

BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

1. Analisis Sentimen

Analisis sentimen merupakan salah satu bidang dari *Natural Language Processing* (NLP) yang membangun sistem untuk mengenali dan mengekstraksi opini dalam bentuk teks informasi yang saat ini banyak terdapat di internet dalam format forum, blog, media sosial, serta situs berisi *review*. Analisis sentimen sendiri adalah proses memahami dan mengelompokkan emosi (positif, negatif, dan netral) yang terdapat dalam tulisan menggunakan teknik analisis teks (Fathonah & Herliana, 2021). Diketahui bahwa analisis sentimen adalah interaksi untuk memutuskan perasaan atau penilaian seseorang yang ditampilkan sebagai pesan dan dapat diurutkan sebagai pendapat baik atau negatif, salah satu algoritma yang digunakan yaitu algoritma *Naïve Bayes* (Sodik & Kharisudin, 2021). Analisis sentimen yang disebut juga *opinion mining* merupakan salah satu cabang ilmu dari data *mining* yang bertujuan untuk menganalisis, memahami, mengolah, dan mengekstrak data tekstual yang berupa opini terhadap entitas seperti produk, servis, organisasi, individu, dan topik tertentu (Bhatia et al., 2018).

2. Text Mining

Text mining adalah bidang yang mengacu kepada pengambilan informasi, data *mining*, pembelajaran mesin, statistik, dan linguistik komputasional. *Text mining* digunakan untuk mendapatkan informasi yang berguna dari

serangkaian dokumen dengan sumber data pada teks yang memiliki format yang tidak terstruktur (Prianto et al., 2019).

3. *Data Preprocessing*

Data Preprocessing merupakan teknik data *mining* yang melibatkan perubahan data mentah menjadi sebuah format yang terstruktur dan dimengerti. Dokumen dalam teks mentah sering kali tidak lengkap, tidak konsisten, dan mungkin mengandung banyak kesalahan. Beberapa teknik yang digunakan untuk *pre-processing* adalah normalisasi, *case folding*, tokenisasi, *stopwords removal*, dan *stemming* (Jumeilah, 2017). Penjelasan terkait beberapa teknik *preprocessing text* dapat dilihat pada Tabel 2.1

Tabel 2.1 Penjelasan Teknik *Data Preprocessing*

Proses	Penjelasan
<i>Case Folding</i>	Mengubah huruf dalam dokumen menjadi huruf kecil. Karakter yang tidak sesuai akan dihilangkan dan dianggap delimiter.
<i>Tokenisasi</i>	Tahap pemotong string input berdasarkan setiap kata yang menyusun.
<i>Filtering</i>	Menghilangkan kata-kata yang sering muncul dalam dokumen, akan tetapi tidak memberikan arti yang signifikan dan tidak berpengaruh secara semantik.
<i>Stemming</i>	Menggabungkan atau memecahkan setiap varian-varian suatu kata menjadi dasar.

4. *Naïve Bayes Classifier*

Algoritma *Naïve Bayes* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. Algoritma *Naïve Bayes* didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan *Decision Tree* dan *Neural Network* (Nooraeni et al., 2020). Algoritma *Naïve Bayes* merupakan salah satu bentuk pengklasifikasian dengan mempertimbangkan nilai probabilitas statistik. Sehingga, prediksi keanggotaan kelas suatu data baru dilakukan dengan memprediksi probabilitas keanggotaan dari data tersebut (Handayani & Pribadi, 2015).

Menurut (D. G. Nugroho et al., 2015) tahapan klasifikasi metode *Naive Bayes* dijelaskan sebagai berikut.

a. Pembentukan Fitur Data Latih

Tahap pertama yang dilakukan yaitu melakukan pelatihan terhadap data latih. Fitur yang dimaksud adalah kata kunci yang menjadi parameter satuan data latih, yaitu untuk diklasifikasikan ke dalam kelas yang telah ditentukan apakah termasuk pada kelas positif, negatif dan netral.

b. Perhitungan Nilai Probabilitas Data Latih

Setelah fitur dibentuk, selanjutnya proses perhitungan nilai probabilitas dari setiap kelas.

c. Menentukan Probabilitas Kelas

Setelah didapat probabilitas dari setiap kelas, selanjutnya menghitung probabilitas setiap fitur pada setiap kelas menggunakan persamaan 2.1.

$$P(ai|vj) = \frac{fk + 1}{n + term} \quad (2.1)$$

Keterangan:

- $P(ai|vj)$: Nilai probabilitas kelas (kata|kelas)
 fk : Frekuensi kata
 n : Probabilitas kata setiap kelas
 $term$: Jumlah kata keseluruhan

d. Perhitungan V_{map}

V_{map} merupakan proses perhitungan untuk menentukan probabilitas tertinggi data uji dari setiap kelas berdasarkan proses *learning*.

Perhitungan V_{map} dilakukan berdasarkan persamaan 2.2.

$$V_{map} = p(c) * p(c) \quad (2.2)$$

Keterangan:

- V_{map} : Nilai probabilitas tertinggi
 $p(c)$: Hasil probabilitas per kelas

5. *TF-IDF (Term Frequency – Inverse Documents Frequency)*

TF-IDF adalah algoritma yang menunjukkan seberapa sering kata-kata muncul dalam dokumen. *Term Frequency* (TF) adalah seberapa sering kata-kata muncul dalam dokumen, sedangkan *Inverse Document Frequency* (IDF) adalah jumlah kemunculan suatu kata pada koleksi dokumen (Rizkia et al., 2019). Tahapan pembobotan kata dengan TF-IDF adalah sebagai berikut.

