BAB II
TINJAUAN PUSTAKA

A. Penelitian Terdahulu

Beberapa penelitian serupa yang telah dilakukan oleh beberapa peneliti yaitu sebagai berikut:

Tabel 2.1 Penelitian Terdahulu

No	Peneliti	Metode	Hasil
1	(Zhafira et al., 2021)	*Naïve Bayes Classifier* dengan pembobotan TF-IDF dan validasi data *k-fold cross validation*	Rata-rata akurasi yang didapat dari 10 iterasi *k-fold cross validation* yaitu sebesar 91.8% dengan nilai *Precision, Recall, f-measure* sebesar 90.35%, 93.6%, 91.95%.
2	(Yulita et al., 2021)	*Naïve Bayes Classifier* dengan pembobotan TF-IDF	Masyarakat indonesia rata-rata memberikan respon positif 60.3%, respons negatif 5.4%, dan respons netral 34.4%. Nilai akurasi yang dihasilkan sebesar 93%.
3	(Ratnawati, 2018)	*Naïve Bayes Classifier* dan *k-fold cross validation*	Akurasi tertinggi didapatkan pada *fold* kedua yaitu 90%, *Precision* 92%, *Recall* 90% dan *F-measure* 90%.
4	(Ruhyana, 2019)	*Naïve Bayes Classifier*	Hasil klasifikasi teks dalam bentuk positif dan negatif untuk penerapan lalu lintas ganjil genap dalam penelitian ini menghasilkan

Analisis Sentimen Masyarakat..., Hafidz Maulana An Nahar, Fakultas Teknik dan Sains, UMP 2023

No	Peneliti	Metode	Hasil
			Accuracy 86.67%, *Precision* 71.43% dan *Recall* 80%.
5	(Arsi et al., 2021)	*Naïve Bayes Classifier* dan *k-fold cross validation*	Hasil penelitian yang didapat dari dataset sejumlah 1007 data adalah nilai *Accuracy* sebesar 94.33%, *Precision 87%*, *Recall* 99% dan *F1-Score 92%* yang berarti sistem ini sudah baik dalam mendeteksi sentimen.
6	(Kurniawan & Susanto, 2019)	*K-Means* dan *Naïve Bayes Classifier*	Dari hasil pengujian 100 dan 150 data uji didapatkan akurasi rata-rata 93.35% dan *error rate* rata-rata 6.66%.
7	(Sundara et al., 2020)	*Naïve Bayes Classifier* dengan pembobotan menggunakan TF-IDF	Pembagian data menjadi data latih dan uji dengan perbandingan 8:2 yang masing-masing sebanyak 440 data latih dan 110 data uji. Pengklasifikasian data tersebut menjadi kelas positi dan negatif sebanyak 219 dan 331. Hasil akurasi dari metode ini dan dikolaborasikan dengan TF-IDF sebesar 86%.
8	(Mahendrajaya et al., 2019)	*Lexicon Based* dan *Support Vector Machine*	Proses pelabelan data dibagi menjadi dua kelas yaitu kelas positif dan negatif dengan menggunakan *Lexicon Based* sebanyak 923 data positif dan 287 data negatif. Penelitian ini menunjukkan metode klasifikasi SVM dengan membandingkan kernel cukup baik.

Analisis Sentimen Masyarakat..., Hafidz Maulana An Nahar, Fakultas Teknik dan Sains, UMP 2023

No	Peneliti	Metode	Hasil
			Untuk kernel linear mendapat akurasi 89.17%, sedangkan kernel polynomial mendapat akurasi 84.38%.
9	(Devita et al., 2018)	*Naïve Bayes* dan *K-Nearest Neighbor*	Tingkat akurasi dengan metode *Naïve Bayes* sebesar 70% sedangkan pada metode *K-Nearest Neighbor* sebesar 40%.

