ANALISIS SENTIMEN PENGGUNA PLATFORM X TERHADAP INDONESIA EMAS 2045 MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

Muhamad Rafli Tanjung, Nono Heryana, Siska
Sistem Informasi, Universitas Singaperbangsa Karawang
Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat, Indonesia
Tanjoeeng19@gmail.com

ABSTRAK

Indonesia Emas 2045 adalah visi negara Indonesia yang genap 100 tahun untuk menjadi negara maju dengan memanfaatkan bonus demografi. Visi tersebut bisa menjadi sebuah anugerah ataupun sebuah bencana, bergantung terhadap persiapan dan kesiapan untuk mencapai Indonesia Emas 2045. Salah satu persiapan yang harus disiapkan adalah usia produktif dalam menyongsong bonus demografi, yaitu Gen-Z. Dengan melakukan analisis sentimen terhadap masyarakat usia produktif bisa membantu kesadaran masyarakat terhadap peluang emas itu dan mewujudkan Indonesia Emas 2045. Analisis sentimen bisa membantu dalam mendapatkan representasi umum opini masyarakat tentang berbagai subjek. Berdasarkan hal tersebut penelitian ini akan melakukan analisis sentimen pengguna platform X terhadap Indonesia Emas 2045, memakai metodologi penelitian Knowledge Discovery Database (KDD) melalui 5 tahap yaitu *Data Selection*, *Data Preprocessing*, *Data Transformation*, *Text* Mining dan *Evaluation*. Dengan memanfaatkan algoritma Support Vector Machine (SVM) menggunakan kernel linear penelitian ini akan melaksanakan 5 skenario pembagian data pelatihan dan data pengujian, yaitu 90:10 pada skenario pertama, 80:20 yang kedua, 70:30 skenario ketiga, 60:40 skenario keempat, dan terakhir 50:50 pada skenario kelima. Confusion Matrix diaplikasikan untuk mengevaluasi performa algoritma yang dipakai, Hasil penelitian menunjukkan bahwa performa terbaik terhadap model yang dapat terbentuk terdapat dari skenario kedua, dengan nilai akurasi sebesar 94%, presisi 95%, recall 98% dan f1-score 97%.

Kata kunci : *Analisis Sentimen, Support Vector Machine, Platform X, Indonesia Emas*

1. PENDAHULUAN

Platform X adalah salah satu media sosial yang berkembang pesat, dimana penggunanya dapat menghasilkan, memposting, serta membaca pesan teks pendek atau yang biasa dianggap sebagai tweet. Melalui tweet tersebut pengguna platform X diberikan kebebasan untuk mengungkapkan pendapat dan pernyataan mengenai layanan, institusi, atau topik politik atau perdebatan.[1].

Platform X adalah tahap blogging skala kecil di mana pengguna menghasilkan 'tweet' yang dikomunikasikan kepada para pengikut mereka atau pengguna lain. Pada tahun 2023, Platform X memiliki lebih dari 372 juta pengguna dinamis dalam satu bulan tertentu, termasuk 64,9 juta pengguna dari Amerika Serikat yang merupakan angka terbesar pengguna twitter yang masih aktif [2]. Sebagai platform media sosial yang digunakan untuk berbagi pendapat, Platform X memiliki jutaan pengguna dari berbagai penjuru dunia. Secara rutin, ada banyak tweet yang mengekspresikan berbagai opini tentang berbagai topik. Dampaknya adalah meningkatnya peningkatan penyebaran dan juga pertumbuhan akan informasi di Platform X. [3].

Platform X banyak digunakan atau dimanfaatkan oleh masyarakat sebagai sumber data untuk menganalisis opini publik dan sentimen. Informasi yang diperoleh dari platform ini dapat memberikan wawasan yang lebih mendalam mengenai pandangan masyarakat terhadap suatu topik. Fokus besar dalam melakukan analisis sentimen di Platform X adalah untuk memilah dan memilih opini terhadap sebuah tweet, apakah itu positif atau sebaliknya [4]. Analisis sentiment banyak digunakan dalam beberapa tahun terakhir, analisis ini tidak hanya digunakan dalam ruang penelitian, analisis sentiment juga dapat digunakan dalam ruang bisnis, ruang pemerintahan dan ruang organisasi [5].

pengaruh Generasi Z terutama di media social cukup besar. Dengan memiliki pengaruh yang besar di media sosial, membuat apapun yang sedang dibahas oleh Generasi Z tentang berbagai hal selalu ramai dan menarik. Apalagi saat ini, tidak sedikit *influencer* media sosial merupakan Generasi anak muda yang bisa menjadikan pembahasan atau topik yang sedang diperbincangkan menjadi perhatian publik. *Influencer* media sosial memainkan peran penting dalam membentuk sebuah hubungan yang positif dan memiliki pengaruh terhadap individu yang menjadi target mereka [6].

Meskipun demikian, Saat ini terdapat segelintir penelitian mengenai sentiment analysis dengan objek Indonesia Emas 2045 sebagai orang yang berpengaruh di media sosial yang dapat membantu kesadaran masyarakat terhadap peluang emas itu. Untuk mengatasi permasalahan yang disebutkan di atas, penulis melakukan analisis sentimen pengguna Platform X terhadap Indonesia Emas 2045 menggunakan metode SVM (*Support Vector Machine).*

Pada penelitian kali ini, sumber data yang dipergunakan berasal pada Platform X. Analisis sentimen dilaksanakan melalui pengumpulan informasi yang berasal dari cuitan para pengguna

Platform X dalam bahasa Indonesia mengenai opini dan sudut pandang mereka terkait Indonesia Emas 2045. Data tersebut akan dikelompokkan ke dalam tiga kelas, yakni berupa sentimen positif, negatif, dan netral. Kemudian, dari tweet-tweet yang ada akan diklasifikasikan dengan memanfaatkan algoritma SVM dalam proses text mining.

