

Analisis Sentimen Pengguna *Twitter* terhadap Konflik antara Palestina dan Israel Menggunakan Metode *Naïve Bayesian Classification* dan *Support Vector Machine*

Haiqal Ramanizar Al Fajri¹, Roy Binsar Sinaga², Hasan Mubarak³, Albet Dwi Pangestu⁴, Desta Sandya Prasvita⁵
Informatika/Departemen
UPN Veteran Jakarta

Jl. RS. Fatmawati Raya, Pd. Labu, Kec. Cilandak, Kota Depok, Daerah Khusus Ibukota Jakarta 12450
haiqalramanizaralfajri@gmail.com¹, roybs@upnvj.ac.id², hasanm@upnvj.ac.id³, albetdp@upnvj.ac.id⁴,
desta.sandya@upnvj.ac.id⁵

Abstrak. Konflik antara Israel dan Palestina merupakan konflik yang menyita perhatian dunia. Hal ini menjadi perbincangan diseluruh dunia termasuk Indonesia, sehingga memunculkan berbagai opini publik salah satunya di media sosial *Twitter*. Tujuan dari penelitian ini adalah untuk membandingkan penggunaan metode klasifikasi dalam mengidentifikasi *sentiment public* dari data *tweet* pengguna *Twitter*. Subjek pada penelitian ini yaitu berupa data *tweet* pengguna yang memuat #israel dan #palestina. Dimana setelah dilakukan *preprocessing*, pembobotan, dan *resampling* terkumpul 508 data *tweet* dengan 453 data *training* dan 55 data *testing*. Metode klasifikasi yang digunakan dalam penelitian ini yaitu metode *Naïve Bayes* dan *Support Vector Machine*. Setelah diklasifikasi, dilakukan evaluasi terhadap dua model klasifikasi tadi. Hasil penelitian menunjukkan bahwa penggunaan metode *Naïve Bayes* menghasilkan akurasi, *precision*, dan *recall* masing-masing sebesar 74%, 74%, 56% sedangkan dari metode *Support Vector Machine* didapatkan akurasi, *precision*, dan *recall* sebesar 80%, 79%, 72%. Dapat disimpulkan bahwa metode klasifikasi *Support Vector Machine* dapat mengklasifikasikan dataset lebih baik dari *Naïve Bayes Classification*.

Kata Kunci: Israel, Palestina, Sentiment, *Naïve Bayes*, *Support Vector Machine*.

1 Pendahuluan

Konflik antara Israel dan Palestina merupakan salah satu konflik yang menyita perhatian dunia internasional hingga saat ini. Konflik ini telah berlangsung selama 73 tahun sejak 1948 hingga sekarang. Konflik kembali memuncak pada tanggal 7 Mei 2021 dengan diluncurkannya serangan 4000 roket oleh *Israel Defence Force* ke kota Gaza, Palestina.

Kembali memuncaknya konflik tersebut menimbulkan berbagai opini publik khususnya masyarakat Indonesia. Opini publik tersebut memiliki berbagai pandangan baik itu pro maupun kontra terhadap kedua belah pihak. Pandangan masyarakat tersebut dituangkan di berbagai media sosial. Salah satu media sosial yang banyak digunakan ialah *Twitter*.

Twitter merupakan salah satu media sosial yang menjadi wadah bagi publik untuk mengutarakan opini mereka terkait masalah tersebut. Sebagai wadah opini publik, terkumpul berbagai data *tweet* pengguna mengenai konflik antara Palestina dan Israel. Data ini dapat diolah dan dianalisis untuk memahami sentimen pengguna *Twitter* terhadap konflik tersebut.

