

BAB II

TINJAUAN PUSTAKA

2.1 Analisis Sentimen

Analisis sentimen merupakan salah satu bagian dari studi *text mining* yang bisa disebut studi komputasi untuk melakukan klasifikasi terhadap pendapat, emosi dan sikap seseorang terhadap entitas (Hudaya, Fakhurroja and Alamsyah, 2019). Analisis Sentimen ini dikenal juga sebagai *opinion mining* atau pun *emotion artificial intelligence* yang berguna untuk memproses bahasa alami, analisis teks dan komputasi linguistik untuk bisa mengidentifikasi, mengekstrak, menghitung serta mempelajari informasi secara terstruktur (Pratama, Andrean and Nugroho, 2019). Analisis sentimen bisa disebut sama dengan *opinion mining* karena pada metode ini berfokus pada pendapat yang bersifat positif, netral atau pun negatif (Samsir *et al.*, 2021). Analisis sentimen banyak digunakan oleh peneliti sebagai cabang penelitian di bidang ilmu komputer.

Metode analisis sentimen ini mengekstrak sebuah data opini dengan memahami dan memproses data yang bersifat *textual* secara otomatis agar bisa dilihat sentimen yang terkandung pada opini tersebut. Analisis sentimen juga dapat dibedakan berdasarkan sumber data. Terdapat beberapa level yang paling sering digunakan pada penelitian analisis sentimen, yaitu diantaranya pada level dokumen dan pada level kalimat. Analisis sentimen ini terbagi menjadi 2 kelompok besar, yaitu diantaranya *Coarse-gained* dan *Fine-gained* (Pertiwi, 2019). Analisis sentimen *Coarse-gained* ini merupakan analisis yang hanya dilakukan pada dokumen yang secara garis besar menganggap seluruh isi dokumen sebagai suatu sentimen positif

dan sentimen negatif. Analisis sentimen *Fine-grained* ini merupakan analisis yang dilakukan hanya pada level kalimat. Fokus utamanya ialah menentukan sentimen pada kalimat saja (Ardiani, Sujaini and Tursina, 2020). Penerapan analisis sentimen ini biasanya digunakan diberbagai hal seperti halnya informasi konsumen, pemasaran, politik serta sosial. Di pemerintahan atau pun kesehatan dan lain-lain, analisis sentimen ini bisa digunakan untuk mengetahui opini masyarakat terhadap suatu isu yang sedang terjadi sehingga pemerintah pun bisa membuat solusi yang tepat berdasarkan data yang telah dikumpulkan (Savitri *et al.*, 2021). Terdapat beberapa kelas dalam analisis sentimen, yaitu diantaranya :

1. Sentimen Positif : Menurut Kamus Besar Bahasa Indonesia atau disingkat menjadi KBBI, sentimen positif merupakan sebuah reaksi atau sikap yang meningkatkan nilai seseorang atau terhadap sesuatu (Ardiani, Sujaini and Tursina, 2020).
2. Sentimen Negatif : Menurut Kamus Besar Bahasa Indonesia atau disingkat menjadi KBBI, sentimen negatif merupakan sebuah reaksi atau sikap yang bisa menurunkan nilai seseorang atau terhadap sesuatu hingga membuat tren menjadi *down*. Kalimat yang bersentimen negatif ditandai dengan penggunaan kata yang negasi. Biasanya negasi ini digunakan untuk mengubah pola dari suatu pernyataan (Ardiani, Sujaini and Tursina, 2020).
3. Sentimen Netral : Menurut Kamus Besar Bahasa Indonesia atau disingkat menjadi KBBI, sentimen negatif merupakan sebuah reaksi yang tidak berpihak. Biasanya kalimat sentimen netral ini tidak memiliki ekspresi kalimat yang tidak bersifat positif maupun kalimat negatif (Ardiani, Sujaini and Tursina, 2020).

2.2 Text mining

Text mining merupakan bagian dari *data mining* yang merupakan proses untuk ekstraksi pengetahuan dan informasi dari pola pada sekumpulan dokumen berupa *text* dengan menggunakan metode analisis tertentu (Eni Tri Handayani and Ari Sulistiyawati, 2021). Meskipun *text mining* merupakan bagian dari *data mining*, namun keduanya berbeda dalam artian bahwa *data mining* lebih melibatkan data yang terstruktur, sementara teks berkaitan dengan fitur tertentu dan relatif tidak terstruktur serta memerlukan *preprocessing* (Pertiwi, 2019). *Text mining* juga memiliki fungsi untuk mencari kata yang dapat mewakili apa saja yang terdapat pada dokumen sehingga dilakukan analisis keterhubungan pada dokumen.

