

BAB II TINJAUAN PUSTAKA

A. Penelitian Terdahulu

Penelitian terdahulu dapat di lihat pada Table 2. 1

Tabel 2. 1 Penelitian Terdahulu

No	Peneliti	Metode	Hasil
1	Hidayatullah <i>et al.</i> (2021)	<i>Naive Bayes Classifier</i>	Pengambilan data menggunakan proses <i>crawling Twitter</i> yang dianalisis menggunakan <i>Naive Bayes</i> . Hasil dari penelitian ini bahwa sentimen kinerja polisi masuk ke klasifikasi positif 39.5%, negatif 58.1% dan netral 75.8% serta model yang dihasilkan pada nilai akurasi 0.929, <i>recall</i> 0.923, <i>precision</i> 0.933 dan <i>f-measure</i> 0.954.
2	Duei <i>et al.</i> (2022)	<i>Naive Bayes Classifier</i>	Penelitian ini menggunakan sebanyak 1546 data <i>tweet</i> , hasil dari penelitian ini didapatkan bahwa DPR mendapatkan 693 tweet netral dengan polaritas 0.82 atau 82%, 95 tweet positif dengan polaritas 0.75 atau 75%, negatif dengan accuracy score 0.82 atau 82% sentimen negatif dengan accuracy score 0.8 atau 80% berdasarkan data testing sebanyak 20%..
3	Martiti & Juliane (2021)	<i>Naive Bayes Classifier</i>	Data yang disiapkan dalam penelitian ini dibagi menjadi dua yaitu data latih dan data uji. Data latih yang digunakan dalam penelitian

No	Peneliti	Metode	Hasil
			ini hingga 475 kalimat, berdasarkan data uji diperoleh akurasi sebesar 86,66%.
4	Harijatio (2019)	<i>Naive Bayes</i>	Hasil uji akurasi tanpa menggunakan <i>k-fold cross validation</i> menghasilkan akurasi yang lebih besar, yaitu 72,941% dibandingkan dengan semua <i>k-fold cross validation</i> dimana 3-fold menghasilkan akurasi 71,601% 5-fold menghasilkan akurasi 70,72% dan 10-fold menghasilkan akurasi 70,68%.
5	Amalia <i>et al.</i> (2021)	<i>Support Vector Machine</i>	Pengambilan data pada analisis sentimen review pelanggan restoran solaria ini menggunakan data <i>Twitter</i> menggunakan teknik <i>crawling</i> dengan kata kunci “nasi goreng dan kwetiau Solaria” yang di ambil dari tanggal 24 Agustus sampai tanggal 12 September data yang telah di ambil sebanyak 238 data. Hasil pengujian menggunakan algoritma <i>Support Vector Machine</i> menghasilkan nilai akurasi sebesar 81.92%.
6	Nurmawati <i>et al.</i> (2020)	<i>Naive Bayes Classifier</i>	1000 data yang diambil <i>tweet</i> yang bernilai negatif sebanyak 342, analisis sentimen dari anggapan pengguna <i>Twitter</i> tentang Army BTS memiliki polaritas sentimen negatif sebanyak 34,2% netral 58% dan positif 7,3% dengan tingkat akurasi sebesar 75%
7	Putra <i>et al.</i> (2020)	<i>Naive Bayes Classifier</i>	100 data diambil dengan rincian 70 data latih dan 30 data uji yang menghasilkan dompet elektronik Dana dengan nilai (positif = 37.50%, negatif = 30.00%) OVO dengan nilai (positif = 16.67%,

No	Peneliti	Metode	Hasil
			negatif = 63.33%) dan dompet elektronik GoPay lebih efektif oleh pengguna <i>Twitter</i> dengan nilai (positif = 46.67%, negatif = 53.33%)
8	Normawati & Prayogi (2021)	<i>Naïve Bayes Classifier</i>	Data yang diteliti sebagai studi kasus sederhana hanya menggunakan sebanyak 8 data cuitan yang dibagi menjadi 5 data latih dan 3 data uji. Penelitian ini menghasilkan pemaparan yang terstruksur pada proses dan hasil implementasi <i>Naïve Bayes</i> dan pengujian performa menggunakan <i>confusion matrix</i> dengan akurasi sebesar 82%, <i>recall</i> sebesar 52%, dan presisi 93%.

