

sadfa

by samsun maarif

Submission date: 11-Dec-2022 09:48AM (UTC+0700)

Submission ID: 1968724898

File name: BAB_2_Landasan_Teori_-_Copy.docx (36.33K)

Word count: 3079

Character count: 19705

2.1 Tinjauan Pustaka

Berdasarkan beberapa jurnal artikel yang mempunyai topik penelitian yang sama yaitu mengenai analisis sentimen naïve bayes, jurnal-jurnal artikel tersebut akan digunakan sebagai acuan atau referensi dalam melakukan penelitian ini. Jurnal-jurnal artikel tersebut ditemukan dengan bantuan situs pencarian jurnal Google Scholar. Dalam situs tersebut, akan ditampilkan beberapa jurnal terkait yang sedang dicari berdasarkan kata kunci yang diinputkan dalam pencarian. Dalam penelitian-penelitian yang telah dilakukan dalam jurnal-jurnal yang ada, terdapat beberapa variasi proses-proses yang digunakan dalam penelitian yang dilakukan. Penelitian di bawah ini adalah beberapa referensi jurnal artikel yang ditemukan disertai juga proses dan hasil penelitian yang ada di setiap jurnal artikel tersebut.

Penelitian pertama adalah penelitian tentang analisis sentimen opini film di twitter yang dilakukan oleh Fajar Ratnawati pada tahun 2018. Penelitian ini berjudul Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film pada Twitter. Penelitian ini mengambil komentar dari masyarakat di twitter mengenai film yang pernah ditonton sebelumnya. Penelitian ini dilakukan untuk mempermudah melihat kecenderungan opini terhadap film apakah opini tersebut positif atau negatif. Algoritma yang digunakan dalam penelitian ini adalah algoritma Naïve Bayes. Tahapan dalam penelitian ini adalah pengumpulan data, *preprocessing* dan klasifikasi data. Teknik *scrapping tweet* digunakan untuk mengumpulkan data *tweet* dengan memanfaatkan *API Search Twitter* dan *library scrapy* di *python*. *Tweet* yang diambil adalah *tweet* yang mengandung *hashtag* judul film yang diinginkan kemudian dilakukan tahap pelabelan dengan label positif dan negatif secara manual. Setelah itu data dibagi 80% *training* dan 20 % data *testing*. Adapun data *training* yang digunakan sekitar 200 *tweet* positif dan 200 *tweet* negatif. Selanjutnya data *training* dilabeli dan melalui proses *preprocessing* dan dijadikan *input* untuk *Naive Bayes Classifier*. Proses dalam tahapan *preprocessing* adalah *case folding* dan normalisasi fitur. Data yang sudah melalui tahapan *preprocessing* selanjutnya masuk ke tahap pengimplementasian *Naive*

Bayes Classifier. Setelah itu pada bagian proses *testing* atau pengujian, data *testing* tidak ada di data *training* dan data *testing* adalah data baru yang tidak dilakukan pelabelan. Pada pengujian ini metode *K-fold*, metode ini adalah metode *Cross Validation* yang populer dengan melipat data sebanyak k dan mengulangi (iterasi) eksperimennya sebanyak k juga. Total data yang digunakan dalam 500 tweet dan dibagi menjadi 5 bagian atau $k=5$ sehingga setiap lipatan masing-masing isinya adalah 100 data. Perbandingan data *training* dan *testing* adalah 80:20 sehingga ada 400 data *training* dan 100 data *testing* disetiap lipatan. Penelitian ini menghasilkan sistem untuk analisis sentimen dengan tingkat akurasi sebesar 90%, dengan nilai *precision* 92%, *recall* 90% dan *f-measure* 90%.

