

BAB II KAJIAN PUSTAKA

A. Peneliti Terdahulu

Hasil Penelitian Terdahulu ditunjukkan pada tabel 2.1

Tabel 2.1 Peneliti Terdahulu

no	Nama	Judul	Metode	Hasil
1	(Yulita et al., n.d.)	Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier	Naïve bayes clasiffier	Hasil yang didapatkan dari penelitian ini dengan persentase data sebesar 60.3% dan jumlah data sebanyak 2278 data. respon negatif adalah 5.4% (203 data), dan respon netral adalah 34.4% (1299 data). Nilai akurasi yang dihasilkan sebesar 0,93 (93 %).
2	(Anjas Ramadhan & Budi Setiawan SSi, n.d.) (2022)	ANALISIS SENTIMEN PROGRAM ACARA DI SCTV PADA TWITTER MENGGUNAKAN METODE NAÏVE BAYES DAN SUPPORT	Naïve bayes clasiffier & svm	Mendapatkan hasil beita :79,81% Entertainment: 89,80% Sinetron :73,68% Ftv : 87,74%

		VECTOR MACHINE		
3	(Fajar et al., n.d.) (2018)	Implementasi Algoritma <i>Naive Bayes</i> Terhadap Analisis Sentimen Opini Film Pada Twitter	Naïve bayes clasiffier	Mendapatkan hasil precision 92% Recall 90% f-measure 90%
4	(Pratama et al., 2019)	Sentiment Analysis of the Indonesian Police Mobile Brigade Corps Based on Twitter Posts Using the SVM And NB Methods	SVM And NB	Best classification model was using the Support Vector Machine (SVM) method with 86.96% accuracy, 91.06% precision, 83.10% recall, and 0.945%. It is different from using the Naïve Bayes (NB) method whose results are 86.48% Accuracy, 88.86% precision, 84.68% recall, and 0.822.
5	(Garbian Nugroho et al., n.d.)	Analisis Sntimen Pada Jasa Ojek Online Menggunakan	Naïve Bayes Classidier	Akurasi yang dihasilkan sebesar 80% berdasarkan 800 data <i>tweet</i> yang terdiri atas 300 data

		Metode Naïve Bayes Clasifier		latih dan 500 data uji.
6	(Irawan et al., 2022)	Analisis Sentimen Terhadap Pengguna Gojek Menggunakan Metode K-nearst Neighbors.	Knearst Neighbros	Mendapatkan akurasi sebesar 79,43% pada k=15, dengan sentimen positif sebesar 47.59%, negatif 8.66%, dan netral 43.75%.
7	(Rozi et al., n.d.) (2020)	Analisis Sentimen Pada Twiter Mengenai Pasca Bencana Menggunakan Metode Naïve Bayes Dengan Fitur N-Gram	Naïve Bayes dan N-gram	hasil akurasi tertinggi pada <i>unigram</i> yaitu sebesar 93.33% dan <i>bigram</i> sebesar 86.67%.
8	Fiktor Imanuel Tanesa, Irawan Smbiring and Hindriyanto Dwi Purnomo (2017)	Sentiment Analysis Model Based On Youtube Comment Using Support Vector Machine	Suport Vector Machine	Value of accuracy 84%, precision 91%, recall 80%, TP rate 91.1% and TN rate 44.8%. TN rate is inversely proportional to TP rate, where TN rate is obtained percentage 44.8%, while TP rate 91.1%.

9	(Byna & Basit, 2020) (2020)	Penerapan etode Adabost Untuk Mengoptimasi Prediksi Penyakit	Naïve Bayes Clasifier Dan Adabost	Hasil penelitian untuk nilai akurasi algoritma Naïve Bayes memiliki nilai 0.976 dengan Split data 80/20, sedangkan untuk nilai akurasi optimasi adaboost dengan naïve bayes senilai 0.981 split data 70/30 kedua model tersebut memiliki diagnosa Excellent Classification.
10	(Yahyadi et al., 2022)	Analisis Sentimen Twitter Terhadap Kebijakan PPKM Di Tengah Pandemi Covid-19 Menggunakan Mode Lstm	Lstm	Hasil pengujian analisis sentimen tweet dengan menggunakan molde LSTM terhadap penerapan PPKM oleh pemerintah Indonesia didapatkan hasil 36,5 % merespos positif, 15,7% merespon netral dan 57.8% merespons negatif,

