

BAB II

LANDASAN TEORI

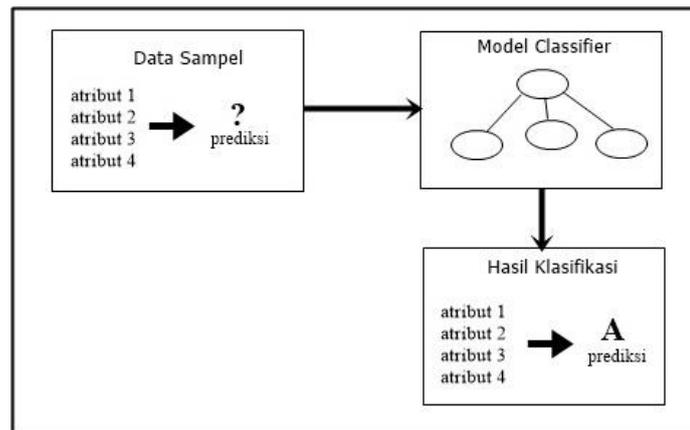
2.1 Analisis Sentimen

Analisis sentimen yang juga disebut dengan *opinion mining* merupakan bidang studi yang menganalisis pendapat, penilaian, evaluasi, sentimen, sikap dan emosi terhadap entitas seperti produk, jasa, organisasi, individu, peristiwa, topik dan atribut lainnya. *Big data* dari situs media sosial seperti Twitter dapat digunakan untuk analisis. (Chakra, Gupta, & Kumar, 2018).

Analisis sentimen dapat digunakan untuk mendapatkan persentase sentimen positif dan sentimen negatif terhadap seseorang, perusahaan, institusi, produk atau pada sebuah kondisi tertentu. Nilai dari analisis sentimen bisa dibagi menjadi 3 yaitu, sentimen positif, sentimen negatif dan sentimen netral atau diperdalam lagi sehingga dapat menemukan siapa atau kelompok yang menjadi sumber sentimen positif atau sentimen negatif (Suyanto, 2017).

2.2 Klasifikasi

Klasifikasi dilakukan untuk menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi terdapat dua pekerjaan yang dilakukan yaitu membangun model sebagai *prototipe* untuk disimpan sebagai memori dan penggunaan model tersebut untuk melakukan prediksi pada suatu objek data lain agar diketahui dikelas mana data tersebut dalam model yang telah disimpan (Prasetyo, 2012).



Gambar 2.1 Bagan proses klasifikasi (Han dan Kamber, 2006)

Menurut Han dan Kamber (2006) terdapat beberapa tahapan yang dilakukan untuk mendapatkan hasil klasifikasi yaitu:

1. Pembersihan Data

Pembersihan data dilakukan untuk mengurangi data yang cacat dalam data *training*, terdapat beberapa metode yang digunakan seperti *smoothing* untuk menghilangkan *noise* data dan melengkapi jika terdapat data yang hilang.

2. Analisis Relevansi

Atribut-atribut yang telah digunakan dalam proses klasifikasi memungkinkan terdapat atribut yang berhubungan antara satu sama lain, kedua atribut yang memiliki kemiripan salah satunya akan di buang sehingga proses klasifikasi bisa lebih optimal.

Hasil klasifikasi dan prediksi dapat dievaluasi menggunakan 2 kriteria yaitu :

1. Akurasi

Akurasi digunakan untuk mengetahui kemampuan model klasifikasi untuk dapat memberikan ketepatan hasil prediksi.

2. Kecepatan

Mengetahui kecepatan iterasi untuk mendapatkan model klasifikasi dan iterasi mendapatkan hasil prediksi.

2.2.1. *Naïve Bayes classifier*

Algoritma *Naïve Bayes classifier* merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema *Bayes* dengan asumsi independensi yang kuat (Suryadi, 2017). Berdasarkan kompleksitasnya, *Naive Bayes* adalah salah satu algoritma paling mudah untuk menerapkan aturan *Bayes*. Algoritma ini juga memiliki beberapa kelebihan yaitu sangat efisien, hanya membutuhkan data uji yang sedikit, mudah untuk diimplementasikan, dan memiliki akurasi yang cukup tinggi. .

Dalam bahasa yang lebih sederhana, *Naïve Bayes classifier* berpendapat bahwa keberadaan sebuah fitur dalam sebuah kelas tidak dikaitkan dengan keberadaan fitur lainnya.

Teorema *Bayes* merupakan teorema yang mengacu pada konsep *probabilitas* bersyarat. Secara umum teorema *Bayes* dapat dinotasikan pada persamaan 2.1 berikut :

$$P(A|B) = \frac{P(A|B)P(A)}{P(B)} \quad (2.1)$$

Setiap ulasan dalam Klasifikasi Naive Bayes dipresentasikan dalam pasangan atribut ($a_1, a_2, a_3, \dots, a_n$) di mana a_1 adalah kata pertama, a_2 adalah kata kedua, dan seterusnya, dan V adalah himpunan kelas. Metode ini akan menghasilkan kategori/kelas dengan probabilitas tertinggi (VMAP) dengan memasukkan atribut ($a_1, a_2, a_3, \dots, a_n$) selama proses klasifikasi. Rumus VMAP dapat dilihat dalam persamaan 2.2 berikut:

$$V_{MAP} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j | a_1, a_2, a_3, \dots, a_n) \quad (2.2)$$

Dengan menggunakan teorema Bayes, maka persamaan (2.2) dapat ditulis menjadi :

$$V_{MAP} = \underset{v_j \in V}{\operatorname{argmax}} \frac{P(a_1, a_2, a_3, \dots, a_n | V_j) P(V_j)}{P(a_1, a_2, a_3, \dots, a_n)} \quad (2.3)$$

$P(a_1, a_2, a_3, \dots, a_n)$ nilainya konstan untuk semua v_j sehingga persamaan (2.3) dapat juga dinyatakan menjadi persamaan 2.4 berikut:

$$V_{MAP} = \underset{v_j \in V}{\operatorname{argmax}} P(a_1, a_2, a_3, \dots, a_n | V_j) P(V_j) \quad (2.4)$$

Naïve Bayes Classifier menyederhanakan hal ini dengan mengasumsikan bahwa didalam setiap kategori, setiap atribut bebas bersyarat satu sama lain. Dengan kata lain,

$$P(a_1, a_2, a_3, \dots, a_n | V_j) = \prod_i P(a_i | V_j) \quad (2.5)$$

Kemudian apabila persamaan (2.4) disubstitusikan ke persamaan (2.5), maka akan menghasilkan persamaan 2.6 berikut :

$$V_{MAP} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j) \times \prod_i P(a_i | v_j) \quad (2.6)$$

$P(v_j)$ dan probabilitas kata a_1 untuk setiap kategori $P(a_1|v_j)$ dihitung pada saat training yang dirumuskan seperti berikut:

$$P(v_j) = \frac{docs_j}{training} \quad (2.7)$$

$$P(a_i|v_j) = \frac{n_i+1}{n + kosakata} \quad (2.8)$$

Dimana $docs_j$ adalah jumlah dokumen pada kategori j dan $training$ adalah jumlah dokumen yang digunakan dalam proses *training*. Sedangkan n_i adalah jumlah kemunculan kata a_i pada kategori v_j , n adalah jumlah kosakata yang muncul pada kategori v_j dan kosakata adalah jumlah kata unik pada semua data *training*. (Rodiyansyah & Winarko, 2013).