2. TINJAUAN PUSTAKA

Dalam penelitian sebelumnya yang dilakukan oleh [7] telah dibandingkan mengenai ketepatan perhitungan yang digunakan SVM dan NB. Dengan menggunakan data dari Google Play Store, hasil penelitian menunjukan bahwa ketepatan perhitungan SVM lebih tinggi daripada ketepatan perhitungan Naïve Bayes. Ketepatan perhitungan SVM mencapai 81,46%, sedangkan ketepatan perhitungan Naïve Bayes mencapai 75,42%, berdasarkan hasil tersebut, algoritma SVM terbukti unggul dalam melakukan klasifikasi jika dibandingkan dengan algoritma lainnya.

Berdasarkan penelitian lainnya [8] Kinerja 4 kernel dalam algoritma SVM dibandingkan, yaitu kernel Linear, Polynomial, Radial Basis Function (RBF), dan Sigmoid pada tugas klasifikasi. Dari penelitian ini didapatkan hasil bahwa kernel linear memberikan hasil terbaik dengan parameter default. Kernel linear mencapai akurasi 81%, presisi 82%, dan recall 81%. Dengan demikian, algoritma SVM menggunakan kernel linear menjadi pilihan untuk diterapkan pada proses klasifikasi data dalam penelitian ini.

Sebuah penelitian lain yang terkait dengan Analisis Sentimen adalah penelitian tentang Opini pengguna Gopay yang dilakukan oleh [9] Dalam penelitian ini, mereka menggunakan metode SVM sebagai pendukung untuk mengklasifikasikan sentimen dan membandingkan dua kernel, yaitu linear dan polynomial. Hasil penelitian menunjukan bahwa kernel linear menghasilkan akurasi yang lebih optimal, yaitu sebesar 89,17%, dibandingkan dengan kernel polynomial yang hanya mencapai akurasi sebesar 84,38%.

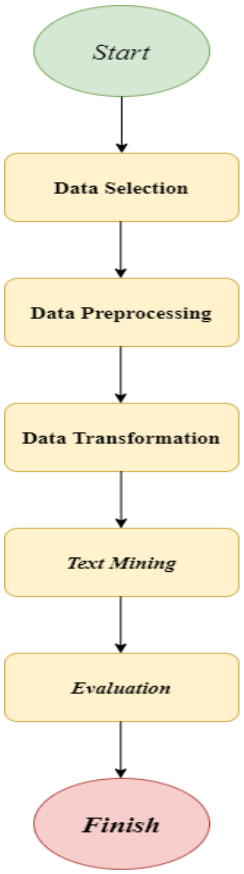
Pada penelitian sebelumnya yang dilakukan oleh [10], analisis sentimen menggunakan algoritma SVM dilakukan dengan hanya satu kali split data. Namun, dalam penelitian ini, data dibagi menjadi tiga bagian dalam mendapatkan akurasi terbaik yang berasal dari model yang terbentuk oleh algoritma dalam setiap skenario yang dijalankan. Evaluasi performa algoritma yang digunakan [11] dalam penelitian ini dilakukan menggunakan Confusion Matrix pada tiga skenario yang berbeda. Terdapat empat faktor yang digunakan dalam mengukur performa menggunakan Confusion Matrix, yaitu Akurasi, Presisi, Recall, dan F1-Score. Dalam klasifikasi, akurasi mewakili ketepatan model dalam mengklasifikasikan data dengan benar. *Precision* menggambarkan akurasi antara data yang diminta dan hasil prediksi yang diberikan oleh model. Sementara itu, recall mewakili keberhasilan model dalam menemukan kembali informasi. Dan F1-Score adalah perbandingan rata-rata tertimbang antara nilai ketepatan dan recall [12]. Berikut adalah rumus *Confusion matrix* untuk menghitung *accuracy*, *precision*, *recall* dan f1-*score.*

1. *Accuracy* $= \frac{TP+TN}{TP+FP+FN+TN}$ (1)
2. *Precision* $= \frac{TP}{TP+FP}$ (2)
3. *Recall* $= \frac{TP}{TP+FN}$ (3)
4. F1-*Score* $= \frac{2\times Recall \times Precision}{Recall + Precision}$ (4)

Keterangan:
TP (*True Positive*) = Jumlah data yang dapat diprediksi dengan benar dari kelas True.
TN (*True Negative*) = Jumlah data yang dapat diprediksi dengan benar dari kelas False.
FP (*False Positive*) = Jumlah data di mana kelas *True* diprediksi salah pada kelas *False*
FN (*False Negative)* = Jumlah data di mana kelas *False* diprediksi salah pada kelas *True*

3. METODE PENELITIAN

Metodologi penelitian dari penelitian ini adalah *Knowledge Discovery Database* (KDD). Gambar 1 menunjukkan alur dari rancangan penelitian yang dilakukan dalam penelitian ini.



Gambar 1. Rancangan Penelitian

3.1. Data Selection

Proses *data selection* adalah fase yang penting dalam analisis data, dikarenakan data yang diambil akan disesuaikan dengan kebutuhan penelitian. Pada tahap awal ini, pengumpulan data dilakukan melalui teknik *crawling* melalui API dari Twitter. Proses ini akan dibantu dengan *library tweepy* yang ada didalam bahasa pemrograman Python. Dengan rentang waktu dari tanggal 25 Januari 2024 sampai dengan 14 Februari 2024 Sebanyak 1518 tweet diraih untuk penelitian ini dengan menggunakan kata kunci "indonesia emas 2045".