Text mining merupakan sebuah bidang yang mengacu terhadap pencarian informasi, pertambangan data, statistik, pembelajaran mesin dan komputasi linguistik tentunya. *Text mining* ini juga digunakan untuk menangani klasifikasi, pengelompokan dan pengambilan informasi. Metode ini bertujuan untuk bisa menghasilkan informasi dari suatu dokumen menjadi sebuah informasi dengan melakukan tahapan pemrosesan, pengelompokan dan analisis data yang sebelumnya tidak terstruktur. *Text mining* ini dalam pengambilan informasinya menghasilkan analisis perasaan yang bisa dirasakan secara emosional serta mengidentifikasi pernyataan antara positif atau pun negatif (Hudaya, Fakhurroja and Alamsyah, 2019). *Text mining* ini memiliki beberapa tahapan, yaitu diantaranya sebagai berikut :

1. *Tokenizing*

Tahapan *Tokenizing* ini merupakan proses pemotongan string yang memasukkan berdasarkan tiap kata yang tersusun. Proses ini memisahkan setiap kata yang menyusun sebuah dokumen (Pratama, Andrean and Nugroho, 2019). *Tokenizing* ini dilakukan untuk menghasilkan kumpulan kata serta menghilangkan delimiter seperti halnya titik (.), koma (,), spasi dan karakter angka yang terdapat pada kata tersebut. Tujuan dari *Tokenizing*, yaitu untuk mengeksplorasi kata dalam sebuah kalimat. *Tokenizing* dilakukan untuk memecahkan sebuah kata dan melakukan penghapusan terhadap delimiter serta karakter angka bersama entitas *tweet* seperti halnya *hashtag*, *retweet* dan *mention* (Ardiani, Sujaini and Tursina, 2020).

2. *Filtering*

Tahapan *Filtering* ini dilakukan setelah tahap *Tokenizing* yang merupakan sebuah proses yang dilakukan pengambilan sebagian data dari data-data tertentu dan membuangnya pada variabel yang lain (Pratama, Andrean and Nugroho, 2019). *Filtering* membuang kata yang tentunya tidak akan digunakan atau bisa disebut *stopword* yang terdapat dalam *bag of words stopwords*. *Stopword* merupakan daftar kata yang tidak mempresentasikan isi dari suatu dokumen teks. *Stopword* ini dilakukan untuk menghilangkan kata yang bisa disebut tidak memiliki arti. Sebuah daftar *stopword* akan dibuat sebelum melakukan proses *stopword removal*. Jika terdapat kata dalam daftar *stopword*, maka kata tersebut akan dihapus sehingga kata yang tersisa hanya dianggap kata yang menandakan isi dari suatu dokumen (Ardiani, Sujaini and Tursina, 2020).

3. *Stemming*

Tahapan *Stemming* ini merupakan sebuah proses yang dilakukan pemetaan dan penguraian dari berbagai varian dari suatu kata menjadi bentuk kata dasar (Pratama, Andrian and Nugroho, 2019). *Stemming* dilakukan untuk menyeragamkan sebuah bentuk kata. Tujuan tahapan *Stemming* ini ialah menghilangkan imbuhan baik itu prefiks, sufiks atau pun konfiks yang terdapat pada setiap kata. Proses ini biasanya menggunakan *library* yang terdapat pada bahasa pemrograman *python*, yaitu *sastrawi* (Ardiani, Sujaini and Tursina, 2020).

4. *Tagging*

Tahapan *Filtering* ini merupakan sebuah proses yang menyusun sebuah informasi berdasarkan kategorinya (Pratama, Andrian and Nugroho, 2019).

5. *Analyzing*

Tahapan *Analyzing* ini merupakan proses untuk mencari seberapa jauh hubungan antara kata dari setiap dokumen (Pratama, Andrian and Nugroho, 2019).

2.3 Naïve Bayes Classifier

Metode *Naïve Bayes Classifier* ini merupakan sebuah metode klasifikasi dalam hal penambangan teks yang terdapat dalam analisis sentimen yang mempelajari sebuah emosi atau pun sikap. Metode ini banyak digunakan dalam teknik klasifikasi salah satunya pada media sosial twitter. Algoritma ini menggunakan perhitungan probabilitas dengan menggunakan statistik berdasarkan teori bayes yang berasumsi keberadaan serta ketiadaan dari suatu kelas (Ahmadi, Gustian and Sembiring,

2021). Proses pada metode ini, yaitu akan mencari probabilitas yang tinggi dari semua kategori yang ada pada dokumen yang diujikan.

Metode *Naïve Bayes* ini bisa mengambil keputusan berdasarkan data yang diperoleh sebelumnya. Walaupun independensi pada tiap kata pada dokumen tidak dapat bekerja secara maksimal, tetapi kinerja *Naïve Bayes* ini dalam melakukan klasifikasi sangatlah baik. Metode *Naïve Bayes* ini memiliki keunggulan, yaitu diantaranya sangat mudah untuk diimplementasikan, cepat dalam melakukan klasifikasi data, serta memiliki akurasi yang cukup tinggi (Rahat, Kahir and Masum, 2020). Secara umum rumus *Naïve Bayes* ini pada dasarnya adalah sebagai berikut (Tuhenay, 2021) :

$$P(C_i/X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

Keterangan :

- a. $P(C_i/X)$: Probabilitas C_i dengan bukti X
- b. $P(C_i)$: Probabilitas C_i
- c. $P(X/C_i)$: Probabilitas X berpengaruh terhadap C_i
- d. $P(X)$: Probabilitas X

2.4 Twitter

Twitter merupakan media sosial yang berguna untuk mengekspresikan sebuah opini serta perasaan bagi penggunanya terhadap isu-isu atau permasalahan yang sedang terjadi atau sebelumnya terjadi. Twitter menjadi media sosial yang membuat banyak orang dari seluruh dunia bisa mengungkapkan pendapat mereka masing-masing. Twitter ini didirikan pada tahun 2006 pada bulan maret oleh Jack Dorsey. Twitter ini menjadi salah satu dari sepuluh situs yang paling sering

dikunjungi pada situs internet serta sempat dijuluki sebagai pesan singkat dari layanan internet (Pratama, Andrean and Nugroho, 2019). Media sosial twitter ini juga menjadi media sosial yang populer digunakan sebagai sumber data penelitian analisis sentimen. Media sosial twitter ini juga akan tetap terhubung dengan yang lain walaupun tidak saling berteman satu sama lain (Pravina, Cholissodin and Adikara, 2019).

