

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 Tinjauan Pustaka

Tahap ini dilakukan dengan mengumpulkan data serta informasi terkait dengan tema penelitian. Selanjutnya data dan informasi yang telah dikumpulkan akan digunakan sebagai data pendukung dan perbandingan dengan penelitian yang sedang dilaksanakan. Penelitian ini memanfaatkan sepuluh jurnal yang relevan sebagai landasan dan perbandingan dari penelitian sebelumnya. Dari keseluruhan jurnal tersebut, terdapat tujuh jurnal nasional dan tiga jurnal internasional. Jurnal terbaru diterbitkan pada tahun 2023, sementara jurnal terlama diterbitkan pada tahun 2019. Sepuluh jurnal tersebut dipilih dua jurnal yang akan menjadi referensi utama untuk penelitian ini. Pertama, untuk pemahaman analisis sentimen pelanggan menggunakan penelitian yang dilakukan oleh Nuraini Ika Pratiwi Kalingara, Oktariani Nurul Pratiwi, dan Hilman Dwi Anggana pada tahun 2021 yang berjudul “Analisis Sentimen *Review Customer* Terhadap Layanan Ekspedisi JNE dan J&T *Express* Menggunakan Metode Naïve Bayes“ [2] sebagai acuannya. Kemudian untuk metode Naïve Bayes menggunakan penelitian yang dilakukan oleh Samsir, Ambiyar, Unung Verawardina, Firman Edi, dan Ronal Watrianthos pada tahun 2021 yang berjudul “Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes“ [6].

Kesepuluh jurnal tersebut disusun ringkas dengan memanfaatkan pendekatan 3C2S yang merupakan singkatan dari konsep *Compare* yaitu menemukan kesamaan; *Contrast* untuk menemukan perbedaan; *Criticize* untuk memberikan evaluasi terhadap penelitian sebelumnya; *Synthesize* untuk menghasilkan ide baru berdasarkan penelitian sebelumnya; dan *Summarize* bertujuan merangkum informasi. Ringkasan tersebut akan disajikan melalui sebuah tabel yang telah mencakup sepuluh jurnal penelitian terdahulu. Informasi yang menampilkan data dapat dilihat melalui Tabel 2.1.

Tabel 2.1 Tinjauan Pustaka

No.	Judul	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
1.	Analisis Sentimen <i>Review Customer</i> Terhadap Layanan Ekspedisi JNE dan J&T <i>Express</i> Menggunakan Metode Naïve Bayes [2]	Melakukan perbandingan sentimen pelanggan terhadap layanan ekspedisi dengan menggunakan metode Naïve Bayes.	Melakukan analisis sentimen kepuasan pelanggan JNE dan J&T, sedangkan penelitian yang akan dilakukan adalah terhadap kepuasan pelanggan J&T <i>Express</i> , JNE dan Sicepat.	Penggunaan metode <i>oversampling</i> perlu dicermati karena dapat mempengaruhi hasil analisis sentimen tergantung pada distribusi data yang digunakan.	Penelitian ini memanfaatkan metode penelitian sebelumnya, khususnya menggunakan metode Naïve Bayes, untuk konsistensi dengan penelitian dalam menganalisis sentimen kepuasan pelanggan.	Penelitian ini membahas analisis sentimen terhadap pelayanan JNE dan J&T <i>Express</i> dalam komentar pelanggan di Twitter menggunakan metode Naïve Bayes. Setiap data akan diberi label secara manual, kemudian melalui tahap preprocessing yang melibatkan <i>case folding</i> , <i>tokenization</i> , penghapusan <i>stopwords</i> , dan <i>stemming</i> . Selanjutnya, data akan dibagi menjadi dua bagian, yakni 75% untuk data pelatihan dan 25% untuk data uji. Setelah itu, dilakukan evaluasi menggunakan <i>confusion matrix</i> .

