

BAB II TINJAUAN PUSTAKA

2.1 Landasan Teori

Dalam penelitian analisis sentimen Twitter tentang hubungan sipil-militer terkait masalah kemanusiaan pengungsi Rohingya di Indonesia dengan menggunakan metode *Support Vector Machine* dan *Naïve Bayes*, terdapat beberapa landasan teori yang relevan, antara lain:

2.1.1 Analisis sentimen

Analisis sentimen adalah salah satu teknik *data mining* yaitu metode yang digunakan untuk mengidentifikasi, mengekstraksi, dan menganalisis sentimen atau opini yang terkandung dalam teks. Pendekatan ini biasanya melibatkan pemrosesan bahasa alami atau NLP (*Natural Language Processing*) untuk mengklasifikasikan teks menjadi kategori sentimen seperti positif, negatif, atau netral. Konsep dasar dalam analisis sentimen adalah untuk menentukan polaritas atau orientasi emosional dari teks yang dianalisis (Munawaroh, 2022).

2.1.2 Social Media Analytics

Social media analytics adalah bidang studi yang berkaitan dengan pengumpulan, analisis, dan interpretasi data dari platform media sosial. Dalam konteks penelitian ini, analisis sentimen Twitter digunakan sebagai bagian dari *social media analytics* untuk memahami opini publik terkait isu Rohingya dan hubungan sipil militer di Indonesia. *Social media analytics* memanfaatkan data dan alat analitik untuk mendapatkan wawasan tentang pola, tren, dan sikap yang berkembang di media sosial (Andryani et al., 2019).

2.1.3 K-Fold Cross Validation

K-Fold Cross Validation adalah teknik validasi model dengan cara membagi dataset secara acak menjadi *k fold* dimana satu *fold* digunakan untuk menguji (*test set*) sedangkan *k-1 fold* lainnya digunakan sebagai data pelatihan (*training set*). Menurut Nada, metode ini dapat mengurangi bias yang ada pada dataset dengan

jumlah *fold* yang dianjurkan adalah 10 (Nada et al., 2023). Dari penelitian yang dilakukan oleh peneliti, dengan menggunakan 10 *fold* sendiri sudah menghasilkan nilai *fold accuracy* diatas 0,5 atau diatas batas valid. Dengan menerapkan *k-fold cross validation*, model dataset yang dibangun dapat memberikan prediksi yang akurat hingga pada data baru yang belum pernah digunakan sebelumnya. Menurut Sarang, jika nilai *Area Under The Curve* (AUC) 0,7 (70%) atau lebih berarti ada kemungkinan model dataset tersebut valid dan mampu membedakan kelas positif dan negatif, lalu apabila nilai AUC dibawah 0,5 maka model tidak memiliki kemampuan membedakan kelas positif dan negatif, hal paling buruk yang dapat terjadi adalah jika nilai AUC 0, maka model malah melakukan *reciprocating* atau memprediksi kelas negatif sebagai positif dan sebaliknya (Narkhede, 2019).

2.1.4 Support Vector Machine (SVM)

SVM adalah salah satu metode *machine learning* yang digunakan untuk klasifikasi dan regresi. SVM bertujuan untuk mencari *hyperplane* terbaik yang memisahkan dua kelas dalam ruang fitur. SVM dapat digunakan dalam klasifikasi sentimen dengan membangun model yang dapat memisahkan teks yang memiliki sentimen positif dan negatif. SVM sering digunakan dalam tugas-tugas klasifikasi yang kompleks, termasuk analisis sentimen (Harun et al., 2023). SVM dapat dinotasikan pada rumus 2.1 :