1. Zhafira et al (2021) melakukan penelitian tentang analisis sentimen kebijakan kampus merdeka menggunakan *Naïve Bayes* dan pembobotan TF-IDF berdasarkan komentar Youtube. Analisis sentimen menggunakan 1000 dataset yang tertampung dalam komentar Youtube kemudian diklasifikasikan kedalam sentimen positif dan negatif. Klasifikasi diimplementasikan pada *Google Collaboratory* yang berbasis bahasa *Python* dan *Jupyter Notebook* dengan algoritma *Naïve Bayes Classifier*, pembobotan kata TF-IDF serta validasi data menggunakan *k-fold cross validation.* Rata-rata akurasi yang didapat dari 10 iterasi *k-fold cross validation* yaitu sebesar 91.8% dengan nilai *Precision, Recall, f-measure* sebesar 90.35%, 93.6%, 91.95%.
2. Yulita et al (2021) melakukan penelitian tentang analisis sentimen terhadap opini masyarakat tentang vaksin Covid-19 menggunakan algoritma *Naïve Bayes.* Analisis dilakukan terhadap 3780 tweet yang berkaitan vaksinasi dengan menggunakan algoritma *Naïve Bayes Classifier* dengan metode pembobotan TF-IDF. Hasil yang didapatkan dari penelitian ini adalah masyarakat Indonesia rata-rata memberikan respon positif terkait kebijakan vaksinasi Covid-19 dengan presentase data sebesar 60.3% dan jumlah data sebanyak 2278 data. respon negatif terhadap kebijakan lebih kecil dibandingkan respon netralnya, sehingga membuktikan hanya sedikit orang yang menentang kebijakan vaksinasi

Analisis Sentimen Masyarakat…, Hafidz Maulana An Nahar, Fakultas Teknik dan Sains, UMP 2023

ini. hal ini ditunjukkan dengan nilai respon negatif adalah 5.4% (203 data), dan respon netral adalah 34.4% (1299 data). Penggunaan algoritma Naïve Bayes Classifier untuk melakukan analisis sentimen ini sudah sangat baik ditunjukkan dengan hasil akurasi tertinggi sebesar 93%.

3. Ratnawati (2018) melakukan penelitian tentang implementasi algoritma *Naïve Bayes* terhadap analisis sentimen opini film pada *Twitter*. *Tweet* yang dikumpulkan adalah *tweet* yang mengandung hastag #judul film. Data diambil sebanyak 500 *tweet* termasuk data uji yang nantinya akan dibagi menggunakan 5-*fold cross validation* atau k=5 sehingga masing-masing berisikan 100 data. Perbandingan data *training* dan data *testing* 80:20 sehingga didapatkan 400 data sebagai data *training* dan 100 data sebagai data *testing*. Semakin banyak data *training* yang digunakan maka akan mempengaruhi kinerja sistem. Hasil akurasi akan semakin tinggi dan itu menandakan sistem berhasil melakukan klasifikasi dengan baik. Akurasi tertinggi didapat pada *fold* kedua yaitu 90%, *Precision* 92%, *Recall* 90% dan *F-measure* 90%.
4. Ruhyana (2019) melakukan penelitian tentang analisis sentimen terhadap penerapan sistem plat nomor ganjil/genap pada *Twitter* dengan metode klasifikasi *Naïve Bayes*. Pengambilan data dilakukan dengan kata kunci #ganjilgenap pada aplikasi *RapidMiner*. Pelabelan dilakukan secara manual dengan mengelompokkan sentimen positif maupun negatif. Penelitian ini menggunakan metode *data mining* untuk klasifikasi dengan algoritma *Naïve Bayes Classifier*. Hasil klasifikasi teks dalam bentuk positif dan negatif untuk penerapan lalu lintas ganjil genap dalam penelitian ini menghasilkan *Accuracy* 86.67%, *Precision* 71.43% dan *Recall* 80%.
5. Arsi et al (2021) melakukan penelitian tentang analisis sentimen pindah Ibu Kota berbasis *Naïve Bayes Classifier*. Pada penelitian ini diusulkan metode *Naive Bayes Classifier* (NBC) untuk menganalisis sentimen terhadap wacana pemerintah memindahkan Ibukota Indonesia di media *Twitter* dengan komentar positif dan negatif. penentuan model klasifikasi

