalam konteks analisis sentimen terhadap komentar berita, permasalahan utama adalah kompleksitas dalam memahami dan menyebarkan opini serta tanggapan masyarakat terhadap berita . Dengan banyaknya komentar yang besar, menilai apakah umpan balik bersifat positif atau negatif, menjadi tugas yang rumit. Oleh karena itu, perlu adanya pendekatan yang canggih dan efektif, seperti menggunakan algoritma SVM, untuk meningkatkan presisi dan efisiensi analisis sentimen.dalam penelitian Data berisi 1000 komentar dari berbagai judul dan topik berita yang diambil dari website berita Detik .Com.komentar tersebut di Analisis menggunakan algoritma SVM untuk menentukan tingkat sentimen negatif dan positif. Performa SVM dalam analisis sentimen diukur dengan perhitungan akurasi, presisi, recall, dan f1-score.dari 400 komentar yang menghasilkan 288 komentar negatif dengan nilai presisi:0.99, recall:0.98, dan f-score:0.99 dan 112 komentar positif dengan nilai presisi :0.95, recall:0.98, dan f-score: 0.96, dengan akurasi sebesar 0.98Hasil penelitian ini menunjukkan bahwa kompleksitas bahasa Indonesia dalam komentar berita di Detik.com mempengaruhi sentimen distribusi. Hasil menunjukkan bahwa sentimen negatif lebih dominan daripada sentimen positif dalam komentar berita di Detik.com Berdasarkan temuan penelitian, disarankan agar pengembangan model analisis sentimen lebih lanjut mempertimbangkan peningkatan dalam mengatasi kompleksitas bahasa Indonesia.

Kata kunci : *Analisis Sentimen, SVM ,Komentar Berita,Detik.com,TF-ID,Kompleksitas Opini*

1. PENDAHULUAN

Dalam era perkembangan teknologi yang pesat, bidang Informatika menjadi pilar utama yang mengakibatkan dampak signifikan pada berbagai aspek kehidupan. Transformasi digital menciptakan paradigma baru di dunia teknologi, bisnis, pendidikan, dan sektor lainnya, khususnya dalam analisis sentimen terhadap komentar berita di platform daring. Penelitian ini berfokus pada analisis sentimen terhadap komentar berita di situs detik.com, sebagai sumber berita utama bagi masyarakat Indonesia. Kompleksitas dalam memahami dan mengevaluasi opini serta tanggapan masyarakat terhadap berita menjadi tantangan utama, terutama dengan jumlah komentar yang besar. Oleh karena itu, penelitian ini mengusulkan pendekatan dengan menggunakan algoritma Support Vector Machine (SVM) untuk meningkatkan ketepatan dan efisiensi analisis sentimen terhadap komentar berita. Tujuan utama penelitian adalah mencari pengaruh kompleksitas bahasa Indonesia dalam konteks komentar berita terhadap akurasi analisis sentimen, dengan harapan dapat mengisi kesenjangan literatur dan memberikan kontribusi pada pengembangan metode analisis sentimen dalam konteks berita daring. Implikasi besar dari hasil penelitian ini melibatkan pemahaman yang lebih baik terhadap sentimen masyarakat terhadap berita di era digital. Jika berhasil, model analisis sentimen berbasis SVM dapat menjadi alat penting bagi praktisi media, peneliti, dan pengambil keputusan, serta berpotensi untuk pengembangan alat serupa pada platform berita lainnya, meningkatkan efektivitas pengelolaan konten berita di era digital.

2. TINJAUAN PUSTAKA

Dalam era perkembangan teknologi yang pesat, bidang Informatika menjadi pilar utama yang memberikan dampak signifikan pada berbagai aspek kehidupan. Transformasi digital telah menciptakan paradigma baru di dunia teknologi, bisnis, pendidikan, dan sektor lainnya, dengan penelitian ini berfokus pada analisis sentimen terhadap komentar berita di situs detik.com sebagai sumber berita utama bagi masyarakat Indonesia. Tantangan utama yang muncul dalam konteks analisis sentimen terhadap komentar berita adalah kompleksitas dalam menilai opini dan tanggapan masyarakat, terutama dengan volume komentar yang besar. Oleh karena itu, diperlukan pendekatan canggih dan efektif, seperti penggunaan algoritma Support Vector Machine (SVM), untuk meningkatkan ketepatan dan efisiensi analisis sentiment.dalam komentar netizen pada portal berita online di Indonesia menggunakan metode Neural Language Processing dengan algoritma Support Vector Machine (SVM). Hasilnya menunjukkan tingkat akurasi sebesar 53,88%, dengan Recall 49,69%, Precision 48,77%, Classification error 46,12%, dan fmeasure 49,23%. Temuan ini dapat menjadi dasar bagi portal berita untuk menerapkan sistem filtering guna meminimalisir kasus Hate Speech di media online[1].pada analisis sentimen judul berita online ekonomi terkait COVID-19 menggunakan metode Support Vector Machine.