Twitter ini juga bisa berguna untuk pengguna bertukar pikiran atau pun ide secara bebas dengan cakupan nasional maupun internasional seperti halnya antara pengguna yang tertarik pada bidang keahlian tertentu yang serupa dan juga pengguna bisa terlibat dalam debat kritis di dalam sebuah komentar (Savitri *et al.*, 2021). Twitter ini lebih banyak digunakan dibandingkan dengan media sosial lainnya karena dengan adanya ketersediaan data, mudah untuk diakses serta memiliki banyak konten. Media sosial ini memiliki jutaan *tweet* dengan berbagai jenis topik. Membuktikan bahwa media sosial menjadi sumber informasi yang sangat penting.

2.5 Rapidminer

Rapidminer merupakan sebuah *platform* perangkat lunak (*software*) data ilmu pengetahuan yang telah dikembangkan oleh perusahaan. Rapidminer ini menyediakan lingkungan terpadu untuk pembelajaran mesin atau bisa disebut *machine learning*, pembelajaran mesin atau bisa disebut *Deep Learning*, penambangan teks atau bisa disebut *text mining* dan analisis prediktif atau bisa disebut *predictive analytics*. Aplikasi Rapidminer digunakan untuk melakukan bisnis, penelitian, pendidikan, pelatihan, membuat *prototype* secara cepat serta

mengembangkan aplikasi dan mendukung semua langkah poses pembelajaran mesin seperti halnya persiapan data, visualisasi hasil, validasi dan optimalisasi (Savitri *et al.*, 2021).

2.6 Support Vector Machine

Algoritma ini merupakan sebuah algoritma *machine learning* yang masuk ke dalam jenis *Supervised Learning* yang bisa memprediksi kelas berdasarkan proses pelatihan. Algoritma ini juga cocok untuk melakukan pengklasifikasian teks, namun harus diubah dulu menjadi format yang lebih terstruktur dan bersifat *polynomial*. Nilai yang dihasilkan pada algoritma ini merupakan sebuah garis pemisah atau yang biasa disebut *hyperplane*, yang nantinya garis tersebut akan berfungsi untuk memisahkan *tweet* antara *tweet* yang berlabel positif, netral dan negatif (Pravina, Cholissodin and Adikara, 2019). Algoritma ini dibutuhkan sebuah kernel untuk bisa meningkatkan akurasi ke ruang dimensi yang lebih tinggi yang sering disebut ruang kernel. Terdapat beberapa macam kernel pada algoritma ini yaitu diantaranya (Rahman Isnain *et al.*, 2021) :

1. Linier

$$K(x, y) = x \cdot y \quad (1)$$

2. Polynomial

$$K(x_i, x_j) = ((x_i \cdot x_j) + c)^d \quad (2)$$

3. Gaussian

$$K(x_i, x_j) = \exp\left(-\frac{|x_i - x_j|^2}{2\sigma^2}\right) \quad (3)$$

4. Sigmoid

$$K(x_i, x_j) = \tanh(\sigma x_i \cdot x_j + \beta) \quad (4)$$

Keterangan :

- a. $K(x_i, x_j)$ = Fungsi Kernel
- b. x_i = Data ke-i
- c. x_j = Data ke-j
- d. exp = Operasi eksponen
- e. d, σ = Parameter kernel
- f. β = Matriks *output weight*

Awalnya algoritma *Support Vector Machine* ini digunakan untuk persoalan klasifikasi 2 kelas, kemudian seiring berjalannya waktu dikembangkan kembali untuk klasifikasi *multi class* (lebih dari 2 kelas). Salah satu metode pendekatan yang terdapat pada *multi class* ialah *One Against All* (OAA). Metode OAA yang artinya satu lawan semua yang digunakan untuk kasus klasifikasi k-kelas, menemukan k *hyperplane* di mana k adalah banyak kelas dan p adalah *hyperplane*. Metode ini p (ℓ) diujikan dengan semua data dari kelas ℓ dengan label +1, dan semua data dari kelas lain dengan label -1 (Fathurrohman, Lambang G. H and Susilo, 2019).