Adapun nilai akurasi dari algoritma *Naïve Bayes* dihitung dengan menggunakan formula berikut :

$$Accuracy = \frac{\text{jumlah data yang terklasifikasi dengan benar}}{\text{jumlah keseluruhan data}} \times 100\% \quad (2.6)$$

2.2.2. Confusion Matrik

Confusion Matrix merupakan metode untuk menghitung tingkat akurasi, dengan menghitung jumlah prediksi benar dan salah dari sebuah metode klasifikasi berbanding dengan data sesungguhnya atau prediksi target (Junaidi, 2015). Menggunakan *Matrix* (NxN) dimana N adalah jumlah kelas. Berikut ini adalah contoh perhitungan akurasi dengan *Confusion Matrix* menggunakan dua kelas yaitu positif dan negatif.

Tabel 2.2 Contoh *Confusion Matrix*

Confusion Matrix		Target			
		Positif	Negatif		
Model	Positif	A	B	<i>Positive Predictive Value</i>	$a/(a+b)$
	Negatif	C	D	<i>Negative Predictive Value</i>	$d/(c+d)$
		<i>Sensitivity</i> $a/(a+c)$	<i>Specificity</i> $d/(b+d)$	Accuracy = $(a+d)/(a+b+c+d)$	

Keterangan :

Accuracy : Jumlah prediksi benar $(a+d)$ / Jumlah seluruh data $(a+b+c+d)$

Positive Predictive Value / Precision : Jumlah prediksi benar positif (a) / Jumlah target data positif $(a+b)$

Negative Predictive Value : Jumlah prediksi benar negatif (d) / Jumlah target data negatif $(b+d)$

Sensitivity / Recall : Jumlah prediksi benar positif (a) / Jumlah seluruh prediksi benar $(a+c)$

Specificity : Jumlah prediksi salah negatif (d) / Jumlah seluruh prediksi salah $(c+d)$

2.2 Shopee

Shopee merupakan platform *marketplace online* yang menjembatani penjual dan pembeli untuk mempermudah transaksi jual beli *online* melalui perangkat ponsel mereka. Shopee menawarkan berbagai macam produk-produk mulai dari produk *fashion* sampai dengan produk untuk kebutuhan sehari-hari. Shopee hadir dalam bentuk aplikasi *mobile* guna untuk menunjang kegiatan berbelanja yang

mudah dan cepat. Shopee hadir dalam bentuk aplikasi *mobile* untuk memudahkan penggunaanya dalam melakukan kegiatan belanja *online* tanpa harus membuka website melalui perangkat komputer.

Shopee mulai masuk ke pasar Indonesia pada akhir bulan Mei 2015 dan Shopee baru mulai beroperasi pada akhir Juni 2015 di Indonesia. Shopee merupakan anak perusahaan dari Garena yang berbasis di Singapura. Shopee telah hadir di beberapa negara di kawasan Asia Tenggara seperti Singapura, Malaysia, Vietnam, Thailand, Filipina, dan Indonesia. Shopee Indonesia beralamat di Wisma 77 Tower 2, Jalan Letjen. S. Parman, Palmerah, Daerah Khusus Ibukota Jakarta 11410, Indonesia.

2.3 Preprocessing

Data ulasan yang telah diambil dari *Google Play Store* masih merupakan data mentah maka dari itu perlu dilakukan *preprocessing* untuk mendapatkan data yang siap untuk diproses selanjutnya. Tahapan dari *preprocessing* ada *cleansing*, *tokenizing*, *case folding*, *filtering / stopword removal* dan *stemming*. Adapun langkah-langkah *preprocessing* yang akan digunakan adalah :

2.3.1. Cleaning

Teks hasil *scrapping* dari Google Play Store setelah tahap akuisisi data selanjutnya akan dilakukan proses *cleansing* untuk menghilangkan karakter yang dirasa tidak perlu seperti *emoticon* dan tanda baca. *Cleansing* bertujuan untuk meringankan beban dalam proses komputasi dan juga menghemat waktu.

2.3.2. Tokenizing

Tokenizing merupakan proses memecah kalimat menjadi kata-kata menjadi lebih berarti dan bermakna. Tahap selanjutnya adalah proses memisahkan teks dalam dokumen menjadi potongan kata yang tidak saling berpengaruh atau independen yang disebut dengan token. *Tokenizing* digunakan untuk mendapatkan potongan kata atau token yang akan menjadi entitas yang memiliki nilai dalam penyusunan matriks dokumen pada proses selanjutnya. Pada proses ini memudahkan proses perhitungan keberadaan kata tersebut dalam dokumen maupun dalam menghitung frekuensi kemunculan kata tersebut dalam *corpus*.

2.3.3. Case Folding

Case folding merupakan proses untuk penyeragaman bentuk huruf atau kata kedalam bentuk huruf kecil. Proses ini juga dapat menghilangkan tanda baca dan menghilangkan spasi yang berlebihan, membuat huruf besar dan huruf kecil tidak terdeteksi memiliki perbedaan arti.

2.3.4. Filtering atau Stopword Removal

Tahap *filtering* merupakan tahap pengambilan kata-kata yang penting dari hasil token. Pada tahap ini, algoritma *wordlist* (menyimpan kata-kata yang dianggap penting) atau *stoplist* (membuang kata-kata yang dianggap kurang penting) digunakan. kata-kata yang kurang penting dan tidak bersifat deskriptif dalam dokumen, sehingga dapat dibuang. Sedangkan untuk *Stopword* adalah kata-kata yang kurang penting dan tidak bersifat deskriptif dalam dokumen, sehingga dapat dibuang. Contohnya adalah frase "dan", "saya", "kamu", dan sebagainya (Putri, 2016).

Dengan menggunakan *stopword* atau *stoplist*, filtrasi ini dapat menghilangkan kata-kata yang sering muncul yang tidak penting dalam dokumen. Hal ini dilakukan tanpa mengubah makna atau arti dokumen yang akan diproses selanjutnya.