- a. Menghitung nilai TF dengan persamaan 2.3.

$$TF(t_i, d) = \frac{\text{frekuensi } t \text{ pada } d}{\text{jumlah } t \text{ pada } d} \quad (2.3)$$

Keterangan:

t_i : kata ke- i
 d : Dokumen

- b. Menghitung nilai *document frequency* untuk mengetahui frekuensi kata pada seluruh dokumen.
- c. Menghitung nilai *inverse document frequency* dengan persamaan 2.4.

$$(IDF(t_i) = \ln \frac{N}{DF(t_i)} \quad (2.4)$$

Keterangan:

t_i : kata ke- i
 \ln : logaritma alami
 N : jumlah seluruh dokumen
 $DF(t_i)$: Jumlah dokumen yang mengandung t_i

- d. Menghitung bobot akhir TF-IDF dengan persamaan 2.5.

$$TFIDF = TF(t_i, d) * IDF(t_i) \quad (2.5)$$

6. Adaboost

Algoritma *Adaboost* pertama kali diperkenalkan pada tahun 1995 oleh Freund dan Schapire, telah banyak memecahkan berbagai masalah praktis dari algoritma *boosting* sebelumnya, pada dasarnya metode *boosting (adaboost)* dapat meningkatkan ketelitian dalam proses klasifikasi dan prediksi dengan

cara membangkitkan kombinasi yang dipilih adalah model yang memiliki nilai bobot paling besar (Afza et al., 2011). *Adaboost* merupakan algoritma *boosting* yang bersifat adaptif, algoritma ini diterapkan secara luas pada model prediksi dalam data *mining* sehingga dapat memberikan suatu bobot lebih pada observasi yang tidak tepat.

7. XGBoost

Extreme Gradient Boosting (XGBoost) merupakan metode *boosting* yang didasarkan pada model pohon keputusan. Model tersebut merupakan algoritma pohon ensemble yang terdiri dari beberapa pohon klasifikasi dan regresi (Liang, 2021). Pada saat melakukan klasifikasi, XGBoost akan memperbarui bobot pada setiap pohon yang dibangun untuk mendapatkan pohon klasifikasi yang kuat (Nasional et al., 2022).

8. Confusion Matrix

Confusion Matrix adalah suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep data *mining*. *Confusion matrix* adalah media yang digunakan untuk menganalisis seberapa baik *classifier* dapat mengenali dari kelas yang berbeda (Sumiah & Mirantika, 2020). Metode ini menggunakan tabel *matrix* yang terdiri atas dua kelas yaitu kelas positif dan negatif. Kerangka *confusion matrix* dapat dilihat pada Tabel 2.2

Tabel 2.2 *Confusion Matrix*

<i>Confusion Matrix</i>		Kelas Aktual		
		Positif	Netral	Negatif
Kelas Prediksi	Positif	TF	FP	FP
	Netral	FNeu	TNeu	FNeu
	Negatif	FN	FN	TN

Berdasarkan Tabel 2.2 *confusion matrix* memiliki 4 nilai keluaran yang dapat dihitung untuk mengukur performa algoritma klasifikasi, yaitu:

a. Akurasi

Akurasi merupakan nilai yang dihasilkan dari perbandingan semua hasil prediksi dengan seluruh data. Rumus menghitung nilai akurasi dapat dilihat pada persamaan (2.6).

$$Akurasi = \frac{TP + TNeu + TN}{TP + TNeu + TN + FP + FNeu + FN} \quad (2.6)$$

Keterangan:

- True Positive (TP)* : Banyaknya nilai data positif yang tergolong positif
- False Positive (FP)* : Banyaknya nilai data positif yang tergolong negatif atau netral
- False Negative (FN)* : Banyaknya nilai data negatif yang tergolong positif atau netral
- True Negative (TN)* : Banyaknya nilai data negatif yang tergolong negatif
- True Neutral (TNeu)* : Banyaknya nilai data netral yang tergolong netral
- False Neutral (FNeu)* : Banyaknya nilai data netral yang tergolong

negatif atau positif

b. Presisi

Presisi merupakan nilai yang dihasilkan dari perbandingan nilai prediksi benar dan hasil prediksi keseluruhan. Rumus menghitung nilai presisi dapat dilihat pada persamaan (2.7).

$$Presisi = \frac{TP}{TP + FP} \quad (2.7)$$

c. Recall

Recall merupakan nilai yang dihasilkan dari perbandingan nilai prediksi benar dan hasil keseluruhan hasil yang benar. Rumus menghitung nilai *recall* dapat dilihat pada persamaan (2.8)

$$Recall = \frac{TP}{TP + FN + FNeu} \quad (2.8)$$

2.2 State Of The Art (SOTA)

Tabel 2.3 menunjukkan perbandingan penelitian yang berhubungan dengan fokus pada permasalahan penelitian dan hasil penelitian yang dihasilkan dalam proses analisis sentimen.