Penelitian kedua adalah penelitian yang dilakukan oleh Elly Indrayuni pada tahun 2019. Penelitian ini melakukan klasifikasi pada data review produk kosmetik dengan salah satu algoritma *text-mining* yaitu algoritma Naive Bayes. Penelitian ini mengelompokkan opini menjadi opini positif atau negatif. Tahapan dalam penelitian ini adalah pengumpulan data, *preprocessing*, *modeling*, dan pengujian. Data dikumpulkan dari review mengenai produk kosmetik di sebuah situs <https://>. Data yang diambil adalah review yang menggunakan bahasa Indonesia. Total data review yang diperoleh adalah 100 positif dan 100 negatif yang dilabeli secara manual. Agar dapat menghilangkan noise pada data teks, tag *HTML*, simbol ataupun tanda baca diperlukan tahap pengolahan awal data agar dapat dilakukan klasifikasi teks atau sentiment. Tahap *preprocessing* dalam penelitian ini adalah *tokenization*, *filter stopword*, *generate n-gram*. Semua tahapan *preprocessing*, tahap klasifikasi dan tahap pengujian menggunakan perangkat lunak Rapidminer. Data di *training* menggunakan Naive Bayes. Untuk *testing*, penelitian ini menggunakan beberapa metode. Metode *Cross Validation* untuk memvalidasi, *confusion matrix* untuk menghitung akurasi dan nilai *training cycle*, sedangkan nilai *Area Under Curve* (AUC) pada *ROC Curve*. Penelitian ini menghasilkan 90,50% untuk nilai terbaiknya. Ada juga nilai sensitivitas 94,00 %, nilai *specificity* 87,00 %, nilai *ppv* 87,55 %, dan nilai *npv* 93,55 %.

Penelitian ketiga adalah penelitian yang dilakukan Dwi Normawati dan Surya Allit Prayogi pada tahun 2021. Penelitian ini membahas bagaimana mengimplementasikan

algoritma *Naive Bayes* untuk melakukan analisis sentimen opini masyarakat di *Twitter* dan mengimplementasikan *confusion matrix*. Opini masyarakat di *Twitter* mengenai berita kontroversi Ahok menjadi bahan analisis sentimen dalam penelitian ini. Opini masyarakat ini diklasifikasi dalam tiga label yaitu positif, netral dan negatif. Penelitian ini berfokus pada tahapan-tahapan proses implementasi dan hasil klasifikasi. Analisis sentimen ini menggunakan *Naive Bayes Classifier* dan evaluasi performa menggunakan *confusion matrix*. Pengumpulan data dilakukan menggunakan teknik *crawling*. Teknik *crawling* adalah teknik yang bertujuan mengumpulkan data dari suatu basis data, misalkan pengumpulan data dengan cara mengunduh data dari server *Twitter*. Teknik *crawling* dalam penelitian ini menggunakan fasilitas *Google Spreadsheet* yaitu *add-one Twitter Archiver*. Data yang dikumpulkan adalah data user dan *tweet* termasuk atributnya. Setelah itu *preprocessing*, tahapan *preprocessing* di penelitian ini adalah *cleaning*, *tokenizing*, *filtering* atau *stopword*, dan *stemming*. Kemudian klasifikasi menggunakan *Naive Bayes Classifier*. Disini dijelaskan secara detail bagaimana perhitungan-perhitungan dalam proses klasifikasi label dengan algoritma *Naive Bayes*. Karena penelitian ini lebih fokus pada perhitungan, data yang digunakan tidak banyak. Penelitian ini hanya menggunakan 5 data latih dan 3 data uji. Setelah training selesai, langkah selanjutnya adalah mengimplementasikan data uji dan evaluasi performasi. Evaluasi performa menggunakan *confusion matrix*. *Confusion matrix* digunakan untuk mengukur *akurasi*, *precision* dan *recall*. Hasil pengujian pada klasifikasi menggunakan metode *Naive Bayes* di penelitian ini adalah 60% untuk akurasi, 100% untuk *precision*, dan 50% untuk *recall*.