2.3.5. Stemming

Dalam pengolahan kata, *stemming* digunakan untuk mendapatkan kata dasar dari kata yang telah mengalami imbuhan. Ini dilakukan dengan asumsi bahwa kata-kata tersebut sebenarnya memiliki arti dan makna yang sama. Algoritma ini bergantung pada struktur morfologi kalimat Indonesia, yang terdiri dari awalan, akhiran, sisipan, dan awalan+akhiran. Inti dari tahap ini memiliki tujuan. Pertama dalam perkara keefisiensian, pada *stemming* dilakukan pengurangan jumlah kata dalam dokumen agar mengurangi kebutuhan dalam ruang penyimpanan dan mempercepat dalam melakukan pencarian. Kedua dalam perkara keefektifan, *stemming* dilakukan untuk mengurangi *recall* dengan pengurangan bentuk-bentuk kata ke dalam bentuk dasarnya (Putri,2016). Sebagai contoh adalah kata “duduk-lah”, “minum-lah”, “jika-pun” dan sebagainya.

2.4 Google Colab

Google Colab adalah sebuah IDE pemrograman *Python* yang diproses oleh server *Google* yang memiliki perangkat keras cepat. *Google Colab* telah menyediakan hampir semua perangkat lunak yang diperlukan. Penelitian ini membutuhkan *library* seperti *nlTK*, *Sastrawi*, *NumPy*, *Pandas*, dan pendukung lainnya. Misalnya, untuk membuat grafik, *library Matplotlib* diperlukan. Dari sisi perangkat keras, *Google Colab* menawarkan layanan media penyimpanan yang

terintegrasi dengan *Google Drive*, serta prosesor CPU, GPU, TPU, dan RAM. Dengan kemampuan servernya yang stabil, hampir semua pemrosesan tidak akan mengalami kendala dengan *Google Colab* selama koneksi jaringan internet yang lancar (Handayanto, 2020).

Google Colab dibuat diatas environment Jupyter sehingga mirip dengan Jupyter Notebook. *Google Colab* dibangun di atas lingkungan Jupyter, sehingga hampir mirip dengan Jupyter Notebook. Penggunaannya juga hampir sama, hanya media penyimpanannya yang berbeda. *Google Colab* menyediakan *Google Drive* sebagai media penyimpanan dan alat ini berjalan pada sistem cloud. *Google Colab* menyediakan *runtime Python 2* dan *3*, yang telah dikonfigurasi sebelumnya dengan berbagai *library*.

2.5 Kajian Penelitian Terkait

Penelitian ini mengumpulkan informasi dan laporan-laporan penelitian sejenis yang dijadikan sebagai acuan dari studi literatur. Berikut Tabel 2.3 adalah beberapa penelitian sejenis yang dijadikan literatur :

Tabel 2.2 Penelitian Terkait

1.	Peneliti (Tahun)	Dedi Darwis ¹ , Nery Siskawati ¹ , Zaenal Abidin (2021)
	Judul	Penerapan Algoritma <i>Naive Bayes</i> untuk Analisis Sentimen <i>Review</i> Data <i>Twitter</i> BMKG Nasional
	Metode/ Algoritma	Naive Bayes Clasifier
	Tujuan	melakukan pencarian komentar negatif, positif, dan netral dengan review data <i>Twitter</i> BMKG Nasional menggunakan algoritma <i>naive bayes</i>
	Pembahasan	Penelitian ini Mengklasifikasikan Review Data Twiter BMKG Nasional, Data Yang Diambil Yaitu 1179 twet. Data Diklasifikasikan Dengan Menggunakan Naive Bayes Clasifier
	Hasil	Tingkat akurasi berdasarkan pengujian yang dilakukan adalah 68,97%
2.	Peneliti (Tahun)	Faizal Fakhri Irfani, Mohamad Triyanto, Anggit Dwi Hartanto (2020)
	Judul	Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine
	Metode/ Algoritma	Algoritma Support Vector Machine
	Tujuan	Untuk Mengklasifikasikan Review atau ulasan terhadap Terhadap Aplikasi Ruangguru Dari situs Google Play Store
	Pembahasan	Penelitian ini Mengklasifikasikan ulasan yang diberikan pengguna Aplikasi Ruangguru yang Diambil Dari situs Google Play Store, Data Yang Diambil Yaitu Data Teks Review Dengan Jumlah 2000 Review. Data Diklasifikasikan Dengan Menggunakan Algoritma Support Vector Machine, Dan Dilakukan Pengujian Menggunakan Kombinasi Dari Pembagian Data Latih Dan Data Uji.

Tabel 2.2 Penelitian Terkait (Lanjutan 1)

2.	Hasil	Tingkat Akurasi Dalam Penelitian Ini Tinggi Berada Di Kisaran 90%. Hasil Dari Beberapa Pengujian Menunjukkan Bahwa Sentimen Masyarakat Terhadap Aplikasi Ruangguru Cenderung Positif.
	Peneliti (Tahun)	Hafzullah I,S, Taner Tuncer (2019)
3.	Judul	<i>Interaction-Based Behavioral Analysis Of Twitter Social Network Accounts</i>
	Metode/ Algoritma	<i>k-nearest neighbor (K-NN), support vector machine (SVM), dan artificial neural network (ANN)</i>
	Tujuan	Penelitian ini bertujuan untuk mengidentifikasi akun sosial media dan mengklasifikasikan dengan tiga kategori yaitu akun populer-aktif, pengamat-pasif, dan spam-bot-berbahaya
	Pembahasan	Dalam penelitian ini, analisis perilaku pengguna media sosial dilakukan, dan mereka diklasifikasikan sesuai dengan karakteristik dan interaksi akun mereka. Kumpulan data terbaru yang terdiri dari 4209 data latih dan pengguna uji telah dibuat. Hasilnya ditentukan bahwa 10 berbeda parameternya memadai dan efektif dalam mengklasifikasikan pengguna.
	Hasil	Menurut hasil yang diperoleh, nilai akurasi klasifikasi ditemukan mencapai maksimum 96,81% dan minimum 92,33%. Hasil klasifikasi penelitian menunjukkan bahwa metode yang diusulkan adalah cocok untuk pemisahan akun populer-aktif, pengamat-pasif, dan spam-bot-berbahaya
4.	Peneliti (Tahun)	Budi Haryanto, Yova Ruldeviyani, Fathur Rohman, Julius Dimas T. N., Ruth Magdalena, Muhamad Yasil F [2019]
	Judul	<i>Facebook Analysis of Community Sentiment on 2019 Indonesian Presidential Candidates from Facebook Opinion Data</i>
	Metode/ Algoritma	<i>Naive Bayes Clasifier</i>
	Tujuan	untuk mengetahui popularitas dan sentimen publik terhadap capres