2.7 Adaboost

Adaboost merupakan sebuah algoritma yang dapat menghubungkan suatu klasifikasi lemah menjadi klasifikasi kuat karena mempelajari kesalahan klasifikasi dari setiap klasifikasi yang lemah dan menyesuaikan bobot kontribusi untuk klasifikasi akhir (Gao *et al.*, 2016). Terdapat rumus perhitungan Adaboost yaitu diantaranya (Suhendra and Jondri, 2022) :

1. *Sample Weight*

$$w(x_i, y_i) = \frac{1}{N}, i = 1, 2, \dots, n \quad (1)$$

2. *Amount of say*

$$\text{amount of say} = \frac{1}{2} \log \frac{1-\text{Total Error}}{\text{Total Error}} \quad (2)$$

3. *Performance of The Stump*

$$\text{performance of the stump} = \frac{1}{2} \log_e \left(\frac{1-\text{Total Error}}{\text{Total Error}} \right) \quad (3)$$

4. *New Sample Weight*

$$\text{new sample weight} = \text{sample weight} * e^{\pm \text{amount of say}} \quad (4)$$

Keterangan :

- a. $w(x_i, y_i)$ = bobot sampel
- b. N = Jumlah sampel data

2.8 Seleksi Fitur

Seleksi fitur merupakan sebuah metode peningkatan kinerja *classifier* yang berguna untuk meningkatkan akurasi serta mengurangi waktu komputasi (Rahman Isnain *et al.*, 2021). Terdapat beberapa macam pada metode seleksi fitur ini yaitu diantaranya TF-IDF, *Chi Square*, *Information Gain* dan lain-lain. Salah satu seleksi fitur yang digunakan pada penelitian ini adalah TF-IDF atau biasa disebut pembobotan. Pembobotan kata ini dilakukan untuk menentukan bobot pada setiap kata yang nantinya akan dijadikan penilaian. Terdapat rumus untuk menghitung bobot masing-masing dokumen terhadap kata kunci yaitu (Wiyanto, Priatna and Hidayat, 2019) :

$$Wdt = tf dt * IDFt$$

Keterangan :

- a. d = dokumen ke-d
- b. t = kata ke-t dari kata kunci

- c. W = bobot dokumen ke-d terhadap kata ke-t
- d. tf = banyaknya kata yang dicari pada sebuah dokumen
- e. IDF = Inversed Document Frequency $IDF = \log_2 (D/df)$
- f. D = total dokumen
- g. df = banyak dokumen yang mengandung kata yang dicari

2.9 Confusion Matrix

Confusion Matrix ini merupakan sebuah metode pengujian dalam penentuan klasifikasi (Fathurrahman1, Yupi Kuspani Putra2, 2020). *Confusion Matrix* ini adalah tabel 4 dimensi atau lebih berbeda dari nilai prediksi dan nilai aktual. Penelitian ini terdapat 6 tabel dimensi yaitu TP (*True Positive*), TN (*True Negative*), TNeu (*True Neutral*), FP (*False Positive*), FN (*False Negative*), FNeu (*False Neutral*) (Wafa *et al.*, 2022). Hasil proses klasifikasi pada *Confusion Matrix* dihitung akurasi, presisi dan *recall*. Terdapat rumus dari akurasi, presisi dan *recall* yaitu diantaranya (Rahman Isnain *et al.*, 2021) :

1. Akurasi

$$\frac{TP+TN+TNeu}{TP+TN+TNeu+FP+FN+FNeu} \quad (1)$$

2. Presisi

$$\frac{TP}{TP+FP} \quad (2)$$

3. *Recall*

$$\frac{TP}{TP+FN+FNeu} \quad (3)$$

Keterangan :

- a. TP = *True Positive*
- b. TN = *True Negative*

- c. TNeu = *True Neutral*
- d. FP = *False Positive*
- e. FN = *False Negative*
- f. FNeu *False Neutral*

2.10 *State Of The Art (SOTA)*

Banyak yang sudah melakukan penelitian sebelumnya mengenai Analisis sentimen dengan studi kasus serta menggunakan beberapa algoritma yang berbeda-beda. Berikut *State Of The Art* yang dapat dilihat pada Tabel 2.1 berikut ini.

Tabel 2.1 *State Of The Art* Penelitian

No	Nama Penulis	Tahun	Judul	Algoritma	Hasil
1	Eni Tri Handayani, Ari Sulistiyawati	2021	Analisis Sentimen Respon Masyarakat Terhadap Kabar Harian Covid-19 Pada Twitter Kementerian Kesehatan Dengan Metode Klasifikasi <i>Naive Bayes</i>	<i>Naïve Bayes</i>	Penelitian ini menghasilkan sentimen negatif sebesar 77%. Keakuratan hasil pengujian akurasi sebesar 78%. dan pengujian precision sebesar 92%, Recall sebesar 85%, F1-Score sebesar 88%. Menggunakan pengujian metode <i>Naive Bayes Classifier</i> .
2	Melisa Winda Pertiwi	2019	Analisis Sentimen Opini Publik Mengenai Sarana Dan Transportasi Mudik Tahun 2019 Pada Twitter Menggunakan Algoritma <i>Naïve Bayes, Neural Network, KNN Dan SVM</i>	<i>Naïve Bayes, K-NN, SVM, Neural Network</i>	Hasil sentimen publik yang diambil dengan kata kunci Mudik Hari Raya 2019 menunjukkan bahwa pengguna Twitter lebih banyak memberikan opini positif. Algoritma K-NN akurasi lebih tinggi dibandingkan algoritma SVM, <i>Naïve Bayes</i> dan <i>Neural Network</i> untuk melakukan klasifikasi sentimen analisis teks berbahasa Inggris dengan nilai accuracy=90.76% dan AUC=0.939.