Adapun penelitian lain yang dilakukan oleh Dedi Darwis dkk. pada tahun 2021. Penelitian ini menerapkan algoritma *Naive Bayes* untuk melakukan analisis sentimen mengenai opini masyarakat terhadap Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Penelitian ini mengambil data dari opini masyarakat di *Twitter*, dimana masyarakat dapat memberikan kritik maupun saran kepada pelayanan yang diberikan BMKG. Tahapan yang dilakukan adalah *crawler Twitter* akun BMKG, *case folding*, *cleansing*, *stopword removal*, dan *tokenisasi*. Metode klasifikasi yang digunakan untuk

analisis sentimen dalam penelitian ini adalah *Naïve Bayes Classifier*. Kalimat dibedakan menjadi opini positif, negatif, dan netral. Penelitian ini menggunakan *Python 3.74*. Di penelitian ini juga dijelaskan bagaimana perhitungan-perhitungan dalam proses klasifikasi kalimat. Penelitian ini menghasilkan akurasi berdasarkan pengujian sebesar 68,97%.

Penelitian kelima adalah penelitian dari Dianati Duei Putri dkk. pada tahun 2021. Penelitian ini melakukan analisis sentimen opini masyarakat di Twitter mengenai kinerja Dewan Perwakilan Rakyat (DPR) dengan menggunakan *Naive Bayes Classifier*. Penelitian ini menggunakan total 1546 data tweet dari Twitter yang diambil dengan crawling, dengan rincian data positif sebanyak 95, data netral 693, dan data negatif 758. Tahapan proses pre-processing dalam penelitian ini adalah case folding, tokenization, filtering, stop removal. Penelitian ini menggunakan bahasa pemrograman python dengan library Tweepy, NLTK, Sastrawi dan Scikit-Learn. Library Tweepy untuk mengambil informasi mengenai DPR yang ada di Twitter. NLTK digunakan untuk tahapan tokenization. Sastrawi digunakan untuk tahapan stop removal. Scikit-Learn digunakan untuk proses machine learning. Setelah proses pre-processing, proses selanjutnya adalah klasifikasi dengan Metode NBC. Hasil penelitian ini adalah *accuracy score* dari 20% dataset yang digunakan sebagai data testing didapatkan nilai 0,8 atau 80%. Kemudian ada *precision*, nilai *precision* 0.75 atau 75% untuk prediksi label positif, 0.79 atau 79% untuk nilai netral, dan 0.82 atau 82% untuk label. Ada juga *recall score* sebesar 0.29 atau 29% untuk label positif, 0.67 atau 67% untuk label netral, dan 84% untuk label negatif. Ada juga *F1-Score* sebesar 0.43 atau 43% untuk label positif, 0.7 atau 70% untuk label netral dan 0.77 atau 77% untuk label negatif.

No.	Peneliti	Tahun	Judul	Metode	Hasil
1	Fajar Ratnawati	2018	Implementasi Algoritma Naive Bayes Terhadap Analisis	Analisis sentimen dilakukan menggunakan Naïve Bayes	Dengan ada 400 data training dan 100 data testing disetiap lipatan <i>K-Flod Cross Validation</i> , penelitian

			Sentimen Opini Film Pada Twitter	Classifier menggunakan library NLTK.	ini menghasilkan sistem untuk analisis sentimen dengan tingkat akurasi sebesar 90%, dengan nilai precision 92%, recall 90% dan f-measure 90%.
2	Elly Indrayuni	2019	Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes	Analisis sentimen menggunakan Naïve Bayes dengan alat bantu software Rapidminer.	Dari 100 review positif dan 100 review negatif, penelitian ini menghasilkan 90,50% untuk nilai terbaiknya. Ada juga nilai sensitiviy 94,00 %, nilai specificity 87,00 %, nilai ppv 87,55 %, dan nilai npv 93,55 %.
3	Dwi Normawati, Surya Allit Prayogi	2021	Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter	Analisis sentimen menggunakan Naïve Bayes Classifier dengan perhitungan manual.	Karena penelitian ini lebih fokus pada perhitungan, data yang digunakan tidak banyak. Penelitian ini hanya menggunakan 5 data latih dan 3 data uji. Setelah training selesai, langkah