11	Jimmy ^{1*} , Eni Heni Hermaliani ² , Laela Kurniawati 2 mei 2023	ANALISIS KLASIFIKASI SENTIMEN PENGGUNA MEDIA SOSIAL TWITTER TERHADAP PENUNDAAN PEMILU PRESIDEN TAHUN 2024	<i>Cross-Industry Standard Process for Data Mining</i>	Hasil pengujian ini menghasilkan True Positif 182, True negatif 34, Class Precision 84.26
12	Fransiska Vina Sari, Arief Wibowo 2 november 2019	ANALISIS SENTIMEN PELANGGAN TOKO ONLINE JD.ID MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER BERBASIS KONVERSI IKON EMOSI	Naïve bayes clasifier	Hasil pengujian ini mendapatkan hasil accuracy 98% , precision 94,64%, recall 100%

B. Landasan teori

1. Analisis sentiment

Analisis sentimen adalah bentuk teknologi baru yang saat ini sedang dikembangkan dan digunakan sebagai alat penelitian untuk menganalisis topik dan opini. Teknik ini sering digunakan untuk menilai dan menganalisis tingkat kepuasan pelanggan atau pengguna terhadap produk atau praktik tertentu (Yahyadi et al., 2022).

Dan menurut (Nurjanah et al., 2017). Analisis sentimen dapat digunakan untuk mengetahui apakah opini atau komentar terkait sedang trend positif atau negatif dan dapat menjadi acuan untuk meningkatkan pelayanan atau meningkatkan kualitas produk.

Tujuan dari analisis sentimen adalah untuk menentukan perilaku atau opini penulis terhadap suatu topik tertentu. Perilaku dapat mendukung pemikiran atau penilaian dan kecenderungan dalam keadaan, analisis sentimen juga dapat mengungkapkan perasaan sedih, gembira, dll.

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

- 1) Tentukan datatest domain
Kumpulan-kumpulan data yang mengelilingi domain, seperti dataset usualn ,produk, suatu kejadian dan sebagainya
- 2) Pre-procesing
Tahap processing awal, umumnya dilakukan dengan *cleaner*, *lowercase*, normalisasi *slang*, menghapus *stopword* dan *stemming* process.
- 3) Transformasi
Proses representasi angka dihitung dari kata tekstual. Representasi biner biasanya di gunakan dan hanya sekedar menghitung ada atau tidaknya kata dalam dokumen. Berapa kali sebuah kata muncul dalam dokumen juga di gunakan adalah *Countvectorizer*.
- 4) Klasifikasi
Proses klasifikasi umumnya menggunakan metode pengklasifikasian yaitu *Naïve Bayes*
- 5) Evaluasi
Tahapan evaluasi biasanya menghitung akurasi dengan melihat *Confusion Matrix*

2. Data Mining

Data mining adalah pemrosesan data yang informatif dan menarik berdasarkan model data besar. Data yang digunakan dapat diperoleh dari database, data warehouse, website, dll. dan diolah menjadi informasi. Data mining dapat digunakan di berbagai industri dan bertujuan untuk banyak hal, yaitu menambah pengetahuan, meningkatkan penjualan di beberapa industri,

dll. Salah satu sub bidang dari data mining adalah klasifikasi. Dalam klasifikasi, menganalisis data pelatihan, ditemukan pola yang menggambarkan dan memisahkan kelas label atau konsep pengetahuan (Yoga Pratama et al., 2021).

3. Text Mining

Text mining adalah proses pengelolaan informasi dalam bentuk teks yang berfungsi sebagai template untuk penggalian informasi penting. Data yang ditangani dalam text mining adalah data teks. Sebelum text mining, ada tahap text preprocessing. Langkah-langkah preprocessing teks adalah text *preprocessing* yaitu *case folding*, *cleansing*, *tokenizing*, *stemming*, dan *stopword removal* (Yoga Pratama et al., 2021).

4. Media sosial

Media sosial adalah media online di mana pengguna dapat dengan mudah berpartisipasi, berbagi, dan membuat konten, seperti blog, jejaring sosial, wiki, forum, dan dunia virtual. Blog, jejaring sosial, dan wiki adalah bentuk media yang paling umum digunakan oleh orang-orang di seluruh dunia. Media sosial juga dapat diartikan sebagai media online yang mendukung interaksi sosial dan media sosial dengan teknologi berbasis web yang mengubah komunikasi menjadi dialog interaktif.