No.	Judul	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
2.	Analisis Sentimen Pembangunan Kereta Cepat Jakarta-Bandung di Media Sosial Twitter Menggunakan Metode Naïve Bayes [1]	Melakukan penelitian untuk menganalisis sentimen dan mengkategorikannya ke dalam kelas positif, negatif, dan netral, sama dengan penelitian yang akan dijalankan.	Melakukan analisis sentimen terhadap pembangunan kereta cepat Jakarta-Bandung, sedangkan pada penelitian yang akan dilakukan adalah menganalisis sentimen kepuasan pelanggan terhadap layanan jasa ekspedisi J&T Exprees, JNE dan Sicepat.	Penelitian hanya memfokuskan pada sentimen negatif terhadap pembangunan kereta cepat, sementara pemahaman terhadap sentimen positif dan netral juga penting untuk pemahaman yang lebih holistik.	Penelitian ini menggunakan penelitian terdahulu, yaitu dengan menggunakan metode Naïve Bayes agar selaras dengan penelitian dalam menganalisis sentimen dan mengelompokkannya kedalam <i>class</i> positif, negatif dan netral.	Penelitian memanfaatkan data yang berisi kiriman tweet dari kata kunci yang telah ditetapkan. Setiap data akan mendapatkan label secara manual satu per satu, selanjutnya akan mengalami proses pra-pemrosesan seperti <i>case folding</i> , <i>tokenization</i> , <i>stopwords removal</i> , dan <i>stemming</i> . Langkah berikutnya melibatkan ekstraksi fitur, <i>oversampling</i> , dan kemudian memasuki tahap klasifikasi. Dataset akan dipisahkan menjadi dua bagian, di mana 75% akan digunakan sebagai data latih dan 25% sebagai data uji. Evaluasi dilakukan menggunakan <i>confusion matrix</i> . Dalam penelitian ini, menggunakan <i>k-fold cross validation</i> dan menghasilkan skor untuk sentimen positif, negatif, dan netral.

No.	Judul	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
3.	Analisis Sentimen Ulasan Ekspedisi J&T <i>Express</i> Menggunakan Algoritma Naïve Bayes [7]	Melakukan penelitian untuk menganalisis sentimen ulasan pelanggan jasa ekspedisi.	Melakukan analisis sentimen terkait kepuasan pelanggan J&T <i>Express</i> , sedangkan pada penelitian yang akan dilaksanakan adalah analisis sentimen terhadap kepuasan pelanggan J&T <i>Express</i> , JNE, dan Sicepat.	Penelitian hanya membagi data menjadi dua kelas, yaitu sentimen negatif dan positif. Hal ini dapat mengabaikan sentimen netral yang mungkin juga relevan dalam analisis kepuasan pengguna penggunaan dataset yang relatif kecil (500 data), sehingga representasi keseluruhan opini masyarakat terhadap J&T <i>Express</i> mungkin tidak sepenuhnya terwakili.	Penelitian ini menggunakan penelitian terdahulu, yaitu dengan menggunakan metode Naïve Bayes agar selaras dengan penelitian dalam menganalisis sentimen ulasan pelanggan jasa ekspedisi dan mengelompokkannya kedalam kelas positif, negatif dan netral.	Penelitian ini menggunakan algoritma Naïve Bayes untuk menganalisis sentimen ulasan ekspedisi J&T <i>Express</i> . Pengelompokan data dalam analisis sentimen di J&T <i>Express</i> dilakukan dengan membaginya ke dalam dua kelas, yakni kelas negatif dan positif. Digunakan sejumlah 500 data yang dibagi menjadi 400 data pelatihan dan 100 data pengujian. Dari total 400 data latihan, 237 data termasuk dalam kategori negatif, sementara 163 data masuk dalam kategori positif. Berdasarkan temuan dari penelitian ini, metode analisis sentimen yang digunakan mencapai tingkat akurasi 87%, dengan tingkat presisi 70% untuk kelas positif, tingkat presisi 95% untuk kelas negatif, dan tingkat error sebesar 13%.