Tabel 2.3 SOTA

No	Penulis	Tahun	Judul	Masalah	Hasil
1.	Pradana, Yanuar Rahmat Yoga, Astiningrum, Mungkin, Hani'ah dan Mamluatul	2020	Analisis Sentimen Tentang Opini Terhadap Performa Timnas Sepak Bola Indonesia Pada Twitter	Tweet dari pengguna juga dapat menjadi evaluasi bagi pengelolaan Timnas Indonesia kedepannya agar performa sesuai yang diharapkan oleh supporter. Namun, pengguna akan mengalami kesulitan apabila melihat tweet secara langsung tanpa ada label tweet tersebut bernilai positif atau	Dari tiga pengujian didapatkan hasil nilai algoritma Naive Bayes pada komposisi data training 371 dan data testing 159 sebesar 78%, komposisi data training 424 dan data testing 106 sebesar 84% dan komposisi data training 477 dan data testing 53 sebesar 87%. Nilai akurasi terendah adalah 78% dan tertinggi adalah 87%.

No	Penulis	Tahun	Judul	Masalah	Hasil
				negatif. Sehingga diperlukan klasifikasi untuk memberikan kemudahan pada pengguna dalam melihat tweet yang bernilai positif atau negatif.	
2.	Putut Abrianto, Nurul Hidayat dan Ratih Kartika Dewi	2021	Implementasi Metode <i>Modified K-Nearest Neighbor</i> (MK-NN) Untuk Klasifikasi Cedera Pada Pemain Futsal	Kesalahan penentuan tingkat keseriusan cedera pada pemain dapat mengakibatkan kekalahan timnya. Perlunya klasifikasi cedera dengan cepat bagi pelatih untuk menentukan pengambilan keputusan tindakan kelanjutan tim. Maka dari itu, diperlukan adanya suatu metode yang bias menjadi terobosan guna membantu pelatih mengklasifikasikan cedera	Nilai akurasi paling tinggi yang diperoleh atas pengujian ini adalah saat $K = 5$ yaitu akurasi sebesar 97.1428571428571%. Dengan demikian dapat ditarik kesimpulan bahwa metode MK - NN berhasil dengan sangat baik melakukan klasifikasi cedera pemain futsal.

No	Penulis	Tahun	Judul	Masalah	Hasil
				yang terjadi pada olahraga futsal.	
3.	Muhammad Naufal Rabbani, Ahmad Yusuf dan Dwi Rolliawati	2021	Komparasi Model Prediksi Daftar Uang Calon Mahasiswa Baru Menggunakan Metode <i>Decision Tree</i> dan <i>Adaboost</i>	Data dari proses penerimaan mahasiswa baru di Universitas Islam Negeri Sunan Ampel Surabaya Tahun 2019 pada Tabel I mendukung pernyataan bahwa banyak calon mahasiswa baru yang tidak melakukan heregistrasi hingga tahap akhir. Hal ini menyebabkan adanya jumlah kursi kosong yang mencapai 324 kursi. Apabila jumlah kursi kosong dapat diminimalisir, maka pemasukan dari uang kuliah tunggal (UKT) bisa	Berdasarkan hasil pengolahan data didapatkan sebaran data yang tidak seimbang (imbalance). Namun sebaran data yang tak seimbang tersebut dapat ditangani lebih baik oleh model Adaboost daripada model Decision Tree. Hasil pelatihan model menunjukkan bahwa Adaboost memiliki peforma yang lebih baik daripada model Decision Tree dengan skor precision 83.66%, recall 99.50%, dan f-measure 90.89%. Hasil riset ini menjadi bukti tambahan bahwa proses boosting dapat menambah

No	Penulis	Tahun	Judul	Masalah	Hasil
				dimaksimalkan.	peforma model.
4.	Samsir, Ambiyar, Unung Verawardina, Firman Edi dan Ronal Watrianthos	2021	Analisis Sentimen Pembelajaran Daring Pada Twitter Di Masa Pandemi Covid-19 Menggunakan Metode <i>Naïve Bayes</i>	Pembelajaran daring yang dianggap sebagai strategi kemudian menjadi kontroversi karena perlu adaptasi. Perbedaan infrastruktur, kualitas koneksi, perangkat yang digunakan, dan masih mahalnya kuota internet menjadi hambatan utama. Perubahan mendadak dari pembelajaran tatap muka ke pembelajaran daring pada skala besar menyebabkan berbagai respons di masyarakat. Penelitian ini bertujuan menganalisis opini publik terhadap pembelajaran daring	Penelitian ini menerapkan algoritma <i>Naïve Bayes</i> untuk klasifikasi sentimen. Analisis sentimen dalam study ini menggunakan data Twitter dengan kata kunci ‘pembelajaran daring’, ‘kuliah’, ‘belajar’, ‘online’, ‘daring’, dan tagar #BelajarDariRumah yang difilter dengan kata kunci ‘online’ dan ‘rumah’ pada twit-twit dalam bahasa Indonesia pada minggu pertama November 2020. Penelitian pada periode tersebut menunjukkan 30% sentimen positif, 69% sentimen negatif, dan 1% netral.

