BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Adapun beberapa penelitian sebelumnya yang memuat masalah sejenis dengan metode analisis yang sama dengan masalah penelitian yang sedang diteliti, yaitu :

Tabel 2. 1 Perbandingan dengan Penelitian Sebelumnya.

No.	Peneliti	Topik	Objek	Metode	Hasil
1.	Murni, Imam Riadi dan Abdul Fadlil (2023)	Analisis Sentimen *HateSpeech* pada Pengguna Layanan *Twitter* dengan Metode *Naïve Bayes Classifier* (NBC)	*HateSpeech* (Ujaran Kebencial)	*Naïve Bayes Classifier*	Berdasarkan evaluasi menggunakan Confusion Matrix diperoleh akurasi tertinggi sebesar 80%, precision 100%, recall 80%, dan F1-Score 89% pada skenario pengujian data training 70% dan testing 30%.
2.	Nurirwan Saputra, Karandi Nurbagja, dan Turiyan (2022)	Sentimen *Analysis Of Presidential Candidates* Anies Baswedan and Ganjar Pranowo *Using Naïve Bayes Method*	Anies Baswedan dan Ganjar Pranowo	*Naïve Bayes*	Dilihat dari data yang dihasilkan 49% komentar negatif, 35% komentar positif dan 16% netral.
3.	Aryo dewandaru dan jati Sasongko	Analisis Sentimen dan Klasifikasi *Tweet* Terkait Mutasi COVID-	*mutase COVID-19*	*Naïve Bayes Classifier*	Berdasarkan dari hasil penelitian, dapat disimpulkan bahwa *Naïve*

	Wibowo (2022)	19 Menggunakan Metode *Naïve Bayes Classifier*			*Bayes* Metode *classifier* dapat digunakan untuk menganalisis data sentimen dari *tweet* tentang mutasi *COVID-19* dengan akurasi 86,67%.
4.	Dianati Duei Putri, Gigih Forda Nama dan Wahyu Eko Sulistiono (2022)	Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) pada *Twitter* menggunakan Metode *Naïve Bayes Classifier*	Kinerja Dewan Perwakilan Rakyat (DPR)	*Naïve Bayes Classifier*	Hasil penelitian menunjukan bahwa DPR memiliki 95 *tweet* positif 75%, 693 *tweet* netral 79%, dan 758 *tweet* negatif 82%). Akurasi pengujian sekitar 80% dari data uji 20%
5.	Ariesta Damayanti, Helda Ludya Safitri dan Rudy Cahyadi (2022)	Analisis Sentimen Tindakan Pemerintah Indonesia Dalam Penanganan Covid-19 Menggunakan Metode Support Vector Machine	Covid-19	Support Vector Machine	Hasil dari penelitian ini mencapai akurasi pelatihan 77% untuk klasifikasi sentimen positif, negatif, dan netral, dengan mayoritas *tweet* memiliki sentimen negatif.
6.	Dini Fakta Sari, Deborah Kurniawati, Edy Prayitno, dan Irfangi (2019)	Sentimen Analysis of Twitter Social Media to Online Transportation in Indonesia Using Naïve Bayes Classifier	Transportation Online in Indonesia	Naïve Bayes Classifier	Hasil ujian menunjukkan akurasi 84%, dengan 11% positif, 14% negatif, dan 75 netral.
7.	Lia Durrotul Mahbubah dan Eri Zuliarso (2019)	Analisis Sentimen *Twitter* pada PILPRES 2019 menggunakan	PILPRES	*Naïve Bayes*	Hasil penelitian ini menunjukan bahwa algoritma naïve bayes classifier memberikan

		Algoritma *Naïve Bayes*			akurasi 73%, dengan precision 78% untuk kelas negaif dan 66% untuk kelas positif.
8.	Rosit Sanusi (2021)	Analisis Sentimen Pada Twitter Terhadap Program Kartu Prakerja Menggunakan *Long Short Term Memory*	Program Kartu Prakerja	*Recurrent Neural Network* dengan L*ong Short Term Memory*	Hasil proses pelatihan model didapatkan tingkat akurasi 95.66% dalam epoch dengan loss 0.1999. namun, pada pengujian, akurasi masih rendah, yaitu 64.48% dengan loss 1.5922, disebebkan oleh tidakseimbangan data sampel, dengan 2460 netral, 689 positif, dan 973 negatif dari total 4122 dataset *Twitter*
9.	Usulan (2023)	Analisis Sentimen pada Media Sosial Twitter terhadap Ganjar Pranowo Sebagai Calon Presiden 2024 menggunakan Metode Naïve Bayes Classifier	Ganjar Pranowo	*Naïve Bayes Classifier*	Mencari sentimen dari setiap *tweet* terhadap Ganjar Pranowo sebagai Calon Presiden 2024