Analisis Sentimen Masyarakat…, Hafidz Maulana An Nahar, Fakultas Teknik dan Sains, UMP 2023

dengan cara melakukan *splitting* data *training* dan data *testing* yang kemudian dikomputasi menggunakan metode *Naïve Bayes* sehingga mendapatkan luaran. Dari data luaran tersebut kemudian dilakukan validasi dan evaluasi dengan menggunakan *5-fold cross validation* dan *confusion matrix*. Hasil penelitian yang didapat dari dataset sejumlah 1007 data adalah nilai *Accuracy* sebesar 94.33%, *Precision 87*%, *Recall* 99% dan *F1-Score 92%* yang berarti sistem ini sudah baik dalam mendeteksi sentimen.

6. Kurniawan & Susanto (2019) melakukan penelitian tentang implementasi metode *K-Means* dan *Naïve Bayes Classifier* untuk analisis sentimen pemilihan presiden 2019. Penelitian ini mempresentasikan analisis sentimen terhadap data opini pada *Twitter* mengenai pemilihan presiden tahun 2019. Metode untuk analisis sentimen menggunakan *K-Means* untuk melakukan *clustering* pada data latih dan menghasilkan bobot positif atau negatif pada setiap dokumen latih, kemudian *Naïve Bayes* digunakan untuk melakukan klasifikasi pada dokumen uji. Untuk menguji *Naïve Bayes* dalam proses klasifikasi dilakukan percobaan sebanyak dua kali menggunakan *confusion matrix*. Dari hasil pengujian 100 dan 150 data uji didapatkan akurasi rata-rata 93.35% dan *error rate* rata-rata 6.66%.
7. Sundara et al (2020) melakukan penelitian tentang *Naïve Bayes Classifier* untuk analisis sentimen isu radikalisme. Data yang diambil dari *Twitter* terkait wacana larangan bercadar dan bercelana cingkrang bagi Aparatur Sipil Negara (ASN) sebanyak 550. Data yang sudah disimpan dalam format CSV, kemudian dilakukan pembagian data menjadi data latih dan uji dengan perbandingan 8:2 yang masing-masing sebanyak 440 data latih dan 110 data uji. Pengklasifikasian data tersebut menjadi kelas positi dan negatif sebanyak 219 dan 331. Hasil akurasi dari metode ini dan dikolaborasikan dengan TF-IDF sebesar 86%. Penelitian ini menghasilkan tingkat kecenderungan masyarakat terhadap ASN yang bercadar dan bercelana cingkrang adalah negatif hal ini disebabkan karena masyarakat

Analisis Sentimen Masyarakat..., Hafidz Maulana An Nahar, Fakultas Teknik dan Sains, UMP 2023

memandang bahwa ASN adalah pelayan publik yang mesti memberikan pelayanan terbaik.