Kemudian pada tahap ini dilakukan seleksi data yang meliputi pemeriksaan postingan berupa teks pada platform X yang cocok dan sama seperti kata kunci, membuang postingan berupa teks pada platform X yang memiliki banyak kesamaan, menyeleksi tweet menggunakan Bahasa asing tidak berbahasa Indonesia, dan pelabelan *tweet* yang positif, netral, dan negatif dengan Natural Language Toolkit dalam bahasa pemrograman Python yaitu *VADER (Valence Aware Dictionary and sentiment Reasoner)*.

3.2. Data Preprocessing

Tahapan Preprocessing ada 6 tahapan [13], meliputi :

a. *Cleansing* : tahapan ini diperuntukan meminimalisir gangguan dalam data tweet tersebut, seperti simbol, tautan URL, tanda hashtag "#", dan tanda "@" saat menyebut nama pengguna, dilakukan pembersihan.
b. *Case* Folding : Dalam tahapan setelah *cleansing* ini kata dalam kalimat yang terdapat huruf kapital (uppercase) diganti dengan huruf kecil (lowercase). Dengan tujuan menghapus redudansi data.
c. Tokenization : tahapan ini dilakukan pengurangan kalimat dalam teks, dibagi dengan setiap kata pada kalimat tersebut. Setiap satu kata dibagi menggunakan tanda petik tunggal serta spasi.
d. Stopword *Removal* : filtering atau *Stopword Removal*, pada tahap ini dilakukan penghapusan kata-kata yang diangap tidak perlu atau tidak memiliki makna pada sebuah kalimat dalam data *tweet.* Contohnya : “dan”, “yang”, “kah”, “gak”, “agar” dan lain lainnya
e. Normalization : memperbarui dan menyempurnakan struktur pada kata dalam sebuah kalimat yang tidak baku diubah menjadi kata-kata yang baku dalam data tweet.
f. Stemming : dilaksanakan langkah-langkah untuk mengonversi kata-kata yang mengandung afiks pada data tweet menjadi kata dasar.

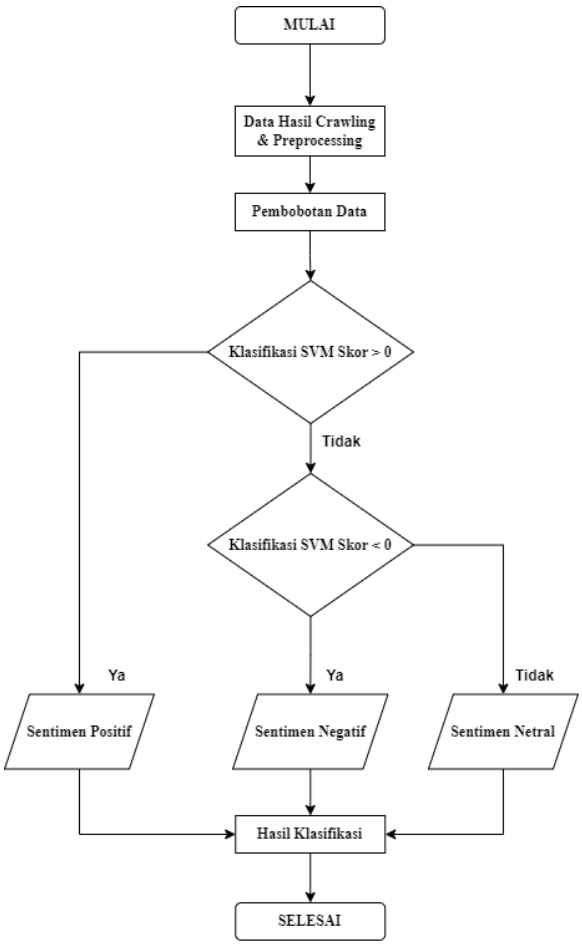
3.3. Data Transformation

Pada tahapan kali ini dilaksanakan proses transformasi data dengan TF-IDF merupakan model statistik untuk mengevaluasi nilai penting dari kata-kata dalam kumpulan dokumen [14] Metode labelEncoder akan digunakan untuk mengubah data teks menjadi data numerik untuk memudahkan jalannya pembelajaran melalui algoritma. Sebelum melakukan transformasi data, data akan dibagi lebih dulu. Pada riset kali ini, data akan dibagi menjadi tiga bagian dengan persentase sebagai berikut.

- Skenario Pertama, komparasi data 90:10
- Skenario Kedua, komparasi data 80:20
- Skenario Ketiga, komparasi data 70:30

3.4. Text Mining

Pada tahap selanjutnya, dilakukan pengolahan data dengan mengklasifikasikan tweet melalui algoritma Support Vector Machine (SVM) berbasis kernel linear. Mengubah data yang belum diolah menjadi informasi yang bermanfaat adalah tujuan yang ingin dicapai dengan mengolah data tersebut. [15]. Flowchart tahapan klasifikasi SVM penelitian ini terdapat pada gambar 2



Gambar 2. Flowchart

3.5. Evaluation

Pada tahap akhir, evaluasi dilakukan untuk menilai validitas terhadap model dari algoritma yang diterapkan pada proses sebelumnya. Evaluasi ini memanfaatkan Confusion Matrix untuk mengukur kinerja algoritma yang didasarkan pada empat faktor, yaitu Akurasi, Presisi, Recall, dan *F1-Score* [12].

Terdapat nilai *True Positive* yang berasal dari data sebenarnya positif dan diprediksi positif oleh model, *True Negative* yang merupakan data sebenarnya negatif dan diprediksi negatif oleh model, lalu *False Positive* merupakan data yang sebenarnya negatif tetapi diprediksi positif oleh model, dan *False Negative* data yang sebenarnya positif tetapi diprediksi negatif oleh model. Nilai-nilai tersebut yang akan dipakai untuk menghitung Akurasi, Presisi, dan Recall.