No.	Judul	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
4.	Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier [8]	Melakukan penelitian dengan menggunakan metode Naïve Bayes yang berfungsi untuk menganalisis sentimen dan mengklasifikasi kannya kedalam kelas positif, negatif dan netral.	Menganalisis sentimen terhadap opini masyarakat tentang vaksin Covid-19 sedangkan pada penelitian yang akan dilakukan adalah menganalisis sentiment kepuasan pelanggan terhadap layanan ekspedisi J&T Express, JNE dan Sicepat.	Peneliti tidak menjelaskan secara rinci mengenai proses pengambilan data dari Twitter, tidak menjelaskan mengenai proses <i>preprocessing</i> data yang dilakukan, serta tidak menjelaskan secara rinci mengenai interpretasi hasil dari penelitian.	Penelitian ini menggunakan penelitian terdahulu, yaitu dengan menggunakan metode Naïve Bayes agar selaras dengan penelitian dalam menganalisis sentimen dan mengelompokkannya kedalam kelas positif, negatif dan netral.	Penelitian ini menganalisis sentimen opini masyarakat tentang vaksin covid-19. Proses klasifikasi data terbagi menjadi 3 kelas, positif, negatif dan netral. Sejumlah 3780 tweet mengenai vaksinasi telah dianalisis menggunakan metode klasifikasi Naïve Bayes. <i>Output</i> analisis menunjukkan bahwa mayoritas <i>tweet</i> menunjukkan pendapat positif sebanyak 60,3%, sementara tweet dengan pendapat netral mencapai 34,4%, yang lebih tinggi daripada jumlah tweet yang menunjukkan pendapat menentang sebesar 5,4%. Tingkat akurasi yang tercapai adalah sebesar 0,93 (93 %).

No.	Judul	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
5.	Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi <i>COVID-19</i> Menggunakan Metode Naïve Bayes [6].	Melakukan penelitian dengan menerapkan metode Naïve Bayes untuk menganalisis sentimen dan mengklasifikasikannya ke dalam kelas positif, negatif, dan netral, sebagaimana yang akan dilakukan dalam penelitian selanjutnya.	Penelitian ini menganalisis sentimen pembelajaran daring, sedangkan penelitian yang akan diteliti adalah menganalisis sentimen kepuasan pelanggan jasa ekspedisi J&T <i>Express</i> , JNE dan Sicepat.	Meskipun penelitian ini memberikan wawasan awal tentang sentimen pembelajaran daring, kesimpulan yang lebih mendalam pada penelitian tersebut, serta kurang detailnya penjelasan pada abstrak terkait tahapan penelitian maupun proses pengambilan datanya.	Penelitian ini menggunakan penelitian terdahulu, yaitu dengan menggunakan metode Naïve Bayes agar selaras dengan penelitian dalam menganalisis sentimen dan mengelompokkannya kedalam kelas positif, negatif dan netral.	Penelitian ini menganalisis sentimen terkait pembelajaran daring di Twitter selama pandemi <i>COVID-19</i> . <i>Output</i> analisis menunjukkan sekitar 30% sentimen bersifat positif, 69% bersifat negatif, dan 1% bersifat netral selama periode tersebut. Persentase tinggi sentimen negatif mencerminkan ketidakpuasan masyarakat terhadap pembelajaran daring.

No.	Judul	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
6.	Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naïve Bayes Classifier [9]	Melakukan penelitian dengan pendekatan metode Naïve Bayes, yang bertujuan untuk melakukan analisis terhadap sentimen dan mengelompokkannya ke dalam kategori positif, negatif, dan netral.	Penelitian ini menganalisis sentimen masyarakat terhadap kinerja Dewan Perwakilan Rakyat (DPR), sedangkan penelitian yang akan dilakukan adalah menganalisis sentimen kepuasan pelanggan jasa ekspedisi J&T Express, JNE dan Sicepat.	Tidak dijelaskan dengan rinci jangka waktu pengambilan data pada Twitter yang digunakan untuk penelitian.	Penelitian ini menggunakan metode Naïve Bayes, agar selaras dengan penelitian sebelumnya dalam menganalisis sentimen dan dapat mengelompokkannya kedalam kelas positif, negatif dan netral.	Penelitian ini memiliki tujuan untuk mengevaluasi sentimen terkait kinerja Dewan Perwakilan Rakyat (DPR) di Twitter dengan menerapkan metode Naïve Bayes pada 1546 tweet. Hasil analisis menunjukkan bahwa DPR menerima 95 tweet dengan sentimen positif, memiliki polaritas sekitar 0.75 atau 75%. Sebanyak 693 twit diidentifikasi sebagai netral, memiliki polaritas sekitar 0,79 atau 79% sentimen netral. Sementara itu, sebanyak 758 twit dianggap sebagai negatif, dengan polaritas 0,82 atau 82% sentimen negatif. Akurasi hasil uji mencapai 0.8 atau 80%, berdasarkan pengujian yang mencakup 20% dari total data.