8. Mahendrajaya et al (2019) melakukan penelitian tentang analisis sentimen pengguna Gopay menggunakan metode *Lexicon Based* dan *Support Vector Machine*. Pengambilan data dari Twitter dana analisa data menggunakan R Studio. Data yang digunakan berupa opini tentang ulasan Go-Pay dari media sosial *Twitter* yang berjumlah 1210 data. Proses pelabelan data dibagi menjadi dua kelas yaitu kelas positif dan negatif dengan menggunakan *Lexicon Based*. Hasil dari pelabelan dengan *Lexicon Base* berjumlah 923 data positif dan 287 data negatif. Dari penelitian ini juga menunjukkan metode klasifikasi SVM dengan membandingkan kernel cukup baik. Untuk kernel linear mendapat akurasi 89.17% dan dapat melakukan klasifikasi dengan benar sebanyak 1109 ulasan, sedangkan kernel polynomial mendapat akurasi 84.38% dan dapat melakukan klasifikasi dengan benar sebanyak 1021 ulasan.
9. Devita et al (2018) melakukan penelitian tentang perbandingan kinerja metode *Naïve Bayes* dan *K-Nearest Neighbor* untuk klasifikasi artikel berbahasa Indonesia. Dataset yang digunakan sebanyak 40 jurnal yang telah dipublikasi 2 tahun terakhir, jurnal tersebut di antaranya yaitu: jurnal Pendidikan Ekonomi, Pendidikan Bisnis dan Manajemen, Akutansi Aktual, dan jurnal Ekonomi Bisnis. Masing-masing jurnal diambil sebanyak 10 jurnal. Setelah menerapkan metode *Naïve Bayes* dan *K-Nearest Neighbor* untuk mengklasifikasikan artikel jurnal berbahasa Indonesia diketahui bahwa kinerja dari metode *Naïve Bayes* lebih unggul dari metode *K-Nearest Neighbor*. Terbukti dari 40 data uji yang digunakan, *Naïve Bayes* mampu mengklasifikasikan artikel jurnal sebanyak 28 dokumen. sedangkan *K-Nearest Neighbor* hanya dapat mengklasifikasikan artikel jurnal sebanyak 16 dokumen. Hal tersebut dapat dipengaruhi jumlah data dan tahapan *preprocessing* yang digunakan. Sehingga didapatkan tingkat akurasi dengan metode *Naïve Bayes* sebesar 70% sedangkan pada metode *K-Nearest Neighbor* sebesar 40%.

Analisis Sentimen Masyarakat…, Hafidz Maulana An Nahar, Fakultas Teknik dan Sains, UMP 2023

B. Landasan Teori

1. Analisis sentimen

Analisis sentimen (*sentiment analysis*) merupakan bidang studi yang menganalisa opini masyarakat, evaluasi, penilaian, sikap dan emosi terhadap sebuah produk, pelayanan, organisasi atau perhimpunan, seorang tokoh dan isu atau masalah serta peristiwa yang terjadi pada masyarakat itu sendiri (Fairuz et al., 2021). Analisis sentimen dilakukan untuk melihat opini atau kecenderungan opini seseorang terhadap suatu masalah atau objek, baik cenderung berpandangan positif maupun negatif (Dewi Utami et al., 2021).

Analisis sentimen merupakan salah satu eksplorasi baru dari NLP atau yang biasa disebut dengan *Natural Language Processing.* NLP merupakan riset ilmiah yang menekuni bagaimana komputer bisa digunakan untuk memanipulasi bahasa alami. NLP bisa membuat bahasa lebih terstandarisasi, mengganti bahasa alami jadi bahasa yang lebih terstruktur, standart, serta terintegrasi sehingga mempermudah buat memastikan penaksiran yang tepat (Nur et al., 2020).

2. Pelayanan Medis

Pelayanan medis adalah perjanjian antara rumah sakit dan pasien untuk memberikan tindakan medis sesuai kebutuhan pasien. Dokter, perawat atau mereka yang dinyatakan memiliki kompetensi dalam melakukan pelayanan kesehatan merupakan tenaga ahli profesional. Upaya maksimal yang dilakukan oleh dokter dalam menangani pasien bertujuan agar masyarakat mendapatkan hak atas kesembuhan dan pemulihan kesehatan (Nuratih et al., 2021). Pasien dapat menggugat dokter apabila perbuatannya melawan hukum, sedangkan gugatan terhadap rumah sakit dapat dilakukan berdasarkan wanprestasi (ingkar janji) disamping perbuatan melawan hukum (Azzahra & Mufidin, 2021).