Penelitian sentimen tentang *hatespech* (Murni, Riadi and Fadlil, 2023) pada media sosial Twitter dilakukan dengan tahapan *preprocessing (case folding, tokenization, stopword removal, normalization,* dan *stemming), labelling*, *Term Weighting* pembobotan dengan *Term Frequency (TF),* dan *Inverse Document*

Frequency (IDF). Teknik K-Flod Cross Validation dilakukan untuk memvalidasi data dengan membagi data *training* dan *testing* ke dalam 3 skenario pengujian yaitu, data *training* 70% dan *testing* 30%, data *training* 30% dan *testing* 70%, dan data *training* 50% dan *testing* 50% terhadap Metode *Naïve Bayes Classifier*. Berdasarkan evaluasi menggunakan *Confusion Matrix* diperoleh tertinggi akurasi tertinggi sebesar 80%, *precision* 100%, *reccal* 80%, dan *F1-Score* 89% pada skenario pengujian data *training* dan *testing* 30%.

Penelitian *sentimen* tentang pemilihan presiden mendatang (Saputra, Nurbagja and Turiyan, 2022) pada media *sosial* Facebook dengan pengambilan data melalui proses *scraping* yang kemudian dibersihkan atau diberisihkan, maka diberi lima label yaitu: 1 (sangat negatif) , 2 (negatif), 3 (netral), 4 (positif), dan 5 (sangat positif). Tujuannya untuk melihat yang mana sentimen tertingi yang diberikan warganet terhadap pemilihan presiden mendatang. Penelitian ini menyimpulkan bahwa *netizen* memiliki sentimen negatif terhadap tokoh dalam pemilihan presiden mendatang. Dilihat dari data yang dihasilkan secara acak 49% komentar negatif, 35% komentar positif dan 16% netral. Selain itu, dari data 510 diambil dengan klasifikasi menggunakan metode *Naïve Bayes*, serta dilakukan pengujian menggunakan metode validasi silang 10 kali lipat dengan tokenisasi *Quadgram* dengan akurasi 42,75%, presisi 42,10%, dan recall 42,70%.

Penelitian sentimen tentang COVID-19 (Dewandaru and Wibowo, 2022) pada media sosial *twitter* penelitian ini menggunakan metode N*aïve Bayes Classifier*. Metode ini dapat mengklasifikasikan data atau opini menjadi dua sentimen, yaitu positif dan negatif. Data diambil menggunakan *API Twitter* dengan

kata kunci "mutasi *covid*", untuk pengolahan data ada beberapa proses dilakukan, yaitu klasifikasi sentimen, pembersihan data, dan *preprocessing* sehingga diperoleh hasil akhir. Hasil tes dari penelitian ini menunjukkan bahwa metode *Naïve Bayes Classifier* memiliki akurasi 86,67% dengan skor f1 82,00% positif sentimen dan 89,00% pada sentimen negatif. Berdasarkan dari hasil penelitian, dapat disimpulkan bahwa *Naïve Bayes* Metode *classifier* dapat digunakan untuk menganalisis data sentimen dari *tweet* tentang mutasi *COVID-19* dengan akurasi 86,67%.

Penelitian sentimen tentang Kinerja Dewan Perwakilan Rakyat (DPR) (Putri, Nama and Sulistiono, 2022) pada media sosial *twitter* dilakukan dengan tahapan pengumpulan data (*Crawling*), *preporcessing data* yang terdiri dari proses *cleaning data, tokenization, stop remova* dan *case folding, spitting data* dan klasifikasi data menggunakan metode *Naïve Bayes Classifier*. Penelitian ini menggunakan sebanyak 1546 data *tweet*. Hasil dari penelitian ini didapatkan bahwa DPR mendapatkan 95 *tweet* positif dengan polaritas 0.75 atau 75% sentimen positif, 693 *tweet* netral dengan polaritas 0.79 atau 79% sentimen netral dan 758 *tweet* negatif dengan polaritas 0.82 atau 82% sentimen negatif dengan *accuracy score* 0.8 atau 80% berdasarkan data *testing* sebanyak 20%.