$$\hat{f}(X) = W^T x + b$$

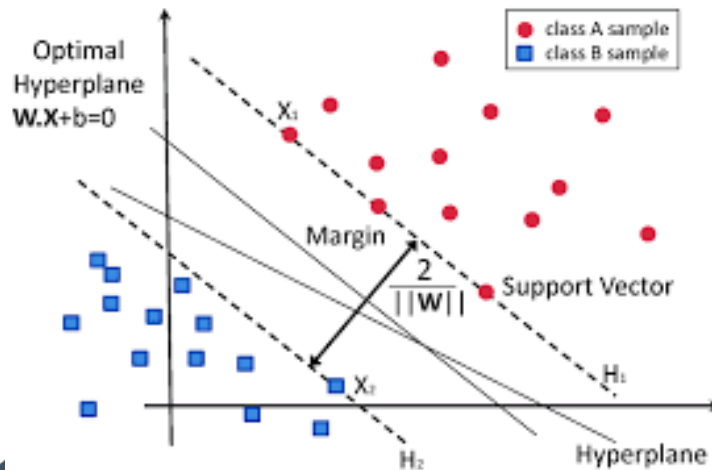
Rumus 2.1 Persamaan SVM

W : vector bobot

x : vektor fitur input

b : bias

Nilai dari $f(x)$ akan digunakan untuk menentukan kelas prediksi berdasarkan tanda dari $f(x)$ (Nugraha, 2022).



Gambar 2.1 Ilustrasi Hyperplane

Sumber: (Siregar et al., 2023)

Hyperplane pada gambar 2.1 berada pada tengah tengah antar tiap kelas. Jarak terdekat ke titik tiap kelas disebut *margin*, sedangkan titik yang jaraknya tepat pada data disebut *support vector*. Persamaan *hyperplane* dinyatakan sebagai berikut :

$$\vec{w} \cdot \vec{x} + b = 0$$

Rumus 2.2 Persamaan Hyperplane

$\vec{w}$ = vektor pembobot $\vec{x}$ = vektor input

b = bias

(Siregar et al., 2023)

2.1.5 Naïve Bayes

Naïve Bayes adalah metode klasifikasi probabilitas yang didasarkan pada teorema Bayes dengan asumsi independensi fitur. Metode ini menghitung probabilitas posterior untuk setiap kelas berdasarkan fitur-fitur yang ada dalam *data training*. *Naïve Bayes* sering digunakan dalam analisis sentimen karena kecepatannya yang tinggi dan sederhana dalam mengklasifikasikan teks menjadi kategori sentimen (Harun et al., 2023). Aturan *Naïve Bayes* dapat dinyatakan sebagai berikut :