Tabel 2.2 *State Of The Art* Penelitian (Lanjutan 1)

No	Nama Penulis	Tahun	Judul	Algoritma	Hasil
3	Arsya Monica Pravina, Imam Cholissodin, Putra Pandu Adikara	2019	Analisis Sentimen Tentang Opini Maskapai Penerbangan Pada Dokumen Twitter Menggunakan Algoritme <i>Support Vector Machine</i> (SVM)	SVM	Metode klasifikasi <i>Support Vector Machine</i> dengan fitur <i>Lexicon Based</i> mendapatkan akurasi paling baik sebesar 40%, precision sebesar 40%, recall sebesar 100%, dan f-measure sebesar 57,14%. Tingkat akurasi tersebut didapatkan dengan jumlah iterasi maksimum sebanyak 50 kali dengan diimplementasikannya fitur <i>Lexicon Based</i> .
4	Muhammad Iqbal Ahmadi, Dudih Gustian, Falentino Sembiring	2021	Analisis Sentiment Masyarakat Terhadap Kasus Covid-19 Pada Media Sosial <i>Tweet</i> Dengan Metode <i>Naive Bayes</i>	<i>Naive Bayes</i>	Tanggapan masyarakat lebih condong ke sisi negatif. dengan jumlah komentar negatif sebanyak 800 komentar dan jumlah komentar positif yang di berikan masyarakat sebanyak 361 komentar. Kinerja dari metode algoritma <i>Naive Bayes</i> berjalan dengan cukup memuaskan dengan menghasilkan tingkat akurasi sebesar 74%.
5	Muhammad Iqbal Ahmadi, Dudih Gustian, Falentino Sembiring	2020	Analisis Sentimen Pro Dan Kontra Masyarakat Indonesia Tentang Vaksin COVID-19 Pada Media Sosial Twitter	<i>Latent Dirichlet Allocation</i> (LDA)	Masyarakat lebih banyak memberikan respon positif terhadap wacana tersebut (30%) dibandingkan dengan respon negatifnya (26%). Model LDA yang dibangun juga dapat menangkap topik yang dibicarakan masyarakat terkait wacana vaksinasi tersebut seperti pembicaraan masyarakat mengenai kontroversi vaksin yang dinilai terlalu terburu-buru.

Tabel 2.3 *State Of The Art* Penelitian (Lanjutan 2)

No	Nama Penulis	Tahun	Judul	Algoritma	Hasil
6	Andris Faesal, Aziz Muslim, Aditya Hastami Ruger, Kusri	2020	Sentimen Analisis Pada Data <i>Tweet</i> Pengguna Twitter Terhadap Produk Penjualan Toko Online Menggunakan Metode <i>K-Means</i>	<i>K-Means</i>	Berdasarkan hasil analisis dari 1.130 <i>tweet</i> yang berhubungan dengan Tokopedia dengan menggunakan metode <i>K-Means</i> mendapatkan tingkat akurasi sebesar 92.86 %. Proses yang sama pada data <i>tweet</i> harian, diperoleh kata <i>tweet</i> “mulaiajadulu” dan “toppers” merupakan kata yang sering digunakan di hampir semua hari dengan tingkat akurasi sebesar 89.80 %.
7	Bagus Muhammad Akbar, Ahmad Taufiq Akbar, Rochmat Husaini	2021	<i>Classification Of Sentiments On Twitter Opinions With The Keyword Sinovac Using Naive Bayes</i>	<i>Naïve Bayes</i>	Analisis sentimen menggunakan metode <i>Naïve Bayes</i> menghasilkan 1433 (71,65%) sentimen positif, 403 (20,15%) sentimen negatif dan 164 (8,2%) sentimen netral. Analisis sentimen menggunakan metode <i>Valence Shifter-Lexicon Based</i> menghasilkan 903 (45,15%) sentimen positif, 437 (21,85%) sentimen negatif, dan 660 (33%).
8	Astiko, Achmad Khodar	2020	Membangun Model <i>Machine learning</i> Untuk Meninjau Layanan Indosat Ooredoo Dari Twitter Menggunakan <i>Naive Bayes Classifier</i>	<i>Naïve Bayes</i>	Klasifikasi analisis konten twitter untuk meninjau kualitas layanan provider Indosat Ooredoo menggunakan sentimen positif dan negatif dilakukan dengan baik. Ada beberapa aspek penting di dalam model yang perlu ditingkatkan seperti penggunaan library, keberagaman <i>stemmer</i> dan <i>stopword list</i> .

Tabel 2.4 *State Of The Art* Penelitian (Lanjutan 3)