No.	Judul	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
7.	Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naïve Bayes Classifier [10].	Melakukan penelitian dengan menerapkan metode Naïve Bayes untuk menganalisis sentimen. Sumber data yang digunakan adalah dari platform Twitter, sebagaimana halnya pada penelitian yang akan dilaksanakan.	Penelitian ini bertujuan untuk menganalisis sentimen Twitter mengenai dampak virus corona, sedangkan penelitian yang akan dilakukan dengan menganalisis sentimen kepuasan pelanggan J&T Express, JNE dan Sicepat.	Hasil klasifikasi dalam penelitian ini belum optimal dalam hal akurasi, tidak memperhitungkan aspek semantik dalam dokumen yang dianalisis, seperti sinonim, homonim, dan polisemi, yang menyebabkan ketidakstrukturan data yang diambil dan mengakibatkan hasil klasifikasi yang kurang optimal.	Penelitian ini menggunakan penelitian yang sudah dilakukan sebelumnya, yaitu menggunakan metode Naïve Bayes agar selaras dengan penelitian dalam menganalisis sentimen dan mengelompokkannya kedalam kelas positif, negatif dan netral.	Penelitian ini menganalisis sentimen dalam dokumen Twitter terkait dampak virus corona menggunakan metode Naïve Bayes. Hasil penelitian menunjukkan tingkat ketepatan sekitar 67% dan tingkat kesalahan sekitar 33%. Eksperimen yang melibatkan tiga variasi jumlah data yang berbeda (100, 200, dan 500) mendapatkan perbedaan nilai akurasi yang minimal, yakni sekitar 0,02.
8.	<i>Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naïve Bayes</i> [11]	Melakukan penelitian dengan menggunakan metode Naïve Bayes, bertujuan untuk menganalisis sentimen pelanggan dan	Penelitian ini menganalisis sentimen pelanggan restoran ulasan pada TripAdvisor sedangkan penelitian yang akan dilakukan adalah menganalisis	Pada bagian abstrak kurang menggambarkan keseluruhan penelitian, seperti tahapan proses <i>preprocessing</i> metode Naïve Bayes yang digunakan dalam penelitian, hasil akurasi metode Naïve	Penelitian ini menggunakan penelitian yang sudah dilaksanakan sebelumnya, yaitu dengan menggunakan metode Naïve Bayes agar selaras dengan penelitian dalam menganalisis sentimen	Penelitian ini mengklasifikasikan kepuasan pelanggan restoran Surabaya menggunakan Naïve Bayes. Pengambilan sampel data dengan menggunakan WebHarvy Tools. <i>Output</i> penelitian menunjukkan bahwa akurasi metode Naïve Bayes mempunyai

No.	Judul	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
8.		Mengklasifikasi data ke dalam kelas positif, negatif dan netral sama seperti penelitian yang akan dilakukan.	sentimen kepuasan pelanggan jasa ekspedisi J&T Express, JNE dan Sicepat pada aplikasi Twitter.	Bayes dan Textblob tidak dijelaskan secara mendetail hanya menjelaskan perbedaan tingkat akurasi saja.	dan mengelompokkannya kedalam kelas positif, negatif dan netral.	nilai sebesar 72,06% akurasi dan sedikit lebih baik (2,94%) dibandingkan TextBlob analisis sentimen.
9.	Sentiment Analysis of Restaurant Reviews in Social Media using Naïve Bayes [12]	Melakukan penelitian dengan menerapkan metode Naïve Bayes, dengan tujuan untuk menganalisis sentimen pelanggan dan mengkategorikan data ke dalam kelas positif, negatif, dan netral, sama seperti penelitian yang akan dilakukan.	Penelitian ini menganalisis ulasan pelanggan restoran sedangkan penelitian yang akan dilakukan menganalisis kepuasan pelanggan jasa ekspedisi J&T Express, JNE dan Sicepat.	kurang mendalamnya penjelasan pada abstrak dalam mewakili gambaran keseluruhan isi penelitian, seperti tidak adanya penjelasan tahapan dalam penelitian, ataupun penjelasan langkah penggunaan metode Naïve Bayes secara singkat.	Penelitian ini menggunakan penelitian yang sudah dilaksanakan sebelumnya, yaitu dengan menggunakan metode Naïve Bayes agar selaras dengan penelitian dalam menganalisis sentimen dan mengelompokkannya kedalam kelas positif, negatif dan netral.	Penelitian ini dilaksanakan untuk menganalisis sentimen masyarakat terhadap restoran, yang diekspresikan melalui platform media sosial Twitter. Dalam penelitian ini, metode Klasifikasi Naïve Bayes digunakan untuk mengelompokkan data. Temuan penelitian mengindikasikan bahwa masyarakat cenderung memberikan sentimen positif terhadap restoran, dengan akurasi sebesar 73%.