Penelitian mengenai analisis sentimen terhadap data *tweet* pengguna *Twitter* telah banyak dilakukan oleh berbagai peneliti. Penelitian yang dilakukan tersebut banyak menggunakan metode *Naïve Bayesian Classification* dan *Support Vector Machine* (SVM), diantaranya penelitian yang dilakukan oleh Ghulam Asrofi Buntoro membahas mengenai klasifikasi opini publik di *twitter* terkait calon gubernur DKI Jakarta tahun 2017 menggunakan algoritma klasifikasi *Support Vector Machine* (SVM) dan *Naïve Bayesian Classification* serta fitur *Lexicon Based*. Hasil dari penelitian menunjukkan akurasi tertinggi SVM sebesar 90% sedangkan akurasi tertinggi *Naïve Bayes* sebesar 95% [1]. Selain itu terdapat juga penelitian yang dilakukan oleh Ronny Julianto, dkk yang membahas terkait analisis sentimen publik terhadap layanan provider telepon di *Twitter* menggunakan metode klasifikasi *Naïve Bayesian Classification*. Hasil penelitian menghasilkan akurasi sebesar 74% [2].

Berdasarkan penelitian tersebut, penulis melakukan pengolahan mengenai analisis sentimen dari data *tweet* pengguna terhadap konflik antara Palestina dan Israel. Metode yang akan digunakan pada penelitian kali ini adalah *Naïve Bayesian Classification* dan *Support Vector Machine (SVM)*. Kedua metode tersebut dibandingkan karena berdasarkan penelitian sebelumnya keduanya menghasilkan akurasi yang tinggi, selain itu kedua metode tersebut telah banyak digunakan dalam analisis sentimen. Sehingga pada penelitian kali ini kedua metode tersebut akan dibandingkan untuk mendapatkan hasil akurasi terbaik.

2 Landasan Teori

2.1 Analisis Sentimen

Analisis sentimen merupakan suatu metode dalam pengolahan kata yang bertujuan untuk menganalisis opini atau penilaian publik mengenai suatu peristiwa, aktivitas, kegiatan, organisasi, dan hal lainnya yang diketahui masyarakat secara umum [3] [4]. Analisis sentimen berfokus dalam mengklasifikasi atau mengelompokkan suatu opini publik pada sebuah kalimat atau dokumen sehingga diketahui opini tersebut apakah masuk kedalam opini yang bersifat negatif, positif, atau netral. Seiring perkembangan jaman, analisis sentimen semakin banyak digunakan. Tidak hanya berguna dalam bidang ilmu komputer, namun juga sudah banyak digunakan dalam bidang politik, ekonomi, bisnis, dan ilmu sosial lainnya. Hal ini terjadi karena semua bidang tersebut juga membutuhkan opini dari publik sebagai penentu arah dan keputusan penting suatu bidang tertentu [5].

2.2 Naïve Bayes

Synthetic Naive Bayes Classification didasarkan pada teorema *Bayesian*. *Naive Bayes Classification* bekerja dengan cara membagi tiap atribut menjadi atribut independen. Atribut yang ada ditentukan nilai probabilitasnya sebagai penentu dalam keputusan akhir. Atribut yang independen berarti atribut tersebut tidak saling mempengaruhi satu sama lain. Dalam penentuan probabilitas akhir digunakanlah rumus teorema bayesian. Berikut merupakan rumus *bayes* [6].

Dimana $P(C|X)$ merupakan *Posterior Probability* dari hasil perkalian $P(X|C)$ sebagai probabilitas berdasarkan kondisi pada hipotesis dengan $P(C)$ sebagai *prior probability*. Kemudian hasil perkalian tersebut dibagi dengan $P(X)$ sebagai *predictor prior probability*.