No	Nama Penulis	Tahun	Judul	Algoritma	Hasil
9	Veny Amilia Fitri, Rachmadita Andreswari, Muhammad Azani Hasibuan	2019	<i>Sentiment Analysis Of Social Media Twitter With Case Of Anti LGBT Campaign In Indonesia Using Naïve Bayes, Decision Tree, And Random Forest Algorithm.</i>	<i>Naïve Bayes, Decision Tree dan Random Forest</i>	Analisis sentimen pada Twitter di Indonesia memberikan komentar yang lebih netral. Penelitian ini didapatkan akurasi sebesar 86,43% dari pengujian data menggunakan algoritma <i>Naive Bayes</i> pada <i>tools</i> RapidMiner dan akurasinya lebih tinggi dibandingkan dengan algoritma <i>Decision Tree</i> dan <i>Random Forest</i> yaitu 82,91%.
10	Abas Sunarya, Rina Refianti, Ahmad Benny Mutiarra, Wiranti Octaviani	2019	<i>Comparison Of Accuracy Between Convolutional Neural Networks And Naïve Bayes Classifiers In Sentiment Analysis On Twitter.Af Id</i>	<i>Naïve Bayes dan Neural Network</i>	Hasil yang diperoleh dari penelitian ini menunjukkan bahwa model <i>Classifier CNN</i> menghasilkan akurasi sebesar 0,88 atau 88% sedangkan model <i>Classifier NBC</i> menghasilkan akurasi sebesar 0,78 atau 78% pada tahap pengujian data.
11	Yazhi Gao, Wenge Rong, Yikang Shen, Zhang Xiong	2016	<i>Convolutional Neural Network Based Sentiment Analysis Using Adaboost Combination</i>	CNN	Hasil ini menunjukkan bahwa model <i>boosted</i> (Adaboost) dapat menghasilkan hasil yang lebih baik untuk ekstraksi informasi sentimen dalam teks. Algoritma CNN membuktikan kemampuannya dalam memanfaatkan informasi dengan memindai kombinasi kata secara berurutan dan menjaga informasi sekuensial tetap utuh pada operasi pool yang menjembatani ruang informasi di kedua ujung kalimat.

Tabel 2.5 *State Of The Art* Penelitian (Lanjutan 4)

No	Nama Penulis	Tahun	Judul	Algoritma	Hasil
12	Mohammad Ehsan Basiri, Shahla Nemati, Moloud Abdar, Erik Cambria, U. Rajendra Acharya	2020	<i>ABCDM: An Attention-Based Bidirectional CNN-RNN Deep Model For Sentiment Analysis</i>	RNN dan CNN	Hasil eksperimen pada kumpulan data ini menunjukkan bahwa algoritma yang digunakan mencapai hasil mutakhir baik pada ulasan panjang maupun klasifikasi <i>tweet</i> pendek. Meskipun demikian, perbandingan hasil yang diperoleh untuk kumpulan data ulasan dan <i>tweet</i> menunjukkan bahwa jumlah perbaikan pada kumpulan data <i>tweet</i> pendek lebih sedikit daripada kasus serupa untuk kumpulan data ulasan panjang.
13	Li Yang, Ying Li, Jin Wang dan R. Simon Sherratt	2020	<i>Sentiment Analysis For E-Commerce Product Reviews In Chinese Based On Sentiment Lexicon And Deep Learning</i>	SLCABG, CNN dan BiGRU	Model tersebut memiliki kinerja klasifikasi yang lebih baik daripada model analisis sentimen lainnya. Menggunakan leksikon sentimen yang berkelanjutan dan peningkatan dataset, akurasi klasifikasi model akan meningkat secara bertahap.
14	Huyen Trang Phan, Van Cuong Tran, Ngoc Thanh Nguyen dan Dosam Hwang	2017	<i>Improving The Performance Of Sentiment Analysis Of Tweets Containing Fuzzy Sentiment Using The Feature Ensemble Model</i>	CNN	Hasil yang diperoleh dengan menggunakan model ini adalah <i>tweet embeddings</i> , yang digunakan sebagai vektor fitur pada lapisan input model CNN. Analisis eksperimen mengungkapkan bahwa metode yang diusulkan secara signifikan meningkatkan kinerja dalam analisis sentimen <i>tweet</i> .

Tabel 2.6 *State Of The Art* Penelitian (Lanjutan 5)

No	Nama Penulis	Tahun	Judul	Algoritma	Hasil
15	Ala Noor dan Mohrima Islam	2019	<i>Sentiment Analysis For Women's E-Commerce Reviews Using Machine learning Algorithms</i>	<i>Naïve Bayes</i> , JRip, J48, dan <i>Sequential Minimal Optimization (SMO)</i>	Pesaing dekat SMO adalah NB, memiliki perbedaan kecil sebesar 4,25% dari SMO. Prosedur yang digunakan oleh NB lebih sederhana daripada semua algoritma lainnya, namun kinerjanya cukup efisien dibandingkan dengan klasifikasi lainnya JRip pada 72,25% dan J48 pada akurasi 71,25% berkinerja paling buruk. Akurasi J48 adalah 1% kurang dari JRip.
16	Ankita Rane dan Anand Kumar	2018	<i>Sentiment Classification System Of Twitter Data For US Airline Service Analysis</i>	<i>Decision Tree</i> , <i>Random Forest</i> , SVM, KNN, <i>Logistic Regression</i> , <i>Gaussian Naïve Bayes</i> dan Adaboost	Teknik klasifikasi yang digunakan antara lain pendekatan ensemble seperti Adaboost memberikan akurasi sebesar 84,5%. Akurasi yang dicapai oleh algoritma cukup tinggi untuk digunakan oleh industri penerbangan untuk melaksanakan investigasi yang memuaskan pelanggan.
17	Bhavitha BK, Anisha P Rodrigues, Niranjan N Chiplunkar	2017	<i>Comparative Study Of Machine learning Techniques In Sentimental Analysis</i>	SVM, <i>Neural Network</i> , <i>Naïve Bayes</i> , Bayesian Network, Maximum Entropy	Hasil akurasi 85% dengan menggunakan <i>Support Vector Machine</i> menghasilkan akurasi yang sangat baik dibandingkan dengan banyak algoritma lainnya. Hal akurasi bahwa dengan set fitur kecil <i>Naïve Bayes</i> berkinerja baik, jika set fitur besar diambil maka SVM akan menjadi pilihan terbaik.