Tabel 2.2 Penelitian Terkait (Lanjutan 2)

4.	Pembahasan	Dalam penelitian ini Data diperoleh melalui komentar di postingan Facebook tiga media berita besar di Indonesia, yaitu Detik (@detikcom), TribunNews (@tribunnews), dan Liputan6 (@liputan6online). Pendapat tersebut diklasifikasikan menggunakan pendekatan data mining menggunakan algoritma Naive Bayes Classifier
	Hasil	Riset ini menemukan bahwa 40,52% komentar berbicara tentang capres-cawapres Joko Widodo dan Maruf Amin sedangkan capres Prabowo Subianto dan pasangannya Sandiaga Uno mendominasi komentar sebesar 59,48%. . Sebaliknya, untuk hasil polaritas sentimen, Jokowi-Maruf mendominasi dengan sentimen positif 56,76% dan sentimen negatif 43,24%, sedangkan Prabowo-Sandi mendapat sentimen positif 24,21% dan sentimen negatif 75,79%.
5.	Peneliti (Tahun)	Normah [2019]
	Judul	<i>Naïve Bayes Algorithm For Sentiment Analysis Windows Phone Store Application Reviews</i>
	Metode/ Algoritma	<i>Naïve Bayes</i>
	Tujuan	Penelitian ini bertujuan untuk menganalisis sentimen pengguna aplikasi Windows Phone Store oleh secara otomatis mengklasifikasikan ulasan ke dalam kategori opini positif atau negatif.
	Pembahasan	Dalam penelitian ini menggunakan algoritma naive bayes. Review data dikumpulkan melalui tahap preprocessing untuk diklasifikasikan. Evaluasi Proses dilakukan dengan menggunakan validasi silang 10 kali lipat. Akurasi pengukuran diukur dengan confusion matriks.
Hasil	Berdasarkan hasil pengujian, akurasi 84,50% diperoleh, dan dapat dikatakan bahwa Naïve Bayes memang metode yang baik dalam mengklasifikasikan teks terutama dalam kasus analisis sentimen seperti dalam penelitian ini.	

Tabel 2.2 Penelitian Terkait (Lanjutan 3)

6.	Peneliti (Tahun)	Euis Oktavianti, Ade Rahma Yuly, Fitria Nugrahani [2019]
	Judul	<i>Implementation Of Naïve Bayes Classification Algorithm On Infant And Toddler Nutritional Status</i>
	Metode/ Algoritma	<i>Naïve Bayes Classifier</i>
	Tujuan	Penelitian ini dilakukan untuk mengklasifikasikan Status Gizi Bayi Dan Balita berdasarkan data tentang berat badan dan Panjang tubuh.
	Pembahasan	pengklasifikasian dilakukan dengan menghitung data probabilitas, yang digunakan dalam hal ini studi untuk mengukur status gizi bayi dan anak-anak. Variabel yang digunakan dalam mengevaluasi gizi status bayi dan balita adalah data tentang berat badan dan Panjang
Hasil	Berdasarkan hasil penelitian, digunakan 46 data tentang bayi dan balita dibagi menjadi 90% data uji dan 10% data latih. keakuratan Naïve Bayes dalam klasifikasi status gizi bayi dan balita adalah 75%	
7.	Peneliti (Tahun)	Yono Cahyono, Saprudin (2019)
	Judul	Analisis Sentiment Tweets Berbahasa Sunda Menggunakan <i>Naive Bayes Classifier</i> Dengan <i>Seleksi Feature Chi Squared Statistic</i>
	Metode/ Algoritma	<i>Naïve Bayes Classifier</i> (NBC) dan metode pemilihan <i>Fitur Chi Squared Statistics</i>
	Tujuan	Tujuan dari penelitian ini adalah menerapkan algoritma klasifikasi <i>Naïve Bayes Classifier</i> (NBC) dan metode pemilihan <i>Fitur Chi Squared Statistics</i> untuk menganalisis sentimen <i>tweets</i> berbahasa Sunda di media social <i>Twitter</i> ke dalam kategori positif, negatif dan netral.
	Pembahasan	Penelitian dilakukan dengan kombinasi metode atau melakukan penentuan jumlah fitur pada proses seleksi fitur. <i>Dataset</i> data yang digunakan sebanyak 316 <i>tweets</i> . Proses klasifikasi di sini untuk menentukan sebuah kalimat terklasifikasi ke 3 kategori, berdasarkan nilai perhitungan probabilitas <i>naïve bayes</i> yang lebih besar.

Tabel 2.2 Penelitian Terkait (Lanjutan 4)

7.	Hasil	Hasil uji fitur <i>Chi Square Statistic</i> dapat mengurangi fitur yang tidak relevan dalam proses klasifikasi <i>Naïve Bayes Classifier</i> pada <i>tweets</i> berbahasa Sunda dengan akurasi 78,48%.
	Peneliti (Tahun)	Veny Amilia Fitri, Rachmadita Andreswari, Muhammad Azani Hasibuan (2019)
8.	Judul	<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>
	Metode/ Algoritma	<i>Naïve Bayes, Decision Tree, dan Random Forest.</i>
	Tujuan	Untuk menganalisis kecenderungan komentar publik tentang kampanye Anti-LGBT di Indonesia, apakah itu positif, negatif, atau netral.
	Pembahasan	Tahapan dalam melakukan analisis sentimen dalam penelitian ini adalah preprocessing data, pengolahan data, klasifikasi, dan evaluasi. Hasil data yang relevan yang dikumpulkan sebanyak 3744 komentar / tweet. Rasio data pelatihan dan pengujian adalah 75:25 di mana 75% digunakan sebagai data pelatihan dan 25% akan digunakan sebagai data pengujian.
	Hasil	Analisis sentimen yang diperoleh dalam penelitian ini menunjukkan bahwa pengguna Twitter di Indonesia memberikan komentar yang lebih netral. Dalam penelitian ini, akurasi 86,43% diperoleh oleh <i>Naïve Bayes</i> pada alat RapidMiner, di mana akurasinya lebih tinggi daripada algoritma lainnya, <i>Decision Tree</i> dan <i>Random Forest</i> adalah 82,91%
9.	Peneliti (Tahun)	Dwimarcahyani, D., Badriyah, T., & Karlita, T. (2019)
	Judul	<i>Classification On Category Of Public Responses On Television Program Using Naïve Bayes Method</i>
	Metode/ Algoritma	<i>Naïve Bayes</i>

Tabel 2.2 Penelitian Terkait (Lanjutan 5)