Analisis Sentimen Masyarakat..., Hafidz Maulana An Nahar, Fakultas Teknik dan Sains, UMP 2023

3. *Twitter*

Twitter merupakan salah satu platform media sosial yang banyak digunakan oleh masyarakat indonesia. *Twitter* hadir dengan format berbeda, dimana *Twitter* memiliki konsep yaitu menyebarkan informasi pesan secara singkat, padat dan *real time* kepada pembacanya di seluruh dunia (Puspitadewi et al., 2016).

Pengguna *Twitter* diberikan kesempatan untuk dapat berbagi pikiran, mencurahkan perasaan mereka kepada *follower* dan melakukan aktifitas lain (Sukendar, 2016). Pengguna juga dibebaskan untuk mengkritisi, mendukung maupun mengekspresikan kemarahan produk atau berita yang diterima oleh orang yang mereka *follow* atau isu yang sedang di bahas di *Twitter* (Fatimatuzzahra et al., 2019).

Twitter berbeda dengan media sosial lainnya karena memiliki keterbukaan terhadap data yang dimilikinya melalui API (Application Programming Interface). API merupakan antarmuka perangkat lunak, dengan API maka aplikasi berbicara satu sama lain tanpa sepengetahuan atau intervensi pengguna (Salim & Mayary, 2020).

4. Web *Scraping*

Web *Scraping* adalah cara yang digunakan dalam mendapatkan data atau informasi dari situs web yang dilakukan secara otomatis. Tujuan dari web *scraping* adalah untuk menggali informasi dari situs web yang berbeda dan tidak terstruktur lalu ditransformasikan menjadi bentuk yang lebih rapi dan terstruktur dalam format *spreadsheets*, basis data, atau *comma separated values* (CSV) (Parasati et al., 2020).

5. *Text Mining*

Text mining merupakan teori tentang pengolahan kumpulan teks dengan tujuan untuk mengetahui dan mengekstrak informasi bermanfaat dari kumpulan teks tersebut. Informasi didapatkan dengan cara identifikasi dan eksplorasi pola yang menarik dari sumber data. *Text mining* merupakan bidang khusus dari *data mining* dimana data yang digunakan

Analisis Sentimen Masyarakat..., Hafidz Maulana An Nahar, Fakultas Teknik dan Sains, UMP 2023

adalah data tekstual yang tidak terstruktur. Bagian-bagian dari *text mining* meliputi *classification, clustering,* dan *association* (Sabrani et al., 2020).

6. *Naïve Bayes Classifier*

Naïve Bayes Classifier adalah teknik prediksi berdasarkan probabilistik sederhana dan penerapan teorema Bayes (aturan Bayes) dengan asumsi independensi yang kuat. *Naïve Bayes Classifier* dikenal lebih baik dari beberapa metode klasifikasi lainnya. Dikarenakan pertama, ciri utama *Naïve Bayes* adalah asumsi independensi yang kuat dari setiap kondisi atau peristiwa. Kedua, modelnya sederhana dan mudah dibuat. Ketiga, model dapat diimplementasikan untuk kumpulan data yang besar (Salmi & Rustam, 2019).

Perhitungan untuk klasifikasi *Naïve Bayes* dapat lakukan dengan langkah-langkah berikut:

a. Peluang kemunculan kategori positif dan negatif (probabilitas kelas)

$$P(c) = \frac{X}{A} \quad (1)$$

Keterangan:

P(c) = probabilitas kelas

X = jumlah data latih pada kelas

A = jumlah semua data latih

b. Perhitungan probabilitas kemunculan *term*

$$P(a/c) = \frac{n_{(a|c)} + 1}{n_c} \quad (2)$$

Keterangan:

P(a|c) = probabilitas *term* dalam kelas

$n_{(a|c)}$ = jumlah *term* dalam kelas

n_c = jumlah semua *term* dalam kelas

c. Perhitungan klasifikasi kategori

P(c) X P(a|c)

(3)

Keterangan:

P(c) = probabilitas kelas

P(a|c) = probabilitas *term* dalam kelas

7. *Confusion Matrix*

Confusion matrix adalah alat yang berguna untuk menganalisis seberapa baik klasifikasi yang telah dibuat, dapat mengenali *tuple* dari kelas yang berbeda (Sundara et al., 2020). Berikut tabel perhitungan *confusion matrix* dapat dilihat pada Tabel 2.2.