Berdasarkan Table 2.1 dapat di jelaskan bahwa penelitian Hidayatullah *et al.* (2021) yang berjudul *Sentiment Analysis of Police On Twitter Users Using Naive Bayes Method*, membahas tentang kinerja polisi pada pengguna *Twitter* yang bertujuan untuk mengukur berapa sentimen kinerja polisi menurut warga *Twitter*. Topik penelitian ini merupakan teks mining yang menggunakan metode *Naïve Bayes*. Langkah-langkah dalam metode ini yaitu *preprocessing*, pembobotan kata seperti *TF-IDF* dan di akhiri dengan klasifikasi serta evaluasi yang menghasilkan 39.5% positif, 75.8% klasifikasi netral, 58.1% negatif dan dengan nilai akurasi 0.929, dan *f-measure* 0.954, *precision* 0.933, *recall* 0.923.

Penelitian Duei *et al.* (2022) yang berjudul Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada *Twitter* Menggunakan Metode *Naïve Bayes Classifier* yang bertujuan untuk menganalisis sentimen masyarakat terhadap kinerja DPR yang diungkap melalui media sosial *Twitter*. Tahap untuk melakukan analisis sentimen ini yaitu *crawling*, *preprocessing* data dan klasifikasi data menggunakan metode *Naïve Bayes Classifier*. Penelitian ini menggunakan sebanyak 1546 data *tweet*, hasil dari penelitian ini didapatkan bahwa DPR mendapatkan 758 *tweet* negatif dengan polaritas 0.82 atau 82%, 693 *tweet* netral dengan polaritas 0.79 atau 79%, 95 *tweet* positif dengan polaritas 0.75 atau 75% dengan *accuracy score* 0.8 atau 80% berdasarkan data testing sebanyak 20%.

Penelitian Martiti & Juliane (2021) yang berjudul *Implementation of Naïve Bayes Algorithm on Sentimen Analysis* yang bertujuan untuk membangun aplikasi analisis sentimen yang memiliki nilai akurasi yang sama dengan

mechin learning Weka, yang dapat mengklasifikasi kalimat yang membingungkan untuk diklasifikasi menjadi positif, negatif atau netral. Klasifikasi sentimen dalam penelitian ini menggunakan metode *Naïve Bayes Classifier* dimana sentimen untuk data latih diambil dari komentar *public* diakun resmi media sosial Grab Indonesia di *Twitter*. Hasil analisis pengujian menunjukkan *Naïve Bayes* memiliki akurasi yang stabil setelah diuji dengan nilai akurasi 99,62% pada data testing, berdasarkan pengujian akurasi 30 karakter menghasilkan nilai akurasi sebesar 86,66%, dengan rata-rata kegagalan klasifikasi sebesar 0,13%.

Penelitian Harijiatno (2019) yang berjudul *Analisis Sentimen Pada Twitter Menggunakan Multinomial Naïve Bayes* yang bertujuan untuk menganalisis sentimen *Twitter* terhadap tokoh *public* menjelang pemilu 2019, yaitu Jokowi dan Prabowo dengan menggunakan metode *Multinomial Naïve Bayes*, hasil uji akurasi tanpa menggunakan *k-fold cross validation* menghasilkan akurasi yang lebih besar, yaitu 72.941% dibandingkan dengan semua *k-fold cross validation* dimana *10-fold* menghasilkan akurasi 70.68%, *5-fold* menghasilkan akurasi 70.72% dan *3-fold* menghasilkan akurasi 71.601%.