Dengan 1000 data judul berita yang diberi label positif dan negatif, uji coba menunjukkan performa terbaik pada rasio 9:1 dengan akurasi 76%, presisi 80,48%, recall 89,18%, dan f-measure 84,61%. Hasil ini menggambarkan efektivitas metode SVM dalam mengukur sentimen terhadap berita ekonomi di era pandemi[2].SVM dan TF-IDF untuk mengklasifikasikan opini masyarakat di Twitter terkait New Normal di Indonesia. Dengan akurasi 76.5%, recall 90.91%, dan presisi 70.80%, penelitian ini berhasil mengidentifikasi opini sebagai positif atau negatif terhadap tatanan New Normal. Metode ini memanfaatkan crawling data melalui RapidMiner untuk mengumpulkan data dari Twitter. Temuan ini dapat menjadi pertimbangan dalam pengambilan keputusan terkait New Normal[3]. mengeksplorasi opini masyarakat Indonesia, khususnya di Twitter, terkait vaksin Covid-19 dengan mengumpulkan 3131 cuitan yang diklasifikasikan sebagai sentimen positif dan negatif. Algoritma Support Vector Machine (SVM) digunakan untuk klasifikasi data, mencapai akurasi terbaik sebesar 94.88% dengan perbandingan data training 90% dan data testing 10%. Analisis asosiasi teks pada sentimen positif menyoroti kata-kata terkait sehat, dukungan, dan ekonomi, sementara sentimen negatif terkait dengan ketakutan, virus, dan efek samping vaksin. Hasil negatif dianalisis menggunakan diagram pohon untuk pemecahan masalah. Studi ini memberikan wawasan tentang pandangan masyarakat terhadap vaksin Covid-19 di Indonesia melalui platform Twitter[4].Penelitian ini menggunakan analisis sentimen dengan algoritma Support Vector Machine (SVM) untuk mengklasifikasikan berita menjadi dua kelas, yaitu positif dan negatif. Dikembangkan sebuah aplikasi berbasis CodeIgniter (PHP) yang menggunakan 100 data berita dari detikcom. Aplikasi mencapai tingkat akurasi sebesar 76%, membantu dalam menilai tingkat opini positif atau negatif dari berita tersebut[5].Penelitian ini menggunakan metode Support Vector Machine (SVM) dengan dataset berisi 1000 tweet untuk mengklasifikasikan sentimen masyarakat terhadap tokoh masyarakat di Twitter. SVM dikombinasikan dengan pembobotan menggunakan TF-IDF. Kernel terbaik adalah RBF dengan akurasi rata-rata sebesar 84%, menggunakan nilai c 1, gamma 0,01, ambang batas 0.3, dan maksimal iterasi 1000 kali. Studi ini bertujuan untuk menganalisis dan mengukur kinerja SVM dalam mengklasifikasikan sentimen terhadap tokoh masyarakat di platform Twitter[6].membahas dampak pandemi COVID-19 terhadap industri perfilman Indonesia dan adaptasi pelaku industri dengan memanfaatkan teknologi streaming online, khususnya aplikasi Iflix. Data ulasan Iflix (4,501 ulasan) dari Januari hingga Maret 2021 dianalisis menggunakan metode Support Vector Machine (SVM) dengan akurasi tertinggi pada skenario 70% data latih dan 30% data uji (93,45%). Visualisasi sentimen positif mencakup kata-kata seperti film, bagus, aplikasi, dan menghibur, sedangkan sentimen negatif mencakup kata-kata seperti jelek, lama, dan pulsa. Ulasan negatif dianalisis menggunakan diagram fishbone untuk mengidentifikasi faktor-faktor seperti price, product, people, process, dan promotion[7].Data tweet diolah dengan text preprocessing dan diklasifikasikan menggunakan algoritma Support Vector Machine dengan empat parameter kernel. Kernel linear memiliki presisi terbaik (80%), kernel sigmoid memiliki recall terbaik (85%), dan kernel sigmoid memiliki akurasi terbaik (81%). Penelitian ini memberikan gambaran tentang sentimen masyarakat terhadap tokoh tersebut di platform Twitter[8].Sentimen analis pada Twitter melibatkan proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam tweet. Pendekatan text mining menjadi alternatif terbaik untuk mengartikan makna dari berbagai macam konten yang terdapat pada tweet. Pengklasifikasian konten positif dan negatif menjadi sangat penting bagi pengguna Twitter untuk menilai seberapa positif atau negatif sentimen dari sebuah tweet[9].Penelitian ini menggunakan teknik web scraping untuk mengumpulkan data ulasan pelanggan Indihome dari Twitter. Analisis sentimen dilakukan dengan algoritma Support Vector Machine (SVM), menghasilkan akurasi tertinggi pada metode kernel Radial Basis Function (RBF) sebesar 88,47% pada Maret dan 98,06% pada April 2021. Mayoritas ulasan pelanggan cenderung negatif, terutama terkait masalah sinyal internet yang lambat, hilang, atau mati. Ulasan positif dan netral menyoroti respons pihak Indihome terhadap keluhan pelanggan dan berisi pertanyaan serta tips terkait layanan Indihome. Penelitian ini memberikan gambaran sentimen pelanggan terhadap Indihome dan evaluasi akurasi algoritma SVM dalam menganalisis ulasan Twitter[10].