Tabel 2.2 *Confusion Matrix*

		Kelas Terprediksi	
		Negative	**Positive**
Aktual Kelas	**Negative**	TN(True Negative)	FP(False Positive)
	Positive	FN(False Negative)	TP(True Positive)

Keterangan:

TP (*True Positive*) : data positif yang terdeteksi dengan benar.

TN (*True Negative*) : data negatif yang terdeteksi dengan benar.

FP (*False Positive*) : data negatif namun terdeteksi sebagai data positif.

FN (*False Negative*) : data positif namun terdeteksi sebagai data negatif.

TP, FN, FP, TN merupakan parameter yang digunakan untuk menghitung nilai *accuracy*, *precision*, *recall*, dan *F1-Score.* Pada Persamaan (4), (5), (6), (7) digunakan untuk menilai hasil dari model yang sedang diuji (Zaenal et al., 2020).

a. *Accuracy*

Merupakan tingkat kedekatan nilai prediksi dengan nilai aktual, untuk menghitung *accuracy* menggunakan Persamaan (4).

Analisis Sentimen Masyarakat..., Hafidz Maulana An Nahar, Fakultas Teknik dan Sains, UMP 2023

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \text{ X } 100\% \quad (4)$$

b. *Precision*

Precision merupakan tingkat ketetapan atau ketelitian dalam pengklasifikasian, untuk menghitung *Precision* menggunakan Persamaan (5).

$$P(positif) = \frac{TP}{TP+FP} \times 100\%$$
$$P(negatif) = \frac{TN}{TN+FN} \times 100\% \quad (5)$$

c. *Recall*

Recall berfungsi untuk mengukur proporsi data aktual yang benar diidentifikasi, untuk menghitung *Recall* menggunakan Persamaan (6).

$$R(positif) = \frac{TP}{TP+FN} \times 100\%$$
$$R(negatif) = \frac{TN}{TN+FP} \times 100\% \quad (6)$$

d. *F1-Score*

F1-Score digunakan untuk mengetahui keseimbangan antara presisi dan recall yang didapat dari sistem yang akan dibangun, untuk menghitung *F1-Score* menggunakan persamaan (7).

$$F1-Score(positif) = 2 \times \frac{TP}{TP+FN} \times 100\%$$
$$F1-Score(negatif) = 2 \times \frac{TN}{TN+FP} \times 100\% \quad (7)$$

8. *Python*

Python adalah bahasa pemrograman interaktif yang dapat dijalankan di berbagai platform dan aplikasi. Apabila membuat program

python untuk mengolah data maupun *machine learning* disarankan menggunakan *python* versi 3.0 (Mujilahwati et al., 2021).

Bahasa pemrograman *Python* dipilih karena memiliki beberapa keunggulan (Retnoningsih & Pramudita, 2020) yaitu:

a. *Python* memliki *source code* yang sederhana sehingga mudah dibaca, mudah ditulis dan mudah diingat.
b. *Python* memliki *library* yang lengkap, *Python code* akan lebih sederhana bila dibandingkan dengan *code* yang ditulis dengan bahasa pemrograman lain.
c. *Python* dikembangkan sebagai *project open source* dan bisa digunakan siapa saja secara gratis.
d. *Python* dapat digunakan pada hampir semua sistem operasi termasuk *Windows, Linux, Mac OS, Unix* dan juga sistem operasi pada perangkat lunak berbasis *mobile* seperti *Android* dan *IOS*.

Analisis Sentimen Masyarakat..., Hafidz Maulana An Nahar, Fakultas Teknik dan Sains, UMP 2023