

BAB II

TINJAUAN PUSTAKA

A. Hasil penelitian terdahulu

Penelitian terdahulu adalah upaya mencari perbandingan dan selanjutnya untuk menemukan inspirasi baru untuk penelitian selanjutnya. Penelitian ini berfokus pada analisis perbandingan terhadap pemilihan calon presiden 2024 dengan menggunakan metode *naive bayes classifier*. Pengujian dilakukan untuk mengetahui sentimen masyarakat terhadap pemilihan presiden 2024 pada media sosial twitter menggunakan metode *naive bayes classifier* dengan polaritas positif, negatif, dan netral.

Tabel 2. 1 Penelitian terdahulu

No	Judul artikel	Tempat penelitian	Metode	Hasil penelitian
1	Perbandingan Metode K-NN, Naive Bayes, Decision Tree untuk Analisis Sentimen Tweet Terkait Opini Terhadap PT PAL Indonesia (Pattihha & Hendry, 2022)	Universitas Kristen Satya Wacana, Salatiga, Indonesia	<i>Naive Bayes Classifier</i>	Berdasarkan hasil penelitian ini menghasilkan nilai akurasi dari setiap metode dengan standar deviasi dan juga nilai presisinya. Nilai akurasi suatu metode dapat dijadikan sebagai suatu tolak ukur untuk melihat kinerja dari metode tersebut karena nilai akurasi merupakan ukuran kedekatan hasil pengukuran dengan nilai sebenarnya.
2	Analisis Sentimen Twitter Debat Calon Presiden Indonesia Menggunakan Metode Fined-Grained Sentiment Analysis (Pratama et al., 2019)	Universitas Narotama Surabaya	Fined-Grained Sentiment Analysis	Berdasarkan penelitian yang dilakukan disimpulkan bahwa <i>tweet</i> , yang menyertakan #JokowiA minMenangDebat dengan sentiment positif berjumlah 8.769 <i>tweet</i> , <i>tweet</i> dengan sentiment negatif berjumlah 4.518, dan <i>tweet</i> yang cenderung netral 1.713.

3	Analisis Sentimen Pengguna Twitter Terhadap Kebijakan Pemberlakuan Pembatasan Sosial Berskala Besar (PSBB) Dengan Metode Naïve Bayes (Mubarok, 2021)	Universitas Siliwangi Tasikmalaya	<i>Naïve Bayes Classifier</i>	Berdasarkan hasil survei menghasilkan tiga kategori yaitu positif, negatif, dan netral. Hasil dari pengujian terhadap algoritma naïve bayes untuk data sentiment psbb menggunakan tools rapid miner menghasilkan rata-rata akurasi sebesar 58.21%.
4	Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier (Yulita et al., 2021)	Institut Teknologi Sumatra	<i>Naïve Bayes Classifier</i>	Dari hasil penelitian ini pendapat masyarakat Indonesia khususnya pengguna media sosial Twitter, rata-rata memberikan respon positif terkait adanya kebijakan vaksinasi Covid-19 di Indonesia dengan presentase respon positif 60.3% (2278 respon). Penggunaan algoritma Naïve Bayes Classifier sudah sangat baik.
5	Analisis Sentimen Publik Terhadap Calon Presiden 2019 Melalui Twitter Menggunakan Metode Naïve Bayes Classifier (Panjaitan & Lhaksana, 2019)	Universitas Telkom Bandung	<i>Naïve Bayes Classifier</i>	Dari hasil penelitian ini menghasilkan klasifikasi sentiment terhadap kandidat calon Presiden pada Pilpres 2019 berdasarkan hasil dari data yang diujikan adalah sentiment yang bersifat positif.

6.	Analisis Sentimen Isu Pilpres 2019 Dengan Menggunakan Algoritma Naïve Bayes (Agusti et al., 2021)	Universitas AMIKOM Yogyakarta	<i>Naïve Bayes Classifier</i>	Hasil penelitian menunjukkan dari pengamatan polaritas sentiment masyarakat terhadap rangkaian proses pemilu, dan terhadap capres dan cawapres adalah sebesar kurang lebih 80,8% menunjukkan sentiment positif 19,2% menunjukkan sentiment negatif. Dengan menggunakan metode naïve bayes mendapatkan hasil yang cukup baik.
7.	Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes (Merinda Lestandy et al., 2021)	<i>University Muhammadiyah Malang</i>	<i>Naïve Bayes</i>	Dari hasil penelitian ini menghasilkan perbandingan kinerja beberapa metode RNN dan Naïve Bayes menggunakan teknik pembobotan TF-IDF. Data pada penelitian ini diperoleh dari 5000 <i>tweet</i> vaksin. Dengan pembagian 3800 <i>tweet</i> positif, 800 <i>tweet</i> negatif dan 400 <i>tweet</i> netral.
8.	Menggabungkan Analisis Suka-Retweet dan Naif Pengklasifikasi Bayes dalam Twitter untuk Sentimen Analisis	<i>University Brawijaya Malang</i>	<i>Naïve Bayes Classifier</i>	Hasil penelitian ini menunjukkan bahwa yang diusulkan menggabungkan metode dapat mencapai ukuran F1 0,838, yang mengatasi nilai textual mining sebesar 0,436. Hasil menunjukkan kombinasi terbaik tekstual dan non-tekstual untuk memprediksi sentiment Twitter masing-masing adalah 0,6 dan 0,4 yang menunjukkan ciri tekstual lebih penting dari pada nontekstual.