9.	Tujuan	Penelitian ini bertujuan mengimplementasikan algoritma <i>Naive Bayes</i> untuk mengkategorikan tanggapan publik dari dua media sosial Twitter dan Facebook.
	Pembahasan	penelitian ini melalui tahapan mulai dari mengambil data dari media sosial, Preprocessing data, menyimpan data ke database, dan proses kategori klasifikasi. Penelitian ini memproses tanggapan data menjadi dua kategori yaitu kategori stasiun tv dan kategori program.
	Hasil	Hasil dari penelitian ini adalah tanggapan publik yang telah dikategorikan untuk memudahkan KPI atau orang-orang yang mengakses dapat menentukan program terbaik. Dari 326 jumlah data yang telah digunakan sebagai dataset dengan prosentase 80% data latih dan 20% data uji dan menghasilkan nilai akurasi 82%.
10.	Peneliti (Tahun)	Tri Herdiawan Apandi, Castaka Agus Sugiant (2019)
	Judul	Algoritma <i>Naive Bayes</i> untuk Prediksi Kepuasan Pelayanan Perekaman e-KTP
	Metode/ Algoritma	<i>Naive Bayes</i>
	Tujuan	Penelitian ini bertujuan untuk mengetahui lebih dini tingkat kepuasan masyarakat terhadap pelayanan perekaman e-KTP dengan dilakukannya data mining
	Pembahasan	Pada penelitian ini akan memprediksi tingkat kepuasan pelayanan perekaman e-KTP dengan mengumpulkan 17 indikator dengan menggunakan skala likert. Data kepuasan yang telah dikumpulkan akan dipisahkan antara data uji dan data latih. Hasil dari model data latih digunakan untuk melihat akurasi pada data uji.
	Hasil	Algoritma <i>Naive Bayes</i> lebih akurat untuk data tingkat kepuasan pelayanan e-ktp di Kecamatan Batujajar dengan akurasi sebesar 91.70% dan nilai pengujian f-measure sebesar 93,92%, sedangkan hasil dari akurasi algoritma <i>decision tree</i> 65.90%.

Tabel 2.2 Penelitian Terkait (Lanjutan 6)

11.	Peneliti (Tahun)	Fithri Selva Jumeilah (2018)
	Judul	Klasifikasi Opini Masyarakat Terhadap Jasa Ekspedisi JNE Dengan <i>Naïve Bayes</i>
	Metode/ Algoritma	<i>Naïve Bayes</i>
	Tujuan	Penelitian ini bertujuan untuk mengklasifikasikan opini masyarakat terhadap jasa ekspedisi JNE.
	Pembahasan	Jumlah data yang digunakan dalam penelitian berjumlah 1 725 <i>tweets</i> . Data akan dibagi dua 70% <i>data training</i> sebanyak 1208 data dan 30% <i>data testing</i> atau sebanyak 517 data. Sebelum proses klasifikasi data sebelumnya harus melalui proses <i>preprocessing</i> .
	Hasil	Dari hasil klasifikasi komentar pelanggan JNE diperoleh bahwa <i>Naïve bayes</i> mampu mengklasifikasikan data dengan baik. Hasil perhitungan ini dibuktikan dari nilai rata-rata presentase akurasi 85%, <i>precision</i> 78% dan <i>recall</i> 67%.
12.	Peneliti (Tahun)	Fatma Saviera ¹ , Andry Alamsyah (2018)
	Judul	Penggunaan Analisis Sentimen Menggunakan Metode <i>Naive Bayes Classifier</i> Untuk <i>Marketing Intelligence</i> Dengan Membandingkan Tingkat Kepuasan Pelanggan Terhadap <i>E-Commerce Di Indonesia</i>
	Metode/ Algoritma	<i>Naive Bayes Classifier</i>
	Tujuan	Mengklasifikasikan sentimen pelanggan dalam media sosial twitter
	Pembahasan	Penelitian ini mengambil dan menggunakan konten dari Twitter dalam Bahasa Indonesia . Dari proses pengambilan data berdasarkan kata kunci selama rentang waktu tersebut terkumpul data sebanyak 69.598 <i>tweets</i> . Setelah data terkumpul dilakukan <i>praprosesing</i> . Sehingga terseleksi data sebanyak 2.974 <i>tweets</i> , dengan 1.243 <i>tweets</i> Bukalapak, 1.178 <i>tweets</i> Tokopedia dan 553 <i>tweets</i> Elevenia.

Tabel 2.2 Penelitian Terkait (Lanjutan 7)

12.	Hasil	Dari hasil pengujian dari ketiga <i>e-commerce</i> menunjukkan bahwa tingkat sentiment negatif lebih mendominasi dibandingkan tingkat sentiment positif di media sosial Twitter. Dari hasil pengujian data terhadap Bukalapak, Tokopedia dan Elevenia, Elevenia mempunyai tingkat sentiment positif paling tinggi di media sosial Twitter yaitu sebesar 46,3%, Tokopedia sebesar 46,2% dan Bukalapak 45,9%.
13.	Peneliti (Tahun)	Yustia Hapsari, Muhammad Fikri Hidayattullah), Dairoh, Mohammad Khambali (2018)
	Judul	<i>Opinion Mining Terhadap Toko Online Di Media Sosial Menggunakan Algoritma Naïve Bayes (Studi Kasus: Akun Facebook Dugal Delivery)</i>
	Metode/ Algoritma	<i>Naïve Bayes</i>
	Tujuan	Untuk untuk mengklasifikasikan hasil sentimen Toko <i>Online Dugal Delivery</i> di Media Sosial Facebook
	Pembahasan	Penelitian ini membuat sebuah sistem <i>opinion mining</i> dengan menerapkan algoritma <i>Naïve Bayes</i> dengan mengambil studi kasus akun facebook Dugal Delivery. <i>Data training</i> yang digunakan ini berjumlah sebanyak 49 komentar positif dan 26 komentar bernilai negatif.
Hasil	pengujian menggunakan pendekatan <i>confusion matrix</i> diperoleh hasil untuk presisi sebesar 88,89%, <i>recall</i> sebesar 80% dan akurasi sebesar 85%	
14.	Peneliti (Tahun)	Suwanda Aditya Saputra, Didi Rosiyadi , Windu Gata, Syepry Maulana Husain [2017]
	Judul	Analisis Sentimen E-Wallet Pada Google Play Menggunakan Algoritma Naive Bayes (Studi Kasus : OVO)
	Metode/ Algoritma	Naive Bayes
	Tujuan	Untuk menganalisis data ulasan aplikasi E-wallet OVO menggunakan Naive Bayes dengan optimasi penggunaan <i>Feature Selection (FS) Particle Swarm Optimization</i> .