3. METODE PENELITIAN

3.1. Sumber Data

Dataset dalam penelitian ini berjumlah 1000 komentar dari berbagai judul berita dan topik berita yang di ambil dari website berita Detik .Com.Setiap komentar tersebut nantinya di analisis mengunakan algoritma Suport Vektor Machine(SVM) untuk menentukan tingkat sentimen dari komentar yang ada di dataset, yaitu menentukan negatif dan positif. Performa algoritma Suport Vektor Machine(SVM) dalam analisis.Data yang digunakan adalah data sekunder yang berjumlah 1000 komentar yang diambil dari platform berita detik.com.proses pengumpulan data dilakukan melalui teknik web scraping dengan menggunakakan bahasa pemrograma python di platform Google Colabotary.

3.2. Teknik Pengumpulan Data

Teknik pengumpulan data tentang penelitian Analisis Sentimen terhadap komentar Berita di

detik.com Menggunakan Algoritma SVM adalah Menggunakan Web Scraping Data di Detik.com.

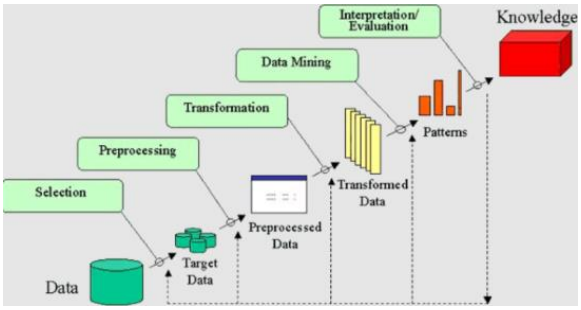
a. Identifikasi sumber data
   Sumber data yang digunakan adalah Detik.Com. platform berita online ,Ini adalah tempat di mana pengguna dapat membaca berita dari berbagai topik dan memberikan komentar tentang berita yang ada di website Detik.Com.
b. Pengumpulan Data dengan Web Scraping
   Untuk memperoleh jumlah komentar yang lebih besar, teknik web scraping digunakan dalam penelitian ini. Ini adalah proses otomatis untuk mengumpulkan data dari halaman web. menyalin link halaman judul berita di detik.com, dan memasukkannya ke dalam kode perintah yang sudah dibuat sebelumnya pada Google Colaboratory. Dengan metode ini, diperoleh data sebanyak 1000 komentar.
c. Pengolahan Data
   Setelah komentar terkumpul, langkah selanjutnya adalah pengolahan data. Data ulasan disimpan dalam format yang sesuai, yaitu spreadsheet, untuk mempermudah analisis dan penggunaan data selanjutnya. Proses pengolahan data mencakup tahapan transformasi data cleansing, tokenizing, stopwords removal, dan stemming. Semua ini bertujuan untuk mempersiapkan data mentah yang belum terstruktur menjadi data yang siap digunakan untuk tahap berikutnya dalam analisis sentimen.

3.3. Discovery of Knowledge in Databases (KDD)

Metode Dalam Teknik Analisis Data pada penelitian ini menggunakan metode KDD (Knowledge Discovery in Databases). Discovery of Knowledge in Databases (KDD) merupakan suatu prosedur yang diterapkan untuk menemukan pola-pola yang berguna dan menggali informasi dari dataset yang besar. Tujuan utama dari proses KDD ini adalah mengekstraksi pengetahuan atau informasi yang tersembunyi dalam data yang sebelumnya tidak dapat diidentifikasi Pemrosesan Awal Data.