					<p>selanjutnya adalah mengimplementasikan data uji dan evaluasi performasi. Evaluasi performa menggunakan confusion matrix.¹⁸ Confusion matrix digunakan untuk mengukur akurasi, precision dan recall. Hasil pengujian pada klasifikasi menggunakan metode Naive Bayes di penelitian ini adalah 60% untuk akurasi, 100% untuk precision, dan 50% untuk recall.</p>
4.	Dedi Darwis, Nery Siskawati, Zaenal Abidin	²⁴ 2021	<p>Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional</p>	¹² Analisis sentimen menggunakan Naïve Bayes Classifier dengan python 3.74.	<p>Dengan 1179 data opini masyarakat mengenai BMKG. Penelitian ini mendapatkan tingkat akurasi berdasarkan pengujian sebesar 68,97%</p>

5.	³ Dianati Duei Putri, Gigih Forda Nama, Wahyu Eko Sulistiono	³ 2022	Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier	¹² Metode klasifikasi yang digunakan dalam penelitian analisis sentimen ini adalah Naïve Bayes Classifier.	Penelitian yang membahas opini masyarakat mengenai DPR dengan jumlah data 1546 menghasilkan accuray ³ score dari 20% dataset yang digunakan sebagai data testing didapatkan nilai 0,8 atau 80%. Kemudian ada precision, nilai precision 0.75 atau 75% untuk prediksi label positif, 0.79 atau 79% untuk nilai netral, dan 0.82 atau 82% untuk label. Ada juga recall score sebesar 0.29 atau 29% untuk label positif, 0.67 atau 67% untuk label netral, dan 84% untuk label negatif. Ada juga F1-Score sebesar 0.43 atau 43% untuk label positif, 0.7 atau 70% untuk label netral dan
----	--	----------------------	---	---	---

					0.77 atau 77% untuk label negatif.
--	--	--	--	--	------------------------------------

Setelah mempelajari beberapa penelitian dari referensi-referensi yang diulas diatas, penulis menyadari bahwa walaupun topik dan algoritma yang digunakan sama namun ada beberapa hal yang berbeda. Tahapan preprocessing yang dilakukan tiap penelitian hampir berbeda-beda. Tahapan preprocessing di penelitian ini adalah cleansing, case folding, tokenization, filtering stopword dan stemming. Setelah itu untuk algoritmanya yang akan digunakan adalah Naïve Bayes. Untuk label sentimen, penulis menggunakan label positif dan negatif seperti beberapa penelitian diatas. Penulis berharap dengan gabungan beberapa tahapan preprocessing dan algoritma yang dipilih dapat menghasilkan peforma yang baik.

2. Tinjauan Pustaka

dpr

Twitter

Twitter adalah media sosial yang digunakan sebagai layanan jejaring sosial dan microblog. Ada 300 juta lebih pengguna aktif di Twitter dengan tweet per hari mencapai total 500 juta. Di Twitter pengguna dapat mengirimkan pesan yang disebut tweet dengan maksimal 280 karakter, di pesan dapat berisi pendapat, kejadian yang sedang terjadi atau bahkan emosi dari pengguna itu sendiri. Pengguna Twitter dapat mengirim foto, grafis, video maupun tautan link website tertentu yang dapat memberikan kemudahan pengguna untuk berinteraksi. Kemudian karena kemudahan beberapa, ada beberapa fitur Twitter yang diadopsi oleh jejaring sosial lainnya seperti Facebook maupun Instagram. Beberapa fitur Twitter yang dapat memudahkan penggunaannya adalah fitur follow, dimana kita dapat mengikuti akun pengguna lain. Ada juga fitur penggunaan ikon "@" dan hastag "#", penggunaan ikon "@" untuk menyebut akun pengguna lain, sedangkan penggunaan tanda pagar/hastag "#" untuk mengumpulkan tweet berdasarkan sebuah kata/frasa/topik kejadian. Selain penggunaan tanda ikon "@" dan hastag "#", ada juga menu trending topic dimana pengguna dapat melihat

tweet yang membahas kejadian atau topik yang sedang ramai ¹⁰ dibicarakan pengguna Twitter di suatu negara maupun pengguna di seluruh dunia pada jangka waktu tertentu. Twitter yang memiliki popularitas yang tinggi membuat layanan ini banyak digunakan penggunanya untuk banyak ⁹ keperluan dalam berbagai aspek kehidupan, misalnya sebagai sarana protes, kampanye politik, sarana pembelajaran dan sebagai media komunikasi darurat.

