

COMPARISON OF NAÏVE BAYES ALGORITHM AND SUPPORT VECTOR MACHINE IN SENTIMENT ANALYSIS OF BOYCOTT ISRAELI PRODUCTS ON TWITTER

Laisha Amilna Hayurian^{*1}, Nirwana Hendrastuty²

^{1,2}Information System, Faculty of Engineering and Computer Science, Universitas Teknokrat Indonesia,
Indonesia

Email: ¹laisha_amilna_hayurian@teknokrat.ac.id, ²nirwanahendrastuty@teknokrat.ac.id

(Article received: January 31, 2024; Revision: February 29, 2024; published: May 27, 2024)

Abstract

The Israeli-Palestinian conflict has captured the attention of Indonesians and even the world for decades, with the death toll reaching 17,000 Palestinians. Indonesians have expressed various opinions, including a proposed boycott of products that allegedly support Israel as a form of protest against the ongoing conflict. This study explores the opinions and sentiments of the Indonesian people regarding the Israel-Palestine conflict and the efforts to boycott Israeli products on social media twitter. This study aims to compare the accuracy of the two algorithms in classifying sentiment towards boycotting Israeli products. A total of 2288 comment data were processed using the Naïve Bayes and Support Vector Machine (SVM) algorithm classification methods. The results show that the Naïve Bayes algorithm has higher accuracy with a data division ratio of 70:30 and 80:30 for training data and testing data. Accuracy results with 70:30 data division reached 84% using the Naïve Bayes algorithm model, while the SVM algorithm model reached 78%. And the accuracy results with 80:20 data division reached 85% using the Naïve Bayes algorithm model, with the SVM algorithm model reaching 84%. This study provides an understanding of the concept of text mining and data mining and can be a reference for similar research.

Keywords: *Boycott of Israeli Products, Naïve Bayes, Support Vector Machine, Public Sentiment, Twitter, Text Mining.*

PERBANDINGAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE DALAM ANALISIS SENTIMEN BOIKOT PRODUK ISRAEL DI TWITTER

Abstrak

Konflik Israel-Palestina menarik perhatian masyarakat Indonesia bahkan dunia selama beberapa dekade, dengan korban jiwa mencapai 17.000 jiwa warga Palestina. Masyarakat Indonesia memberikan berbagai pendapat, termasuk usulan boikot produk yang diduga mendukung Israel sebagai bentuk protes terhadap konflik yang sedang berlangsung. Penelitian ini menggali opini dan sentimen masyarakat Indonesia terkait konflik Israel-Palestina serta upaya boikot produk Israel di media sosial twitter. Penelitian ini bertujuan membandingkan tingkat akurasi kedua algoritma dalam mengklasifikasikan sentimen terhadap boikot produk Israel. Sebanyak 2288 data komentar diolah menggunakan metode klasifikasi algoritma Naïve Bayes dan Support Vector Machine (SVM). Hasilnya menunjukkan bahwa algoritma Naïve Bayes memiliki akurasi lebih tinggi dengan perbandingan pembagian data 70:30 dan 80:30 untuk data training dan data testing. Hasil akurasi dengan pembagian data 70:30 mencapai 84% menggunakan model algoritma Naïve Bayes, sedangkan model algoritma SVM mencapai 78%. Dan hasil akurasi dengan pembagian data 80:20 mencapai 85% menggunakan model algoritma Naïve Bayes, dengan model algoritma SVM mencapai 84%. Studi ini memberikan pemahaman tentang konsep text mining dan data mining serta dapat menjadi referensi untuk penelitian serupa.

Kata kunci: *Boikot Produk Israel, Naïve Bayes, Support Vector Machine, Sentimen Publik, Twitter, Text Mining.*

1. PENDAHULUAN

Konflik antara Israel dan Palestina, yang sudah berlangsung selama 100 tahun yang lalu dan menjadi perbincangan tren saat ini[1]. Dalam konflik ini telah memakan korban jiwa mencapai 17.000 warga

Palestina[2]. Masyarakat Indonesia turut memberikan berbagai pendapat terkait konflik ini, baik yang mendukung maupun menentang. Beberapa diantaranya mengusulkan untuk melakukan boikot terhadap produk yang dianggap mendukung Israel

sebagai bentuk protes terhadap konflik yang sedang berlangsung. *Twitter* menjadi salah satu *platform* di mana masyarakat menyampaikan pendapat dan sentimen terkait pemboikotan produk Israel[1].

Twitter merupakan sebuah layanan jejaring sosial yang didirikan pada Maret 2006 oleh Jack Dorsey, Noah Glass, Biz Stone, dan Evan Williams[3]. Pengguna *twitter* di Indonesia pada Juli 2023 sekitar 25,25 juta pengguna dan menduduki peringkat ke-4 terbanyak di dunia[4]. Dalam konteks boikot produk Israel, analisis sentimen publik di *twitter* menjadi kunci untuk memahami pandangan dan perasaan masyarakat. Meskipun informasi dari *twitter* seringkali tidak terstruktur dan mungkin terganggu (*noise*), penggunaan *text mining* menjadi esensial dalam data mining. [1].