4. HASIL DAN PEMBAHASAN

Di dalam penelitian kali ini, dilakukan analisis sentimen pengguna Platform X terhadap Indonesia Emas 2045. Metode yang digunakan adalah algoritma Support Vector Machine menggunakan kernel linear. Hasilnya dikategorikan ke dalam tiga kelas sentimen, yaitu positif, negatif, dan netral. Selain itu, penelitian ini juga memanfaatkan bahasa pemrograman Python.

4.1. Data Selection

Tabel 1 menampilkan dataset awal dari proses pengumpulan data

Tabel 1. Dataset awal

No	full_text
1	Udah mah pendidikan belum merata ditambah lagi dengan masalah besar begini. Bingung dah gw. 2045 Indonesia emas? Gak yakin kalau gak ada niatan untuk berubah.
2	membalas data dan fakta dengan argumentasi manipulatif. Entah apa yg dimaksud dengan term Indonesia emas 2045 ini.
3	@RR_Raggy4 Tabrak tabrak masuk Indonesia Emas 2045

Dataset awal hanya terdiri atas atribut full_text, tabel 2 memuat keterangan mengenai atribut tersebut.

Tabel 2. Atribut Dataset awal

No	Atribut	Penjelasan
1	full_text	Merupakan kolom yang berisikan tweet/cuitan dari pengguna platform X

Selanjutnya data akan dipilah terlebih dahulu, mulai dari menghapus tweet duplikat (yang mempunyai isi dan makna yang sama), memilih tweet yang berbahasa Indonesia dan menghapus tweet berbahasa asing, dan juga hanya menggunakan tweet yang sesuai dengan kata kunci.

Kemudian data yang telah dipilih akan diberi label positif, negatif, dan netral menggunakan Natural Language Toolkit dalam bahasa pemrograman Python yang dikenal sebagai VADER (Valence Aware Dictionary and sentiment Reasoner). Dari total 679 data tweet yang ada, terdapat 580 sentimen netral, 54 tweet yang memiliki sentimen positif dan 45 tweet yang memiliki sentimen negatif. Tabel 3 menunjukan dataset yang sudah diberi label

Tabel 3. Sampel Penlabelan Dataset

No	full_text	Label
1	Udah mah pendidikan belum merata ditambah lagi dengan masalah besar begini. Bingung dah gw. 2045 Indonesia emas? Gak yakin kalau gak ada niatan untuk berubah.	netral
2	membalas data dan fakta dengan argumentasi manipulatif. Entah apa yg dimaksud dengan term Indonesia emas 2045 ini.	netral
3	@RR_Raggy4 Tabrak tabrak masuk Indonesia Emas 2045	netral

4.2. Data Preprocessing

1. Cleansing

Tabel hasil cleansing bisa dilihat pada tabel 4

Tabel 4. Perbandingan sampel cleansing

Sebelum	Sesudah
@Icanisa31 @Bull_winner Udh mah darurat baca ini ditambah darurat nonton sampe kelar. Berat.. Berat #IndonesiaEmas2045	Udh mah darurat baca ini ditambah darurat nonton sampe kelar. Berat Berat
@tubirfess Menuju Indonesia emas 2045 melalui bonus demografi 2030.	Menuju Indonesia Emas 2045 melalui bonus demografi 2030.
@idextratime Dari masa kegelapan menuju Indonesia emas 2045	Dari masa kegelapan menuju Indonesia emas 2045

Setelah melalui proses pembersihan data, gangguan yang ada pada data berhasil diatasi sehingga data menjadi lebih terstruktur dan bersih. Hal ini membuat proses selanjutnya menjadi lebih mudah untuk dilakukan.

2. Case Folding

Pada tabel 5 menunjukan data proses *case folding*

Tabel 5. Perbandingan sampel *case folding*

Sebelum	Sesudah
Udh mah darurat baca ini ditambah darurat nonton sampe kelar. Berat Berat	udh mah darurat baca ini ditambah darurat nonton sampe kelar. berat berat
Menuju Indonesia Emas 2045 melalui bonus demografi 2030	menuju indonesia emas 2045 melalui bonus demografi 2030
Dari masa kegelapan menuju Indonesia emas 2045	dari masa kegelapan menuju Indonesia emas 2045

Dalam dataset, setiap kata yang sebelumnya ditulis dengan huruf besar akan diubah menjadi huruf kecil melalui proses case folding.

3. Tokenization

Tabel 6 menunjukan data proses *tokenization*

Tabel 6. Perbandingan sampel *tokenization*

Sebelum	Sesudah
udh mah darurat baca ini ditambah darurat nonton sampe kelar. berat berat	['udh', 'mah', 'darurat', 'baca', 'ini', 'ditambah', 'darurat', 'nonton', 'sampe', 'kelar', 'berat', 'berat']
menuju indonesia emas 2045 melalui bonus demografi 2030	['menuju', 'indonesia', 'emas', 'melalui', 'bonus', 'demografi']
dari masa kegelapan menuju Indonesia emas 2045	['dari', 'masa', 'kegelapan', 'menuju', 'indonesia', 'emas']

Setelah melakukan tokenisasi pada dataset, setiap kata yang membentuknya dipisahkan menggunakan tanda petik tunggal dan spasi.