No.	Judul	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
10.	<i>Naïve Bayes Twitter Sentiment Analysis In Visualizing The Reputation Of Communication Service Providers: During Covid-19 Pandemic</i> [13]	Melakukan penelitian dengan menerapkan metode Naïve Bayes guna menganalisis sentimen pelanggan serta mengklasifikasi data ke dalam kelas positif, negatif, dan netral. Sumber data diambil melalui platform Twitter, sebagaimana yang direncanakan dalam penelitian yang akan datang.	Penelitian ini menganalisis sentimen reputasi penyedia layanan komunikasi saat pandemic Covid-19 sedangkan penelitian yang akan dilakukan berfokus pada analisis sentimen kepuasan pelanggan jasa ekspedisi J&T Express, JNE dan Sicepat.	Pada penelitian ini meskipun disebutkan bahwa visualisasi dilakukan melalui <i>word clouds</i> , tidak dijelaskan secara rinci bagaimana <i>word clouds</i> digunakan untuk mencapai visualisasi reputasi merek yang disajikan dengan angka negatif, seperti -56%, -42%, dan -43%.	Penelitian ini menggunakan penelitian yang sudah diteliti sebelumnya, yaitu dengan menggunakan metode Naïve Bayes agar selaras dengan penelitian dalam menganalisis sentimen dan mengelompokkannya ke dalam kelas positif, negatif dan netral.	Penelitian ini menggunakan metode Naïve Bayes untuk menganalisis sentimen reputasi merk, menghasilkan skor pengujian akurasi sebesar 89%, keluar dengan visualisasi melalui kata cloud dan menyajikan -56%, -42%, dan -43% Reputasi Merek Bersih untuk Celcom, Maxis, dan Digi.

Dari data yang ada pada Tabel 2.1, dapat dinilai bahwa metode Naïve Bayes dapat diterapkan dalam analisis sentimen berdasarkan teks atau data yang dianalisis dan memiliki kemampuan untuk memberikan hasil klasifikasi dengan akurasi yang cukup tinggi dan menghasilkan hasil yang memuaskan. Persamaan penelitian ini dari penelitian yang sebelumnya yaitu menganalisis sentimen kepuasan. Perbedaan penelitian tugas akhir ini terletak pada studi kasus yang diterapkan. Dapat disimpulkan bahwa sebelumnya belum ada penelitian yang melibatkan **Analisis Sentimen Kepuasan Pelanggan JNT *Express*, JNE, dan Sicepat Pada Twitter Menggunakan Metode Multinomial Naïve Bayes**. Kontribusi yang diberikan dalam penelitian ini memberikan wawasan yang berharga kepada penyedia layanan ekspedisi JNT *Express*, JNE dan Sicepat serta pemangku kepentingan untuk meningkatkan kualitas layanan mereka, dengan memahami sentimen yang dinyatakan pelanggan secara *online*.