$$P(W_i | C) = \frac{\text{count}(w_i.C) + 1}{\text{count}(C) + |V|}$$

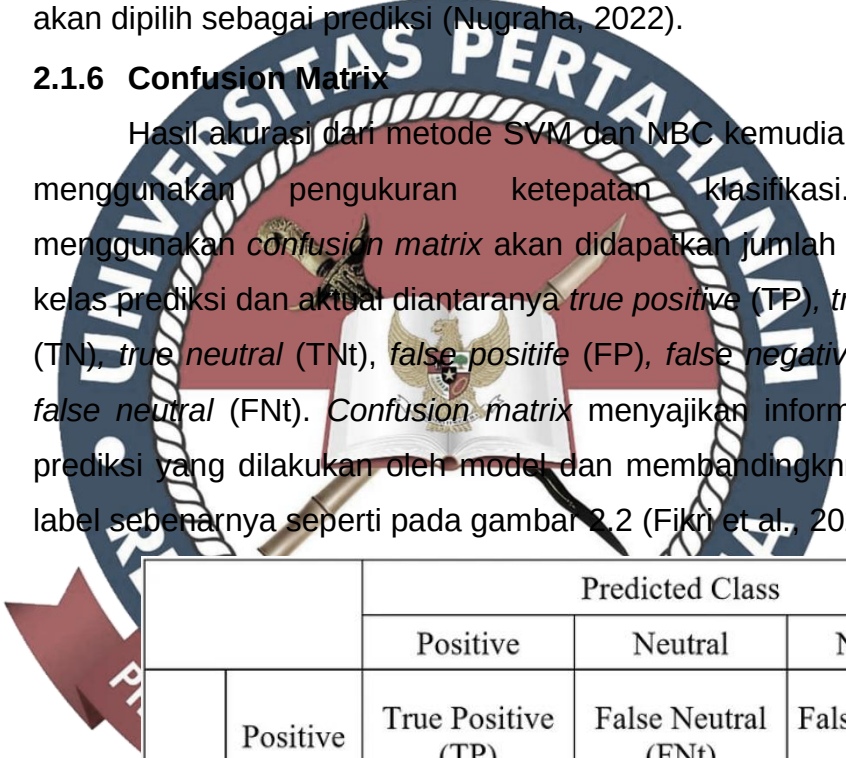
Rumus 2.3 Persamaan Naïve Bayes

- PerWi : adalah kata ke i
 Wi.C : jumlah kata w_i dalam $C(count)$
 $|V|$: total kata pada data *testing*
 Count (C) : jumlah kata pada berdasarkan kategori

Pada tahap pengujian atau prediksi, algoritma menggunakan rumus di atas untuk menghitung probabilitas posterior kelas untuk setiap kelas yang mungkin, dan kelas dengan probabilitas tertinggi akan dipilih sebagai prediksi (Nugraha, 2022).

2.1.6 Confusion Matrix

Hasil akurasi dari metode SVM dan NBC kemudian dievaluasi menggunakan pengukuran ketepatan klasifikasi. Dengan menggunakan *confusion matrix* akan didapatkan jumlah pada setiap kelas prediksi dan aktual diantaranya *true positive* (TP), *true negative* (TN), *true neutral* (TNt), *false positive* (FP), *false negative* (FN), dan *false neutral* (FNt). *Confusion matrix* menyajikan informasi tentang prediksi yang dilakukan oleh model dan membandingkannya dengan label sebenarnya seperti pada gambar 2.2 (Fikri et al., 2020).



		Predicted Class		
		Positive	Neutral	Negative
True Class	Positive	True Positive (TP)	False Neutral (FNt)	False Negative (FN)
	Neutral	False Positive (FP)	True Neutral (TNt)	False Negative (FN)
	Negative	False Positive (FP)	False Neutral (FNt)	True Negative (TN)

Gambar 2.2 Confusion Matrix

Sumber: (Malsi & Jatikusumo, 2022)

True Positive = data positif yang terklasifikasi dengan benar.

True Negative = data negatif yang terklasifikasi dengan benar.

True Neutral = data netral yang terklasifikasi *dengan* benar.

False Positive = data negatif yang terklasifikasi menjadi positif.

False Negative = data positif yang terklasifikasi menjadi negatif.

False Neutral = data netral yang terklasifikasi menjadi negatif.

Nilai yang dihasilkan melalui metode *confusion matrix* diatas nantinya akan menjadi evaluasi sebagai berikut:

- Akurasi

Presentasi jumlah data yang dilakukan pada klasifikasi prediksi secara benar dengan rumus :

$$\frac{TP + TN_t + TN}{TP + TN_t + TN + FP + FN_t + FN}$$

Rumus 2.4 Persamaan Akurasi

- Presisi

Tingkat ketepatan model dalam mengklasifikasikan data sebagai positif dengan cara mengukur sejauh mana hasil positif yang diprediksi oleh model adalah benar menggunakan rumus:

$$\frac{TP}{TP + FP}$$

Rumus 2.5 Persamaan Presisi

- Recall

Tingkat kelengkapan model dalam mengidentifikasi data positif dengan cara mengukur sejauh mana model mampu menemukan atau mengklasifikasikan benar semua kasus yang positif menggunakan rumus :

$$\frac{TP}{TP + FN_t + FN}$$

Rumus 2.6 Persamaan Recall

(Ramlan et al., 2023).

- F1-score

Rata-rata dari nilai presisi dan *recall*

$$\frac{2(\text{presisi} * \text{recall})}{\text{presisi} + \text{recall}}$$

Rumus 2.7 Persamaan F-1 Score

Dengan menggunakan *confusion matrix*, penulis dapat menganalisis akurasi klasifikasi, mengidentifikasi bias, menemukan

kelemahan pada masing-masing metode, dan memilih metode terbaik yang paling sesuai untuk masalah klasifikasi yang dihadapi.