Text mining merupakan proses melakukan kategorisasi teks, ekstraksi pola dari data teks yang tidak terstruktur untuk mendapatkan tren, pola data, dan pengetahuan potensial[5], *text mining* bertujuan untuk menemukan informasi atau tren terbaru yang belum terungkap[6]. Selain itu juga, dapat digunakan untuk mengidentifikasi masalah data-data dalam teks dengan cara mengumpulkan informasi dari kumpulan data yang *unsupervised* pada suatu topik tertentu berkekuatan tinggi[7]. Dalam *text mining*, ada beberapa tahapan pengolahan sumber data, salah satunya *Text Preprocessing*. Dalam proses ini terdapat beberapa tahapan diantaranya *Case Folding*, *Cleaning*, *Tokenizing*, *Stopword Removal*, dan *Stemming*[8]. Fokus utama *text mining* adalah untuk melakukan analisis sentimen.

Analisis sentimen merupakan proses mengklasifikasi umpan balik positif, negatif, netral dari opini publik untuk mempercepat dan mempermudah dalam memahami reaksi, perspektif dan perasaan masyarakat terhadap topik tertentu[9]. Dalam penelitian ini tentang analisis sentimen opini publik, metode klasifikasi yang digunakan adalah algoritma *naïve bayes machine* dan algoritma *support vector* [7]. Penerapan metode klasifikasi *naïve bayes* dan *support vector machine* ini telah dilakukan oleh beberapa penelitian sebelumnya[10].

Algoritma *naïve bayes* adalah metode prediksi probabilistik sederhana yang bergantung pada penerapan teorema *bayes* dengan asumsi independensi yang kuat[11]. Kelebihan dari algoritma ini memiliki tingkat efektif dan efisiensi untuk *machine learning* dan *data mining*, algoritma ini juga dapat mengolah data dengan cepat dalam jumlah yang besar[10].

Metode klasifikasi lain yang digunakan dalam penelitian ini adalah *support vectore machine*. *support vector machine* diperkenalkan pertama kali oleh Vapnik pada tahun 1992 sebagai rangkaian harmonis konsep konsep unggulan dalam bidang *pattern recognition*[12]. Algoritma *support vectore machine* merupakan[13]. Algoritma *support vectore machine* memiliki kelebihan yaitu mampu mengatasi masalah regresi dan klasifikasi, baik linear maupun

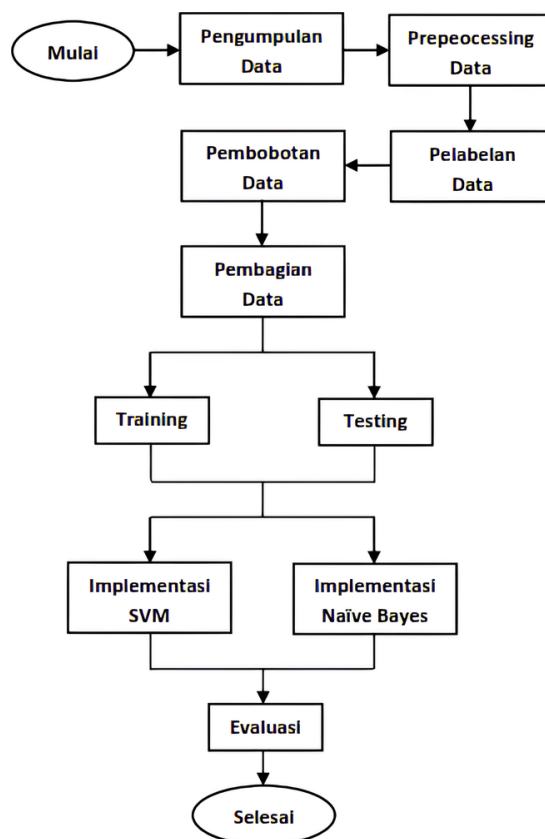
non-linear, dan memiliki tingkat akurasi cenderung baik[14].

Tujuan dari penelitian ini adalah untuk melakukan perbandingan algoritma mana yang memiliki tingkat akurasi paling tinggi dalam klasifikasi boikot produk Israel diantara dua algoritma *naïve bayes* dan *support vector machine*. Manfaat dari penelitian ini, yaitu peneliti akan memperoleh pemahaman tentang konsep *text mining* dan data mining lebih baik, serta dapat juga digunakan sebagai referensi untuk penelitian lain yang membahas topik yang sama[15].

2. METODE PENELITIAN

Penelitian ini menggunakan aplikasi *Google Colaboration* dengan menggunakan bahasa *Phyton*. Selanjutnya, data yang diperoleh dari tahapan *scrapping* akan dianalisis dengan menggunakan perhitungan algoritma *naïve bayes* dan *support vectore machie*.

Dalam metode penelitian ini diuraikan kedalam beberapa bagian tahapan penelitian seperti yang ada pada gambar 1 berikut.



Gambar 1. Alur Tahapan Penelitian

Untuk melakukan penelitian analisis sentimen ini memerlukan beberapa tahapan, seperti yang ditunjukkan pada gambar 1 di atas. Proses ini mencakup pengumpulan data, *preprocessing* data, pelabelan manual, pembobotan dengan TF-IDF, pembagian data, dan kemudian klasifikasi data menggunakan metode *naïve bayes classifier* dan

support vector machine. Selanjutnya, hasil penelitian dievaluasi dengan menentukan nilai akurasi, precision, recall dan f-score dan terakhir proses pengujian model akan dilakukan dengan cross validation untuk menentukan metode mana yang memiliki akurasi yang lebih tinggi[16].