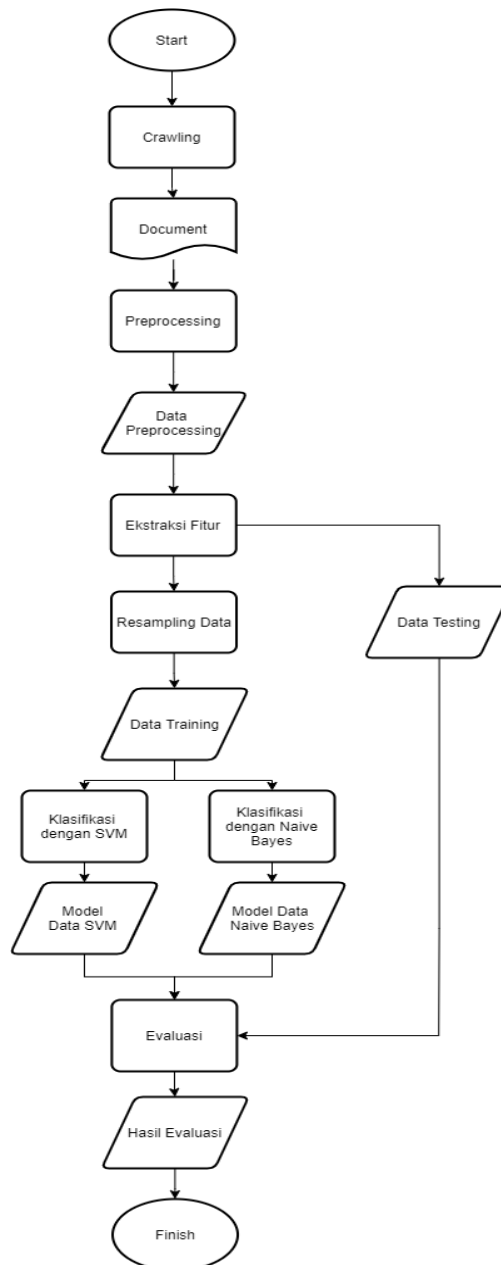
$$P(C|X) = \frac{P(x|c)P(c)}{P(x)} \quad \text{—————} \quad (1)$$

2.3 Support Vector Machine

Support Vector Machine merupakan satu dari beberapa metode berjenis *supervised learning*. SVM merupakan metode yang menggunakan hipotesis dalam bentuk beberapa fungsi linear dari fitur berdimensi tinggi. Fitur ini dilatih dengan menggunakan algoritma pembelajaran berdasarkan teori optimasi. SVM ini dapat digunakan dalam pengklasifikasian dua kelas serta digunakan dalam klasifikasi data linear dan non-linear. Dalam pengklasifikasian non-linear, digunakan fungsi linear untuk memaksimalkan margin dalam fitur berdimensi tinggi [7].

3 Metode Penelitian

Dalam melakukan analisis sentimen, terdapat beberapa tahapan penelitian, yang perlu dilakukan untuk mendapatkan hasil pengujian terakurat. Tahapan ini digunakan sebagai alur penelitian untuk memudahkan pelaksanaan penelitian. Tahapan penelitian seperti pada Gambar 1 dibawah ini merupakan tahapan atau alur penelitian.



Gambar. 1. Tahap - tahap Metode Penelitian

3.1 Crawling

Pada tahap penelitian ini yang pertama yakni tahap *crawling*. *Crawling* merupakan tahap pengumpulan dataset.

Data yang digunakan merupakan data *tweet* berbahasa Indonesia dari *twitter* mengenai konflik yang terjadi antara Palestina dan Israel yang akhir-akhir mulai memanas kembali dengan #Israel dan #Palestina. Data yang digunakan berjumlah 316 data yang selanjutnya akan diolah dengan dilakukan *preprocessing*. Pada tabel 1 dibawah merupakan contoh dataset yang berhasil *dicrawling*.

Tabel 1. Data Tweet Hasil *Crawling*

Data set
Orang ga punya etika berbicara ialah menghina agama! Dan jgn umbar" ke Calon Presiden!! #PalestinaIsrael
Faktanya Israel itu negara; Palestina pun negara. Pasti ada yang tak ingin kedua bangsa itu hidup rukun berdampingan #PalestinaIsrael
Konoha di serang? Terus harus diam saja atas nama kasih sayang. BELA DIRI, tegakkan keadilan! #NarutoTheMovie #PalestinaIsrael #Perang
Semoga Tuhan kita semua menghentikan perang #PalestinaIsrael #SuriahIrak

3.2 Preprocessing

Setelah pengumpulan dataset yang berupa *tweet*, tahap selanjutnya yang harus dijalani adalah *preprocessing*. Tahapan ini bertujuan untuk menyiapkan data secara terstruktur dan siap untuk dilakukan proses selanjutnya. Tahapan ini sangat penting dikarenakan dataset yang diambil dari *Twitter* masih belum terstruktur atau masih banyak *noise*. *Preprocessing* ini dibagi menjadi beberapa bagian, diantaranya *Case Folding*, *Cleaning*, *Stemming*, *Stopword*, dan *Tokenized*.