2.1.7 Word Cloud

Pada penelitian ini, penulis merepresentasikan data yang telah diambil menjadi *word cloud* dengan menggunakan aturan *matrix* sebelumnya. Menurut R. Kusmawati, *word cloud* sendiri merupakan visualisasi tekstual yang menggambarkan frekuensi kemunculan kata-kata pada sebuah dokumen (Kusumawati et al., 2019). Kata yang akan direpresentasikan dibagi menjadi 3 kelas yaitu positif, negatif dan netral. Hasil kata positif yang muncul inilah yang dapat menjadi analisis dalam membuat rekomendasi keputusan.

2.2 Hasil Penelitian Terdahulu

Penelitian terdahulu yang menjadi sumber sekaligus bahan ajar untuk penulis dicantumkan pada tabel 2.1, yang pertama penelitian oleh Arafat Hossain bertujuan untuk menggambarkan pola kata-kata di halaman depan surat kabar dan mengaitkannya dengan konteks sosial dan politik pada waktu itu menggunakan teknik *text mining* (Hossain et al., 2021).

Pada penelitian oleh Shriphani Palakodety, dilakukan analisis komentar yang mendukung komunitas Rohingya di media sosial, khususnya komentar yang ada di platform YouTube. Penelitian tersebut bertujuan untuk mempelajari *Artificial Intelligent* (AI) dalam membantu komunitas yang terpinggirkan (Palakodety et al., 2020). Berbeda dengan penelitian Nahian Ahmed yang cenderung mempelajari penggunaan *machine learning* dan *remote sensing* dalam membuat metode estimasi populasi pengungsi Rohingya (Ahmed et al., 2019).

Menurut Hassan Araghizadeh, penerapan faktor-faktor koordinasi dalam konteks organisasi militer dan sipil dapat meningkatkan respon yang lebih baik terhadap bencana alam (Araghizadeh et al., 2020). Maka dari itu, penulis mengembangkan penelitian dengan konteks hubungan sipil militer (*CIMIC*). Pengambilan konteks tersebut akan berdampak pada keputusan

pertahanan seperti penelitian oleh Bunbunan Hesty (Hesty FMT et al., 2021) dan Andrew K. (Baker, 2016) untuk mendukung pertahanan negara.

Pengambilan metode SVM dan *Naïve Bayes* berpedoman pada penelitian oleh Dinar Ajeng (Kristiyanti et al., 2019), penelitian oleh Uning (Lestari et al., 2022), penelitian oleh Dwi Normawati dan Prayogi (Normawati & Prayogi, 2021), penelitian oleh Mujaddid Izzul Fikri (Fikri et al., 2020), serta penelitian oleh Kusmawati (Kusumawati et al., 2019) yang menjelaskan implementasi kedua algoritma tersebut secara terstruktur dalam mengklasifikasi teks di Twitter. Selain itu adapula penelitian terdahulu oleh Oryza Habibie Rahman (Oryza Habibie Rahman et al., 2021), Mohammad Kamrul (*Syntax Admiration*, Vol. 4, No. 9, September 2023, 2023), dan Diva Durnotun Nada (Nada et al., 2023) yang secara keseluruhan menjelaskan tentang beberapa metode analisis sentimen sebagai referensi penulis.