No	Penulis	Tahun	Judul	Masalah	Hasil
				pada masa pandemi COVID-19 di Indonesia pada bulan November 2020.	
5.	Muhammad Andi Ramadhan dan Muhammad Iwan Wahyudin	2022	Analisis Sentimen Mengenai Keberhasilan Indonesia di ajang Thomas Cup 2020 (Studi Kasus Media Sosial Twitter) Menggunakan Metode <i>Naïve Bayes</i> dan <i>Decision Tree</i>	Twitter dapat dimanfaatkan sebagai pencarian berita terbaru atau kabar yang terupdate di masyarakat luas dan juga dapat memberikan lebih banyak informasi tentang sesuatu yang kita sukai. Oleh karena itu, sangat disayangkan jika seseorang memiliki akun twitter tetapi tidak bisa menggunakannya secara baik dan bijak.	Hasil penelitian menggunakan metode <i>Naïve Bayes</i> dan <i>Decision Tree</i> diatas ialah telah ditemukan hasil dari analisis sentimen masyarakat mengenai keberhasilan Indonesia di ajang <i>Thomas Cup 2020</i> lebih ke arah netral. Jika dilihat dari hasil menggunakan proses pencarian data dengan API <i>tweets</i> menggunakan metode <i>naïve bayes</i> dan <i>decision tree</i> hasilnya sudah jelas. Analisis klasifikasi dokumen <i>tweets</i> dengan metode <i>naïve bayes</i> nilai ujinya mencapai

No	Penulis	Tahun	Judul	Masalah	Hasil
					akurasi 83.00% +/-10.5% (<i>micro average</i> : 83.00%) sedangkan untuk metode <i>decision tree</i> nilai ujinya mencapai akurasi 77.00% +/- 4.58% (<i>micro average</i> : 77.00%).
6.	Tobby Wiratama Putra, Agung Triayudi dan Andriansyah	2022	Analisis Sentimen Pembelajaran Daring Menggunakan Metode <i>Naïve Bayes</i> , KNN Dan <i>Decision Tree</i>	Dengan diberlakukannya kebijakan ini mengakibatkan dunia pendidikan belum kembali sepenuhnya, sehingga mengharuskan mahasiswa melakukan pembelajaran secara online/daring. Banyak pendapat mahasiswa yang positif dan negatif terhadap pemberlakuan kuliah online. Selain itu jejaring sosial merupakan tempat untuk	Hasil accuracy dari tiga metode tersebut adalah 55.49%, 61.47%, dan 61.92%. Hasil Precision dari tiga metode tersebut adalah 46.45%, 53.90%, dan 73.63%. Dan hasil Recall dari tiga metode tersebut adalah 58.91%, 40.07%, dan 11.42%. Dapat disimpulkan dari hasil tersebut bahwa metode Decision Tree memiliki tingkat accuracy yang paling tinggi dari kedua metode lainnya sebesar

No	Penulis	Tahun	Judul	Masalah	Hasil
				menyampaikan segala aspirasi, kritik, dan pendapat. Media sosial yang cukup sering digunakan adalah Twitter.	61.92% accuracy, 73.63% precision dan 11.42% recall.
7.	Rani Nooraeni, Amirah Balqis S, Aulia Fatin A, Krisna Dwi Agung, Nada Nabila	2019	Analisis Sentimen Publik Terhadap Sistem Zonasi Sekolah Menggunakan Data Twitter Dengan Metode <i>Naïve Bayes Classification</i>	Pada pertengahan tahun ini, Twitter diramaikan dengan isu tentang penerapan sistem zonasi. Pengguna Twitter banyak yang menuliskan pendapatnya mengenai isu tersebut berupa dukungan, masukan, dan juga kritikan. Kritikan masyarakat terhadap sistem zonasi merupakan hal yang harus diperhatikan untuk menyempurnakan sistem zonasi yang masih terbilang baru. Kritikan-kritikan ini	Dari hasil analisis sentimen dapat disimpulkan bahwa respon masyarakat terhadap sistem zonasi sekolah secara mayoritas adalah netral yaitu sebanyak 61%, sedangkan respon positif sebanyak 21% dan respon negatif sebanyak 15%. Metode klasifikasi <i>naïve bayes</i> cukup baik untuk diterapkan untuk memprediksi sentimen masyarakat terhadap sistem zonasi sekolah karena memiliki akurasi yang cukup tinggi yaitu sebanyak 0,8079

No	Penulis	Tahun	Judul	Masalah	Hasil
				dapat ditemukan baik berupa lisan maupun tulisan di media sosial seperti Twitter, Facebook dan media sosial lainnya.	
8.	Tazkia Shabrina Az-Zahra	2021	Analisis Sentimen Terhadap Belajar Daring Menggunakan Optimasi <i>Naïve Bayes Classifier</i> Dengan <i>Adaboost</i>	Kecenderungan sentimen masyarakat mengenai suatu isu dapat diketahui dengan melakukan analisis sentimen pada twitter. Metode yang digunakan untuk analisis sentimen adalah <i>naive bayes classifier</i> (NBC) dengan <i>adaboost</i> . Penelitian ini bertujuan untuk mengoptimasi NBC dengan <i>adaboost</i> untuk analisis sentimen, mengetahui performa NBC dengan	Hasil penelitian menunjukkan <i>adaboost</i> tidak dapat meningkatkan akurasi NBC, didapatkan akurasi NBC sebesar 78%, dan akurasi NBC dengan <i>adaboost</i> sebesar 64%. Hal ini disebabkan karena NBC merupakan pengklasifikasi yang stabil sehingga sulit untuk <i>adaboost</i> dapat meningkatkan akurasi. Kombinasi antara data pada penelitian ini yang <i>imbalance</i> dengan NBC dan