2.1. Pengumpulan Data

Pengumpulan data dilakukan melalui proses scaraping data, yang memanfaatkan *Application Programming Integration (API) Twitter* dengan menggunakan kata kunci “Boikot Produk Israel”. Proses scaraping data dilakukan dengan menggunakan bahasa pemrograman *Phyton*, untuk mendapatkan akses ke *twitter* dengan melakukan konfigurasi *API secret*, *API key*, *access token secret* dan *access token*[17]. Data yang berhasil dikumpulkan pada penelitian ini sebanyak 2288 data. Data tersebut masih merupakan data kotor[16]. Tabel 1 menampilkan hasil pengumpulan data.

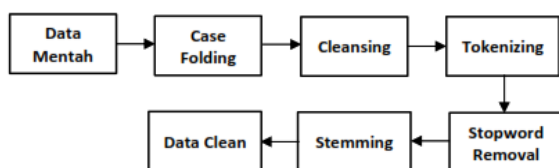
Tabel 1. Contoh dari data twitter

username	Full_text
Affandian2	Semoga konflik segera berakhir di gaza tetap semangat buat lebih teliti dalam boikot produk Israel pastikan bener yg akan di boikot jangan sampai salah target
sixteenboyzzz	@BogdanoffReal @MrOngDedy @ataritafaran Gue sendiri udah gak beli beberapa produk kaya unilever tapi masih ke McD karena ya pemegang merknya di Indo kan gak support Israel, selain itu mereka mempekerjakan difabel yang kalo di tempat kerja orang yang teria
silamay62	Yg mau boikot karena produk Israel harap di cari keabsahan nya ya banyak cara kok jangan sampai salah boikot malah merugikan orang lain

2.2. Preprocessing Data

Preprocessing merupakan tahapan proses awal terhadap teks untuk mempersiapkan teks menjadi data yang akan diolah selanjutnya. Sekumpulan karakter yang bersambungan (teks) harus dipecah-pecah menjadi unsur yang lebih berarti, yang dapat dilakukan dalam tingkatan yang berbeda[18].

Pada tahap *preprocessing* ini dimana data yang telah diperoleh sebanyak 2288 akan dilakukan *preprocessing* untuk membersihkan *noise* yang ada pada data. Tahapan *preprocessing* dapat dilihat pada gambar 2.



Gambar 2. Tahapan Pra-Proses

2.2.1. Case Folding

Case Folding merupakan tahapan untuk mengubah huruf besar menjadi huruf kecil. Proses ini dilakukan dengan tujuan untuk membuat proses selanjutnya lebih mudah[19]. Tabel 2 menampilkan hasil dari tahapan *case folding*.

Tabel 2. Hasil data setelah proses *case folding*

Data full text	Hasil Case Folding
Semoga konflik segera berakhir di gaza tetap semangat buat lebih teliti dalam boikot produk Israel pastikan bener yg akan di boikot jangan sampai salah target	semoga konflik segera berakhir di gaza tetap semangat buat lebih teliti dalam boikot produk israel pastikan bener yg akan di boikot jangan sampai salah target
@BogdanoffReal @MrOngDedy @ataritafaran Gue sendiri udah gak beli beberapa produk kaya unilever tapi masih ke McD karena ya pemegang merknya di Indo kan gak support Israel, selain itu mereka mempekerjakan difabel yang kalo di tempat kerja orang yang teria	@bogdanoffreal @mrongdedy @ataritafaran gue sendiri udah gak beli beberapa produk kaya unilever tapi masih ke mcd karena ya pemegang merknya di indo kan gak support israel, selain itu mereka mempekerjakan difabel yang kalo di tempat kerja orang yang teria
Yg mau boikot karena produk Israel harap di cari keabsahan nya ya banyak cara kok jangan sampai salah boikot malah merugikan orang lain	yg mau boikot karena produk israel harap di cari keabsahan nya ya banyak cara kok jangan sampai salah boikot malah merugikan orang lain

2.2.2. Cleansing

Cleansing merupakan salah satu tahapan dari pre-processing, padatahapan ini data yang tadi sudah melalui tahapan *case folding* yang hasilnya akan digunakan dalam tahapan ini, yaitu untuk menghapus data yang didalamnya terdapat *mantion* (@username), *hastag* (#), *link URL* (http://\situs.com), *retweet*, karakter bukan huruf, dan *emoticon*[20]. Tabel 3 merupakan tampilan dari hasil *cleansing*.

Tabel 3. Hasil data setelah proses *cleansing*

Hasil Case Folding	Hasil Cleansing
semoga konflik segera berakhir di gaza tetap semangat buat lebih teliti dalam boikot produk israel pastikan bener yg akan di boikot jangan sampai salah target	semoga konflik segera berakhir di gaza tetap semangat buat lebih teliti dalam boikot produk israel pastikan bener yg akan di boikot jangan sampai salah target
@bogdanoffreal @mrongdedy @ataritafaran Gue sendiri udah gak beli beberapa produk kaya unilever tapi masih ke mcd karena ya pemegang merknya di indo kan gak support israel, selain itu mereka mempekerjakan difabel yang kalo di tempat kerja orang yang teria	gue sendiri udah gak beli beberapa produk kaya unilever tapi masih ke mcd karena ya pemegang merknya di indo kan gak support israel selain itu mereka mempekerjakan difabel yang kalo di tempat kerja orang yang teria
yg mau boikot karena produk israel harap di cari keabsahan nya ya banyak cara kok jangan sampai salah boikot malah merugikan orang lain	yg mau boikot karena produk israel harap di cari keabsahan nya ya banyak cara kok jangan sampai salah boikot malah merugikan orang lain

2.2.3. Tokenizing

Tokenizing merupakan pemotongan kalimat menjadi satu kata, memisahkan kata-kata, dan mengidentifikasi struktur setiap kata[18]. Tokenizing juga digunakan untuk memeriksa kalimat tweet secara menyeluruh. Kemudian dilakukan pemenggalan kata berdasarkan karakter pemisahannya, sehingga kata yang bukan karakter pemisah akan digabungkan dengan karakter selanjutnya[19]. Tabel 4 menampilkan hasil dari tokenizing.