Penelitian Amalia *et al.* (2021) yang berjudul *Analisis Sentimen Review Pelanggan Restoran Menggunakan Algoritma Support Vector Machine Dan K-Nearest Neighbor* yang bertujuan untuk mengetahui opini ulasan pelanggan restoran solaria di media sosial *Twitter* dengan membandingkan dua algoritma *Support Vector Machine* dan *K-Nearest Neighbor* yang akan dibagi menjadi 2 kelas yaitu kelas negatif dan kelas positif, pengumpulan data dilakukan

menggunakan teknik *crawling* dengan kata kunci “nasi goreng dan kwetiau Solaria” yang diambil pada tanggal 24 Agustus 2021 dan tanggal 12 September 2021, data yang telah diambil sebanyak 238 data. Perhitungan dengan menggunakan algoritma *SVM* memiliki *accuracy* sebesar 81.92% sedangkan algoritma *k-NN* memiliki hasil *accuracy* sebesar 59.03% yang dapat disimpulkan algoritma *SVM* lebih baik dibanding algoritma *k-NN*.

Penelitian Nurmawati *et al.* (2020) yang berjudul Analisis Sentimen Terhadap Penggemar K-POP Media Sosial *Twitter* Menggunakan *Naïve Bayes* (Studi Kasus Penggemar Grup BTS) yang bertujuan untuk mengetahui banyaknya cuitan *Twitter* yang mengutarakan pendapat mengenai grup BTS, menggunakan sebuah program yang dibuat dengan bahasa pemrograman *python*, analisis dapat dilakukan dengan efektif, dari analisis yang dilakukan menggunakan klasifikasi algoritma *Naïve Bayes*, terdapat polaritas sentimen negatif sebanyak 34.2%, netral 58.5%, dan positif 7.3% dari 1000 data yang diambil sesuai hasil polaritas *tweet* yang bernilai negatif sesuai hasil polaritas sebanyak 324 dengan tingkat akurasi sebesar 75%.

Penelitian Putra *et al.* (2020) yang berjudul Analisis Sentimen Dompot Elektronik Pada *Twitter* Menggunakan Metode *Naïve Bayes Classifier* yang bertujuan untuk menganalisa tanggapan masyarakat atas hadirnya berbagai aplikasi dompet elektronik. Data diperoleh secara langsung dari situs *microblogging Twitter* dengan hastag pencarian #GoPay, #OVO dan #Dana dari data tersebut diambil 100 data dengan rincian 70 data latih dan 30 data uji. Hasil penelitian ini menunjukkan bahwa dompet elektronik GoPay lebih dinilai

positif dengan nilai akurasi sebesar 46,67% diikuti oleh Dana sebesar 37,50% dan OVO sebesar 16,67%. Dompot elektronik OVO memiliki nilai negatif sebesar 53,33% dan Dana sebesar 30,00%. Penelitian ini memberikan bukti empiris dan elektronik yang tersedia berdasarkan hasil uji coba sekaligus bisa dijadikan bahan analisis sebelum menjatuhkan pilihan kepada salah satu dompet elektronik tersebut.

Penelitian Normawati & Prayogi (2021) yang berjudul Implementasi *Naïve Bayes Classifier* Dan *Confusion Matrix* Pada Analisis Sentimen Berbasis Teks Pada *Twitter* yang bertujuan untuk menganalisis tingkat sentimen terhadap data yang dilakukan dengan teknik *crawling* pada *Twitter*. Data yang diteliti sebagai studi kasus sederhana hanya menggunakan sebanyak 8 data cuitan yang dibagi menjadi 5 data latih dan 3 data uji. Data diolah dengan tahap *preprocessing*, lalu diklasifikasikan menggunakan metode *Naïve Bayes* dan didapatkan akurasi sebesar 82%, *presisi* 93% dan *recall* sebesar 52%.

B. Landasan Teori

1. Sentimen

Sentimen dapat diekspresikan dengan cara yang berbeda, bisa jadi diekspresikan dengan berbagai macam emosi, dengan melewati penilaian, wawasan, visi atau pandangan orang, jika mempelajari emosi dalam format teks maka dapat dilihat dari dua cara yang berbeda, pertama dapat dipelajari sebagai dampak pada penulis, cara dia memilih kata untuk mengekspresikan emosi tertentu, dan dampak kedua pada cara pembaca menafsirkan konten yang

ditulis tergantung pada keadaan pikirannya dan kemampuannya dalam menganalisa (Mehta & Pandya, 2020).