Terdapat 5 proses text mining tahapan pada KDD yaitu:

a. Data Selection
b. Text Preprocessing
c. Transformation
d. Data Mining
e. Evaluation



Gambar 1. Tahapan KDD

3.4. Algoritma Suport Vektor Machine

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang digunakan untuk tugas-tugas klasifikasi dan regresi. Tujuannya adalah membangun model yang dapat memisahkan data ke dalam kelas atau memprediksi nilai kontinu. Berikut adalah beberapa konsep dasar tentang algoritma SVM:

a. Pemisahan Marginal Maksimum
   SVM bertujuan untuk menciptakan "hyperplane" yang memaksimalkan margin, yaitu jarak antara hyperplane dan titik-titik terdekat dari kedua kelas. Margin ini merupakan batas kepercayaan klasifikasi.
b. Vektor Dukungan (Support Vectors)
   SVM hanya bergantung pada sebagian kecil dari data yang disebut "support vectors", yaitu titik-titik data yang mendefinisikan margin dan mempengaruhi penentuan hyperplane.
c. Kernel
   SVM dapat mengatasi data yang tidak linier dengan menggunakan fungsi kernel. Fungsi kernel mentransformasikan data ke dalam dimensi yang lebih tinggi, di mana pemisahan linier menjadi mungkin. Contoh kernel umum termasuk kernel linear, polynomial, dan radial basis function (RBF).
d. Regularisasi
   SVM menggabungkan elemen regularisasi dalam fungsinya untuk menghindari overfitting. Parameter C digunakan untuk mengontrol tingkat regularisasi; nilai C yang lebih tinggi cenderung memberikan margin yang lebih kecil dan lebih ketat.
e. Klasifikasi
   Setelah pelatihan, SVM dapat digunakan untuk klasifikasi data baru. Data baru diklasifikasikan berdasarkan posisinya terhadap hyperplane yang telah ditentukan selama pelatihan.

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

adalah metode statistik yang digunakan untuk mengevaluasi seberapa penting suatu kata dalam suatu dokumen terhadap kumpulan dokumen yang lebih besar. Tujuannya adalah untuk memberikan bobot kepada kata-kata berdasarkan seberapa sering kata tersebut muncul di suatu dokumen dan seberapa jarang kata tersebut muncul di seluruh kumpulan dokumen. TF-IDF sering digunakan dalam pemrosesan teks, pengelompokan dokumen, dan analisis sentimen.

4. HASIL DAN PEMBAHASAN

4.1. Data Collection

Pada penelitian ini data yang digunakan bertipe sekunder. Data diambil dari komentar pengguna website berita online detik.Com. Proses pengumpulan data dilakukan dengan menggunakan teknik Web Scraping . Hasil dari proses scraping dihasilkan data

berjumlah 1000 komentar yang berasal dari bebagai judul berita dari setiap topik, hasil scraping kemudian disimpan ke dalam format xlxs file.

4.2. Data Selection

Data yang dipilih pada penelitian ini adalah data komentar berita onnline pada platfor berita detik.com dari berbagai judul dan topik berita. Setelah proses scraping dilakukan, menghasilkan sebanyak 1000 data komentar dengan 2 atribut yaitu Nama,dan Komentar . Dari jumlah 2 atribut tersebut hanya akan digunakan 1 atribut saja.

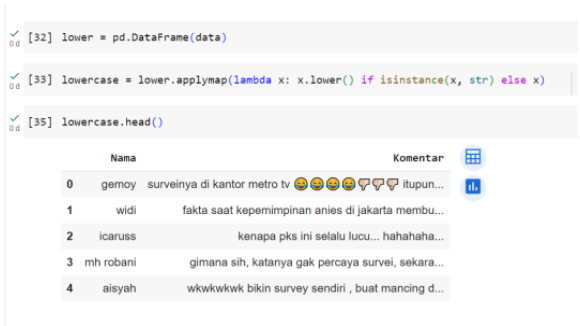
Gambar 2. Data awal

4.3. Text Preprocessing

Text preprocessing adalah serangkaian langkah atau teknik yang digunakan untuk membersihkan, mentransformasi, dan mempersiapkan data teks sebelum diolah lebih lanjut oleh algoritma atau model. Data yang diperoleh melalui Data Scraper tidak selalu berada dalam kondisi ideal untuk diproses. Terkadang, data tersebut menghadapi berbagai masalah yang dapat memengaruhi hasil dari proses penambangan itu sendiri, seperti nilai yang hilang dan data yang berlebihan. Sehinnga langkah pra-pemrosesan dibutuhkan untuk untuk meningkatkan kualitas data dan memfasilitasi analisis atau pemrosesan lebih lanjut. Adapun tahap dalam preprocessing dilakukan sebagai berikut :

a. Lowercasing

Tahapan lowercasing adalah proses mengubah huruf kapital pada dataset agar menjadi huruf kecil. Tujuan utama dari lowercasing adalah membuat teks lebih seragam dalam hal kapitalisasi, sehingga meminimalkan perbedaan kapitalisasi yang mungkin ada dalam data teks.