4. Normalization

Tabel 7 menunjukan data proses *normalization*

Tabel 7. Perbandingan sampel *normalization*

Sebelum	Sesudah
['udh', 'mah', 'darurat', 'baca', 'ini', 'ditambah', 'darurat', 'nonton', 'sampe', 'kelar', 'berat', 'berat']	['sudah', 'juga', 'darurat', 'baca', 'ini', 'ditambah', 'darurat', 'menonton', 'sampe', 'kelar', 'berat', 'berat']
['menuju', 'indonesia', 'emas', 'melalui', 'bonus', 'demografi']	['menuju', 'indonesia', 'emas', 'melalui', 'bonus', 'demografi']
['dari', 'masa', 'kegelapan', 'menuju', 'indonesia', 'emas']	['dari', 'masa', 'kegelapan', 'menuju', 'indonesia', 'emas']

Dalam melakukan normalisasi, dataset dapat diubah dengan menggunakan kamus normalisasi [16] untuk mengganti kata-kata yang tidak baku menjadi kata-kata yang baku.

5. *Stopword Removal*

Tabel 8 merupakan hasil data proses *stopword removal*

Tabel 8. Perbandingan sampel *stopword removal*

Sebelum	Sesudah
['sudah', 'juga', 'darurat', 'baca', 'ini', 'ditambah', 'darurat', 'menonton', 'sampe', 'kelar', 'berat', 'berat']	['darurat', 'baca', 'ditambah', 'darurat', 'menonton', 'sampe', 'kelar', 'berat', 'berat']
['menuju', 'indonesia', 'emas', 'melalui', 'bonus', 'demografi']	['indonesia', 'emas', 'bonus', 'demografi']
['dari', 'masa', 'kegelapan', 'menuju', 'indonesia', 'emas']	['kegelapan', 'indonesia', 'emas']

Dengan menggunakan modul stopword yang ada di perpustakaan nltk dan kamus stopword, kata-kata yang dianggap tidak relevan dalam kumpulan data akan dihapus.

6. *Stemming*

Pada tabel 9 merupakan hasil *stemming*

Tabel 9. Perbandingan sampel stemming

Sebelum	Sesudah
['sudah', 'juga', 'darurat', 'baca', 'ini', 'ditambah', 'darurat', 'menonton', 'sampe', 'kelar', 'berat', 'berat']	['darurat', 'baca', 'tambah', 'darurat', 'nonton', 'sampe', 'kelar', 'berat', 'berat']
['menuju', 'indonesia', 'emas', 'melalui', 'bonus', 'demografi']	['indonesia', 'emas', 'bonus', 'demografi']
['dari', 'masa', 'kegelapan', 'menuju', 'indonesia', 'emas']	['kegelapan', 'indonesia', 'emas']

Pada fase ini, dalam melakukan stemming pada data teks berbahasa Indonesia, langkah-langkahnya didukung oleh pustaka sastrawi.

4.3. Data Transformation

Sebelum melakukan tahap transformasi, langkah pertama yang dilakukan adalah membagi data menjadi tiga bagian menggunakan metode split data. Pada Skenario 1, data training dan data testing memiliki perbandingan 90% dan 10% secara berturut-turut.

Pada Skenario 2, Perbandingan data training dan data testing adalah 80% dan 20%. Sedangkan pada Skenario 3, perbandingan data training dan data testing adalah 70% dan 30%. Setelah itu, dilakukan proses pembobotan kata menggunakan metode TF-IDF dengan menggunakan modul *TfidfVectorizer* yang terdapat dalam library *Scikit-Learn* [17]. Berikut ini adalah contoh implementasi program Python yang membagi data dan mengubah data.

```
# Membaca data
data = pd.read_csv('dataset-test1-indonesiameas.csv', sep =',',
                   usecols=['sentiment', 'tweet_stem']).dropna()

# Menampilkan data untuk memastikan tidak ada masalah
print(data.head())

# Memisahkan fitur dan label
X = data['tweet_stem']
y = data['sentiment']

# Fungsi untuk mengatur vektorisasi dan label encoding
def prepare_data(X, y, test_size, random_state=123):
    # Split data
    X_train, X_test, y_train, y_test = train_test_split
    (X, y, test_size=test_size, random_state=random_state)

    # Label encoding
    le = LabelEncoder()
    y_train = le.fit_transform(y_train)
    y_test = le.fit_transform(y_test)

    # Vectorization
    vectorizer = TfidfVectorizer()
    X_train = vectorizer.fit_transform(X_train)
    X_test = vectorizer.transform(X_test)

    return X_train, X_test, y_train, y_test, vectorizer
```

Gambar 3. metode TF-IDF

4.4. Text Mining

Tahapan text mining pada penelitian kali ini adalah melakukan clustering pada sebuah data tweet melalui algoritma SVM dengan kernel linear. Algoritma SVM dengan bobot C sebesar 1 diimplementasikan dengan memanfaatkan library *sklearn* pada bahasa pemrograman *Python*.

```
from sklearn.svm import LinearSVC, SVC
from sklearn.metrics import accuracy_score,confusion_matrix

svm1 = LinearSVC(C = 1)
svm1.fit(X_train1, y_train1)

print('Accuracy Score Model : %s ' %accuracy_score
      (y_test1, svm1.predict(X_test1)))
```

Gambar 4. Metode SVM

Pada Tabel 10 menampilkan hasil akurasi dari 3 skenario yang dilaksanakan dalam penelitian ini.

Tabel 10. Akurasi 3 skenario

Akurasi	Skenario
93%	Pertama
94%	Kedua
93%	Ketiga

Setelah menerapkan algoritma SVM untuk melakukan klasifikasi, hasil yang didapatkan pada skenario 1 dengan perbandingan data 90:10 adalah 93%. Kemudian, pada skenario 2 dengan perbandingan data 80:20, didapatkan akurasi terbaik sebesar 94%. Sementara itu, pada skenario 3 dengan perbandingan data 70:30, akurasi yang diperoleh adalah 93%.