2.2 Landasan Teori

Dalam tahapan ini, akan diuraikan dasar teori yang berkaitan dengan topik penelitian ini. Beberapa teori yang akan diterapkan meliputi:

2.2.1 Analisis Sentimen

Analisis sentimen adalah tahapan ekstraksi, pengolahan, dan pemahaman data teks yang tidak terstruktur secara otomatis, dengan tujuan mendapatkan informasi mengenai sentimen yang terdapat dalam suatu kalimat pendapat atau opini [14]. Umumnya, dalam penelitian analisis sentimen, sentimen hanya dibagi menjadi dua atau tiga kelas, yakni positif, negatif, dan netral [15].

2.2.2 Kepuasan Pelanggan

Kepuasan pelanggan merupakan persentase sejauh mana produk ataupun layanan memenuhi harapan dan kebutuhan konsumen. Berbagai faktor yang berpengaruh pada kepuasan pelanggan melibatkan kualitas produk, layanan, harga, dan pengalaman secara keseluruhan [16].

2.2.3 JNE



Gambar 2.1 Logo perusahaan JNE

PT Tiki Jalur Nugraha Ekakurir atau JNE adalah perusahaan ekspedisi terkemuka di Indonesia saat ini. Lebih dari 150 lokasi PT Tiki JNE telah terhubung melalui sistem komunikasi *online*, dengan dukungan sistem dan akses informasi melalui situs yang efektif dan efisien, memungkinkan konsumen untuk memantau status terbaru pengiriman paket atau dokumen [17]

2.2.4 JNT Express



Gambar 2.2 Logo perusahaan JNT Express

JNT Express adalah perusahaan ekspedisi Indonesia yang berada di peringkat kedua sebagai merek jasa ekspedisi terkemuka di Indonesia. Perusahaan ini relatif lebih baru daripada JNE. Perusahaan ini memiliki beberapa keunggulan, termasuk sebagai penyedia layanan pengiriman barang yang menjadi pelopor dengan operasional 365 hari tanpa mengenal hari libur, dan mencakup seluruh wilayah Indonesia tanpa melibatkan pihak lain. Perusahaan ini juga menjadi yang pertama di Indonesia yang secara khusus berfokus pada bisnis online dalam layanan pengiriman, dan sejak tahun 2018 telah berhasil meraih Penghargaan *Top Brand*. Selain itu, perusahaan berhasil memperluas cakupan pengiriman internasional ke beberapa negara tetangga seperti Malaysia, Vietnam, Filipina, Thailand, Singapura, dan Kamboja [18].

2.2.5 Sicepat



Gambar 2.3 Logo perusahaan Sicepat

SiCepat yaitu salah satu perusahaan pengiriman barang yang menyediakan tingkat kecepatan dalam proses pengiriman di Indonesia berdiri pada tahun 2014, didirikan oleh The Kim Hai dan Rudy Darwin Swigo. Kecepatan dan keamanan bantuan yang disediakan oleh Sicepat telah memperoleh perhatian banyak perusahaan bisnis *internet* yang tertarik untuk menjalin kerjasama, memberikan SiCepat reputasi yang luas dan dapat dipercaya [19].

2.2.6 Twitter

Twitter adalah *platform* media sosial yang terkenal dan gratis, menyajikan layanan jaringan yang memungkinkan pengguna untuk berbagi opini dalam bentuk teks melalui pesan singkat, yang dikenal sebagai *tweet*. *Platform* gratis yang dinamis dan terus berkembang, memiliki dampak signifikan dalam berbagai aspek kehidupan [20].

2.2.7 Multinomial Naïve Bayes

Multinomial Naïve Bayes termasuk dalam kategori metode *supervised learning* [21]. Multinomial Naïve Bayes menggunakan cabang matematika yang dikenal sebagai teori probabilitas untuk menemukan peluang klasifikasi yang paling besar dengan melihat frekuensi masing-masing klasifikasi dalam data latihan [22]. Rumus Naïve Bayes adalah:

$$P(H|X) = \frac{P(H|X)P(H)}{P(X)} \quad (2.1)$$

Keterangan:

X = Data kelas yang belum teridentifikasi.