No	Penulis	Tahun	Judul	Masalah	Hasil
				<i>adaboost</i> , dan mengetahui kecenderungan sentimen masyarakat mengenai belajar daring.	<i>adaboost</i> juga menyebabkan penurunan akurasi. NBC dengan <i>adaboost</i> melabelkan semua data dengan sentimen positif, hal tersebut menunjukkan bahwa NBC dengan <i>adaboost</i> tidak dapat melakukan klasifikasi. Hasil analisis sentimen menggunakan NBC menunjukkan kecenderungan sentimen masyarakat terhadap belajar daring adalah positif dengan presentasi 89,5% sedangkan sentimen negatif sebesar 10,5%.
9.	Reino Prajamukti, Jayanta dan Mayanda Mega Santoni	2021	Klasifikasi Dan Analisis Sentimen Pada Data Twitter Menggunakan	Timnas Sepakbola Indonesia pada jejaring sosial Twitter. Pada awal tahun 2021 sampai Mei 2021 Timnas Sepakbola	Algoritma Naïve Bayes pada klasifikasi tweet tentang sentimen terhadap Timnas Indonesia berjalan dengan baik dengan nilai

No	Penulis	Tahun	Judul	Masalah	Hasil
			Algoritma <i>Naïve Bayes</i> (Studi Kasus: Timnas Indonesia Senior, U-23 Dan U-19)	Indonesia senior, usia 23 dan usia 19 tahun sedang banyak menjalani training camp di luar negeri, uji coba, hingga klasifikasi piala dunia 2022. Banyak dari pengguna media sosial Twitter yang membahas tentang Timnas Sepakbola Indonesia ini.	akurasi 83%, kemudian nilai precision nilai positif sebesar 86%, nilai precision nilai negatif sebesar 81%, nilai recall sebesar 78%, dan nilai specificity penelitian ini 87,5% dengan menggunakan metode confusion matrix berdasarkan data yang diambil pada Februari sampai Mei 2021.
10.	Abdi Rahim Damanik, Sumijan dan Gunadi Widi Nurcahyo	2021	Prediksi Tingkat Kepuasan Dalam Pembelajaran Daring Menggunakan Algoritma <i>Naïve Bayes</i>	Pembelajaran daring akan memberikan kesempatan bagi mahasiswa untuk dapat mengikuti suatu pembelajaran atau mata kuliah tertentu tanpa harus melaksanakan tatap muka langsung. Dengan menggunakan media	Dari hasil pengujian akhir yang dilakukan dari data sampel kuesioner dengan atribut atau indikator komunikasi, suasana pembelajaran, penilaian mahasiswa dan penyampaian materi menggunakan metode <i>Naïve Bayes</i> didapatkan tingkat

No	Penulis	Tahun	Judul	Masalah	Hasil
				<p>pembelajaran daring yaitu Whatsapp, Google Meet, Cloud x dan Zoom Meeting.</p> <p>Perguruan tinggi mengharapkan mahasiswa dapat mengikuti pendidikan secara efektifserta mendapatkan ilmu yang telah diberikan oleh dosen pengajar.</p>	<p>akurasi sebesar 100% dengan nilai precision sebesar 100% dan nilai recall sebesar 100%.</p>
11	Dery Anjas Ramadhan, Erwin Budi Setiawan S.Si., M.T.	2020	ANALISIS SENTIMEN PROGRAM ACARA DI SCTV PADA TWITTER MENGGUNAKAN METODE NAIVE BAYES DAN SUPPORT VECTOR	Kepuasan penonton merupakan suatu faktor yang menjadi bahan pertimbangan dalam menentukan keberhasilan program acara televisi. Dengan kepuasan penonton perusahaan tahu harus membuat jadwal tayang yang baik dan menentukan berapa durasi	Setelah melakukan kegiatan analisis sentimen aplikasi online shop di google play store menggunakan metode support verctor machine maka dapat disimpulkan bahwa aplikasi yang memiliki tingkat akurasi tertinggi di tempati oleh Tokopedia yang dimana memiliki tingkat akurasi

No	Penulis	Tahun	Judul	Masalah	Hasil
			MACHINE	acara maupun berapa episode yang harus dibuat. Selain itu perusahaan dapat meminimalisir kerugian karena jumlah penonton yang tidak sesuai harapan, maka dari itu dapat dilakukan suatu Analisis Sentimen pada Tweet Twitter SCTV yang berupaya untuk mengetahui sentimen dari setiap Tweet.	sebesar 90,67%, Shopee 70,00%, Lazada sebesar 69,00%, JD.ID 75,33%, Sedangkan marketplace yang memiliki tingkat akurasi terendah dalam penelitian ini jatuh kepada Lazada dengan tingkat akurasi 69,00%. Sedangkan dalam kategori Area Under Curve (AUC) tertinggi di raih oleh JD.ID dengan perolehan 0.857, Shopee 0,784, Lazada 0.748, JD.ID 0.857, Blibli 0.770 dan tingkat Area Under Curve (AUC) terendah di dapatkan oleh marketplace lazada dengan perolehan Area Under Curve (AUC) 0,748.
12	Yazhi Gao, Wenge	2016	Convolutional Neural	Media Twitter merupakan	Berdasarkan hasil penelitian,