¹ 2.2.1 Text Mining

Daring, polisi

Text mining adalah proses mendapatkan informasi dari data sumber yang tidak memiliki struktur yang mengacu pada teknik penambangan data untuk menganalisis dan ⁵ memproses data. Text mining bertujuan menghasilkan informasi dari sebuah set dokumen ⁵ melalui pemrosesan, pengelompokan, dan analisis data-data tidak terstruktur dalam jumlah besar. Text mining dapat mengambil informasi yang dapat menghasilkan analisis perasaan yang secara emosional mengidentifikasi pernyataan positif atau negatif.

Mekanisme algoritma-algoritma teks mining secara umum hampir sama ⁷ dengan algoritma-algoritma data mining. Perbedaan utama dari teks mining dan data mining adalah pada tipe data yang menjadi objek dalam prosesnya. Objek dalam data mining adalah data yang sudah ⁸ terstruktur sedangkan text mining adalah data yang berasal ³⁶ dari dokumen yang tidak terstruktur atau semi terstruktur. ²³ Dokumen yang tidak atau kurang terstruktur, seperti dokumen, word, PDF, kutipan teks dan lain-lain.

³⁴ 2.2.2 Analisis sentimen

Analisis sentimen adalah cabang dari text-mining yang bertujuan untuk menentukan persepsi atau subjektivitas publik terhadap suatu topik pembahasan, kejadian ataupun permasalahan. Vaksin. Hasil pengambilan data dari suatu teks yang dilakukan oleh analisis sentimen adalah informasi mengenai sikap, pendapat dan emosi. Pokok dari analisis sentimen adalah klasifikasi berdasarkan polarisasi. Klasifikasi dalam analisis sentimen adalah menjadi klasifikasi kedalam positif atau negatif. Analisis sentimen dan opinion mining dapat

disamakan karena kedua hal ini memiliki pokok pembahasan mengenai pendapat yang apakah cenderung positif atau negatif. Proses utama dalam analisis sentimen adalah menganalisis, memproses dan mengekstrak data tekstual dalam suatu hal, seperti layanan, produk, individu, fenomena atau topik tertentu.

Cleansing

Bmkgk vaksin.

Cleansing adalah tahapan untuk mengubah semua data-data yang tidak perlu atau memperbaiki data yang memiliki format data yang kurang benar atau salah untuk menghasilkan data yang memiliki kualitas tinggi. Contoh cleansing adalah menghilangkan URL, tag(#), tanda titik(.), koma(,) dan tanda baca lainnya yang tidak diperlukan. Contoh kalimat yang melalui proses cleansing yaitu

Case Folding

film

Case Folding adalah tahap penyeragaman semua huruf dengan cara mengubahnya menjadi lowercase atau huruf kecil. Contoh

2.2.7 Tokenization,

Tokenization adalah proses yang digunakan untuk membagi atau memisah kalimat menjadi kata-kata. Hasil pemisahan kata ini disebut token. Tanda baca yang tidak diperlukan dihilangkan dalam proses tokenization. Contoh

2.2.8 Filtering Stopword

Filtering stopwords adalah tahapan yang digunakan untuk menghilangkan kata-kata yang tidak memiliki makna atau arti(stopword). Stopword dihilangkan dalam text mining karena penggunaan kata yang terlalu umum, sehingga proses text mining dapat lebih fokus pada kata lain yang lebih penting.

Stemming.