H = Hipotesis bahwa data X termasuk dalam suatu kelas tertentu

$P(H|X)$ = Peluang hipotesis H pada kondisi X

$P(H)$ = Peluang hipotesis H

$P(X|H)$ = Peluang X pada kondisi H

$P(X)$ = Peluang dari X

2.2.8 Google Colaboratory

Google colaboratory adalah *tools* yang dikembangkan oleh tim penelitian internal Google untuk mendukung peneliti dalam mengolah data dengan menggunakan bahasa pemrograman Python, baik untuk keperluan pembelajaran maupun eksperimen, terutama dalam bidang *Machine Learning* [23]. Dari segi perangkat lunak, *google colaboratory* telah menyediakan sebagian besar pustaka yang diperlukan.

2.2.9 Text Mining

Text mining ialah tahap eksplorasi dan analisis data teks tidak terstruktur dengan bantuan *software* yang dapat mengenali ide, pola, subjek, kata kunci, dan karakteristik lainnya dalam dataset tersebut [24]. *Text mining* bisa disebut sebagai hasil penggabungan antara *analysis teks* dan *data mining* karena tujuannya untuk memperoleh representasi dari sebuah dokumen atau teks melalui penerapan teknik-teknik yang umumnya terdapat dalam aspek data mining [25]. Objek *text mining* adalah dokumen yang tidak terstruktur atau semi-terstruktur [26]. Tugas-tugas spesifik dari *Text Mining* meliputi kategorisasi teks dan pengelompokan teks.

2.2.10 Text Processing

Text Mining memanfaatkan data yang berasal dari dokumen yang memiliki sifat tidak terstruktur. Karena data tersebut tidak terstruktur, diperlukan tahap *text processing* untuk mengorganisir teks dan meningkatkan struktur. *Text processing* merupakan tahapan aplikasi melakukan pemilihan data yang akan diproses pada setiap dokumen [27]. Tahapan *text processing* antara lain:

- a. *Cleaning* merujuk pada proses dokumen dibersihkan dari kata yang tidak berhubungan dan tidak diperlukan dengan maksud untuk meminimalisir gangguan. Selain penghapusan kata-kata, tindakan penghapusan tanda baca, seperti titik (.), koma (,), dan simbol lain, turut dilaksanakan.
- b. *Case folding* adalah langkah untuk mengonversi semua huruf dalam dokumen menjadi huruf kecil.

- c. *Normalization* adalah perbaikan kata yang tidak baku menjadi baku.
- d. *Tokenizing* adalah langkah di mana kata-kata dipotong atau teks dipecah menjadi potongan kata atau kalimat, yang sering dikenal sebagai token. Kata-kata dalam dokumen dipisahkan berdasarkan kalimat yang dibatasi oleh spasi.
- e. *Stopward removal* ialah proses eliminasi kata yang tidak memiliki arti atau tidak penting, seperti "aku", "atau", "saja", serta sejenisnya.
- g. *Stemming* adalah langkah untuk mengembalikan kata-kata yang memiliki imbuhan kembali ke bentuk aslinya.

2.2.11 Word Cloud

Word Cloud merupakan visualisasi yang menggambarkan frekuensi kemunculan kata-kata dalam sebuah teks. Dalam representasi ini, ukuran huruf mencerminkan frekuensi kemunculan kata tersebut, sehingga hurufnya semakin besar, semakin tinggi frekuensinya, dan sebaliknya, semakin kecil hurufnya, semakin rendah frekuensinya [28].

2.2.12 Fitur dan Pembobotan

Pendekatan pembobotan yang diterapkan adalah *Term Frequency-Inverse Document Frequency* (TF-IDF) [29]. TF-IDF merupakan sebuah metode pembobotan yang terbentuk dari dua nilai yang dihasilkan oleh dua metode pembobotan yang berbeda, yakni *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). Dengan menerapkan pendekatan ini, dapat ditarik kesimpulan bahwa kata yang banyak muncul dalam suatu dokumen akan memiliki nilai TF-IDF yang tinggi, sedangkan kata yang umum di banyak dokumen akan memiliki nilai yang rendah. Proses TF-IDF memiliki rumus sebagai berikut:

- a. Rumus *Term Frequency* (TF)

$$tf = tf_{ij} \tag{2.2}$$

Keterangan:

tf = *Term Frequency*

i = *Term*

j = *Dokumen*

b. Rumus Inverse Document Frequency (IDF)

$$idf_i = \log \frac{N}{df_i} \quad (2.3)$$

Keterangan:

idf_i = Inverse Document Frequency

N = Total dokumen

df_i = Total dokumen yang terdapat term (f_i)

c. Nilai bobot (W)

$$w_{ij} = tf_{ij} \times idf_i \quad (2.4)$$

Keterangan:

w_{ij} = Bobot TF-IDF

tf_{ij} = Term Frequency

idf_i = Invers Document Frequency

2.2.13 Classification

Klasifikasi adalah proses memisahkan objek atau data ke dalam *class* yang telah ditentukan sebelumnya secara keseluruhan, klasifikasi melibatkan dua tahap, yakni proses klasifikasi data pelatihan untuk menghasilkan suatu model. Proses pembelajaran algoritma klasifikasi melibatkan pemrosesan data latihan guna membentuk model. Setelah model melalui uji dan dianggap memadai, Pada tahap klasifikasi, model tersebut dipergunakan untuk melakukan prediksi terhadap kelas dari kasus baru, yang mendukung proses pengambilan keputusan [30]

2.2.14 Confusion Matrix

Confusion matrix adalah tabel yang mencerminkan total data pengujian yang dikelompokkan dengan tepat dan jumlah data pengujian yang dikelompokkan secara tidak tepat [31]. Dengan adanya *confusion matrix*, kinerja suatu model dapat diukur berdasarkan seberapa sering model tersebut berhasil melakukan prediksi dengan benar. Adapun Tabel 2.2 sebagai contoh tabel *confusion matrix*:

Tabel 2.2 Contoh tabel *confusion matrix*

		<i>Predicted Class</i>		
		<i>Positive</i>	<i>Neutral</i>	<i>Negative</i>
<i>True Class</i>	<i>Positive</i>	<i>True Positive (TP)</i>	<i>False Neutral (FNt)</i>	<i>False Negative (FN)</i>
	<i>Neutral</i>	<i>False Positive (FP)</i>	<i>True Neutral (TNt)</i>	<i>False Negative (FN)</i>
	<i>Negative</i>	<i>False Positive (FP)</i>	<i>False Neutral (FNt)</i>	<i>True Negative (TN)</i>

Apabila sebuah model berhasil memprediksi dengan benar bahwa suatu komentar memiliki label positif, hal tersebut dapat diidentifikasi sebagai *True Positive* (TP) dalam Tabel 2.2. Demikian pula, apabila model berhasil menebak label negatif dengan benar, hal tersebut dapat dianggap sebagai *True Negative* (TN) dalam perhitungan. Sebaliknya, jika model memprediksi sebuah komentar dengan label positif namun sebenarnya tidak, itu bisa dianggap sebagai *False Positive* (FP). *False Negative* (FN) juga mencakup prediksi yang keliru terhadap label negatif. *True Neutral* (TNt) terjadi ketika model berhasil mengidentifikasi komentar yang sebenarnya netral sebagai netral. *False Neutral* (FNt) terjadi ketika model salah mengidentifikasi komentar yang sebenarnya netral sebagai positif atau negatif

Perhitungan *confusion matrix* memiliki empat rumus yaitu untuk menghitung *accuracy*, *precision*, *recall* dan *f1-score* [32], antara lain:

- a. *Accuracy*, menghitung seberapa baik model dapat mengklasifikasikan data dengan benar. Rumus *Accuracy* :

$$Accuracy = \frac{(TP+TN)}{TP+TN+FP+FN} \quad (2.5)$$

- b. *Precision* mengacu pada perhitungan ketika kelas aktual dan kelas yang diprediksi sama (positif), yang dikenal sebagai kondisi benar positif untuk semua prediksi yang diposisikan sebagai positif. Rumus *Precision*:

$$Precision = \frac{TP}{TP+FP} \quad (2.6)$$

- c. *Recall* adalah penghitungan ketika kelas yang diprediksi positif sesuai dengan kondisi aktual positif dalam semua kasus yang sebenarnya positif.

$$Recall = \frac{TP}{TP+FN} \quad (2.7)$$

- d. Skor F1 merupakan metrik yang menggabungkan nilai rata-rata dari *recall* dan *precision* sebagai penilaian kualitas model.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.8)$$