2. Analisis Sentimen

Analisis sentimen atau penambangan opini merupakan proses identifikasi dan mengkategorikan emosi atau opini untuk mengendalikan apa pun. Sumber dari analisis ini merupakan saluran komunikasi situs web, seperti ulasan, diskusi forum, *blog*, *Twitter*, mikro, dll. Bidang penelitian ini sangat populer, karena data yang dipentingkan di mana pengguna dapat menemukan penilaian layanan yang bermanfaat untuk kehidupan sehari-hari, dengan penyimpanan data yang berbentuk digital (Mehta & Pandya, 2020).

Penelitian analisis sentimen bekerja dengan emosi perasaan, ringkasan, atau ekstraksi pikiran. Analisis sentimen juga dikenal sebagai penambangan opini yang menggunakan *Neuro Linguistic Programming* – natural pemrosesan bahasa untuk mengikuti emosi, perasaan pendapat masyarakat mengenai topik tertentu untuk jasa ataupun produk, dikarenakan analisis sentimen sangat terkenal, analisis sentimen berguna dalam banyak hal dalam kampanye survei dan iklan untuk mendapatkan tingkat keberhasilan produk atau layanan apa pun dengan pendapat atau saran orang, hal itu juga memberikan informasi mengenai orang yang suka maupun tidak suka dan perusahaan menjadi lebih jelas memikirkan tentang fitur produknya (Mehta & Pandya, 2020).

Analisis sentimen dapat juga disamakan dengan opini mining karena terfokus pada kaidah yang menyatakan positif atau negatif. Analisis sentimen mengerjakan penambangan data untuk menganalisis topik tertentu. Proses

analisis dapat mencakup teks ulasan, forum, *tweet*, atau *blog*, dengan data *processing* mencakup proses *tokenization*, *stopword*, penghapusan, *stemming*, identifikasi sentimen, dan klasifikasi sentimen (Prabowo & Wiguna, 2021).

3. *Text Mining*

Sekumpulan dokumen yang digunakan untuk membuat sebuah informasi merupakan tujuan dari *text mining*. Pemrosesan, pengelompokan dan analisis data dari sebuah informasi dihasilkan dari *text mining* dengan hasil data pada sebuah teks dengan format tidak terstruktur. Proses pengutipan sitasi didalam *text mining* menghasilkan sebuah analisis berupa perasaan emosional dalam menggali pernyataan berupa pernyataan positif maupun pernyataan negatif. Objek *text mining* adalah dokumen semi terstruktur (tidak terstruktur), *text mining* dalam artian terstruktur menjadi data material terstruktur pada *text mining* dan selanjutnya disimpan dalam basis data terstruktur (Prabowo & Wiguna, 2021).

Ketidak teraturan pada teks, maka diperlukan proses *text mining* melalui beberapa tahap dengan tujuan untuk mempersiapkan teks agar diubah menjadi data teks yang lebih terstruktur (Imron, 2019).

4. *Twitter*

Twitter merupakan situs web yang dioperasikan dan dimiliki oleh Twitter Inc., yang mempromosikan jaringan sosial media berupa *microblog* sehingga penggunaanya dapat membaca dan mengirim pesan *tweets*. *Tweets* dapat dilihat oleh siapa saja atau bersifat *public*, namun pengirim juga dapat

membatasi pengiriman *tweets* ke daftar teman-teman dekat saja, dalam *Twitter* terdapat istilah *follower*, dimana pengguna memiliki sebuah pengikut pengguna lain. (Harijianto, 2019).

Tweet memiliki batasan pesan dengan maksimal 140 karakter. *Tweet* merupakan sebuah pesan yang dikirim pengguna yang berbeda, dengan batasan maksimal 140 karakter, *tweets* dapat digunakan untuk mengutarakan sebuah ulasan atau opini untuk masa yang akan datang dimana kita dapat menghasilkan hasil jajak pendapat (Mehta & Pandya, 2020).