3.2.1 Case Folding

Pada tahap penelitian ini, akan dilakukan perubahan pada dataset *tweet* yang dimana merubah huruf kapital (*uppercase*) menjadi huruf kecil (*lowercase*). Tujuan dari tahapan ini adalah untuk membuat format huruf menjadi sama, demi menghindari terjadinya kata yang tidak terdeteksi atau *case sensitive*. Contoh *case folding* dapat dilihat pada Tabel 2 berikut ini:

Tabel 2. Hasil *Case Folding*

Sebelum	Sesudah
Orang ga punya etika berbicara ialah menghina agama! Dan jgn umbar" ke Calon Presiden!! #palestinaisrael	orang ga punya etika berbicara ialah menghina agama! dan jgn umbar" ke calon presiden!! #palestinaisrael

3.2.2 Cleaning

Pada tahapan ini, semua data *tweet* akan dibersihkan dengan cara menghilangkan beberapa komponen yang tidak memiliki *value* seperti tanda baca, *username*, angka, *hashtag*, dan URL atau *link*. Tujuan dilakukan tahapan ini adalah mengurangi *noise* yang berapa pada dataset. Contoh *cleaning* dapat dilihat pada Tabel 3 berikut ini:

Tabel 3. Hasil *Cleaning*

Sebelum	Sesudah
orang ga punya etika berbicara ialah menghina agama! Dan jgn umbar" ke calon presiden!! #palestinaisrael	orang ga punya etika berbicara ialah menghina agama dan jgn umbar ke calon presiden

3.2.3 Stemming

Pada tahapan ini, data *tweet* akan diubah dengan menghilangkan imbuhan sehingga setiap kata akan menjadi kalimat dasar. Dalam kamus Bahasa Indonesia ini merupakan hal yang kompleks karena harus menghilangkan semua imbuhan di setiap kalimat [8]. Contoh *stemming* dapat dilihat pada Tabel 4 berikut ini:

Tabel 4. Hasil *Stemming*

Sebelum	Sesudah
orang ga punya etika berbicara ialah menghina agama dan jgn umbar ke calon presiden	orang ga punya etika bicara ialah hina agama agama dan jgn umbar calon presiden

3.2.4 Stopword

Pada tahapan ini, data yang tidak terlalu penting dan tidak akan mengurangi performa klasifikasi akan dihilangkan demi mengoptimalkan data. Melakukan penyaringan kata-kata yang sering maupun jarang muncul, biasa disebut dengan *stopword*. Contoh *stopword* dapat dilihat pada Tabel 5 berikut :

Tabel 5. Hasil *Stopword*

Sebelum	Sesudah
orang ga punya etika bicara ialah hina agama agama dan jgn umbar calon presiden	orang ga punya etika bicara hina agama jgn umbar calon presiden

3.2.5 Tokenized

Pada tahapan ini, melakukan pemisahan setiap *tweet* menjadi sebuah kata tunggal (*term*). Tujuan tahapan ini adalah memisahkan setiap kata, dimana biasanya setiap kata dipisahkan oleh spasi. Contoh *tokenized* dapat dilihat pada Tabel 6 berikut ini:

Tabel 6. Hasil *Tokenized*

Sebelum	Sesudah
orang ga punya etika bicara hina agama jgn umbar calon presiden	['orang', 'ga', 'punya', 'etika', 'bicara', 'hina', 'agama', 'jgn', 'umbar', 'calon', 'presiden,']

3.3 Ekstraksi Fitur

3.3.1 Pembobotan TF-IDF

Setelah tahapan *preprocessing* berhasil dilakukan pada data , maka tahapan selanjutnya adalah ekstraksi fitur menggunakan skema *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF merupakan proses pembobotan yang dilakukan dengan cara menghitung frekuensi kemunculan suatu kata pada tiap dokumen (TF) dan frekuensi kemunculan suatu kata pada keseluruhan dokumen (IDF). Pembobotan TF-IDF bertujuan untuk mereduksi ukuran panjang dokumen yang awalnya beragam sehingga berubah ke ukuran yang tetap [9].