B. Landasan Teori

1) Analisis Sentimen

Analisis sentimen merupakan perpaduan dari *data mining* dan *text mining*, atau sebuah cara yang digunakan untuk mengolah berbagai opini yang diberikan oleh konsumen atau para pakar melalui berbagai media mengenai sebuah produk, jasa ataupun sebuah instansi. Analisis sentimen merupakan sebuah metode yang digunakan untuk memahami, mengekstrak data opini, dan mengolah data tekstual secara otomatis untuk mendapatkan sebuah sentimen yang terkandung dalam sebuah opini. Pada Analisis sentimen terdiri dari 3 jenis opini yaitu opini positif, opini negatif, dan opini netral, sehingga dengan *analisis sentiment* perusahaan atau instansi yang terkait dapat mengetahui respon masyarakat terhadap suatu pelayanan atau produk, melalui *feedback* masyarakat ataupun para ahli. (Sipayung et al., 2016)

Langkah-langkah yang bisa ditemukan di klasifikasi teks dari analisis sentiment adalah :

1) Tentukan data test domain

Kumpulan-kumpulan data yang mengelilingi domain, seperti dataset usulan, produk, suatu kejadian dan sebagainya

2) *Pre-processing*

Tahap processing awal, umumnya dilakukan dengan *tokenize*, *stopword removal* dan *stemming process*.

3) Transformasi

Proses representasi angka di hitung dari kata tekstual. Representasi biner biasanya digunakan dan hanya sekedar menghitung ada atau tidaknya kata dalam dokumen. Berapa kali sebuah kata muncul dalam dokumen juga digunakan adalah TF-IDF, *binary transformation* dan *frequency transformation*.

4) Klasifikasi

Proses klasifikasi umumnya menggunakan metode pengklasifikasian yaitu *naïve bayes*, *support vectore* dan sebagainya.

5) Evaluasi

Tahapan evaluasi biasanya menghitung *accuracy*, *recall*, *precision*.

2) Data Mining

Data mining merupakan pengolahan suatu pengetahuan yang informatif dan menarik yang didasarkan pada pola sebuah data besar. Data-data yang digunakan diperoleh dari database, *warehouse* data, web, yang dapat diolah menjadi sebuah informasi. Data mining dapat digunakan dalam berbagai sector dan tujuan untuk berbagai hal yaitu meningkatkan pengetahuan untuk beberapa sektor bisa digunakan dalam meningkatkan penjualan, dan lain sebagainya. Salah satu bagian data mining adalah klasifikasi. Klasifikasi adalah proses menemukan model melalui analisis terhadap sekumpulan data pelatihan yang menggambarkan dan membedakan kelas label atau konsep data. (Komputer, 2021)

3) Text Mining

Text mining merupakan salah satu Teknik yang dapat digunakan untuk melakukan klasifikasi pada dokumen dimana *text mining* merupakan variasi dari data mining yang berusaha menemukan pola menarik dari sekumpulan data tekstual yang berjumlah besar. Tujuan dari *text mining* adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen. Jadi, sumber data yang digunakan dalam *text mining* adalah sekumpulan text yang memiliki format yang tidak terstruktur, tugas utama dari *text mining* adalah pengklasifikasi teks. (Agusti et al., 2021)

4) Media Sosial

Media sosial adalah sebuah media online, dengan para penggunanya bisa dengan mudah berpartisipasi, berbagi sesama, dan menciptakan isi. meliputi contohnya blog, jaringan sosial, wiki, forum dan dunia visual. Blog, jaringan sosial dan wiki merupakan bentuk media sosial yang paling umum digunakan oleh masyarakat diseluruh dunia. Pendapat lain mengatakan bahwa media sosial adalah media online yang mendukung interaksi sosial menggunakan teknologi berbasis web yang mengubah komunikasi menjadi dialog interaktif. (Rafiq, 2015)