Tabel 4. Hasil data setelah proses tokenizing

Hasil Cleansing	Hasil Tokenizig
semoga konflik segera berakhir di gaza tetap semangat buat lebih teliti dalam boikot produk israel pastikan bener yg akan di boikot jangan sampai salah target	semoga,konflik,segera,berak hir,di,gaza,tetap,semangat,bu at,lebih,teliti,dalam,boikot,pr oduk,israel,pastikan,bener,yg ,akan,di,boikot,jangan,sampa i,salah,target
gue sendiri udah gak beli beberapa produk kaya unilever tapi masih ke mcd karena ya pemegang merknya di indo kan gak support israel selain itu mereka mempekerjakan difabel yang kalo di tempat kerja orang yang teria	gue,sendiri,udah,gak,beli,beb erapa,produk,kaya,unilever,ta pi,masih,ke,mcd,karena,ya,pe megang,merknya,di,indo,kan, gak,support,israel,selain,itu, mereka,mempekerjakan,difab el,yang,kalo,di,tempat,kerja,o rang,yang,teria
yg mau boikot karena produk israel harap di cari keabsahan nya ya banyak cara kok jangan sampai salah boikot malah merugikan orang lain	yg,mau,boikot,karena,produk ,israel,harap,di,cari,keabsaha n,nya,ya,banyak,cara,kok,jan gan,sampai,salah,boikot,mala h,merugikan,orang,lain

2.2.4. Stopword Removal

Stopword removal merupakan proses penghapusan kata-kata umum yang dianggap tidak memiliki makna dalam menentukan sentimen teks[21]. Hal ini dilakukan untuk meningkatkan akurasi dan relevansi analisis sentimen dengan mempertahankan hanya kata-kata penting yang mengandung makna emosional atau evaluatif. Contoh stopwords yang dihapus pada tahapan ini meliputi "yang", "dan", "pada", "ini", "itu", "untuk", "dari", "dengan", "di", "ke", "oleh", "atau", "juga", "sudah", "masih", "akan", "maka", "seperti", "dalam", "bahwa". Tabel 5 menampilkan hasil dari stopword removal.

Tabel 5. Hasil data setelah proses stopword removal

Hasil Tokenizing	Hasil Stopword
semoga,konflik,segera,berak hir,di,gaza,tetap,semangat,bu at,lebih,teliti,dalam,boikot,pr oduk,israel,pastikan,bener,yg ,akan,di,boikot,jangan,sampa i,salah,target	Semoga konflik segera berakhir gaza tetap semangat buat lebih teliti boikot produk Israel pastikan bener yg boikot jangan salah target
gue,sendiri,udah,gak,beli,beb erapa,produk,kaya,unilever,ta pi,masih,ke,mcd,karena,ya,pe megang,merknya,di,indo,kan, gak,support,israel,selain,itu, mereka,mempekerjakan,difab el,yang,kalo,di,tempat,kerja,o rang,yang,teria	gue sendiri udah gak beli beberapa produk kaya unilever mcd pemegang merknya indo kan gak support Israel mempekerjakan difabel kalo tempat kerja orang teria

yg,mau,boikot,karena,produk ,israel,harap,di,cari,keabsaha n,nya,ya,banyak,cara,kok,jan gan,sampai,salah,boikot,mala h,merugikan,orang,lain	yg mau boikot produk Israel harap cari keabsahannya banyak cara kok jangan salah boikot malah merugikan orang
---	---

2.2.5. Stemming

Stemming merupakan proses proses mengubah setiap kata dari kata berimbuhan menjadi kata dasar[20]. Tujuannya adalah menyamakan kata-kata yang memiliki akar kata sama, meningkatkan konsistensi, dan membantu model dalam mengenali pola serta makna kata secara lebih efektif. Tabel 6 merupakan tampilan hasil dari tahap stemming.

Tabel 6. Hasil data setelah proses stemming

Hasil Stopword Removal	Hasil Stemming
Semoga konflik segera berakhir gaza tetap semangat buat lebih teliti boikot produk Israel pastikan bener yg boikot jangan salah target	moga konflik segera akhir gaza tetap semangat buat lebih teliti boikot produk Israel pasti bener yg boikot jangan salah target
gue sendiri udah gak beli beberapa produk kaya unilever mcd pemegang merknya indo kan gak support israel mempekerjakan difabel kalo tempat kerja orang teria	gue sendiri udah gak beli beberapa produk kaya unilever mcd pegang merknya indo kan gak support Israel mempekerjakan difabel kalo tempat kerja orang ria
yg mau boikot produk israel harap cari keabsahannya banyak cara kok jangan salah boikot malah merugikan orang	yg mau boikot produk Israel harap cari abash nya banyak cara kok jangan salah boikot malah rugi orang

2.3. Pelabelan Manual

Sebelum memulai tahap klasifikasi data, peneliti melakukan pelabelan dataset secara manual dengan melibatkan ahli bahasa untuk menentukan label pada setiap tweet. Kategori label dalam penelitian ini terbagi menjadi tiga, yaitu label positif, label negatif, dan label netral[20]. Tabel 7 menampilkan hasil dari pelabelan manual denagn melibatkan pakar Bahasa.