Berikut merupakan proses perhitungan TF-IDF pada 5 dokumen yang telah melalui tahapan *preprocessing* yang ditunjukkan pada Tabel 7 dan Tabel 8 dibawah ini.

Tabel 7. Data Awal

Dokumen	Tweet
D1	orang punya etika bicara hina agama umbar calon presiden
D2	fakta israel negara palestina negara pasti kedua bangsa hidup rukun damping
D3	konoha serang diam nama kasih sayang beladiri tegak adil
D4	semoga tuhan kita semua henti perang

Tabel 8. Perhitungan TF-IDF

Term	TF	IDF
------	----	-----

	D ₁	D ₂	D ₃	D ₄	
etika	1	0	0	0	0.602
bicara	1	0	0	0	0.602
hina	1	0	0	0	0.602
agama	1	0	0	0	0.602
umbar	1	0	0	0	0.602
calon	1	0	0	0	0.602
presiden	1	0	0	0	0.602
israel	0	1	0	0	0.602
negara	0	1	0	0	0.602
palestina	0	1	0	0	0.602
rukun	0	1	0	0	0.602
konoha	0	0	1	0	0.602
serang	0	0	1	0	0.602
kasih	0	0	1	0	0.602
sayang	0	0	1	0	0.602
beladiri	0	0	1	0	0.602
adil	0	0	1	0	0.602
tuhan	0	0	0	1	0.602
henti	0	0	0	1	0.602
perang	0	0	0	1	0.602

Keterangan:

TF(D) = jumlah kata di setiap dokumen-D
 IDF = $\log(D/df)$

Setelah dilakukan proses perhitungan pada setiap kata di tiap dokumen maka data tersebut dibagi menjadi *data training* dan *testing*, dimana kedua data tersebut akan dijadikan *input* proses *training* dan *testing* pada tahapan klasifikasi berikutnya.

3.3.2 Pembobotan Lexicon Based Features

Lexicon Based Features adalah fitur kata yang didalamnya terdapat sentimen positif dan sentimen negatif yang didasarkan pada kamus/*lexicon*. Kamus *Lexicon Based* berisikan himpunan kata sentimen yang telah dikenali. Metode *Lexicon Based* diimplementasikan dengan tiga proses utama yaitu pertama meng-*import* kamus *Lexicon Based* positif dan negatif. Kemudian dilanjutkan dengan pelabelan kata pada sebuah kalimat sehingga dapat teridentifikasi mana saja kata yang berlabel positif, negatif, dan netral. Langkah terakhir dilakukan proses penghitungan skor sentimen positif, negatif, dan netral pada setiap kata pada sebuah kalimat [10].

Untuk menghitung skor sentimen pada *Lexicon Based* digunakan rumus (2), (3), (4):

$$\text{If } \sum lb \text{ Score } (lb) > 0 \text{ then positive (2)}$$

$$\text{If } \sum lb \text{ Score } (lb) = 0 \text{ then netral (3)}$$

$$\text{If } \sum lb \text{ Score } (lb) < 0 \text{ then negative (4)}$$

3.4 Resampling Data

3.4.1 SMOTE

Synthetic Minority Over-sampling Technique (SMOTE) merupakan metode replikasi data sintetis dataset yang *imbalance*. SMOTE ini bekerja dengan mereplikasi data sintetis pada kelas minoritas. Replikasi yang

dilakukan menggunakan teori *K-Nearest Neighbors* dengan mencari ketetanggaan terdekat dari data minoritas, lalu akan dilakukan replikasi berdasarkan data minor dengan data tetangga terdekat [11].

3.5 Klasifikasi

Tahap selanjutnya yang akan dilakukan adalah pengklasifikasian, peneliti menggunakan 2 metode pengklasifikasian. Metode yang digunakan tersebut ialah *Naive Bayes Classification* dan *Support Vector Machine*. Metode ini akan menghasilkan dua model yang kemudian akan dibandingkan nilai akurasinya.