Tabel 2.2 Penelitian Terkait (Lanjutan 8)

15.	Pembahasan	Dalam penelitian dilakukan pengkategorian positif atau negatif terhadap ulasan pengguna.. Metode klasifikasi <i>Naive Bayes</i> digunakan pada analisis sentimen ulasan aplikasi <i>E-wallet OVO</i> .
	Hasil	Berdasarkan penelitian dari <i>cross validation</i> NB tanpa Feature Selection menghasilkan 82.30 % untuk <i>accuracy</i> dan 0.780 untuk AUC. Sedangkan hasil <i>Naive Bayes</i> dengan <i>Feature Selection</i> menghasilkan akurasi adalah 83.60 % dan 0.801 untuk AUC. Peningkatan sangat signifikan dengan penggunaan <i>Feature Selection (FS) Particle Swarm Optimization</i> .
16.	Peneliti (Tahun)	Syahmia Gusriani, Kartina Diah Kusuma Wardhani, Muhammad Ihsan Zul (2017)
	Judul	Analisis Sentimen Terhadap Toko Online di Sosial Media Menggunakan Metode Klasifikasi Naïve Bayes (Studi Kasus: Facebook Page BerryBenka)
	Metode/ Algoritma	Algoritma Naïve Bayes
	Tujuan	Penelitian ini bertujuan untuk mengklasifikasikan opini atau sentimen publik terhadap pelayanan dan produk suatu toko <i>online</i> yaitu BerryBenka di Facebook
	Pembahasan	Pada penelitian ini langkah yang digunakan untuk melakukan analisis sentimen dimulai dari <i>collecting data, preprocessing, feature selection</i> , klasifikasi dan pengukuran akurasi.
	Hasil	Hasil analisis pengujian menunjukkan kestabilan akurasi setelah diuji dengan <i>k-Fold Cross Validation</i> dan <i>Confusion Matrix</i> dengan akurasi 93.7% dimana <i>minimum support</i> untuk <i>Frequent Itemset Mining</i> adalah 0.014

2.3 Penelitian Terdekat

Berikut Tabel 2.3 adalah beberapa penelitian terdekat :

Tabel 2.3 Penelitian Terdekat

1.	Peneliti (Tahun)	Dwimarcahyani, D., Badriyah, T., & Karlita, T. (2019)
	Judul	<i>Classification On Category Of Public Responses On Television Program Using Naive Bayes Method</i>
	Metode/ Algoritma	<i>Naive Bayes</i>
	Tujuan	Penelitian ini bertujuan mengimplementasikan algoritma naive bayes metode untuk mengkategorikan tanggapan publik dari dua media sosial Twitter dan Facebook.
	Pembahasan	Dalam penelitian ini melalui tahapan mulai dari mengambil data dari media sosial, Preprocessing data, menyimpan data ke database, dan proses kategori klasifikasi. Penelitian ini memproses tanggapan data menjadi dua kategori yaitu kategori stasiun tv dan kategori program.
	Hasil	Hasil dari penelitian ini adalah daftar informasi tentang tanggapan publik yang telah dikategorikan untuk memudahkan KPI atau orang-orang yang mengakses dapat menentukan program terbaik. Dari 326 jumlah data yang telah digunakan sebagai dataset dengan prosentase 80% data latih dan 20% data uji dan menghasilkan akurasi 82%.
2.	Peneliti (Tahun)	Fithri Selva Jumeilah (2018)

Tabel 2.3 Penelitian Terdekat (Lanjutan 2)

	Judul	Klasifikasi Opini Masyarakat Terhadap Jasa Ekspedisi JNE Dengan <i>Naïve Bayes</i>
	Metode/ Algoritma	<i>Naïve Bayes</i>
	Tujuan	Penelitian ini bertujuan untuk mengklasifikasikan opini masyarakat terhadap jasa ekspedisi JNE.
	Pembahasan	Jumlah data yang digunakan dalam penelitian berjumlah 1725 <i>tweets</i> . Data akan dibagi dua 70% <i>data training</i> dan 30% <i>data testing</i> . Tahapannya adalah proses <i>preprocessing</i> , <i>pelabelan</i> dan <i>training</i> untuk mendapatkan model probabilitas untuk setiap kategori. Probailitas diperoleh dengan menggunakan algoritma <i>Naïve bayes</i> .
	Hasil	Dari hasil klasifikasi komentar pelanggan JNE diperoleh bahwa <i>Naïve bayes</i> mampu mengklasifikasikan data dengan baik. Hal ini dibuktikan dari nilai rata-rata presentase akurasi 85%, <i>precision</i> 78% dan <i>recall</i> 67%.
3.	Peneliti (Tahun)	Suwanda Aditya Saputra, Didi Rosiyadi , Windu Gata, Syepry Maulana Husain [2017]
	Judul	Analisis Sentimen E-Wallet Pada <i>Google Play</i> Menggunakan Algoritma Naive Bayes (Studi Kasus : OVO)
	Metode/ Algoritma	<i>Naive Bayes</i>
	Tujuan	Untuk menganalisis data ulasan aplikasi E-wallet OVO menggunakan Naive Bayes dengan optimasi penggunaan <i>Feature Selection (FS) Particle Swarm Optimization</i> .

Tabel 2.3 Penelitian Terdekat (Lanjutan 3)

	Pembahasan	Dalam penelitian Algoritma <i>Naive Bayes</i> digunakan untuk mengkategorikan ulasan pengguna, apakah termasuk positif atau negatif.
3	Hasil	Berdasarkan penelitian dari <i>cross validation</i> NB tanpa Feature Selection menghasilkan 82.30 % untuk <i>accuracy</i> dan 0.780 untuk AUC. Sedangkan hasil <i>Naive Bayes</i> dengan <i>Feature Selection</i> menghasilkan akurasi adalah 83.60 % dan 0.801 untuk AUC. Peningkatan sangat signifikan dengan penggunaan <i>Feature Selection (FS) Particle Swarm Optimization</i> .

Tabel 2.3 menerangkan beberapa penelitian yang dijadikan acuan terkait analisis sentimen komentar terhadap aplikasi shopee dengan menggunakan algoritma Naive Bayes. Penelitian berjudul “*Classification On Category Of Public Responses On Television Program Using Naive Bayes Method*” yang dilakukan oleh Dwimarcayani dkk (2019). Penelitian tersebut melakukan klasifikasi tanggapan publik terhadap program di televisi. Hasil penelitian tersebut menunjukkan bahwa dari 326 jumlah data yang telah digunakan sebagai dataset dengan prosentase 80% dari data latih dan 20% dari data uji menghasilkan nilai akurasi 82%.

Penelitian lain yang menggunakan *algoritma Naive Bayes* dengan objek penelitian yang hampir serupa adalah penelitian berjudul “*Klasifikasi Opini Masyarakat Terhadap Jasa Ekspedisi JNE Dengan Naive Bayes*” yang dilakukan oleh Fithri (2018) membahas tentang klasifikasi opini masyarakat terhadap jasa

ekspedisi JNE menggunakan algoritma *Naive Bayes* yang menghasilkan nilai akurasi 85%, *precision* 78% dan *recall* 67%.