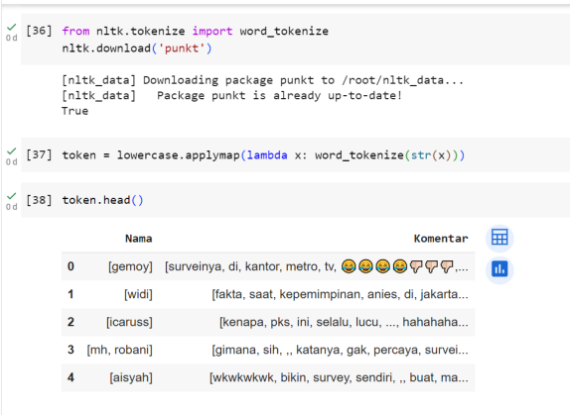
```
[32] lower = pd.DataFrame(data)

[33] lowercase = lower.applymap(lambda x: x.lower() if isinstance(x, str) else x)

[35] lowercase.head()
```

	Nama	Komentar
0	gemoy	surveinya di kantor metro tv 😂😂😂😂👎👎👎 itupun...
1	widi	fakta saat kepemimpinan anies di jakarta membu...
2	icaruss	kenapa pks ini selalu lucu... hahahaha...
3	mh robani	gimana sih, katanya gak percaya survei, sekara...
4	aisyah	wkwkwkwk bikin survey sendiri , buat mancing d...

Gambar 3. Proses dan hasil lowercasing

b. Tokenizing

Pada tahap ini teks atau kalimat akan dipecah menjadi unit-unit yang lebih kecil, yang disebut dengan token,Mengunakan modul nltk tokenize dan package ‘punkt’.



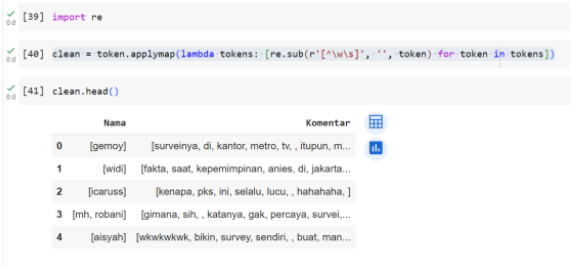
```
[36] from nltk.tokenize import word_tokenize
     nltk.download('punkt')

     [nltk_data] Downloading package punkt to /root/nltk_data...
     [nltk_data]   Package punkt is already up-to-date!
     True

[37] token = lowercase.applymap(lambda x: word_tokenize(str(x)))

[38] token.head()
```

	Nama	Komentar
0	[gemoy]	[surveinya, di, kantor, metro, tv, 😂😂😂😂👎👎👎,...
1	[widi]	[fakta, saat, kepemimpinan, anies, di, jakarta...
2	[icaruss]	[kenapa, pks, ini, selalu, lucu, ..., hahahaha...
3	[mh, robani]	[gimana, sih, ,, katanya, gak, percaya, survei...
4	[aisyah]	[wkwkwkwk, bikin, survey, sendiri, ,, buat, ma...

Gambar 4. Proses dan hasil tokenizing

c. Cleaning Data

Pada tahap cleansing data dibersihkan dari elemen-elemen yang tidak relevan atau tidak diinginkan, seperti tanda baca, angka, karakter khusus, atau spasi ganda,menggunakan yang namanya modul re. Selain itu, tahap ini juga mencakup penghapusan data yang duplikat atau redundant. Penghapusan tersebut dilakukan karena dianggap sebagai gangguan (Noise) dalam data dan tidak memiliki relevansi yang diperlukan dalam analisis sentimen. Proses ini bertujuan untuk menyempurnakan kualitas data, memastikan fokus pada informasi yang memiliki dampak signifikan terhadap evaluasi sentimen, dan menghilangkan duplikasi yang dapat memengaruhi akurasi hasil analisis



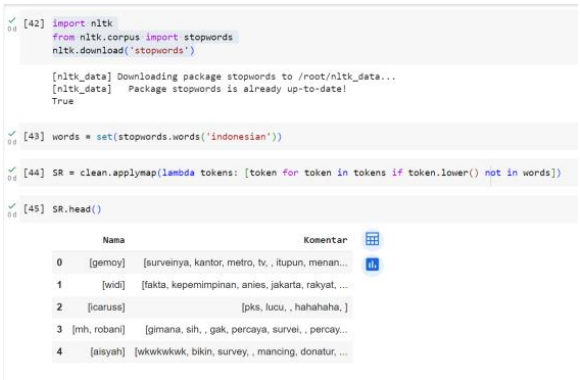
```
[39] import re

[40] clean = token.applymap(lambda tokens: [re.sub(r'[^\w\s]', '', token) for token in tokens])

[41] clean.head()
```

	Nama	Komentar
0	[gemoy]	[surveinya, di, kantor, metro, tv, , itupun, m...
1	[widi]	[fakta, saat, kepemimpinan, anies, di, jakarta...
2	[icaruss]	[kenapa, pks, ini, selalu, lucu, , hahahaha,]
3	[mh, robani]	[gimana, sih, , katanya, gak, percaya, survei,...
4	[aisyah]	[wkwkwkwk, bikin, survey, sendiri, , buat, man...