5) Twitter

Twitter adalah layanan komunikasi daring yang menyediakan teks yang singkat sehingga pengguna dapat menuliskan opininya secara jelas, singkat, dan padat serta dapat dipublikasi ke semua orang atau kelompok tertentu. Kelebihan twitter diantaranya adalah

komunikasi terjadi sangat cepat, publikasi yang luas, dan twitter membantu penyebaran informasi secara lebih cepat yang kemudian akan menjadi sebuah topik yang dibahas oleh para penggunanya. (Nurul et al., 2019)

- a. *Tweet*, fitur *tweet* ini adalah fitur utama di twitter. Twitter ialah kicauan, yaitu untuk mengirim dan melihat kicauan setiap pengguna twitter.
- b. *Following*, fitur *following* adalah fitur untuk mengikuti teman atau kerabat di Twitter, fitur ini merupakan salah satu fitur utama di twitter.
- c. *Followers*, fitur ini adalah fitur dimana kita bisa melihat siapa yang mengikuti kita di twitter.
- d. *Bio*, Fitur ini adalah fitur yang digunakan untuk mengetahui pesan akun twitter anda yang terdapat di profil.

6) *Preprocessing*

Preprocessing merupakan proses untuk mempersiapkan data mentah sebelum dilakukan proses lain. Pada umumnya, *preprocessing* data dilakukan dengan cara mengeliminasi data yang tidak sesuai atau mengubah data menjadi bentuk yang lebih mudah yang diproses oleh sistem. (Erfina, 2021)

7) *Naïve bayes classifier*

Naïve bayes classifier adalah metode yang sering digunakan dalam pengkategorian teks, dengan menggabungkan probabilitas kata-kata dan kategori untuk memprediksi probabilitas kategori dari sebuah dokumen. Selain itu *naïve bayes classifier* merupakan Teknik *machine learning* yang populer untuk pengklasifikasian teks, serta memiliki performa yang optimal pada banyak domain.

Teori ditulis dengan rumus sebagai berikut :

$$p(c|x) = \frac{p(x|c)p(c)}{p(x)}$$

X : Data dengan kelas yang belum diketahui

C : Hipotesis data merupakan suatu kelas spesifik

$p(c|x)$: Probabilitas hipotesis berdasarkan keadaan

$P(c)$: Probabilitas hipotesis

$P(x|c)$: Probabilitas berdasarkan keadaan pada hipotesis

(x) : Probabilitas C

8) *Confusion Matrix*

Menurut (Nugroho, 2018) *confusion Matrix* adalah tool berguna untuk menganalisis seberapa baik *classifier* mengenali kelas yang berbeda. *True Positive* (TP) dan *True Negative* (TN) menjelaskan ketika pengklasifikasi mendapatkan sesuatu dengan benar, sementara *False Positive* (FP) dan *false Negative* (FN) menjelaskan ketika pengklasifikasi mendapatkan hal yang salah. Dibawah ini menunjukkan tabel *Confusion Matrix*.

Tabel 2.2 *Confusion Matrix*

<i>Confusion Matrix</i>		<i>Predict Class</i>	
		Positif	Negatif
<i>Actual Class</i>	Positif	<i>TruePositif</i> (TP)	<i>FalseNegative</i> (FN)
	Negatif	<i>FalsePositife</i> (FP)	<i>TrueNegative</i> (TN)

Evaluasi dengan *Confusion Matrix* menghasilkan nilai *accuracy*, *precision*, dan *recall*. Dijelaskan sebagai berikut :

Accuracy merupakan jumlah dokumen yang diklasifikasi dengan benar, baik benar *Positif* maupun benar *Negatif*. Menghitung nilai *accuracy* dapat menggunakan persamaan.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

Precision menggambarkan jumlah data kategori positif yang diklasifikasi secara benar dibagidengan total data yang diklasifikasi positif. Dengan kata lain, presisi adalah klasifikasi *TruePositive* dan semua data dprediksi sebagai kelas positif. Perhitungan nilai *precision* dapat menggunakan persamaan.

$$precision = \frac{TP}{TP + FP} \times 100\%$$

Recall menunjukkan beberapa persen data kategori positif yang terklasifikasi dengan benar oleh sistem. *Recall* jumlah dokumen yang memiliki klasifikasi

TruePositive dari semua dokumen yang benar-benar positif (termasuk *FalseNegatif*)
Perhitungan nilai *accuracy* dapat menggunakan persamaan.

$$recall = \frac{TP}{TP + FN} \times 100\%$$

9) Evaluasi

Evaluasi performansi dilakukan untuk menguji hasil dari klasifikasi dengan mengukur nilai performansi dari sistem yang telah dibuat. Parameter pengujian yang digunakan untuk evaluasi yaitu akurasi yang perhitungannya dari table *confusion matrix* (matriks klasifikasi atau tabel kontigensi), Evaluasi dilakukan dengan menggunakan metode *confusion matrix*, adalah alat penting dalam metode visualisasi untuk menunjukkan performa *accuracy*, *precision*, dan *recall*.