Penelitian berjudul “Analisis Sentimen E-Wallet Pada *Google Play* Menggunakan Algoritma *Naive Bayes*“ yang dilakukan Suwandar dkk (2017). Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna E-wallet OVO di situs *Google Play* . Dari hasil pengujian, Hasil dari *cross validation* Naive Bayes dengan FS adalah 83.60 % untuk *accuracy* dan 0.801 untuk AUC.

Berdasarkan ketiga penelitian yang telah disebutkan, maka diusulkan penelitian yang berjudul “Penerapan Algoritma *Naive Bayes Classifier* Untuk Analisis Sentimen Data Ulasan Aplikasi *E-Commerce* Shopee Pada *Google Play Store*”.

2.5.2 Matrik Penelitian

Berikut Tabel 2.4 merupakan matrik penelitian terdahulu:

Tabel 2.4 Matrik Penelitian

No	Penulis/Tahun	Judul	Ruang Lingkup													
			Algoritma								Tujuan			Objek		
			SVM	SGD	Neural Network	Decision Tree	Naive Bayes	Rando Forest	LSTM	Linguistic	KNN	Deteksi	Klasifikasi	Review	Facebook	Twitter
1.	Dedi Darwis1,*, Nery Siskawati1, Zaenal Abidin (2021)	Penerapan Algoritma <i>Naive Bayes</i> untuk Analisis Sentimen <i>Review Data Twitter</i> BMKG Nasional	-	-	-	-	√	-	-	-	-	-	-	-	-	-
2.	Faizal Fakhri Irfani, Mohamad Triyanto, Anggit Dwi Hartanto (2020)	Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma <i>Support Vector Machine</i>	√	-	-	-	-	-	-	-	-	√	-	-	-	√
3.	Hafzullah I,S, Taner Tuncer (2019)	<i>Interaction-Based Behavioral Analysis Of Twitter Social Network Accounts</i>	√	-	√	-	-	-	-	-	√	-	-	-	√	-

Tabel 2.4 Matrik Penelitian (Lanjutan 1)

No	Penulis/Tahun	Judul	Ruang Lingkup														
			Algoritma							Tujuan			Objek				
			SVM	SGD	Neural Network	Decision Tree	Naive Bayes	Rando Forest	LSTM	Linguistic	KNN	Deteksi	Klasifikasi	Review	Facebook	Twitter	Website
4.	Budi Haryanto, Yova Ruldeviyani, Fathur Rohman, Julius Dimas T. N., Ruth Magdalena, Muhamad Yasil F [2019]	<i>Facebook Analysis of Community Sentiment on 2019 Indonesian Presidential Candidates from Facebook Opinion Data</i>	-	-	-	-	√	-	-	-	-	-	√	-	√	-	-
5.	Normah [2019]	<i>Naïve Bayes Algorithm For Sentiment Analysis Windows Phone Store Application Reviews</i>	-	-	-	-	√	-	-	-	-	-	√	-	-	-	√
6.	Euis Oktavianti, Ade Rahma Yuly, Fitria Nugrahani [2019]	<i>Implementation Of Naïve Bayes Classification Algorithm On Infant And Toddler Nutritional Status</i>	-	-	-	-	√	-	-	-	-	-	√	-	-	-	-

Tabel 2.4 Matrik Penelitian (Lanjutan 2)

No	Penulis/Tahun	Judul	Ruang Lingkup														
			Algoritma								Tujuan			Objek			
			SVM	SGD	Neural Network	Decision Tree	Naive Bayes	Rando Forest	LSTM	Linguistic	KNN	Deteksi	Klasifikasi	Review	Facebook	Twitter	Website
7.	Yono Cahyono, Saprudin (2019)	Analisis Sentiment Tweets Berbahasa Sunda Menggunakan Naive Bayes Classifier Dengan Seleksi Feature Chi Squared Statistic	-	-	-	-	√	-	-	-	-	-	√	-	-	√	-
8.	Veny Amilia Fitri, Rachmadita Andreswari, Muhammad Azani Hasibuan (2019)	Sentiment Analysis Of Social Media Twitter With Case Of Anti-lgbt Campaign In Indonesia Using Naive Bayes, Decision Tree, And Random Forest Algorithm	-	-	-	√	√	-	√	-	-	-	√	-	-	√	-

Tabel 2.4 Matrik Penelitian (Lanjutan 3)

No	Penulis/Tahun	Judul	Ruang Lingkup													
			Algoritma							Tujuan			Objek			
			SVM	SGD	Neural Network	Decision Tree	Naive Bayes	Rando Forest	LSTM	Linguistic	KNN	Deteksi	Klasifikasi	Review	Facebook	Twitter
9.	Dwimarcahyani, D., Badriyah, T., & Karlita, T. (2019)	Classification On Category Of Public Responses On Television Program Using Naive Bayes Method	-	-	-	-	√	-	-	-	-	√	-	√	√	-
10.	Tri Herdiawan Apandi, Castaka Agus Sugiant (2019)	Algoritma Naive Bayes untuk Prediksi Kepuasan Pelayanan Perekaman e-KTP	-	-	-	-	√	-	-	-	-	√	-	-	-	√

Tabel 2.4 Matrik Penelitian (Lanjutan 4)

No	Penulis/Tahun	Judul	Ruang Lingkup													
			Algoritma							Tujuan			Objek			
			SVM	SGD	Neural Network	Decision Tree	Naive Bayes	Rando Forest	LSTM	Linguistic	KNN	Deteksi	Klasifikasi	Review	Facebook	Twitter
11.	Fithri Selva Jumeilah (2018)	Klasifikasi Opini Masyarakat Terhadap Jasa Ekspedisi JNE Dengan Naive Bayes	-	-	-	-	√	-	-	-	-	√	-	-	√	-
12.	Fatma Saviera1, Andry Alamsyah (2018)	Penggunaan Analisis Sentimen Menggunakan Metode <i>Naive Bayes Classifier</i> Untuk <i>Marketing Intelligence</i> Dengan membandingkan Tingkat Kepuasan Pelanggan Terhadap <i>E-Commerce Di Indonesia</i>	-	-	-	-	√	-	-	-	-	√	-	-	√	-

Tabel 2.4 Matrik Penelitian (Lanjutan 5)