Gambar 5. Proses dan hasil clensing

d. Stopwords Removal

Tahap stopword removal adalah langkah dalam pra-pemrosesan teks yang melibatkan penghapusan kata-kata penghenti (stopwords) dari suatu teks. Stopwords adalah kata-kata umum yang sering muncul namun cenderung tidak memberikan kontribusi signifikan pada pemahaman atau analisis konten. Contoh stopwords kata-kata seperti seperti di, ke, ini, dan dari. Penghapusan kata sambung ini bertujuan untuk menyusutkan kalimat ulasan sehingga lebih ringkas.



Gambar 6. Proses dan hasil stopwords

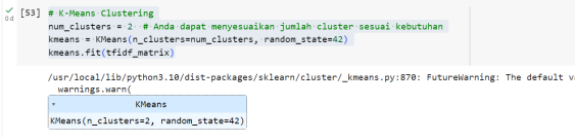
e. Stemming

Stemming adalah proses menghilangkan awalan atau akhiran kata sehingga hanya menyisakan akar kata atau bentuk dasar kata. Dalam bahasa Indonesia, Sastrawi adalah salah satu library yang menyediakan algoritma stemming untuk Bahasa Indonesia. Sastrawi menggunakan algoritma berbasis aturan (rule-based) dan kamus untuk melakukan proses stemming dan mendownload hasil text preprocessing. aturan (rule-based) dan kamus untuk melakukan proses stemming dan mendownload hasil text preprocessing.

Gambar 7. Proses dan hasil stemming

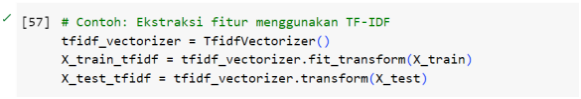
f. Labelling

Sebelum Memberikan label kita ubah menjadi clustering pada data dengan jumlah cluster (num_clusters) sebanyak 2 menggunakan K-Means untuk memudahkan memberi label pada data. Pada tahap pelabelan menggunakan algoritma K-Means, data diubah menjadi clustering,denngan cluster 1 adalah positif dan cluster 0 adalah negatif.



Gambar 8. Proses pengclusteran K-Means

Gambar 9. Proses dan hasil pelabelan



Gambar 10. Proses TF-IDF

4.4. Transformation

Proses transformasi melibatkan penerapan algoritma pembobotan kata yang dikenal dengan Term Frequency-Inverse Document Frequency (TF-IDF). Setiap kata yang terdapat dalam ulasan akan diberikan nilai bobot yang dihasilkan melalui perhitungan menggunakan algoritma TF-IDF. Hal ini bertujuan untuk mengevaluasi sejauh mana suatu kata dianggap penting dalam suatu dokumen. Dimana setiap kata dalam data akan dihitung berpa kali kata tersebut muncul.Hasil data yang sudah di tokenisasi akan di ekstrak kata yang sring muncul. Pada data uji akan di konversi dari data uji kedalam refrentasi TF-IDF yang sesuai dengan kosakata yang telah di vektorizer.

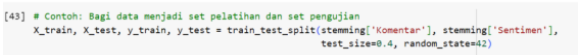
4.5. Data Mining

Setelah melalui proses preprocessing dan ekstraksi data komentar, langkah selanjutnya adalah melakukan klasifikasi. Pada tahap klasifikasi ini, algoritma yang digunakan adalah Suport Vektor Machine(SVM). Sebelum memasuki tahap klasifikasi, dataset akan dibagi menjadi dua bagian utama, yaitu data training dan data testing dengan pembagian rasio 20:80,30:70 , dan 40:60 .

Pada skenario Pertama Pembagian data pelatihan dan data uji dengan rasio 80 : 20 dimana data latihan berjumlah 800 komentar dan data uji berjumlah 200 komentar.

Gambar 11. Proses pembagian data 1

Pada skenario kedua Pembagian data pelatihan dan data uji dengan rasio 70 : 30 dimana data latihan berjumlah 700 komentar dan data uji berjumlah 300 komentar.