4.5. Evaluation

Berikut adalah penjelasan mengenai hasil evaluasi dari tiga skenario yang telah dilakukan dalam penelitian ini.

1. Skenario Pertama

```
Confusion Matrix (90:10) :
[[ 2  3  0]
 [ 0 58  0]
 [ 0  2  3]]
==================================================
              precision    recall  f1-score   support

           0       1.00      0.40      0.57         5
           1       0.92      1.00      0.96        58
           2       1.00      0.60      0.75         5

    accuracy                           0.93        68
   macro avg       0.97      0.67      0.76        68
weighted avg       0.93      0.93      0.91        68

==================================================
```

Gambar 5. Hasi *confusion matrix scenario pertama*

Berdasarkan hasil *confusion matrix*, data training yang digunakan sebesar 90% dan mendapatkan hasil accuracy sebesar 93%, dengan jumlah dari suatu kelas *true* yang bisa diperkirakan dengan benar pada kelas *true* atau disebut *True* Positif (TP) adalah sebanyak 58. Kemudian jumlah dari kondisi di mana kelas *true* yang diprediksi salah pada kelas *false* atau yang disebut *False* Positif (FP) sebanyak 5, dan jumlah dari kondisi di mana pada kelas *false* yang diprediksi salah pada kelas *true* atau disebut *False* Negatif (FN) sebanyak 0. Berdasarkan angka-angka di pembahasan sebelumnya maka diperoleh hasil pada akurasi bernilai 93%, *precision* 92%, *recall* 100% dan f1-*score* 96%. Hasil pada akurasi menandakan efisiensi model dalam mengklasifikasikan data dengan benar, pada skenario pembagian data 90:10 memiliki nilai accuracy yang tinggi, kemudian nilai *precision* yang didapatkan pada skenario ini juga memiliki nilai yang tinggi, hal itu menunjukkan bahwa model ada tingkat efisiensi yang signifikan antara informasi yang diberikan dan informasi yang diminta. Sedangkan untuk nilai *recall* pada data pelatihan tersebut memiliki nilai yang besar, Ini menunjukkan bahwa data pelatihan yang digunakan dalam model ini telah berhasil dalam mengambil informasi

2. Skenario Kedua

```
Confusion Matrix (80:20) :
[[  3   3   0]
 [  1 119   1]
 [  0   3   6]]
==================================================
              precision    recall  f1-score   support

           0       0.75      0.50      0.60         6
           1       0.95      0.98      0.97       121
           2       0.86      0.67      0.75         9

    accuracy                           0.94       136
   macro avg       0.85      0.72      0.77       136
weighted avg       0.94      0.94      0.94       136

==================================================
```

Gambar 6. Confusion matrix scenario 2

Lalu pada hasil c*onfusion matrix* kedua data training yang digunakan sebesar 80% dan mendapatkan hasil accuracy sebesar 94%, dengan jumlah dari suatu kelas *true* yang bisa diperkirakan dengan benar pada kelas *true* atau disebut *True* Positif (TP) adalah sebanyak 119. Kemudian jumlah dari kondisi di mana kelas *true* yang diprediksi salah pada kelas *false* atau yang disebut *False* Positif (FP) sebanyak 6, dan jumlah dari kondisi di mana pada kelas *false* yang diprediksi salah pada kelas *true* atau disebut *False* Negatif (FN) sebanyak 2. Berdasarkan angka-angka di pembahasan sebelumnya maka diperoleh hasil pada akurasi bernilai 94%, lalu *precision* 95%, pada *recall* 98% dan juga f1-*score* 97%. Dari perolehan evaluasi yang didapatkan memperlihatkan bahwa performa algoritma terbaik didapatkan pada skenario 2. Dilihat dari nilai akurasi pada skenario 2 yang memiliki nilai cukup tinggi, kemudian nilai *precision* yang didapatkan pada skenario ini juga memiliki nilai yang cukup tinggi, hal itu menunjukkan bahwa model memiliki tingkat efisiensi yang signifikan antara informasi yang diberikan dan informasi yang diminta. Selain itu nilai

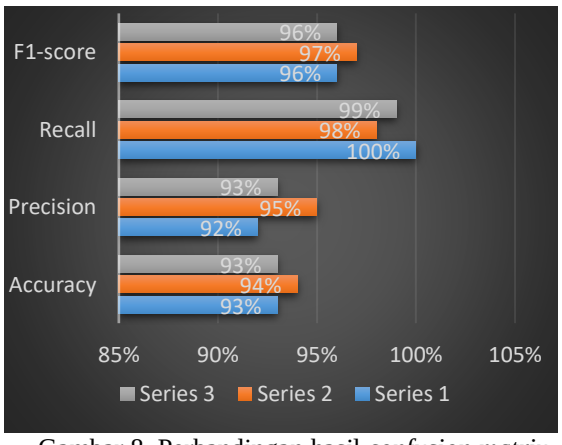
recall dan nilai f1-*score* juga memiliki nilai yang tidak buruk atau tidak rendah.