No	Penulis/Tahun	Judul	Ruang Lingkup														
			Algoritma								Tujuan			Objek			
			SVM	SGD	Neural Network	Decision Tree	Naive Bayes	Rando Forest	LSTM	Linguistic	KNN	Deteksi	Klasifikasi	Review	Facebook	Twitter	Website
13.	Yustia Hapsari, Muhammad Fikri Hidayattullah), Dairoh, Mohammad Khambali (2018)	<i>Opinion Mining Terhadap Toko Online Di Media Sosial Menggunakan Algoritma Naïve Bayes (Studi Kasus: Akun Facebook Dugal Delivry)</i>	-	-	-	-	√	-	-	-	-	-	√	-	√	-	-
14.	Suwanda Aditya Saputra, Didi Rosiyadi, Windu Gata, Syepny Maulana Husain [2017	Analisis Sentimen E-Wallet Pada Google Play Menggunakan Algoritma Naive Bayes	-	-	-	-	√	-	-	-	-	-	√	-	-	-	√

Tabel 2.4 Matrik Penelitian (Lanjutan 6)

No	Penulis/Tahun	Judul	Ruang Lingkup													
			Algoritma							Tujuan			Objek			
			SVM	SGD	Neural Network	Decision Tree	Naive Bayes	Rando Forest	LSTM	Linguistic	KNN	Deteksi	Klasifikasi	Review	Facebook	Twitter
15.	Syahmia Gusriani, Kartina Diah Kusuma Wardhani, Muhammad Ihsan Zul (2017)	Analisis Sentimen Terhadap Toko Online di Sosial Media Menggunakan Metode Klasifikasi Naïve Bayes (Studi Kasus: Facebook Page BerryBenka)	-	-	-	-	√	-	-	-	-	√	-	√	-	-

2.6 State of The Art

Penelitian yang ditulis oleh Fatma Saviera dkk. (2018), "*Sentiment Analysis Using Naive Bayes Classifier Method For Marketing Intelligence Compare Customer Satisfaction Level In Indonesia's E-Commerce*" membahas proses klasifikasi sentimen pelanggan di media sosial Twitter terhadap e-commerce di Indonesia dengan menggunakan algoritma Naive Bayes. Tabel berikut menunjukkan perbandingan persentase sentimen pelanggan terhadap e-commerce di Indonesia pada media sosial Twitter..

Tabel 2.5 persentase sentimen dari penelitian Fatma dkk (2018)

	Jumlah data uji	Sentiment	
		Positif	Negatif
Bukalapak	993	348 (45,9%)	410 (54,1%)
Tokopedia	928	350 (46,2%)	408 (53,8%)
Elevenia	300	163 (46,3%)	189 (53,7%)

Penelitian lain yang menggunakan algoritma Naive Bayes dengan objek penelitian yang hampir serupa adalah penelitian berjudul "*Classification On Category Of Public Responses On Television Program Using Naive Bayes Method*" yang dilakukan oleh Dwimarcayani (2019). Penelitian tersebut melakukan klasifikasi data dengan cara mengelompokkan data menjadi tujuh kategori, yakni *news*, *sport*, *drama*, *children's program*, *entertainment*, *reality show*, dan *talkshow*. Hasil penelitian tersebut digambarkan pada tabel 2.6 yakni tabel mengenai confusion matrix.

Tabel 2.6 Confussion Matrik dari penelitian Dwimarcahyani dkk (2019)

Category	Precission x 100%	Recall x 100%	f1- score x 100%
News	62%	83%	71%
Sport	100%	100%	100%
Drama	82%	100%	90%
Children's Program	75%	60%	67%
Entertainment	91%	83%	87%
Reality show	80%	80%	80%
Talkshow	67%	100%	80%
Average	80%	84%	81%

Suwandar (2017) melakukan penelitian berjudul "Analisis Sentimen *E-Wallet* Pada *Google Play* Menggunakan Algoritma Naive Bayes Berbasis *Particle Swarm Optimization*". Tujuan dari penelitian ini adalah untuk melihat ulasan pengguna OVO E-wallet di situs web *Google Play*. Tabel pengujian yang dibuat oleh Suwandar dkk (2017) dapat dilihat di Tabel 2.7. Model *Naive Bayes Classifier* tanpa *feature selection* menghasilkan akurasi sebesar 82.30% dan nilai AUC 0.780 jika dibandingkan dengan penggunaan *feature selection* akurasi meningkat secara signifikan menjadi 83.60% dan nilai AUC 0.801.

Tabel 2.7 Perbandingan Akurasi dari Penelitian Suwandar (2017)

	NBC tanpa FS	NBC + FS	Selisih
<i>Accuracy</i>	82.30 %	83.60 %	1.3%
<i>AUC</i>	0.780	0.801	0.021

Berdasarkan kajian dari ketiga penelitian yang telah dijelaskan, maka diusulkan penelitian yang dapat mengimplementasikan *Naive Bayes* untuk Analisis Sentimen data ulasan Aplikasi *E-Commerce Shopee* pada situs *Google*

Play terhitung tanggal 1 November 2022. Dalam penelitian ini klasifikasi dilakukan dalam tiga kategori kelas yaitu kelas positif, negatif dan netral. Dari penelitian ini juga akan mengukur tingkat akurasi metode *Naïve Bayes Classifier* dalam Pengklasifikasian Sentimen. *State of The Art* diperoleh dari perbandingan antara penelitian ini dengan penelitian terdekat. Berikut merupakan state of the art dari penelitian ini yang digambarkan melalui tabel relevansi penelitian.

Tabel 2.8 Relevansi Penelitian

Peneliti	Dwimarcahyani, dkk (2019)	Adilia Tri Rizkya (2023)
Judul	<i>Classification On Category Of Public Responses On Television Program Using Naive Bayes Method</i>	Penerapan <i>Naive Bayes Classifier</i> Untuk Analisis Sentimen Data Ulasan Aplikasi <i>E-Commerce</i> Shopee Pada <i>Google Play Store</i>
Masalah Penelitian	<i>Persaingan</i> industri pertelevisian dalam negeri semakin ketat sehingga terkadang kualitas konten jarang diperhatikan. Masalah ini, membuat berbagai tanggapan dari publik dimedia sosial.	Tahun 2022 Indonesia jadi negara dengan pertumbuhan e-commerce tercepat di dunia. Salah satu <i>e-commerce</i> yang sering di akses adalah shopee. Shopee berada di peringkat pertama berdasarkan ranking Playstore. Ulasan yang diberikan pengguna Shopee pun beragam.
Objek Penelitian	Klasifikasi respon publik terhadap program televisi di Twitter	Klasifikasi sentimen pengguna shopee di <i>Google Play Store</i>
Algoritma	<i>Naive Bayes classifier</i>	<i>Naive Bayes classifier</i>
State of The Art	Klasifikasi respon publik terhadap program televisi menjadi 7 kategori menggunakan algoritma <i>Naive Bayes</i>	Penerapan algoritma <i>Naive Bayes</i> dan <i>web scraper</i> untuk klasifikasi data ulasan menjadi 3 kategori