Gambar 12. Proses pembagian data 2

Pada skenario ketiga Pembagian data pelatihan dan data uji dengan rasio 60 : 40 dimana data latihan berjumlah 600 komentar dan data uji berjumlah 400 komentar.

```
# Contoh: Bagi data menjadi set pelatihan dan set pengujian
X_train, X_test, y_train, y_test = train_test_split(stemming['Komentar'], stemming['Sentimen'],
                                                    test_size=0.3, random_state=42)
```

Gambar 13. Proses pembagian data 3

4.6. Klarifikasi Algoritma Suport Vektor Machine(SVM)

Setelah proses pembagian dataset selesai, langkah selanjutnya adalah menerapkan dataset ke dalam model. Pembuatan model SVM dilakukan dengan menggunakan modul SVC Kernel linear untuk melatih model pada data yang sudah diekstraksi fiturnya menggunakan metode TF-IDF. Kelebihan dari kernel linear adalah sederhana dan efisien, terutama ketika data secara intrinsik dapat dipisahkan secara linear. Disini kelas 'SVC' di impor dari pustaka scikit-learn dan menginisialisasi objek model SVM dengaan kernel linear kemudian menggunakan metode .fit() untuk melaatih model.

```
[45] # Contoh: Pelatihan model SVM
     svm_model = SVC(kernel='linear')
     svm_model.fit(X_train_tfidf, y_train)
```

```
SVC
SVC(kernel='linear')
```

Gambar 14. Proses pemodelan SVM kernel linear

4.7. Evaluasi

Evaluasi model klasifikasi pada tahap ini dilakukan untuk mengukur seberapa baik performa Suport Vektor Machine(SVM) pada tugas analisis sentimen. Beberapa parameter evaluasi yang digunakan mencakup akurasi, presisi, recall, dan f-score. Hasil evaluasi akan ditampilkan dibawah.

Uji data pertama dari 200 komentar yang menghasilkan 142 komentar negatif dengan nilai dari precision,recall,dan f-score :0.99 dan 58 komentar positif dengan nilai dari precision, recall, dan f-score: 0.98,yang menghasilkan nilai accuracy sebesar: 0.99

```
[33] # Prediksi pada set pengujian
     y_pred = svm_model.predict(X_test_tfidf)

     # Evaluasi performa model
     accuracy = accuracy_score(y_test, y_pred)
     print(f'Accuracy: {accuracy}')
     print(classification_report(y_test, y_pred))

Accuracy: 0.99
              precision    recall  f1-score   support

     Negatif       0.99      0.99      0.99       142
     Positif       0.98      0.98      0.98        58

    accuracy                           0.99       200
   macro avg       0.99      0.99      0.99       200
weighted avg       0.99      0.99      0.99       200
```

Gambar 15. Proses dan hasil uji evaluasi 1

Uji data kedua dari 300 komentar yang menghasilkan 217 komentar negatif dengan nilai dari precision:0.99, recall:0.98, dan f-score:0.99 dan 83 komentar positif dengan nilai dari precision:0.95, recall:0.98, dan f-score: 0.96, yang menghasilkan nilai accuracy sebesar 0.98

```
[42] # Prediksi pada set pengujian
     y_pred = svm_model.predict(X_test_tfidf)
     # Evaluasi performa model
     accuracy = accuracy_score(y_test, y_pred)
     print(f'Accuracy: {accuracy}')
     print(classification_report(y_test, y_pred))

Accuracy: 0.98
              precision    recall  f1-score   support

     Negatif       0.99      0.98      0.99       217
     Positif       0.95      0.98      0.96        83

    accuracy                           0.98       300
   macro avg       0.97      0.98      0.98       300
weighted avg       0.98      0.98      0.98       300
```

Gambar 16. Proses dan hasil uji evaluasi 2

Uji data ketiga dari 400 komentar yang menghasilkan 288 komentar negatif dengan nilai dari precision:0.99, recall:0.98, dan f-score:0.99 dan 112 komentar positif dengan nilai dari precision:0.95, recall:0.98, dan f-score: 0.96, yang menghasilkan nilai accuracy sebesar 0.98

```
[46] # Prediksi pada set pengujian
     y_pred = svm_model.predict(X_test_tfidf)
     # Evaluasi performa model
     accuracy = accuracy_score(y_test, y_pred)
     print(f'Accuracy: {accuracy}')
     print(classification_report(y_test, y_pred))

Accuracy: 0.98
              precision    recall  f1-score   support

     Negatif       0.99      0.98      0.99       288
     Positif       0.95      0.98      0.96       112

    accuracy                           0.98       400
   macro avg       0.97      0.98      0.98       400
weighted avg       0.98      0.98      0.98       400
```