Tabel 7. Hasil pelabelan manual

No	Data Clean	Kelas
1	semoga konflik segera berakhir gaza tetap semangat buat lebih teliti boikot produk israel pastikan bener yg boikot jangan salah target	Positif
2	gue sendiri udah gak beli beberapa produk kaya unilever mcd pemegang merknya indo kan gak support israel mempekerjakan difabel kalo tempat kerja orang teria	Netral
3	yg mau boikot produk israel harap cari keabsahan nya banyak cara kok jangan salah boikot malah merugikan orang	Negatif

2.4. Pembobotan

TF-IDF (Term Frequency-Inverse Document Frequency) adalah algoritma yang memberikan nilai bobot pada teks[21]. TF-IDF Vectorizer, library python, digunakan untuk tahap pembobotan kata. Menurut nilai pembobotan, semakin berat kata dalam kalimat, semakin layak kata tersebut digunakan sebagai keyword[22]. Berikut persamaan TF-IDF:

$$TF - IDF(d, t) = TD(d, t) * IDF(t) \tag{1}$$

dimana:

$$TF(d, t) = \frac{\text{jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d}$$

$$IDF(t) = \frac{\text{total dokumen}}{\text{jumlah dokumen mengandung kata } t}$$

Dan:

t = kata

d = dokumen

2.5. Naïve Bayes

Naïve Bayes merupakan salah satu algoritma yang digunakan dalam *text mining*. Klasifikasi *Naïve Bayes* ini biasanya digunakan untuk menganalisis data tesk dalam ukuran besar dan cepat untuk mendapatkan pengetahuan yang terkandung dalam data yang digunakan.

Klasifikasi *naïve bayes* merupakan suatu metode klasifikasi yang menggunakan perhitungan probabilitas. Konsep dasar yang digunakan pada klasifikasi *naïve bayes* adalah *teorema bayes* yang dinyatakan pertama kali oleh Thomas Bayes[23]. Nilai probabilitas yang digunakan dapat digambarkan sebagai berikut.

$$P(C_k|X) = \frac{P(X|C_k).P(C_k)}{P(X)} \tag{2}$$

Keterangan:

$P(C_k|X)$ = Nilai probabilitas C_k berdasarkan kondisi X

$P(X|C_k)$ = Probabilitas X berdasarkan kondisi pada C_k

$P(C_k)$ = Probabilitas hipotesis C_k

$P(X)$ = Probabilitas dari nilai X

2.6. Support Vector Machine

Support Vector Machine (SVM) adalah algoritma *machine learning* yang menggunakan fungsi *hyperplane* untuk memisahkan daerah-daerah kelas pada data. *Hyperplane* merupakan fungsi yang berfungsi sebagai pemisah antar kelas. SVM memprediksi kelas suatu data dengan melabelinya berdasarkan daerah kelas tempat data tersebut berada. Algoritma ini umumnya digunakan pada dataset besar, terutama diambil dari situs online, dan populer dalam klasifikasi teks. Prinsip SVM adalah membangun *hyperplane* dengan margin yang sama, tidak mendekati daerah kelas tertentu, dan dilakukan dengan mengukur margin serta mencari titik maksimalnya[21]. Penyelesaian klasifikasi tiga kelas dapat menggunakan persamaan berikut.

$$\min_{w^{ij}} \frac{1}{2} (w^{ij})^T w^{ij} + c \sum_r t_r^{ij} \tag{3}$$

subject to:

$$w^{ij} x_r + b^{ij} \geq t_r^{ij}, \text{ jika } y_r = i$$

$$w^{ij} x_r + b^{ij} \leq t_r^{ij}, \text{ jika } y_r \neq i$$

$$t_r^{ij} \geq 0$$

SVM awalnya digunakan untuk mengklasifikasikan data dalam dua kelas, di mana pemisahannya berupa garis dalam dua dimensi, plane dalam tiga dimensi, dan *hyperplane* dalam dimensi lebih dari tiga. Seiring perkembangannya, SVM dapat diperluas untuk klasifikasi multi kelas, di mana konsep *hyperplane* juga diterapkan. SVM multi kelas melibatkan pendekatan yang berbeda dari kasus dua kelas. Salah satu metode SVM multi kelas adalah metode *One-Against-One*, di mana sejumlah model SVM biner dibangun untuk membandingkan satu kelas dengan kelas lainnya. Untuk mengklasifikasikan data ke k-kelas, perlu dibangun sejumlah $k(k-1)/2$ model SVM biner[20].

3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini merupakan opini atau komentar dalam media sosial *twitter* dengan menggunakan kata kunci boikot produk Israel dalam bahasa indonesia sebanyak 2288 data. Setelah melalui proses *cleaning* data yang tersisa sebanyak 1603 data. Dalam penelitian ini data dibagi menjadi 70:30 yaitu, 70% data *training* dan 30% data *testing*. Kemudian untuk menjadi perbandingan maka dilakukan perbandingan dengan pembagian data 80:20 yaitu, 80% data *training* dan 20% data *testing*. Data tersebut digunakan sebagai data *machine learning* dan data uji untuk mengevaluasi kinerja sistem. Adapun langkah-langkah untuk melakukan analisis sentimen sebagai berikut[24].