Penelitian sentimen tentang Penanganan Covid-19 (Ariesta Damayanti, Helda Ludya Safitri, and Rudi Cahyadi, 2022) penelitian ini dilakukan dengan menganalisis sentimen menggunakan metode Support Vector Machine (SVM) dengan Kernel Radial Basis Function (RBF). *Tweet* akan diklasifikasi menjadi sentimen positif, negatif, dan netral, sehingga dapat diketahui seberapa banyak persentase dari masing-masing kategori opini. Penelitian menggunakan data

sebanyak 600 *tweet* yang diperoleh dari hasil scraping menggunakan twitterscraper. Hasil dari penelitian ini adalah tingkat akurasi pelatihan sebesar 77% dalam melakukan klasifikasi sentimen positif, negatif, dan netral. Dari hasil klasifikasi data, diperoleh sebagai besar *tweet* terdiri dari sentimen negatif.

Penelitian sentimen tentang transportasi online di Indonesia (Dini Fakta Sari, Deborah Kurniawati, Edy Prayitno, and Irfangi, 2019) Metode yang digunakan adalah metode klasifikasi naïve bayes classifier untuk menganalisis sentimen media sosial Twitter terhadap trasportasi online di Indonesia. Studi ini dilakukan dengan mengolah 1009 data, yang terdiri dari 900 data pelatihan dan 109 data uji. Hasil pengujian, dengan akurasi sebesar 84%, menunjukkan 11% nilai positif, 14% nilai negatif, dan sisanya 75% adalah nilai netral.

Penelitian sentimen tentang PILPRES (Mahbubah and Zuliarso, 2019) pada media sosial *twitter* dilakukan dengan tahapan pengambilan tweets dari twitter dengan kata kunci pencarian #pilpres2019 dan #prabowo untuk diolah dan mengklasifikasikan teks dengan menggunakan metode analisis sentimen. Untuk proses klasifikasi teks dibagi menjadi dua kelas yaitu kelas sentimen positif dan kelas sentimen negatif. Data yang digunakan berjumlah 300 *tweets* yang terdiri dari 240 data latih dan 60 data uji. Data untuk pelatihan sudah diketahui sentimennya sedangkan data untuk pengujian belum diketahui nilai sentimennya. Dari 240 data latih terdiri dari 134 sentimen negatif dan 106 sentimen positif. Pada studi ini menunjukkan bahwa klasifikasi data tweets menggunakan algoritma *naive bayes classifier* memberikan akurasi sebesar 73%. Precision kelas negatif sebesar 78% dan *precision* kelas positif sebesar 66%.

Penelitian sentimen tentang Program Kartu Prakerja (Rosit Sanusi, 2021) pada media sosial *Twitter* penelitian ini menggunakan metode Long Short Term Memory. Metode ini gunakan untuk melakukan analisis sentimen pada *tweet* dengan topik program kartu prakerja. Hasil proses pelatihan model didapatkan tingkat akurasi sebesar 0.9566 atau 95,66% dengan tingkat loss 0.1999, hasil ini didapat dari 36 epoch, dan untuk pengujian /testing tingkat akurasi dapat dibilang masih belum memuaskan dimana akurasi yang didapat sebesar 0.6448 atau 64,48% dengan loss sebesar 1.5922. untuk proporsi data sampel yang digunakan dalam penelitian juga kurang seimbang dimana dari 4122 dataset *twitter* sebanyak 2460 diantaranya masuk label netral, 689 masuk klasifikasi positif, dan 973 sisanya masuk klasifikasi negatif.

Penelitian ini, menggunakan algoritma *Naïve Bayes Classifier* yang memiliki tingkat kepercayaan dan akurasi yang optimal dibandingkan dengan metode *classifier* lainnya. Penelitian ini bertujuan untuk mengklasifikasikan frasa atau sentimen terhadap Ganjar Pranowo sebagai calon presiden potensial tahun 2024 ke dalam tiga kategori, yaitu positif, negatif, dan netral, menggunakan metode *Naïve Bayes Classifier* untuk mengimplementasikan sentimen terhadap topik Ganjar Pranowo sebagai calon presiden 2024 di plaform media sosial *Twitter*.