3. Skenario Ketiga

```
Confusion Matrix (70:30) :
[[  4   7   0]
 [  1 177   1]
 [  0   6   8]]
==================================================
              precision    recall  f1-score   support

           0       0.80      0.36      0.50        11
           1       0.93      0.99      0.96       179
           2       0.89      0.57      0.70        14

    accuracy                           0.93       204
   macro avg       0.87      0.64      0.72       204
weighted avg       0.92      0.93      0.92       204

==================================================
```

Gambar 7. *Confusion matrix scenario 3*

Hasil confusion matrix yang terakhir data training yang digunakan sebesar 70% dan mendapatkan hasil accuracy sebesar 93%, dengan jumlah dari suatu kelas *true* yang bisa diperkirakan dengan benar pada kelas *true* atau disebut *True* Positif (TP) adalah sebanyak 177. Kemudian jumlah dari kondisi di mana kelas *true* yang diprediksi salah pada kelas *false* atau yang disebut *False* Positif (FP) sebanyak 13, dan jumlah dari kondisi di mana pada kelas *false* yang diprediksi salah pada kelas *true* atau disebut *False* Negatif (FN) sebanyak 2. Berdasarkan angka-angka di pembahasan sebelumnya maka diperoleh hasil pada akurasi bernilai 93%, *precision* 93%, *recall* 99% dan f1-*score* 96%. Evaluasi menunjukkan bahwa akurasi pada skenario 3 sama dengan skenario 1, sementara nilai precision pada skenario tersebut juga tinggi. Ini menunjukkan bahwa model memiliki tingkat efisiensi yang signifikan dalam memproses informasi yang diminta. Selain itu, nilai recall dan f1-score juga menunjukkan kinerja yang baik.



Gambar 8. Perbandingan hasil *confusion matrix*

Pada Gambar 8, terdapat perbandingan tingkat akurasi, presisi, recall, dan f1-score dalam setiap skenario pengujian. Skenario 2 menunjukkan akurasi tertinggi sebesar 94%. Selain itu, skenario 2 juga memiliki presisi tertinggi dengan nilai 95%. Namun, skenario 1 memiliki recall tertinggi sebesar 100%. Terakhir, skenario 2 juga memiliki f1-score tertinggi yaitu 97%. Berdasarkan hasil penelitian ini, yang menunjukkan ketidakseimbangan antara kelas positif, negatif, dan netral, serta berdasarkan perhitungan confusion matrix, maka disimpulkan jika model memiliki kinerja terbaik terdapat pada skenario 2.

Dari analisis data tweet pengguna Twitter mengenai Indonesia Emas 2045, terdapat beragam tanggapan sentimen, termasuk sentimen positif, sentimen negatif, dan sentimen netral. Visualisasi kemunculan kata dari sentimen positif dapat dilihat pada Gambar 9 yang menggunakan teknik wordcloud.

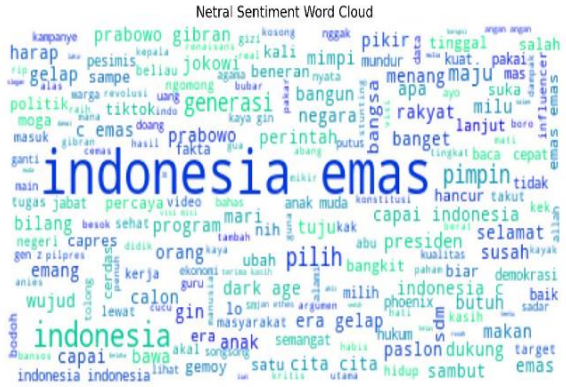
Dalam 9. *Wordcloud* sentimen positif

Dalam data sentimen positif pada Gambar 4.18, terdapat kata-kata seperti 'indonesia emas', 'bonus demografi', 'makan siang', 'prabowo', hilir, dan 'gratis' yang mencerminkan berbagai topik yang penting bagi publik. Kata-kata ini memunculkan emosi positif seperti rasa bangga, optimisme, kebahagiaan, dan kenyamanan. Analisis lebih lanjut dari data ini dapat memberikan wawasan tentang bagaimana topik-topik ini mempengaruhi persepsi dan emosi publik, serta membantu merumuskan strategi komunikasi yang lebih efektif dan responsif.

Gambar 10. *Wordcloud* sentimen negatif

Dalam data sentimen negatif, terdapat beberapa kata yang sering muncul seperti 'indonesia emas', 'propaganda', 'dirty vote', 'prabowo', era gelap, dan 'jokowi'. Kata-kata ini digunakan untuk mengkritik

pemerintah, manipulasi informasi, kecurangan dalam pemilihan umum, tindakan politik, dan masa pemerintahan yang dianggap buruk.



Gambar 11. Wordcloud sentimen netral

Dalam Gambar 11, data sentimen netral mendominasi. Kata-kata yang sering digunakan adalah 'indonesia emas', 'generasi', 'bangsa', 'prabowo', pemimpin, dan 'indonesia'. 'Indonesia emas' mengacu pada harapan untuk melihat Indonesia maju. 'Generasi' mengacu pada kelompok masyarakat yang lahir pada waktu yang sama. 'Bangsa' merujuk pada kesatuan dan identitas negara. 'Prabowo' adalah nama seorang tokoh politik. 'Pemimpin' merujuk pada individu yang menduduki posisi kepemimpinan. 'Indonesia' merujuk pada negara secara keseluruhan. Dominasi sentimen netral menunjukkan diskusi tentang Indonesia Emas 2045 masih bersifat normatif dan informatif.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis sentimen terhadap Indonesia emas 2045 yang dilakukan pada pengguna Platform X menggunakan algoritma Support Vector Machine (SVM) dengan kernel linear dan metodologi KDD melalui 5 tahap, yaitu tahapan seleksi data, *preprocessing*, transformasi data, *text mining*, dan juga evaluasi didapatkan sejumlah 54 data tweet positif, 45 data tweet negatif, dan 580 data tweet netral dari total 679 data tweet yang telah diproses. Dominasi sentiment netral pada penelitian ini menunjukkan bahwa mayoritas pengguna Twitter cenderung belum memiliki emosi positif ataupun negatif terhadap visi Indonesia emas 2045. Oleh karena itu, diperlukan peningkatan informasi dan edukasi agar masyarakat dapat lebih memahami dan aktif berpartisipasi dalam mewujudkan visi tersebut. Visualisasi kata-kata dalam kalimat yang sering muncul pada sentimen netral diantaranya 'indonesia emas', 'generasi', 'bangsa', 'prabowo', pemimpin, dan 'indonesia'. Hasil evaluasi terhadap model yang dibentuk oleh SVM melalui Confusion Matrix menunjukkan bahwa dari 3 skenario yang dilakukan, ketiga nya memiliki performa algoritma yang baik. Skenario pertama memiliki akurasi 94%, precision 95%, recall 98%, dan f1-score 97%. Skenario kedua memiliki akurasi 93%, precision 92%, recall 100%, dan f1-score 96%. Skenario 3 juga memiliki akurasi 93%, precision 93%, recall 99%, dan f1-score 96%.