3.1. Implementasi Algoritma Naïve Bayes

Implementasi *Naïve Bayes* dalam penelitian ini menggunakan bahasa pemrograman *python*. *Python* memiliki *library Sklearn* yang dapat digunakan untuk mengimplementasikan *naïve bayes*[20]. Contoh Baris *code* implementasi algoritma *naïve bayes* dapat dilihat pada gambar 3.

```
nb_model = make_pipeline(CountVectorizer
(stop_words=stopwords.words('indonesian')),
MultinomialNB())
nb_model.fit(X_train, y_train)
nb_pred = nb_model.predict(X_test)
```

Gambar 3. Implementasi Algoritma Naïve Bayes

Dimulai dari baris *code* `nb_model` digunakan untuk membuat model *naïve bayes* dengan menggunakan *count vectorizer* untuk mengonversi teks menjadi representasi *vector* dan *MultinomialNB* sebagai model klasifikasi *naïve bayes*. Lalu menggunakan `stop_words = stopwords.words('indonesian')` daftar *stopwords* yang digunakan dalam bahasa Indonesia. `nb_model.fit(X_train, y_train)` untuk melatih model dengan data pelatihan. `nb_pred = nb_model.predict(X_test)` untuk memprediksi label kelas untuk data uji.

Selanjutnya untuk menampilkan hasil evaluasi *naive bayes* ditampilkan dengan menggunakan metrik akurasi dan laporan hasil klasifikasi dapat dilihat pada gambar 4.

```
nb_accuracy = accuracy_score(y_test, nb_pred)
print("Akurasi Naive Bayes:", nb_accuracy)
print("Laporan Klasifikasi Naive Bayes:\n",
      classification_report(y_test, nb_pred))
```

Gambar 4. Code menampilkan hasil akurasi NB

3.2. Implementasi Algoritma SVM

Implementasi *Support Vector Machine* (SVM) dalam penelitian ini menggunakan bahasa pemrograman *python*. *Python* memiliki *library Sklearn* yang dapat digunakan untuk mengimplementasikan SVM[20]. Contoh Baris *code* implementasi algoritma SVM dapat dilihat pada gambar 5.

```
svm_model = make_pipeline(TfidfVectorizer
                          (stop_words=stopwords.words('indonesian')),
                          SVC())
svm_model.fit(X_train, y_train)
svm_pred = svm_model.predict(X_test)
```

Gambar 5. Implementasi Algoritma SVM

Dimulai dari baris *code* *svm_model* digunakan untuk membuat model *support ectore machine* dengan menggunakan *count vectorizer* untuk mengonversi teks menjadi representasi vector dan *SVC* sebagai model klasifikasi *support vector machine*. Lalu menggunakan *stop_words = stopwords.words ('indonesian')* daftar *stopwords* yang digunakan dalam bahasa Indonesia. *svm_model.fit(X_train, y_train)* untuk melatih model dengan data pelatihan. *nb_pred = svm_model.predict(X_test)* untuk memprediksi label kelas untuk data uji.

Selanjutnya untuk menampilkan hasil evaluasi SVM ditampilkan dengan menggunakan metrik akurasi dan laporan hasil klasifikasi dapat dilihat pada gambar 6.

```
svm_accuracy = accuracy_score(y_test, svm_pred)
print("\nAkurasi SVM:", svm_accuracy)
print("Laporan Klasifikasi SVM:\n",
      classification_report(y_test, svm_pred))
```

Gambar 6. Code menampilkan hasil akurasi SVM

3.3. Hasil Perbandingan

Tabel 8 menampilkan hasil yang diperoleh dari perbandingan algoritma *naive bayes* dan SVM dengan pembagian data 70% data *training* dan 30% data *testing*.

Tabel 8. Hasil evaluasi dengan pembagian 70:30

Model	Accuracy	Precision	Recall	F1-score
Naive Bayes	84%	86%	84%	79%
SVM	78%	78%	83%	78%

Tabel 9 menampilkan hasil yang diperoleh dari perbandingan algoritma *naive bayes* dan SVM dengan pembagian data 80% data *training* dan 20% data *testing*.

Tabel 9. Hasil evaluasi dengan pembagian 80:20

Model	Accuracy	Precision	Recall	F1-score
Naive Bayes	85%	87%	85%	52%
SVM	84%	61%	44%	47%

Maka dalam penelitian ini nilai performa untuk akurasi dengan pembagian data 70% data *training* dan 30% data *testing* dengan model algoritma *naive bayes* memiliki tingkat akurasi yang lebih baik dengan nilai akurasi sebesar 84% dengan selisih 6% jika dibandingkan dengan model algoritma SVM dengan nilai yang lebih rendah sebesar 78%. Kemudian nilai performa untuk akurasi dengan pembagian data 80% data *training* dan 20% data *testing* dengan model algoritma *naive bayes* tetap memiliki tingkat akurasi yang lebih baik dengan nilai akurasi sebesar 85% dan tingkat akurasi menggunakan model algoritma SVM mendapatkan nilai yang lebih rendah sebesar 84% pada konfigurasi dataset sebanyak 1603 data. Maka data ditarik kesimpulan bahwa model algoritma *naive bayes* pada penelitian ini memiliki tingkat performa yang lebih baik dari pada model algoritma SVM.