4 Hasil dan Pembahasan

Data yang digunakan pada penelitian ini diambil dari *Twitter* menggunakan fungsi *tweepy* dengan *hashtag* israel dan palestina. Sebanyak 273 data yang digunakan pada penelitian dengan *tweet* berbahasa indonesia mengenai konflik perseteruan antara Israel dan Palestina. Pada penelitian ini data dibagi menjadi 80% data *training* dan 20% data *testing* menggunakan *Hold Out Validation*. Data *training* akan diberikan pembobotan dengan ekstraksi fitur TF-IDF. Gambar 2 di bawah ini menunjukkan hasil pembobotan dengan menggunakan metode TF-IDF.

	ambil	bantu	bekal	caterpillar	dunia	jentera	laku	musnah	palestin	rumah
1	0	0	0	0	0	0	0	0	0	0
2	1	1	1	2	1	1	1	1	1	1
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	1	1	1	2	1	1	1	1	1	1

Gambar. 2. Hasil pembobotan dengan metode TF-IDF

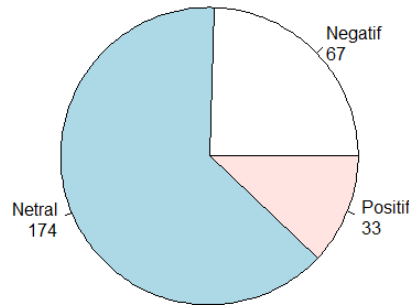
Kemudian data tersebut akan dilakukan pelabelan dengan metode *Lexicon Based* menjadi tiga kelas yaitu positif, netral dan negatif. Pada Gambar 3 di bawah ini ditampilkan data hasil analisis sentimen menggunakan metode *Lexicon Based*.

	klasifikasi	score	text
1	Netral	0	malaysia sentiasa tindak negaranegara jalin hubung diplo...
2	Netral	0	malaysia sentiasa tindak negaranegara jalin hubung diplo...
3	Netral	0	hhi rupa bekal jentera laku bersih etnik palestin utama teb...
4	Positif	1	bds kecam sekeraskerasnya tindak uae terima kunjung ras...
5	Negatif	-1	rupa jentera runtuh buah kedai kawasan jiran albustan sil...
6	Netral	0	institusibisnespersatuan zon bebas apartheid apartheid fre...
7	Negatif	-1	emiriah arab uae khianat juang palestin baca terang penu...
8	Netral	0	warga palestin hadap bersih etnik tindas dahsyat seseteng...
9	Netral	0	serang gaza barubaru bunuh ramai doktor serang hospital...
10	Negatif	-1	caterpillar inc bekal jentera berat israel musnah rumahru...

Gambar. 3. Hasil pelabelan data dibagi menjadi tiga kelas

Berdasarkan analisis sentimen menggunakan *Lexicon Based* yang telah dilakukan didapatkan hasil yaitu dari total 273 data *tweet*, sebanyak 67 *tweet* diidentifikasi sebagai kelas negatif, 173 diidentifikasi sebagai kelas netral, dan 33 data sebagai kelas positif. Gambar 4 merupakan visualisasi jumlah frekuensi tiap kelas pada data.

Frekuensi Tiap Kelas



Gambar. 4. Diagram lingkaran frekuensi kelas pada dataset

Untuk menggambarkan sentimen seberapa sering kata yang muncul pada penelitian ini menggunakan *Word Cloud*. Semakin sering suatu kata digunakan pada *tweet* maka akan besar kata tersebut ditampilkan pada visualisasi *Word Cloud*. Gambar 5 di bawah ini merupakan *Word Cloud* data yang telah diproses.



Gambar. 5. Grafik *Word Cloud*

Berdasarkan Gambar 5 di atas dapat dilihat bahwasanya ada beberapa kata yang sangat menonjol seperti kata israel, palestina, warga, palestina, otoritas dan beberapa kata lainnya. Itu menunjukkan bahwasanya kata tersebut sering digunakan pada *tweet*.