5. *Scraping*

Scraping adalah salah satu teknik yang digunakan untuk mendapatkan sebuah informasi ataupun data di sebuah website dengan otomatis. Informasi tersebut dapat berupa dokumen, *text*, audio, video maupun sebuah tautan (Flores *et al.*, 2020).

6. *Python*

Python adalah bahasa pemrograman bersifat *open source* (sumber terbuka), sehingga tidak dibutuhkan sebuah lisensi untuk dikembangkan secara gratis sehingga dapat dikembangkan sesuai kebutuhan penggunanya. Fleksibilitas dalam penulisan sintaks pada bahasa *python* mempermudah dalam mempelajari bahasa tersebut, selain itu struktur datanya juga sangat efisien, pemrograman berorientasi objek yang efektif namun sederhana, dapat bekerja pada *multiplatform* serta dapat menggabungkan pemrograman lain sehingga dapat menghasilkan sebuah aplikasi yang mumpuni. Pengeksekusi data pada bahasa pemrograman *python* menggunakan cara *interpreter*, dimana terdapat

dua cara menggunakannya modul *script* dan mode baris. Mode baris digunakan dengan cara mengetik perintah pada sebuah *command line* atau *shell* kemudian dieksekusi langsung oleh *python*, pada mode *script*, *command python* dapat ditulis pada sebuah *text editor*(notepad), kemudian disimpan dalam format ".py" lalu dieksekusi oleh *python*(Hasanuddin, 2020).

7. *Machine Learning*

Rangkaian teknik yang berfungsi membantu dalam memprediksi dan menangani sebuah data besar dengan menggambarkan data-data tersebut kedalam algoritma pembelajaran merupakan pengertian dari *Machine Learning*. Arthul Samuel merupakan penemu istilah *machine learning* pertama kalinya pada tahun 1959. *Machine learning* menurut Arthur Samuel adalah salah satu bidang dalam ilmu komputer dengan memberikan kemampuan pembelajaran pada komputer dengan tujuan mengetahui sesuatu tanpa sebuah pemrograman yang jelas (García Reyes, 2020).

8. *Klasifikasi Naïve Bayes Multinomial*

Naive bayes adalah metode pembelajaran mesin probalistik, metode ini memperkirakan setiap atribut dari sebuah data tidak tergantung satu dengan lainnya. Metode *Naïve Bayes* berlawanan dengan keadaan aslinya, ini dikarenakan pada teks atau sebuah dokumen harus memiliki kata yang saling terhubung satu dengan lainnya sehingga dokumen tersebut memiliki metode dan makna. Bidang klasifikasi teks adalah salah satu pemanfaatan yang tepat dalam penggunaan metode *Naive Bayes* (Sabrani *et al.*, 2020).

Multinomial Naive Bayes merupakan salah satu model atau jenis dalam *Naive Bayes*. *Multinomial Naive Bayes* merupakan metode *Supervised Learning*, oleh karena itu setiap data harus diberi label terlebih dahulu sebelum proses *training* (Sabrani *et al.*, 2020). Menurut Harijatio (2019) klasifikasi *Multinomial Naive Bayes* diawali dengan proses pengambilan jumlah kata yang muncul dalam tiap dokumen dengan anggapan dokumen tersebut memiliki beberapa probabilitas kejadian dalam kata dengan panjang yang tidak tergantung pada kelasnya.

9. *Natural Language Tool Kit (NLTK)*

Platform untuk menulis sebuah bahasa pemrograman *python* yang tersusun dari rangkaian program pengolahan bahasa simbolik serta statistik alami dan perpustakaan merupakan pengertian *Natural language Tool Kit (NLTK)*. *NLTK* menggabungkan alat pemrosesan bahasa (analisa semantik, parser sintaksis, *stemmer*, *tokenizer* maupun *taggers*) (Flores *et al.*, 2020).