Stemming adalah tahapan yang dilakukan untuk mengubah kata yang biasanya memiliki imbuhan di depan atau di belakang menjadi kata dasarnya sesuai kaidah indonesia.

vaksin

Naïve Bayes

Naive Bayes adalah suatu algoritma text mining. Algoritma Naive bayes didasarkan pada Teorema Bayes. Algoritma ini menghasilkan prediksi untuk kejadian dimasa depan menggunakan pengalaman-pengalaman dari kejadian dimasa dari kejadian dimasa lalu atau masa sebelumnya.

Naive bayes Classifier melakukan klasifikasi menggunakan nilai kemungkinan atau probabilitas yang dikatakan benar sebelum melakukan percobaan (probabilitas prior) dari setiap label yang merupakan ukuran jumlah terjadinya sebuah peristiwa dari masing-masing label pada data training dan pengaruh dari fitur pada data. Naive bayes dapat digunakan lebih efisien dalam bentuk learning supervised atau terawasi dimana label pada data training telah diketahui sebelumnya. Naive bayes dalam prakteknya menggunakan kemiripan tertinggi untuk dijadikan parameter dalam perhitungan di model. Rumus umum klasifikasi Naive Bayes adalah $P(H|X)$ dimana peluang bahwa anggapan atau hipotesa benar(valid) untuk data sample X dapat diterapkan pada persamaan :

$$P(H|X) = \frac{p(X|H)p(H)}{p(X)} \quad (1)$$

Keterangan:

X = Data sample dengan klas (label) yang tidak diketahui

H = Hipotesa bahwa X adalah data dengan klas (label) C

$P(H|X)$ = Peluang bahwa hipotesa benar (valid) untuk data sampel X yang diamati

$P(X|H)$ = Peluang data sample X , bila diasumsikan bahwa hipotesa benar (valid).

$P(H)$ = Peluang dari hipotesa H

$P(X)$ = Peluang data sample yang diamati

Dpr

Metode Naïve Bayes Classifier menggunakan tahap pelatihan dan tahap klasifikasi. Tahap pelatihan adalah proses analisis terhadap sampel dokumen untuk menghasilkan vocabulary yaitu kata yang ada di dokumen dan menjadi representasi dokumen. Kemudian dilakukan menentukan probabilitas prior bagi setiap kategori berdasarkan sampel dokumen. Pada tahap klasifikasi ditentukan nilai kategori dari suatu dokumen berdasarkan term yang muncul dalam dokumen yang diklasifikasi. Dalam algoritma Naïve Bayes Classifier setiap dokumen direpresentasikan dengan pasangan atribut “ $x_1, x_2, x_3, \dots, x_n$ ” dimana x_1 adalah kata pertama, x_2 adalah kata kedua dan seterusnya. Sedangkan V adalah himpunan kategori Tweet.

Pada saat klasifikasi algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (VMAP), dimana persamaannya adalah sebagai berikut :

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} \frac{p(X_1, X_2, X_3, \dots, X_n | V_j) p(V_j)}{p(X_1, X_2, X_3, \dots, X_n)} \quad (1)$$

Untuk $P(x_1, x_2, x_3, \dots, x_n)$ nilainya konstan untuk semua kategori (V_j). Persamaan diatas dapat disederhanakan menjadi sebagai berikut :

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} \prod_{i=1}^{n_6} P(x_i | V_j) P(V_j)$$

Keterangan :

V_j = Kategori tweet $j = 1, 2, 3, \dots, n$.