No	Penulis	Tahun	Judul	Masalah	Hasil
	Rong, Yikang Shen, Zhang Xiong		Network Based Sentiment Analysis Using Adaboost Combination	media online yang populer di Indonesia. Banyak public figure yang menjadi terkenal dari media sosial. Setiap kontroversi yang ditimbulkan oleh figur publik menimbulkan pro dan kontra dari netizen di media sosial seperti Twitter. Model dan artis Indonesia Anya Geraldine baru-baru ini menjadi incaran netizen dan semakin populer saat berperan sebagai orang ketiga di miniseri "Layangan Putus". Dari informasi tersebut, beberapa netizen di Twitter mencetuskan kata-kata yang mengandung ujaran kebencian	metode terbaik untuk penelitian di semua skenario adalah sama, yaitu akurasi XGBoost. XGBoost bekerja dengan baik di semua skenario dengan 93,4% pada desain sebelum oversampling, 93,1% pada desain setelah oversampling, 62,2% pada desain setelah undersampling, dan 95% pada skenario dengan penyesuaian hyperparameter. Namun, performa akurasi Adaboost untuk analisis sentimen di semua paket relatif sama dengan performa akurasi XGBoost. Meskipun kinerja XGBoost lebih baik daripada Adaboost, dibutuhkan waktu lebih lama daripada

No	Penulis	Tahun	Judul	Masalah	Hasil
				<p>terhadap Anya. Jumlah tweet yang tersebar di timeline dapat diklasifikasikan berdasarkan sentimen dari klasifikasi sentimen ujaran kebencian dapat memberikan informasi apakah kalimat tersebut merupakan ujaran kebencian atau bukan.</p>	<p>Adaboost untuk melatih model. Dari hasil penelitian ini ditemukan bahwa algoritma XGBoost lebih tinggi dari algoritma Adaboost yaitu XGBoost yang menggunakan metode tune hyperparameter terbaik dalam melakukan analisis sentimen. Tingkat akurasi, presisi, daya ingat, dan skor f1 mencapai 95%.</p>
13	Muhamad Azhar, Noor Hafidz, Biktra Rudianto dan Windu Gata	2020	Marketplace Sentiment Analysis Using Naive Bayes And Support Vector Machine	<p>Digitalisasi secara konsisten telah mengubah kebiasaan dan budaya manusia di hampir setiap aktivitas sehari-hari. Teknologi ini juga mempengaruhi kebiasaan belanja masyarakat dan</p>	<p>Setelah melakukan penelitian terhadap dataset review aplikasi dan produk Klikindomaret di GooglePlay dengan Naive Bayes Classifier dibandingkan dengan Support Vector Machine, dapat disimpulkan bahwa dengan</p>

No	Penulis	Tahun	Judul	Masalah	Hasil
				<p>membantu memenuhi kebutuhan manusia. Indomaret, sebuah perusahaan walaraba, kini memiliki aplikasi berbasis digital bernama Klikindomaret, tempat belanja online Indomaret yang merupakan salah satu inovasi dari PT Indomarc Prismatama yang menyediakan berbagai produk dalam satu situs atau aplikasi. Teknologi di marketplace, khususnya aplikasi klikindomaret, menjadi perhatian banyak penelitian tentang analisis sentimen. Halaman aplikasi Klikindomaret di GooglePlay</p>	<p>menggunakan Particle Swarm Optimization (PSO) pemilihan fitur dapat meningkatkan nilai Akurasi dan AUC untuk dataset. Peningkatan akurasi cukup signifikan pada nilai akurasi pemodelan naive bayes dimana akurasi awal 69,74% menjadi 75,21% dan nilai AUC dari 0,518 menjadi 0,598. Selain itu, perbandingan antara kedua algoritma Naive Bayes dengan Support Vector Machine baik yang menggunakan optimasi PSO maupun tidak menggunakan optimasi PSO menghasilkan kesimpulan bahwa Support vector Machine memiliki akurasi dan</p>

No	Penulis	Tahun	Judul	Masalah	Hasil
				berisi informasi pendataan review. Untuk mendapatkan informasi dari opini atau review aplikasi, ada beberapa halhal yang perlu dilakukan, termasuk analisis sentiment.	nilai AUC yang lebih tinggi jika dibandingkan dengan Naive Bayes dengan 11,47% f atau akurasi dan 0.
14	Nuraeni Herlinawati, Yuri Yuliani, Siti Faizah, Windu Gata dan Samudi.	2020	Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine	Pada Google Play Store terdapat beberapa fitur salah satunya adalah fitur rating dan ulasan dari para pengguna aplikasi atau layanan yang tersedia. Review atau ulasan merupakan suatu teks atau kalimat yang berisi penilaian atau komentar terhadap suatu hasil karya seseorang. Pentingnya ulasan tersebut sering digunakan sebagai tolak	Dalam penelitian ini setelah dilakukan preprocessing dan modelling dengan membandingkan dua metode algoritma data mining yaitu Naïve Bayes (NB) dan Support Vector Machine (SVM), dari hasil evaluasi diketahui bahwa nilai akurasi untuk mengklasifikasikan label positif dan label negatif pada ulasan para pengguna aplikasi zoom cloud meetings di Google