Gambar 17. Proses dan hasil uji evaluasi 3

5. KESIMPULAN DAN SARAN

Temuan utama penelitian ini mengungkapkan bahwa kompleksitas bahasa Indonesia dalam komentar berita Detik.com mempengaruhi distribusi sentimen, dengan dominasi sentimen negatif daripada positif. Hasil ini memberikan pemahaman yang berharga terkait respons pembaca terhadap berita dan menyoroti pentingnya penanganan kompleksitas bahasa dalam meningkatkan performa algoritma analisis sentimen. Selain itu, sentimen negatif yang lebih dominan,dari uji data pertama saja yang berjumlah 200 komentar ada 142 komentar negatif dan 58 komentar positif. ini mengindikasikan potensi perbaikan konten berita untuk mengurangi respons negatif, sambil tetap mempertimbangkan perlunya perhatian khusus terhadap kompleksitas bahasa Indonesia guna meningkatkan akurasi analisis

sentimen secara keseluruhan.Jadi dalam Penelitian selanjutnya dapat memperluas cakupan dataset dan menerapkan teknik-teknik pemrosesan bahasa alami yang lebih canggih untuk meningkatkan akurasi analisis sentimen. Selain itu, perlu dilakukan kolaborasi dengan pihak berita untuk memahami konteks spesifik yang mungkin mempengaruhi sentimen pembaca.

DAFTAR PUSTAKA

[1] A. N. U. A. N. Ulfah and ..., "Analisis Sentimen Hate Speech Pada Portal Berita Online Menggunakan Support Vector Machine (SVM)," *JATISI (Jurnal …*. Universitas Multi Data Palembang, 2020.

[2] A. M. Effendi, *Analisis sentimen pada judul berita online ekonomi dengan menggunakan Support Vector Machine*. etheses.uin-malang.ac.id, 2023. [Online]. Available: http://etheses.uin-malang.ac.id/id/eprint/52250

[3] A. Gormantara, "Analisis Sentimen Terhadap New Normal Era di Indonesia pada Twitter Menggunakan Metode Support Vector Machine," *Konferensi Nasional Ilmu Komputer (KONIK)*. researchgate.net, 2020. [Online]. Available: https://www.researchgate.net/profile/Alfredo-Gormantara/publication/342986951_Analisis_Sentimen_Terhadap_New_Normal_Era_di_Indonesia_pada_Twitter_Menggunakan_Metode_Support_Vector_Machine/links/5f104f2a299bf1e548ba4c49/Analisis-Sentimen-Terhadap-New-Norma

[4] A. J. PUTRA, *Implementasi Metode Support Vector Machine Dalam Analisis Sentimen Pada Data Ulasan Twitter Vaksin Covid-19*. dspace.uii.ac.id, 2021. [Online]. Available: https://dspace.uii.ac.id/handle/123456789/34613

[5] R. T. Adek, M. Fikry, and U. Khalil, "News Opinion Classification Application With Support Vector Machine Algorithm Using Framework Codeigniter," *J. Informatics …*, 2021, [Online]. Available: http://www.ojs.uma.ac.id/index.php/jite/article/view/5189

[6] D. ARIFAH, *PENERAPAN METODE SUPPORT VECTOR MACHINE DALAM MENGKLASIFIKASIKAN TWEET UJARAN KEBENCIAN TERHADAP TOKOH PUBLIK PADA …*. repository.uin-suska.ac.id, 2020. [Online]. Available: http://repository.uin-suska.ac.id/25130/

[7] S. K. HASNA, *Analisis Sentimen Data Ulasan menggunakan Algoritma Support Vector Machine (Studi Kasus: Aplikasi Iflix)*. dspace.uii.ac.id, 2021. [Online]. Available: https://dspace.uii.ac.id/handle/123456789/34393

[8] I. Taufik, *Analisis Sentimen Terhadap Tokoh Publik Menggunakan Algoritma Support Vector Machine (SVM)*. repository.uinjkt.ac.id, 2018. [Online]. Available: https://repository.uinjkt.ac.id/dspace/handle/123456789/59802

[9] U. Makhmudah, *Analisis Sentimen Terhadap Tweet Kaum Homoseksual Indonesia Menggunakan Metode Support Vector Machine*. repository.unej.ac.id, 2019. [Online]. Available: https://repository.unej.ac.id/handle/123456789/93827

[10] D. R. RAMADHANTY, *Implementasi Algoritma Support Vector Machine Pada Analisis Sentimen Data Twitter (Studi Kasus: Ulasan Tentang Indihome Pada Platform Twitter)*. dspace.uii.ac.id, 2021. [Online]. Available: https://dspace.uii.ac.id/handle/123456789/36015