Tabel 2.7 *State Of The Art* Penelitian (Lanjutan 6)

No	Nama Penulis	Tahun	Judul	Algoritma	Hasil
18	Abdul Mohaimin Rahati, Abdul Kahir, Abu Kaisar Mohammad Masum	2019	<i>Comparison Of Naive Bayes And SVM Algorithm</i>	<i>Naive Bayes dan SVM</i>	Penelitian ini digunakan untuk melakukan evaluasi terhadap komentar yang diperoleh dari penumpang. Akurasi yang didapatkan SVM adalah 82,48 serta akurasi <i>Naive Bayes</i> adalah 76,56. Nilai <i>Precision</i> dan <i>Recall</i> untuk SVM adalah 90,33 dan 81,79 serta 89,00 dan 83,75 untuk <i>Naive Bayes</i> .
19	Meylan Wongkar, Apriandy Angdresey	2020	<i>Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler Twitter</i>	<i>Naive Bayes, SVM dan KNN</i>	Hasil penelitian ini dilakukan perbandingan menggunakan metode <i>Naive Bayes</i> , SVM dan K-Nearest Neighbor (K-NN) yang diuji menggunakan RapidMiner dengan menghasilkan nilai akurasi <i>Naive Bayes</i> sebesar 75,58%, nilai akurasi SVM sebesar 63,99% dan K- Nilai akurasi NN sebesar 73,34%.
20	Achmad Bayhaqy, Sfenrianto Sfenrianto, Kaman Nainggolan, Emil R. Kaburuan	2018	<i>Sentiment Analysis About E-Commerce From Tweets Using Decision Tree, K-Nearest Neighbor, And Naive Bayes</i>	<i>Decision Tree, KNN dan Naive Bayes</i>	Hasil menunjukkan Akurasi Pohon Keputusan, K-NN, dan <i>Naive Bayes</i> sebesar 80%, 78%, dan 77%. Hasil untuk <i>Precision of Decision Tree</i> , K-NN, dan <i>Naive Bayes</i> adalah sebesar 79,96%, 85,67%, dan 88,50%. Hasil juga menunjukkan bahwa <i>Recall</i> dari <i>Decision Tree</i> , K-NN, dan <i>Naive Bayes</i> adalah 84%, 70%, dan 64%. Kesimpulannya bahwa <i>Naive Bayes Classifier</i> merupakan <i>Classifier</i> terbaik.

Tabel 2.8 *State Of The Art* Penelitian (Lanjutan 7)

No	Nama Penulis	Tahun	Judul	Algoritma	Hasil
21	Nabil Ramdhani, Rifky Haekal Al- Fadillah	2021	Analisis Sentimen Pengguna Twitter Terhadap Belajar Daring Selama Pandemi Covid-19 Dengan <i>Deep Learning</i>	<i>Deep Learning</i> , <i>Naïve Bayes</i> , KNN	Hasil Penelitian yang dengan akurasi sebesar 100%. Metode yang lainnya seperti <i>Naïve Bayes</i> mendapatkan akurasi 99% dan KNN mendapatkan akurasi 82%. Prediksi yang didapat pada penelitian ini ialah banyaknya komentar positif sebesar 73% dibandingkan dengan komentar negatif 14% dan komentar netral 13%.
22	Ainun Rizki, Yuliant Sibaroni	2021	Analisis Sentimen Untuk Pengukuran Tingkat Depresi Pengguna Twitter Menggunakan <i>Deep Learning</i>	CNN	Penelitian ini membahas analisis sentimen pengukuran tingkat depresi pada pengguna twitter dengan menggunakan algoritma <i>Deep Learning</i> yaitu model CNN serta dilakukan menggunakan <i>Confusion Matrix</i> yang mendapatkan akurasi mencapai sebesar 82.90%.
23	Fawwaz Zaini Ahmad, Muhammad Fauzi Satria Arifandy, Muhammad Rasyad Caesarardhi, Nur Aini Rakhmawati	2021	Bagaimana Masyarakat Menyikapi Pembelajaran Tatap Muka: Analisis Komentar Masyarakat Pada Media Sosial <i>Tweet</i> Menggunakan Algoritma <i>Deep Learning</i> Sekuensial Dan LDA	LSTM dan LDA	Data yang digunakan pada penelitian ini komentar dari 9 video di <i>Tweet</i> . Persiapan data berupa <i>upsampling</i> , <i>case folding</i> , <i>cleansing</i> , <i>labelling</i> dan <i>tokenizing</i> . Algoritma yang digunakan adalah <i>Deep Learning</i> dengan model LSTM dengan Optimasi ADAM yang mendapatkan akurasi sebesar 78%.