4. DISKUSI

Pada penelitian ini dilakukan perbandingan algoritma *naive bayes* dan *support vector machine* dengan pembagian data 70:30 dan 80:20 terhadap komentar publik yang diambil dari media sosial twitter. Dari hasil pengujian yang telah dilakukan untuk akurasi algoritma *naive bayes* dengan pembagian data 70% data *training* dan 30% data *testing* akurasi sebesar 84% dan tingkat akurasi pembagian data 80% data *training* dan 20% data *testing* sebesar 85%. Selanjutnya untuk model algoritma *support vector machine* dengan pembagian data 70% data *training* dan 30% data *testing* memiliki tingkat akurasi sebesar 78% dan tingkat akurasi pembagian data 80% data *training* dan 20% data *testing* sebesar 84%. Hasil yang didapatkan dalam penelitian ini bahwa model algoritma SVM lebih rendah.

Penelitian ini memiliki hasil yang berbeda dengan penelitian sebelumnya yang dilakukan oleh Ade T.S dkk. membahas mengenai analisis sentimen masyarakat terhadap boikot produk Israel hanya menggunakan metode *naive bayes*. Sedangkan pada penelitian ini menggunakan perbandingan 2 algoritma. Jumlah data yang digunakan pada penelitian ini juga berbeda dengan penelitian sebelumnya.

5. KESIMPULAN

Berdasarkan penelitian yang sudah dilakukan dengan membandingkan metode klasifikasi algoritma

naïve bayes dan *support vector machine* mendapatkan kesimpulan bahwa model algoritma *naïve bayes* memiliki tingkat performa yang lebih baik dibandingkan dengan model algoritma *support vector machine*.

Hasil evaluasi yang dilakukakn pada nilai akurasi dengan pembagian data 70% data *training* dan 30% data *testing* dengan model algoritma *naïve bayes* memiliki tingkat akurasi sebesar 84% dan model algoritma SVM dengan nilai sebesar 78%. Selanjutnya, hasil evaluasi yang dilakukakn pada nilai akurasi dengan pembagian data 80% data *training* dan 20% data *testing* dengan model algoritma *naïve bayes* tetap memiliki tingkat akurasi sebesar 85% dan tingkat akurasi menggunakan model algoritma SVM sebesar 84%.

Beberapa saran yang dapat digunakan sebagai proses penelitian selanjutnya yang akan mengambil penelitian hampir serupa dengan penelitian ini yaitu, Dapat menambahkan pembagian data training dan data testing dengan perbandingan 60:40 dan 90:10, Selanjutnya juga dapat melakukan perbandingan dengan menggunakan proses *stemming* dan tidak dalam melakukan *preprocessing* data.

DAFTAR PUSTAKA

- [1] A. Tiara Susilawati, A. H. Tiara Susilawati Universitas Muhammadiyah Kalimantan Timur Nur Anjeni Lestari Universitas Muhammadiyah Kalimantan Timur Puput Alpria Nina Universitas Muhammadiyah Kalimantan Timur Jl Ir Juanda No, K. Samarinda Ulu, K. Samarinda, and K. Timur, "Analisis Sentimen Publik Pada Twitter Terhadap Boikot Produk Israel Menggunakan Metode Naïve Bayes," *J. Ilm. Mhs.*, vol. 2, no. 1, pp. 26–35, 2024, [Online]. Available: <https://doi.org/10.59603/niantanasikka.v2i1.240>
- [2] cnbcindonesia.com, "9 Update Gaza, Korban Tewas Tembus 17 Ribu-Israel Makin Gila," cnbcindonesia.com. Accessed: Jan. 01, 2024. [Online]. Available: <https://www.cnbcindonesia.com/news/20231208190322-4-495782/9-update-gaza-korban-tewas-tembus-17-ribu-israel-makin-gila>
- [3] pji.uma.ac.id, "Pengertian, Sejarah dan Manfaat Twitter untuk Kaum Muda Milenial," pji.uma.ac.id. Accessed: Jan. 02, 2024. [Online]. Available: <https://pji.uma.ac.id/index.php/2021/11/05/pengertian-sejarah-dan-manfaat-twitter-untuk-kaum-muda-milenial/>
- [4] C. M. Annur, "Jumlah Pengguna Twitter Indonesia Duduki Peringkat ke-4 Dunia per Juli 2023," databoks.katadata.co.id. Accessed: Jan. 02, 2024. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2023/11/01/jumlah-pengguna-twitter-indonesia-duduki-peringkat-ke-4-dunia-per-juli-2023>
- [5] A. Firdaus and W. I. Firdaus, "Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi: (Sebuah Ulasan)," *J. JUPITER*, vol. 13, no. 1, p. 66, 2021.
- [6] Mwendwa, "A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data," *Physiol. Behav.*, vol. 176, no. 3, pp. 139–148, 2019, doi: 10.1016/j.ijmedinf.2019.02.008.A.
- [7] A. Nursalim, R. Novita, I. Systems, and S. Program, "Sentiment Analysis of Comments on Google Play Store , Twitter and Youtube To the My Pertamina Application With Support Vector Analisis Sentimen Komentar Pada Google Play Store , Twitter Dan Youtube Terhadap Aplikasi My Pertamina Dengan Support," vol. 4, no. 6, pp. 1305–1312, 2023.
- [8] V. A. Flores, L. Jasa, and L. Linawati, "Analisis Sentimen untuk Mengetahui Kelemahan dan Kelebihan Pesaing Bisnis Rumah Makan Berdasarkan Komentar Positif dan Negatif di Instagram," *Maj. Ilm. Teknol. Elektro*, vol. 19, no. 1, p. 49, 2020, doi: 10.24843/mite.2020.v19i01.p07.
- [9] K. A. Rokhman, B. Berlilana, and P. Arsi, "Perbandingan Metode Support Vector Machine Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online," *J. Inf. Syst. Manag.*, vol. 3, no. 1, pp. 1–7, 2021, doi: 10.24076/joism.2021v3i1.341.
- [10] A. Tejawati, A. Septiarini, R. Rismawati, and N. Puspitasari, "Perbandingan Metode K-Nearest Neighbor dan Naive Bayes untuk Klasifikasi Konten Berita," *J. Tek. Inform.*, vol. 4, no. 2, pp. 401–412, 2023.
- [11] M. Ardiansyah, A. Sunyoto, and E. T. Luthfi, "Analisis Perbandingan Akurasi Algoritma Naïve Bayes Dan C4.5 untuk Klasifikasi Diabetes," *Edumatic J. Pendidik. Inform.*, vol. 5, no. 2, pp. 147–156, 2021, doi: 10.29408/edumatic.v5i2.3424.
- [12] A. R. Isnain, A. I. Sakti, D. Alita, and N. S. Marga, "Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm," *J. Data Min. dan Sist. Inf.*, vol. 2, no. 1, p. 31, 2021, doi: 10.33365/jdmsi.v2i1.1021.
- [13] H. Hermanto, A. Mustopa, and A. Y. Kuntoro, "Algoritma Klasifikasi Naive Bayes Dan Support Vector Machine Dalam Layanan Komplain Mahasiswa," *JITK (Jurnal Ilmu Pengetah. dan Teknol.*