j_1 = Kategori tweet sentimen negatif

j_2 = Kategori tweet sentimen positif

3 $P(x_i|V_j)$ = Probabilitas x_i pada kategori

$V_jP(V_j)$ = Probabilitas dari V_j

Untuk $P(V_j)$ dan $P(x_i|V_j)$ dihitung pada saat pelatihan dimana persamaannya adalah sebagai berikut :

$$P(V_j) = \frac{|docs\ j|}{|contoh|}$$

$$P(X_j|V_j) = \frac{nk + 1}{n + |kosakata|}$$

9 Keterangan :

$|docs\ j|$ = jumlah dokumen setiap kategori j

$|contoh|$ = jumlah dokumen dari semua kategori

nk = jumlah frekuensi kemunculan setiap kata

n = jumlah frekuensi kemunculan kata dari setiap kategori

$|kosakata|$ = jumlah semua kata dari semua kategori

Confusion matrix

Polisi ahok

11 Confusion Matrix adalah sebuah tabel yang memperlihatkan bagaimana performa dari model atau algoritma yang digunakan secara spesifik. Confusion matrix juga menyatakan jumlah prediksi data uji yang benar dan prediksi data uji yang salah dari klasifikasi. Setiap baris dari tabel confusion matrix merepresentasikan kelas aktual dari data dan setiap kolom setiap kolom merepresentasikan kelas prediksi dari data (atau sebaliknya). Keterangan:

	17 Predicted Negative	Predicted Positive
--	-----------------------	--------------------

Actual Negative	True Negative (TN)	False Positive (FP)
Actual Positive	False Negative (FN)	True Positive (TP)

⁶
TP (True Positive) = Jumlah data dari kelas positif yang benar diklasifikasikan sebagai kelas positif.

TN (True Negative) = Jumlah data dari kelas negative yang benar diklasifikasikan sebagai kelas negatif.

FP (False Positive) = Jumlah data dari kelas negatif yang salah diklasifikasikan sebagai kelas positif

FN (False Negative) = Jumlah data dari kelas positif yang salah diklasifikasikan sebagai kelas negatif.

Data di tabel confusion matrix digunakan untuk melihat performa model dengan cara menghitung akurasi, presisi, dan recall.

Accuracy adalah jumlah keseluruhan dari berapa sering model melakukan klasifikasi secara benar. Rumus untuk presentasi akurasi adalah

$$Akurasi = \frac{\sum data\ benar}{n\ dokumen} \times 100\%$$

Sedangkan rumus akurasi berdasarkan tabel cconfusion matrix adalah

$$Akurasi = \frac{TP + TN}{Total}$$

Presisi adalah jumlah seberapa sering model memprediksi positif dengan benar. Rumus untuk presentasi presisi adalah

$$Presisi = \frac{\sum data\ positif}{n\ dokumen\ positif} \times 100\%$$

Sedangkan rumus presisi berdasarkan tabel cconfusion matrix adalah

$$Presisi = \frac{TP}{FP + TP}$$

Recall adalah jumlah seberapa sering model memprediksi positif dari data yang kelas aktualnya positif. Rumus untuk presentasi recall adalah

$$Recall = \frac{\sum \text{doc relevan dan terambil}}{\sum \text{seluruh doc relevan}} \times 100\%$$

Sedangkan rumus presisi berdasarkan tabel cconfusion matrix adalah

$$Recall = \frac{TP}{FN + TP}$$

Kerangka

BAB 3

8

Metode pengumpulan data

Dalam penelitian ini, data yang digunakan diambil dari tweet pengguna twitter yang mengandung kata “Tentara Indonesia”. Selain itu tweet yang diambil juga berasal dari komentar-komentar di akun twitter TNI seperti akun @Puspen_TNI, @TNI_AD, @TNI_AL, @TNI_AU. Sehingga ada 5 kata kunci dalam tweet yang didapat. Tweet tersebut didapatkan dengan menggunakan library python yaitu sunscape. Data yang diperoleh perkata kunci ada sebanyak 1000. Kemudian data disimpan di file csv. Data sekunder tersebut tidak akan langsung digunakan melainkan akan diseleksi dan dipilih terlebih dahulu data tweet mana saja yang sesuai.