2.2 Dasar Teori

2.2.1 Twitter

Twitter adalah platform media sosial yang memungkinkan pengguna untuk berbagi pesan singkat yang disebut "*tweet*". Dalam konteks analisis sentimen, *Twitter* dapat menjadi sumber data yang berharga untuk memahami opini dan sentimen publik terhadap berbagai topik, salah satu media yang banyak digunakan adalah *Twitter*. *Twitter* merupakan media social untuk bertukar pikiran dan pendapat. Pengguna *twitter* dapat mengirim dan menerima pesan *tweet* berupa teks, gambar, ataupun video. Perbedaan dengan media social lain, di *twetter* karakter untuk menulis pesan dibatasi sampai 280 karakter, sedangkan media social lainnya tidak dibatasi. *Twitter* bersifat public sehingga status yang dibagikan dapat dilihat oleh orang lain meskipun bukan pengikutnya. Namun, pengirim *tweet* juga dapat dibagikan hanya kepada temannya saya atau *followers*. *Twitter* mempunyai kelebihan yaitu jangkauan yang luas, dapat menjangkau public figure, media promosi lebih luas, banyak jaringan, dan lebih mudah diukur kemampuannya. Berikut yang ada pada *twitter* (I. Taufik dan S.A.Pamungkas, 2018):

1. *Tranding topic* adalah fitur yang menampilkan topik atau pembahasan teratas berupa *hastag* yang banyak dibicarakan pengguna *twitter*.
2. *Hastag* adalah fitur yang dapat mengelompokkan *tweet* atau pesan.
3. *Retweet* adalah fitur untuk membagikan *tweet* dari pengguna lain.
4. *Following* adalah fitur untuk menghubungkan antara pengguna atau sering disebut teman.

2.2.2 Analisis Sentimen

Analisis sentimen atau *opinion mining* merupakan proses dalam mengolah data dan mengekstrak data secara otomatis pada kalimat opini. Analisis sentimen bertujuan untuk mempublikasikan suatu ide penelitian, yang akan mempresentasikan sebuah analisis terkait dengan menklasifikasian sebuah data berupa teks dengan menggunaka metode *Naïve Bayes Classifier*. Analisis sentimen mengelompokkan kalimat atau pendapat berdasarkan popularitas teks didalamnya. Popularitas tersebut merupakan pendapat yang memiliki aspek positif yang berkaitan dengan hal baik atau keadaan yang normal dan aspek negatif yang berkaitan dengan hal buruk atau keadaan yang tidak diinginkan. Analisis sentimen pada *Twitter* menjadi suatu alat yang dapat menganalisis dari persepsi masyarakat. Pendekatan dengan analisis pada *Twitter* ini berfokus pada indetifikasi sentimen *tweet* setiap individu (Arfina Shella Meilany, 2022).

2.2.3 Ganjar Pranowo

H. Ganjar Pranowo, S.H, M.IP lahir di Karanganyar, 28 Oktober 1968, Jawa Tengah. Ganjar Pranowo merupakan seorang Gubernur Jawa Tengah yang dikenal sikap kepemimpinannya cerdik dan tegas. Ganjar salah satu tamatan Universitas Gajah Mada tepatnya Falkultas Hukum dan Pascasarjana Ilmu Politik dari Universitas Indonesia. Profesi seorang Ganjar dalam bidang politik bermula pada tahun 2009 sampai saat ini. Rekam jejaknya cukup memuaskan sehingga banyak disenangi oleh Masyarakat karena beliau merupakan salah satu pejabat

pemerintah yang merakyat. Pengalaman seorang Ganjar dalam dunia politik bisa dilihat sudah sangat matang (Ritonga *et al.*, 2023).

2.2.4 *Python*

Python merupakan bahasa pemrograman yang dibuat oleh Guido van Rossum yang berorientasi pada objek dengan tingkat tinggi. Fitur di dalam python adalah mendukung dalam pemrograman yang berorientasi objek, imperative, dan fungsional. Berikut ini merupakan beberapa *library* pada python sebagai berikut (Arfina Shella Meilany, 2022).

1) *Pandas*

   Pandas merupakan *library* yang digunakan untuk menganalisis dan mengolah data terstruktur.

2) *Snscrape*

   Snscrape merupakan program yang digunakan untuk menyaring layanan sosial media yang peruntukan untuk mencari hal-hal yang diperlukan seperti profil pengguna, tagar, atau penvarian.

3) *Numerical Python*

   Numerical Python merupaka *library* yang digunakan dalam proses komputasi dalam operasi vector dan maktriks.

4) CSV

   Comma Separated Value (CSV) merupakan *library* yang digunakan dalam menyimpan data dengan format csv.