Setelah berhasil melakukan *preprocessing* data dan pelabelan dengan metode *Lexicon Based* maka selanjutnya dilakukan pembagian data menjadi data *training* dan *testing* menggunakan metode *Hold Out Validation*. Setelah dilakukan pembagian data, lalu data *training* dilakukan *resampling* dengan menggunakan metode SMOTE. Jumlah data *training* sebanyak 453 data, dan data *testing* sebanyak 55 data. Lalu data *training* akan dibentuk menjadi model klasifikasi SVM dan *Naive Bayes* dimana model tersebut akan kita gunakan untuk melakukan uji dengan data *testing* menggunakan metode klasifikasi *Naive Bayes* dan SVM.

Tahap selanjutnya dilakukan evaluasi menggunakan metode *Confusion Matriks*. Model *Confusion Matriks* akan membentuk sebuah matriks akurasi, *true positif*, *true negatif*, *precision* dan *recall*. Berikut merupakan hasil *Confusion Matriks* dari kedua metode yakni metode *Naive Bayes* dan SVM yang ditunjukkan pada Tabel 9 dan Tabel 10.

Tabel 9. *Confusion Matriks Naive Bayes Classifier.*

Prediction	Referensi		
	Negatif	Netral	Positif
Negatif	6	0	0
Netral	7	33	5
Positif	2	0	2

Tabel 10. *Confusion Matriks Support Vector Machine*

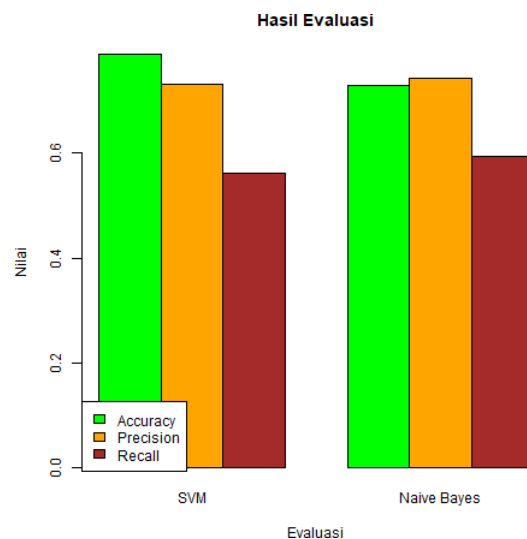
Prediction	Referensi		
	Negatif	Netral	Positif
Negatif	8	0	0
Netral	5	31	2
Positif	2	2	5

Berdasarkan *Confusion Matriks* untuk dua metode di atas dihasilkan akurasi, *precision*, dan *recall* untuk mengetahui metode manakah yang cocok digunakan pada penelitian ini. Pada Tabel 11 dibawah ini merupakan perbandingan hasil proses klasifikasi dari metode klasifikasi *Naive Bayes Classifier* dengan *Support Vector Machine*.

Tabel 11. Perbandingan Hasil Klasifikasi

	Accuracy %	Precision %	Recall %
Naive Bayes Classification	74	74	56
SVM	80	79	72

Tabel 11 berisikan informasi mengenai nilai akurasi, *precision* dan *recall* dari masing-masing kedua metode klasifikasi. pada baris pertama berisikan informasi mengenai metode *Naive Bayes* dengan nilai akurasi sebesar 74%, *precision* 74% dan *recall* 56%. sedangkan pada baris kedua berisikan informasi mengenai metode SVM dengan nilai akurasi sebesar 80%, *precision* 79% dan *recall* 72%. Dari proses klasifikasi tersebut kita dapat divisualisasikan dalam bentuk *bar plot* seperti pada Gambar 6 dibawah ini.


Gambar. 6. Diagram *Bar Plot Accuracy, Precision dan Recall*.

Dari gambar diatas dapat dilihat nilai akurasi tertinggi terdapat pada metode *Support Vector Machine* sebesar 80% sedangkan metode *Naive Bayes* akurasinya 74%. Untuk nilai *precision* metode *Support Vector Machine* juga unggul yaitu sebesar 79% sedangkan metode *Naive Bayes* didapat *precision* yaitu 74%. Terakhir, untuk nilai *recall* untuk metode *Support Vector Machine* nilainya 72% dan metode *Naive Bayes* nilainya 56%.