No	Penulis	Tahun	Judul	Masalah	Hasil
				ukur suatu aplikasi apakah Zoom Cloud Meetings recommended atau tidak bagi para pengguna baru.	Play Store, dapat dibuktikan dengan nilai akurasi dan nilai AUC dari masing-masing algoritma yaitu untuk NB nilai akurasi = 74,37% dan nilai AUC = 0,659. Sedangkan untuk algoritma SVM nilai akurasi = 81,22% dan nilai AUC = 0,886. Dalam penelitian ini dapat diketahui bahwa tingkat akurasi yang didapatkan algoritma Support Vector Machine (SVM) lebih unggul 6,85% dibandingkan algoritma Naïve Bayes (NB) dengan 1.007 record dataset.
15	Muhammad Iqbal Ahmadi, Fuji Apriani, Mia	2020	SENTIMENT ANALYSIS ONLINE SHOP ON THE	Di Indonesia sendiri begitu banyak online marketplace yang bisa di akses atau di	Di Indonesia sendiri begitu banyak online marketplace yang bisa di akses atau di download

No	Penulis	Tahun	Judul	Masalah	Hasil
	Kurniasari, Siti Handayani dan Dudih Gustian		PLAY STORE USING METHOD SUPPORT VECTOR MACHINE (SVM)	download melalui google playstore, seperti Lazada, Tokopedia, Shopee, JD.ID dan masih banyak lagi. Dari masing-masing online shop tersebut mempunyai kelebihan dan kekurangannya masing - masing. Kekurangannya seperti banyaknya penipuan pada saat transaksi, penjual tidak mengirimkan barang yang sesuai dengan konsumen, serta kemasan yang sudah rusak. Masalah tersebut sangat berpengaruh pada kepuasan pelanggan yang seharusnya menjadi tolak ukur kualitas pelayanan dari online shop	melalui google playstore, seperti Lazada, Tokopedia, Shopee, JD.ID dan masih banyak lagi. Dari masing-masing online shop tersebut mempunyai kelebihan dan kekurangannya masing - masing. Kekurangannya seperti banyaknya penipuan pada saat transaksi, penjual tidak mengirimkan barang yang sesuai dengan konsumen, serta kemasan yang sudah rusak. Masalah tersebut sangat berpengaruh pada kepuasan pelanggan yang seharusnya menjadi tolak ukur kualitas pelayanan dari online shop tersebut.

No.	Penulis/Tahun	Judul	Algoritma/Metode					
			Naïve Bayes	KNN	Decision Tree	SVM	CNN	AdaBoost
1	Pradana, Yanuar Rahmat Yoga Astiningrum, Mungki dan Hani'ah, Mamluatul, 2020	Analisis Sentimen Tentang Opini Terhadap Performa Timnas Sepak Bola Indonesia Pada Twitter	✓					
2	Putut Abrianto, Nurul Hidayat Dan Ratih Kartika Dewi, 2021	Implementasi Metode <i>Modified K-Nearest Neighbor (MK-NN)</i> Untuk Klasifikasi Cedera Pada Pemain Futsal		✓				
3	Muhammad Naufal Rabbani, Ahmad Yusuf Dan Dwi Rolliawati, 2021	Komparasi Model Prediksi Daftar Uang Calon Mahasiswa Baru Menggunakan Metode <i>Decision Tree</i> dan <i>Adaboost</i>			✓			

No.	Penulis/Tahun	Judul	Algoritma/Metode					
			Naïve Bayes	KNN	Decision Tree	SVM	CNN	AdaBoost
4	Muhammad Andi Ramadhan Dan Muhammad Iwan Wahyudin, 2022	Analisis Sentimen Mengenai Keberhasilan Indonesia Di Ajang <i>Thomas Cup</i> 2020 (Studi Kasus Media Sosial Twitter) Menggunakan Metode <i>Naïve Bayes</i> Dan <i>Decision Tree</i>	✓					
5	Abdi Rahim Damanik, Sumijan Dan Gunadi Widi Nurcahyo, 2021	Prediksi Tingkat Kepuasan Dalam Pembelajaran Daring Menggunakan Algoritma <i>Naïve Bayes</i>	✓					
6	Tobby Wiratama Putra, Agung Triayudi Dan Andriansyah, 2022	Analisis Sentimen Pembelajaran Daring Menggunakan Metode <i>Naïve Bayes</i> , KNN Dan <i>Decision Tree</i>	✓	✓	✓			