- Komputer*), vol. 5, no. 2, pp. 211–220, 2020, doi: 10.33480/jitk.v5i2.1181.
- [14] D. A. Otchere, T. O. Arbi Ganat, R. Gholami, and S. Ridha, “Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models,” *J. Pet. Sci. Eng.*, vol. 200, no. December 2020, p. 108182, 2021, doi: 10.1016/j.petrol.2020.108182.
- [15] H. Apriyani and K. Kurniati, “Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus,” *J. Inf. Technol. Ampera*, vol. 1, no. 3, pp. 133–143, 2020, doi: 10.51519/journalita.volume1.issue3.year20.page133-143.
- [16] H. Sujadi, “Analisis Sentimen Pengguna Media Sosial Twitter Terhadap Wabah Covid-19 Dengan Metode Naive Bayes Classifier Dan Support Vector Machine,” *INFOTECH J.*, vol. 8, no. 1, pp. 22–27, 2022, doi: 10.31949/infotech.v8i1.1883.
- [17] H. Setiawan, E. Utami, and S. Sudarmawan, “Analisis Sentimen Twitter Kuliah Online Pasca Covid-19 Menggunakan Algoritma Support Vector Machine dan Naive Bayes,” *J. Komtika (Komputasi dan Inform.)*, vol. 5, no. 1, pp. 43–51, 2021, doi: 10.31603/komtika.v5i1.5189.
- [18] Rina Noviana and Isram Rasal, “Penerapan Algoritma Naive Bayes Dan Svm Untuk Analisis Sentimen Boy Band Bts Pada Media Sosial Twitter,” *J. Tek. dan Sci.*, vol. 2, no. 2, pp. 51–60, 2023, doi: 10.56127/jts.v2i2.791.
- [19] A. C. Khotimah and E. Utami, “Comparison Naive Bayes Classifier, K-Nearest Neighbor, and Support Vector Machine in the classification of individual on twitter account,” *J. Tek. Inform.*, vol. 3, no. 3, pp. 673–680, 2022, [Online]. Available: <https://doi.org/10.20884/1.jutif.2022.3.3.254>
- [20] S. Styawati, N. Hendrastuty, and A. R. Isnain, “Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine,” *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, pp. 150–155, 2021, doi: 10.30591/jpit.v6i3.2870.
- [21] M. I. Fikri, T. S. Sabrila, Y. Azhar, and U. M. Malang, “Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter,” *SMATIKA J. STIKI Inform. J.*, vol. 10, no. 2, pp. 71–76, 2020.
- [22] I. Yunanto and S. Yulianto, “Twitter Sentiment Analysis Pedulilindungi Application Using Naïve Bayes and Support Vector Machine,” *J. Tek. Inform.*, vol. 3, no. 4, pp. 807–814, 2022, doi: 10.20884/1.jutif.2022.3.4.292.
- [23] A. Z. Amrullah, A. Sofyan Anas, and M. A. J. Hidayat, “Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square,” *Jurnal*, vol. 2, no. 1, pp. 40–44, 2020, doi: 10.30812/bite.v2i1.804.
- [24] I. Kurniawan, A. Lia Hananto, S. Shofia Hilabi, A. Hananto, B. Priyatna, and A. Yuniar Rahman, “Perbandingan Algoritma Naive Bayes Dan SVM Dalam Sentimen Analisis Marketplace Pada Twitter,” *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 1, pp. 731–740, 2023, [Online]. Available: <http://jurnal.mdp.ac.id>.