Teknik analisis

Metode yang diusulkan

Text Preprocessing

Metode klasifikasi

ORIGINALITY REPORT

38%

SIMILARITY INDEX

35%

INTERNET SOURCES

30%

PUBLICATIONS

%

STUDENT PAPERS

PRIMARY SOURCES

1	Dianati Duei Putri, Gigih Forda Nama, Wahyu Eko Sulistiono. "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier", Jurnal Informatika dan Teknik Elektro Terapan, 2022 Publication	4%
2	ejournal.polbeng.ac.id Internet Source	4%
3	journal.eng.unila.ac.id Internet Source	4%
4	ejournal.poltektegal.ac.id Internet Source	3%
5	ejurnal.stmik-budidarma.ac.id Internet Source	2%
6	digilib.uinsby.ac.id Internet Source	2%
7	journal.institutpendidikan.ac.id Internet Source	1%

8	repository.uin-suska.ac.id Internet Source	1 %
9	adoc.pub Internet Source	1 %
10	Mas Agus Firmansyah, Siti Karlinah, Suwandi Sumartias. "Kampanye Pilpres 2014 dalam Konstruksi Akun Twitter Pendukung Capres", Jurnal The Messenger, 2017 Publication	1 %
11	repository.uinjkt.ac.id Internet Source	1 %
12	123dok.com Internet Source	1 %
13	docplayer.info Internet Source	1 %
14	Elly Indrayuni. "Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes", Jurnal Khatulistiwa Informatika, 2019 Publication	1 %
15	zombiedoc.com Internet Source	1 %
16	ojs.uniska-bjm.ac.id Internet Source	1 %
17	www.mdpi.com Internet Source	

1 %

18

repository.unmuhjember.ac.id

Internet Source

1 %

19

ejournal.unkhair.ac.id

Internet Source

1 %

20

repository.dinamika.ac.id

Internet Source

1 %

21

sista.polindra.ac.id

Internet Source

<1 %

22

Ahmad Fauzi, Muhammad Faittullah Akbar, Yudhi Ferdi Andri Asmawan. "Sentimen Analisis Berinternet Pada Media Sosial dengan Menggunakan Algoritma Bayes", Jurnal Informatika, 2019

Publication

<1 %

23

dspace.uui.ac.id

Internet Source

<1 %

24

iocscience.org

Internet Source

<1 %

25

libraryproceeding.telkomuniversity.ac.id

Internet Source

<1 %

26

jim.teknokrat.ac.id

Internet Source

<1 %

www.researchgate.net

27

Internet Source

<1 %

28

de.scribd.com

Internet Source

<1 %

29

id.123dok.com

Internet Source

<1 %

30

repository.ub.ac.id

Internet Source

<1 %

31

sipora.polije.ac.id

Internet Source

<1 %

32

doaj.org

Internet Source

<1 %

33

ejurnal.teknokrat.ac.id

Internet Source

<1 %

34

ijns.org

Internet Source

<1 %

35

repository.widyatama.ac.id

Internet Source

<1 %

36

Stefanus Nico Soenardjo, Gunawan Gunawan.
"Information Extraction Berbasis Rule Untuk
Soal Ujian", Journal of Intelligent System and
Computation, 2021

Publication

<1 %

37

citisee.amikompurwokerto.ac.id

Internet Source

<1 %

38	ejournal.unma.ac.id Internet Source	<1 %
39	ejurnal.tunasbangsa.ac.id Internet Source	<1 %
40	id.scribd.com Internet Source	<1 %
41	pdfs.semanticscholar.org Internet Source	<1 %
42	text-id.123dok.com Internet Source	<1 %
43	ojs.unikom.ac.id Internet Source	<1 %
44	Retno Sari. "Analisis Sentimen Review Restoran menggunakan Algoritma Naive Bayes berbasis Particle Swarm Optimization", Jurnal Informatika, 2019 Publication	<1 %
45	Vianti Widyasari, Arief Senja Fitrani. "Application of Data Mining with Classification Methods for Promotion of New Student Admissions at Muhammadiyah University of Sidoarjo Using Web-Based Naïve Bayes Algorithm", Procedia of Engineering and Life Science, 2021 Publication	<1 %

46

Internet Source

<1 %

47

jurnal.umk.ac.id

Internet Source

<1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On