2.2.5 *Google Colab*

Google Colab adalah platform pengembangan berbasis cloud yang memungkinkan pengguna untuk melakukan pemrograman dan analisis data menggunakan Python (Diartono et al., 2022). Google Colab dilengkapi dengan berbagai Pustaka Python, Matplotlib, dan Plotly yang dapat digunakan untuk memproses data dan membuat visualisasi data. Google Colab memudahkan pengguna dalam menghasilkan visualisasi data karena tidak perlu melalukan instalasi perangkat lunak pada komputer local.

2.2.6 Klasifikasi *Naïve Bayes Classifier*

Bayesian classification berdasarkan pada teorima bayes yang memiliki kemampuan hampir serupa dengan *decision tree* dan *neural network*. Teorema Bayes adalah teorema yang digunakan dalam statistika untuk menghitung peluang suatu hipotesis. Bayes merupakan teknis prediksi berbasis *probabilistic* sederhana yang berdasarkan pada penerapan teorema Bayes (aturan Bayes) dengan asumsi independensi (ketidaktergantungan) yang kuat atau naif (Febriant, 2017). Dengan kata lain, dalam Naïve Bayes, model yang digunakan adalah "model fitur independent"

Dalam Bayes (terutama Naïve Bayes), maksud independensi yang kuat pada fitur adalah bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama.

Dalam klasifikasi menggunakan *Naïve Bayes* dibagi menjadi 2 proses, yaitu proses training dan testing. Proses training digunakan untuk menghasilkan model analisis sentimen yang nantinya akan digunakan sebagai acuan untuk mengklasifikasikan sentimen dengan data testing atau data mentah yang baru (Yuna Sophia Dewi Febriant, 2017). Adapun persamaan dari Teorema Bayes dapat dilihat dibawah ini.

$$P(A\backslash B) = \frac{P(B\backslash A)P(A)}{P(B)}$$

B = Data dengan kelas yang belum diketahui

A = Hipotesis data B merupakan kelas spesifik

P(A|B) = Probabilitas hipotesis A berdasarkan kondisi B (*posterior probability*)

P(A) = Probabilitas Hipotesis A (*prior probability*)

P(B|A) = Probabilitas B berdasarkan kondisi hipotesis A

P(B) = probabilitas B

2.2.7 *Confusion Matrix*

Confusion matrix adalah suatu metode yang umumnya digunakan untuk melakukan perhitungan tingkat akurasi pada data mining. *Confusion matrix* memuat informasi tentang klasifikasi yang diprediksi dengan benar oleh sebuah sistem klasifikasi. Terdapat tiga parameter yang akan dihitung, yaitu *accuracy*,

recall, dan *precision*. Contoh *confusion matrix* untuk klasifikasi biner ditunjukkan pada Tabel 2.2.

		Kelas Prediksi	
		1	2
Kelas Sebenarnya	1	TP	FN
	0	FP	TN

Tabel 2. 2 Klasifikasi Biner

Keterangan :

TP (*True positive*) = Jumlah dokumen dari kelas 1 ya/ng benar diklasifikasikan sebagai kelas 1

TP (*True Negatif*) = Jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0

FP (*False Positive*) = Jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1

FN (*False Negatif*) = Jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0

Rumus *confusion matrix* untuk menghitung *accuracy*, *precision*, dan *recall* seperti berikut.

a. Akurasi

Akurasi adalah metode yang didasari tingkat kedekatan antara nilai prediksi dengan nilai sebenarnya. Akurasi adalah hasil dari penjumlahan nilai

diagonal dibagi dengan jumlah total keseluruhan data dan selanjutnya dikalikan 100%.

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

b. Presicion

Presisi adalah metode yang dipakai untuk menghitung nilai proporsi kelas positif yang diberhasil diprediksi dengan benar dari keseluruhan hasil prediksi kelas positif. Presisi menunjukkan jumlah data kategori positif yang diklasifikasi secara benar dibagi dengan total data yang diklasifikasi positif.

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

c. Recall

Recall adalah metode yang digunakan untuk menghitung presentase kelas data positif yag berhasil diprediksi benar dari keseluruhan data kelas positif.

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

d. F1-score

F1-score adalah rata-rata harmonic dari presisi dan recall. F1-score juga biasa disebut F1-measure.

$$F1 - Score = 2\,X\,\frac{Presisis\ \ x\ Recall}{Presisi + Recall}$$