5. Twitter

Twitter adalah layanan komunitas online dan layanan microblogging yang memungkinkan pengguna mengirim dan membaca pesan teks hingga 140 karakter, yang disebut tweet. Twitter didirikan oleh Jack Dorsey pada Maret 2006 dan situs jejaring sosial diluncurkan pada Juli. Sejak awal, Twitter telah menjadi salah satu dari sepuluh halaman yang paling banyak dikunjungi di internet (Maksun et al., 2021)

Fitur yang terdapat pada twiter adalah sebagai berikut :

- 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 anda di *twitter*.
- d. *Bio*, Fitur ini adalah fitur yang di gunakan untuk engetahui pesan akun *twitter* anda yang terdapat di profil.
- e. *Prfile*, merupakan fitur utama dari *Twitter* yang berfungsi untuk melihat avatar *twitter*, bio *twitter* dan lainnya.

6. Processing

Pra-pemrosesan adalah proses di mana data disiapkan sebelum diproses lebih lanjut. Data awal yang diperoleh biasanya masih mengandung bagian-bagian yang tidak diperlukan dalam analisis sentimen, sehingga perlu dilakukan pre-processing, terutama data media sosial yang banyak mengandung kata-kata informal dan tidak terstruktur yang menimbulkan banyak kegaduhan.

7. Naïve bayes classifier

Pengklasifikasi Naive Bayes adalah algoritma yang biasa digunakan dalam klasifikasi teks yang menggabungkan probabilitas kata dan kategori untuk memprediksi probabilitas kelas suatu dokumen. Selain itu, Naive Bayes adalah teknik pembelajaran mesin yang populer untuk klasifikasi teks dan menunjukkan performa optimal di banyak area.

Teori bayes di tulis dengan rumus sebagaiberikut :

1. Teori Bayes

Rumus Teori Bayes ditunjukkan pada Persamaan 2.1

$$p(c|x) = \frac{p(x|c)p(c)}{p(x)} \quad (2.1)$$

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

P(x) : Probabilitas C

8. Confusion Matrix

Menurut (Nugroho, 2018), Confusion Matrix adalah alat yang berguna untuk menganalisis seberapa baik classifier mengenali kelas yang berbeda. True positive (TP) dan true negative (TN) menjelaskan kapan classifier melakukan sesuatu yang benar, sedangkan false positive (FP) dan false negative (FN) menjelaskan saat classifier melakukan sesuatu yang salah. Di bawah ini adalah Tabel Confusion Matrixs. Rumus Confusion Matrix ditunjukkan pada Tabel 2.2.

Tabel 2.2 Confusion Matrix

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

Evaluasi dengan *Confusion Matrix* menghasilkan nilai *accuracy*, *precision*, dan *recall*. Berdasarkan (Syarifuddin, 2020) dijelaskan sebagai berikut:

Nilai persisnya adalah jumlah dokumen yang diklasifikasikan dengan benar, baik TruePositive maupun TrueNegative. Nilai ekstrak dapat dihitung dengan menggunakan persamaan. Rumus Accuracy ditunjukkan pada Persamaan 2.2.

Accuracy

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2.2)$$

Nilai akurasi menggambarkan jumlah kategori data positif yang diklasifikasikan dengan benar dibagi dengan jumlah total data yang diklasifikasikan positif. Dengan kata lain, akurasinya adalah klasifikasi TruePositive dan semua data diprediksi sebagai kelas positif. Dengan persamaan tersebut nilai presisi dapat dihitung. Rumus Precision ditunjukkan pada Persamaan 2.3.

Precision

$$precision = \frac{TP}{(TP + FP)} \times 100\% \quad (2.3)$$

Nilai *recall* menunjukkan beberapa persen data kategori positif yang terklasifikasi dengan benar oleh sistem. Dengan kata lain, *recall* adalah jumlah dokumen yang memiliki klasifikasi *TruePositive* dari semua dokumen yang benar-benar positif (termasuk *FalseNegatif*) Perhitungan nilai *Recall* ditunjukkan pada Persamaan 2.4

Recall

$$recall = \frac{TP}{(TP + FN)} \times 100\% \quad (2.4)$$

1. Evaluasi

Evaluasi performasi dilakukan untuk menguji hasil klasifikasi dengan mengukur nilai performasi dari sistem yang dibuat. Akurasi yang dihitung dari tabel matriks kebingungan (matriks klasifikasi atau tabel kontingensi) digunakan sebagai parameter uji dalam evaluasi.