DAFTAR PUSTAKA

[1] A. Supian, B. Tri Revaldo, N. Marhadi, L. Efrizoni, dan R. Rahmaddeni, "PERBANDINGAN KINERJA NAÏVE BAYES DAN SVM PADA ANALISIS SENTIMEN TWITTER IBUKOTA NUSANTARA," *J. Ilm. Inform.*, vol. 12, no. 01, hal. 15–21, 2024, doi: 10.33884/jif.v12i01.8721.

[2] datareportal, "TWITTER USERS, STATS, DATA & TRENDS," datareportal.com. Diakses: 27 Juni 2024. [Daring]. Tersedia pada: https://datareportal.com/essential-twitter-stats

[3] L. Yue, W. Chen, X. Li, W. Zuo, dan M. Yin, "A survey of sentiment analysis in social media," *Knowl. Inf. Syst.*, vol. 60, no. 2, hal. 617–663, 2019, doi: 10.1007/s10115-018-1236-4.

[4] T. Safitri, Y. Umaidah, dan I. Maulana, "Analisis Sentimen Pengguna Twitter Terhadap Grup Musik BTS Menggunakan Algoritma Support Vector Machine," *J. Appl. Informatics Comput.*, vol. 7, no. 1, hal. 28–35, 2023, doi: 10.30871/jaic.v7i1.5039.

[5] J. F. Sánchez-Rada dan C. A. Iglesias, "Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison," *Inf. Fusion*, vol. 52, hal. 344–356, Des 2019, doi: 10.1016/j.inffus.2019.05.003.

[6] R. K. Lianne Lim, M. V Zhayvier Rosales, A. V Therese Salazar, E. E. Pantoja, dan C. Author, "Journal of Business and Management Studies Perception of Filipino Skincare Product Users on the Effectiveness of Social Media Influencers vs Celebrity Endorsers as Brand Ambassadors," 2022, doi: 10.32996/jbms.

[7] L. B. Ilmawan dan M. A. Mude, "Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store," *Ilk. J. Ilm.*, vol. 12, no. 2, hal. 154–161, Agu 2020, doi: 10.33096/ilkom.v12i2.597.154-161.

[8] A. Liani, U. Enri, dan Y. Umaidah, "Analisis Perbandingan Kernel Algoritma Support Vector Machine dalam Mengklasifikasikan Skripsi Teknik Informatika berdasarkan Abstrak," *JOINS (Journal Inf. Syst.*, vol. 5, no. 2, hal. 240–249, Nov 2020, doi: 10.33633/joins.v5i2.3715.

[9] R. Mahendrajaya, G. A. Buntoro, dan M. B. Setyawan, "Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine.," *KOMPUTEK*, vol. 3, no. 2, hal. 52, 2019, doi: 10.24269/jkt.v3i2.270.

[10] A. Kulsumarwati, I. Purnamasari, dan B. A. Darmawan, "Penerapan SVM dan Information Gain Pada Analisis Sentimen Pelaksanaan Pilkada Saat Pandemi," *J. Teknol. Inform. dan Komput.*, vol. 7, no. 2, hal. 101–109, 2021, doi: 10.37012/jtik.v7i2.641.

[11] J. Xu, Y. Zhang, dan D. Miao, "Three-way confusion matrix for classification: A measure driven view," *Inf. Sci. (Ny).*, vol. 507, hal. 772–794, 2020, doi: 10.1016/j.ins.2019.06.064.

[12] D. Putra dan A. Wibowo, "Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes," *Pros. Semin. Nas. Ris. Dan Inf. Sci.*, vol. 2, hal. 84–92, 2020.

[13] L. Hickman, S. Thapa, L. Tay, M. Cao, dan P. Srinivasan, "Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations," *Organ. Res. Methods*, vol. 25, no. 1, hal. 114–146, 2022, doi: 10.1177/1094428120971683.

[14] M. Das, S. Kamalanathan, dan P. Alphonse, "A Comparative Study on TF-IDF feature weighting method and its analysis using unstructured dataset," *CEUR Workshop Proc.*, vol. 2870, hal. 98–107, 2021.

[15] V. Giordano, E. Cervelli, dan F. Chiarello, "Defining definitions: a Text mining approach to automatically define innovative technological fields," hal. 1–22, 2021, [Daring]. Tersedia pada: https://etd.adm.unipi.it/t/etd-04152019-163159/

[16] F. Zuhad dan N. Wilantika, "Perbandingan Penggunaan Kamus Normalisasi dalam Analisis Sentimen Berbahasa Indonesia," *J. Linguist. Komputasional*, vol. 5, no. 1, hal. 13–23, 2022, [Daring]. Tersedia pada: http://inacl.id/journal/index.php/jlk/article/view/60

[17] A. Averina, H. Hadi, dan J. Siswantoro, "Analisis Sentimen Multi-Kelas Untuk Film Berbasis Teks Ulasan Menggunakan Model Regresi Logistik," *Teknika*, vol. 11, no. 2, hal. 123–128, 2022, doi: 10.34148/teknika.v11i2.461.