5 Kesimpulan dan Saran

Berdasarkan hasil penelitian mengenai analisis sentimen pengguna *Twitter* terhadap konflik antara Palestina dan Israel menggunakan metode *Naive Bayes Classification* dan *Support Vector Machine* dapat disimpulkan bahwa pengujian menggunakan metode *Support Vector Machine* memiliki tingkat akurasi lebih tinggi dengan nilai sebesar 80%. Tak hanya itu didapatkan bahwa metode *Support Vector Machine* memiliki tingkat *precision* yang lebih tinggi yaitu 79%. Terakhir, untuk *recall* juga dihasilkan nilai tertinggi pada metode *Support Vector Machine* sebesar 72%. Nilai akurasi, *precision*, dan *recall* dari kedua metode tersebut menunjukkan bahwa metode *Support Vector Machine* dapat mengklasifikasikan dataset lebih baik dari *Naive Bayes Classification*. Untuk kedepannya, diharapkan penelitian ini dapat dikembangkan lebih baik dengan membandingkan menggunakan metode lain dan ekstraksi fitur lain, dapat menggunakan lebih banyak dataset dan kamus kata yang lebih lengkap.

Referensi

- [1] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," *INTEGER J. Inf. Technol.*, vol. 1, no. 1, pp. 32–41, 2017, [Online]. Available: https://www.researchgate.net/profile/Ghulam_Buntoro/publication/316617194_Analisis_Sentimen_Calon_Gubernur_DKI_Jakarta_2017_Di_Twitter/links/5907eee44585152d2e9ff992/Analisis-Sentimen-Calon-Gubernur-DKI-Jakarta-2017-Di-Twitter.pdf.
- [2] R. Julianto, E. D. Bintari, and Indrianti, "Analisis Sentimen Layanan Provider Telepon Seluler pada Twitter menggunakan Metode Naïve Bayesian Classification," *J. Big Data Anal. Artif. Intell.*, vol. 3, no. 1, 2017.
- [3] Hartanto, "Text Mining Dan Sentimen Analisis Twitter Pada Gerakan Lgbt," *Intuisi J. Psikol. Ilm.*, vol. 9, no. 1, pp. 18–25, 2017, doi: 10.15294/intuisi.v9i1.9561.
- [4] Y. Lin, X. Wang, and A. Zhou, "Opinion spam detection," *Opin. Anal. Online Rev.*, no. May, pp. 79–94, 2016, doi: 10.1142/9789813100459_0007.
- [5] J. A. Septian, T. M. Fahrudin, and A. Nugroho, "Journal of Intelligent Systems and Computation 43," pp. 43–49, [Online]. Available: <https://t.co/9W1oaWpfD5>.
- [6] A. L. Fairuz, R. D. Ramadhani, and N. A. Tanjung, "Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial," *J. DINDA*, vol. 1, no. 1, pp. 10–12, 2021, [Online]. Available: <http://journal.itelkom-pwt.ac.id/index.php/dinda/article/view/180>.
- [7] I. Ahmad, M. Basher, M. J. Iqbal, and A. Rahim, "Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection," *IEEE Access*, vol. 6, pp. 33789–33795, 2018, doi: 10.1109/ACCESS.2018.2841987.
- [8] P. P. A. Arsy Monica Pravina, Imam Cholissodin, "Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2789–2797, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4793>.
- [9] M. Astiningrum, K. S. Batubulan, and L. A. Sias, "Implementasi Analisis Sentimen Twitter Mengenai Opini Masyarakat Terhadap Rkuhp Tahun 2019," *Semin. Inform. Apl. POLINEMA 2020*, 2020.
- [10] U. Rofiqoh, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexion Based Feature," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 12, pp. 1725–1732, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/628>.
- [11] R. Siringoringo, "Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan k-Nearest Neighbor," *J. ISD*, vol. 3, no. 1, pp. 44–49, 2018.