No.	Penulis/Tahun	Judul	Algoritma/Metode					
			Naïve Bayes	KNN	Decision Tree	SVM	CNN	AdaBoost
7	Rani Nooraeni, Amirah Balqis S, Aulia Fatin A, Krisna Dwi Agung, Nada Nabila, 2019	Analisis Sentimen Publik Terhadap Sistem Zonasi Sekolah Menggunakan Data Twitter Dengan Metode <i>Naïve Bayes Classification</i>	✓					
8	Tazkia Shabrina Az-Zahra, 2021	Analisis Sentimen Terhadap Belajar Daring Menggunakan Optimasi <i>Naïve Bayes Classifier</i> Dengan <i>Adaboost</i>	✓					
9	Samsir, Ambiyar, Unung Verawardina, Firman Edi Dan Ronal Watrianthos, 2021	Analisis Sentimen Pembelajaran Daring Pada Twitter Di Masa Pandemi Covid-19 Menggunakan Metode <i>Naïve Bayes</i>	✓	✓	✓			

No.	Penulis/Tahun	Judul	Algoritma/Metode					
			Naïve Bayes	KNN	Decision Tree	SVM	CNN	AdaBoost
10	Reino Prajamukti, Jayanta Dan Mayanda Mega Santoni, 2021	Klasifikasi Dan Analisis Sentimen Pada Data Twitter Menggunakan Algoritma <i>Naïve Bayes</i> (Studi Kasus: Timnas Indonesia Senior, U-23 Dan U-19)	✓					
11	Dery Anjas Ramadhan, Erwin Budi Setiawan S.Si., M.T., 2020	ANALISIS SENTIMEN PROGRAM ACARA DI SCTV PADA TWITTER MENGGUNAKAN METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE	✓			✓		
12	Yazhi Gao, Wenge Rong, Yikang Shen, Zhang Xiong, 2016	Convolutional Neural Network Based Sentiment Analysis Using Adaboost Combination					✓	✓

No.	Penulis/Tahun	Judul	Algoritma/Metode					
			Naïve Bayes	KNN	Decision Tree	SVM	CNN	AdaBoost
13	Muhamad Azhar, Noor Hafidz, Biktra Rudianto dan Windu Gata, 2020	Marketplace Sentiment Analysis Using Naive Bayes And Support Vector Machine	✓			✓		
14	Nuraeni Herlinawati, Yuri Yuliani, Siti Faizah, Windu Gata dan Samudi, 2020	Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine	✓			✓		
15	Muhammad Iqbal Ahmadi, Fuji Apriani, Mia Kurniasari, Siti Handayani dan Dudih Gustian, 2020	SENTIMENT ANALYSIS ONLINE SHOP ON THE PLAY STORE USING METHOD SUPPORT VECTOR MACHINE (SVM)				✓		

2.4 Relevansi Penelitian

Relevansi Penelitian merupakan keterkaitan antar variabel pada penelitian terkait dengan penelitian yang dilakukan. Relevansi Penelitian dapat dilihat pada Tabel 2.5.

Tabel 2.5 Relevansi Penelitian

Peneliti	(Mungki Astiningrum, Mamluatul Hani'ah, Yanuar Rahmat Yoga Pradana, 2020)	(Devira Arnetalya, 2023)
Judul	Analisis Sentimen Tentang Opini Terhadap Performa Timnas Sepak Bola Indonesia Pada Twitter	Optimasi Metode <i>Naïve Bayes Classifier</i> Menggunakan <i>Adaboost</i> Pada Analisis Sentimen Timnas Futsal Indonesia Di Twitter
Masalah Penelitian	Tweet dari pengguna juga dapat menjadi evaluasi bagi pengelolaan Timnas Indonesia kedepannya agar performa sesuai yang diharapkan oleh supporter. Karena dukungan dari supporter sangat berpengaruh kepada Timnas mulai dari dukungan semangat, moril maupun pemasukan Timnas itu sendiri. Namun,	Dalam opini dan komentar masyarakat terdapat penyampaian komentar yang beragam apabila Timnas Futsal Indonesia mendapatkan hasil yang baik maka akan ada opini atau komentar bahagia maupun pujian dan jika Timnas Timnas Futsal Indonesia mendapatkan hasil yang buruk maka akan ada opini dan komentar

Peneliti	(Mungki Astiningrum, Mamluatul Hani'ah, Yanuar Rahmat Yoga Pradana, 2020)	(Devira Arnetalya, 2023)
	pengguna akan mengalami kesulitan apabila melihat tweet secara langsung tanpa ada label tweet tersebut bernilai positif atau negatif.	kritikan bahkan cacian. Untuk mengidentifikasi sentimen masyarakat terkait Timnas Futsal Indonesia memerlukan suatu analisis yang melalui proses komputasi, yaitu analisis sentimen.
Objek Penelitian	Klasifikasi data opini terhadap performa Timnas Sepak Bola Indonesia pada Twitter	Klasifikasi data opini terhadap performa Timnas Futsal Indonesia pada Twitter
Algoritma/Metode	<i>Naïve Bayes</i>	<i>Naïve Bayes</i> dan <i>Boosting Adaboost</i>
Dataset	Total dataset yang digunakan pada aplikasi analisis sentimen opini terhadap performa Timnas Sepakbola Indonesia pada Twitter sebanyak 530 tweet.	Dataset yang digunakan pada analisis sentimen opini terhadap performa Timnas Futsal Indonesia pada Twitter.