Tabel 2.14 Matriks Penelitian (Lanjutan 5)

No	Penulis/Tahun	Judul	Ruang Lingkup															
			Algoritma													Tujuan		
			NB	RF	KNN	SVM	CNN	DT	KM	RNN	BiGRU	SLCABG	LSTM	Jrip	J48	SMO	LDA	Klasifikasi
13	(Yang <i>et al.</i> , 2020)	<i>Sentiment Analysis For E-Commerce Product Reviews In Chinese Based On Sentiment Lexicon And Deep Learning</i>					✓										✓	
14	(Phan <i>et al.</i> , 2020)	<i>Improving The Performance Of Sentiment Analysis Of Tweets Containing Fuzzy Sentiment Using The Feature Ensemble Model</i>					✓											✓
15	(Noor and Islam, 2019)	<i>Sentiment Analysis For Women's E-Commerce Reviews Using Machine learning Algorithms</i>	✓										✓	✓	✓			✓

Penelitian yang akan dilakukan ini memiliki beberapa keterkaitan dengan penelitian-penelitian sebelumnya mengenai analisis sentimen pada media sosial twitter. Penelitian ini bertujuan untuk membandingkan beberapa algoritma yaitu diantaranya algoritma *Naïve Bayes Classifier* dan *Support Vector Machine* dengan menggunakan *boosting* Adaboost yang dapat dilihat dari hasil akurasi sekaligus menganalisis dan memprediksi pendapat dari mahasiswa terkait pembelajaran daring pada media sosial twitter. Penelitian yang dilakukan oleh Veny, dkk 2019 dalam penelitiannya menggunakan algoritma *Naïve Bayes Classifier*, *Decision Tree* dan *Random Forest* ini menunjukkan akurasi sebesar 86,43% dari pengujian data menggunakan *Naive Bayes* yang akurasinya lebih tinggi dibandingkan dengan algoritma lainnya yaitu *Decision Tree* dan *Random Forest* yaitu 82,91%.

Algoritma *Support Vector Machine* juga digunakan pada penelitian ini. Seperti pada penelitian yang dilakukan oleh Bhavitha, dkk 2017 yang dalam penelitiannya menghasilkan akurasi 85%. *Support Vector Machine* memberikan akurasi yang sangat baik dibandingkan dengan banyak algoritma lainnya. Akurasi dengan set fitur kecil *Naive Bayes* berkinerja baik, jika set fitur besar diambil maka SVM akan menjadi pilihan terbaik.

Penelitian yang dilakukan oleh Ankita, dkk 2018 yang dalam penelitiannya menghasilkan akurasi yang baik. Teknik klasifikasi yang digunakan antara lain pendekatan ensemble seperti AdaBoost yang menggabungkan beberapa *Classifier* lain untuk membentuk satu *Classifier* yang kuat dan memberikan akurasi sebesar 84,5%.

Berdasarkan penelitian yang dilakukan sebelumnya, penelitian ini akan menggunakan algoritma *Naïve Bayes Classifier* dan *Support Vector Machine* dengan ditambah *boosting* Adaboost sebagai keterbaruan dari penelitian ini yang berfungsi untuk meningkatkan akurasi dari tiap algoritma yang digunakan.

2.12 Relevansi Penelitian

Tabel 2.18 Relevansi Penelitian

Peneliti	(Melisa Winda Pertiwi 2020)	(Rizal 2022)
Judul	Analisis Sentimen Opini Publik Mengenai Sarana dan Transportasi Mudik Tahun 2019 Pada Twitter Menggunakan Algoritma <i>Naive Bayes</i> , <i>Neural Network</i> , KNN dan SVM	Analisis Sentimen Terhadap Opini Mahasiswa Terkait Pembelajaran Daring Dengan Menggunakan Metode <i>Naive Bayes Classifier</i> dan SVM dengan Adaboost
Masalah Penelitian	Banyak masyarakat yang berpendapat pada media sosial Twitter mengenai mudik pada tahun 2019 dengan menggunakan transportasi yang akan diklasifikasi terhadap twit yang mengandung nilai positif dan negatif mengenai mudik hari raya menggunakan algoritma k-Nearest Neighbor (KNN), <i>Support Vector Machine</i> (SVM), <i>Naive Bayes</i> dan <i>Neural Network</i> .	Banyak mahasiswa yang pro dan kontra terkait munculnya kebijakan pembelajaran daring ini di media sosial twitter yang akan dilakukan klasifikasi terkait respon dari mahasiswa tersebut dengan menggunakan metode <i>Naive Bayes Classifier</i> dan SVM dengan Adaboost serta membandingkan mana algoritma yang lebih baik.

Tabel 2.19 Relevansi Penelitian (Lanjutan 1)

Peneliti	(Melisa Winda Pertiwi 2020)	(Rizal 2022)
Objek Penelitian	Klasifikasi respon publik terhadap sarana dan transportasi mudik tahun 2019 dengan membandingkan antara algoritma membandingkan hasil akurasi yang lebih tinggi dari algoritma yang digunakan yaitu <i>K-Nearest Neighbor</i> (k-NN), <i>Support Vector Machine</i> (SVM), <i>Naïve Bayes</i> dan <i>Neural Network</i> .	Hasil analisis sentimen terhadap respon mahasiswa terkait kebijakan pembelajaran daring dan hasil performa dari algoritma <i>Naïve Bayes Classifier</i> dan SVM dengan Adaboost.
Algoritma	<i>K-Nearest Neighbor</i> (K-NN), <i>Support Vector Machine</i> (SVM), <i>Naïve Bayes</i> dan <i>Neural Network</i>	<i>Naïve Bayes Classifier</i> (NBC) dan <i>Support Vector Machine</i> (SVM)
Sumber Dataset	Twitter	Twitter