Selain dari penelitian terdahulu, penulis juga mendapatkan referensi dari ahli yang lebih menguasai terkait tema penelitian ini yaitu Bapak Mohammad Nazar, S.Si., M.T. Sebelum melaksanakan penelitian, peneliti berdiskusi terkait judul yang akan diambil pada tanggal 15 Desember 2023 dengan bukti *chat* yang terlampir pada lampiran 18 guna mendapatkan wawasan tentang kasus terkait. Selanjutnya peneliti berdiskusi kembali terkait penggantian topik sekaligus pertimbangan-pertimbangan yang harus diperkirakan pada tanggal 14 Januari 2024 melalui obrolan Whatsapp dengan bukti *chat* yang terlampir pada lampiran 18. Sesi pertemuan terakhir pada tanggal 2 April 2024 peneliti melaksanakan wawancara di Kampus Universitas Pertahanan untuk membahas tentang hasil penelitian sementara dan contoh rumusan keputusan yang dapat diambil dari hasil penelitian ini. Dengan mempertimbangkan banyak faktor, narasumber memberikan contoh rumusan keputusan dari hasil penelitian yang dapat dilihat pada lampiran 19. Dari wawancara ini, penulis mendapatkan semakin banyak ilmu, sudut pandang, dan validasi terkait metode, hasil, dan penelitian yang sedang dikerjakan.

Tabel 2.1 Penelitian Terdahulu

No	Nama dan Tahun Penelitian	Ringkasan	Research Gap
1	<i>Text Mining and Sentiment Analysis of Newspaper Headlines</i> (Hossain et al., 2021)	Menggambarkan pola dan prioritas kata-kata yang muncul dalam <i>headline</i> surat kabar di Bangladesh (tempat pengungsian Rohingya) menggunakan kombinasi teknik analisis teks.	- Menggunakan 2 analisis teks yaitu analisis sentimen dan analisis klaster - Hanya menggunakan <i>Word Clouds</i>
2	<i>Voice for the voiceless: Active sampling to detect comments supporting the Rohingyas</i> (Palakodety et al., 2020)	Menganalisis peran AI dalam konteks krisis pengungsi Rohingya menggunakan komentar Youtube untuk membangun sebuah klasifikasi yang dapat mendeteksi komentar-komentar yang membela Rohingya di antara komentar-komentar yang mengejek dan netral.	- Menggunakan metode <i>Nearest-neighbor active sampling</i> - Data yang diambil yaitu komentar dari Youtube
3	<i>Artificial Neural Network and Machine Learning Based Methods for Population Estimation of Rohingya Refugees: Comparing Data-Driven and Satellite Image-</i>	Estimasi populasi pengungsi Rohingya menggunakan <i>Artificial Neural Network</i> dan <i>machine learning (linear regression)</i> untuk mengeksplorasi pendekatan estimasi populasi menggunakan penginderaan jauh dan <i>machine learning</i>	- Mengidentifikasi masalah geospasial - Menggunakan metode <i>Artificial Neural Network (ANN)</i>

No	Nama dan Tahun Penelitian	Ringkasan	Research Gap
	<i>Driven Approaches</i> (Ahmed et al., 2019)		
4	<i>Civil-Military Coordination (CIMIC) Model in Natural Disasters in Iran</i> (Araghizadeh et al., 2020)	Penelitian ini bertujuan untuk mengembangkan model koordinasi sipil-militer dalam menghadapi bencana alam di Iran.	• Tidak membahas tentang data mining • Menganalisis sentimen para ahli dengan metode Delphi 2 tahap • Konflik yang diteliti tentang bencana alam
5	Implementasi <i>Naïve Bayes Classifier</i> dan <i>Confusion Matrix</i> Pada Analisis Sentimen Berbasis Teks pada <i>Twitter</i> (Normawati & Prayogi, 2021)	Analisis sentimen pada <i>Twitter</i> menggunakan metode NBC untuk mengetahui pendapat atau opini masyarakat terhadap berita kontroversi Ahok berdasarkan komentar positif, netral, dan negatif	• Hanya menggunakan metode NBC • Konflik yang diteliti terkait berita kontroversi
6	Klasifikasi Ujaran Kebencian pada Media Sosial <i>Twitter</i> Menggunakan <i>Support Vector Machine</i> (Oryza Habibie Rahman et al., 2021)	Mengumpulkan data <i>tweet</i> yang ingin dianalisa berdasarkan topik dan kurun waktu tertentu, lalu diklasifikasikan berdasarkan kelasnya pada sistem	• Hanya menggunakan metode SVM • Hanya mencari kelas negatif untuk ujaran kebencian • Konflik yang diteliti tidak spesifik

No	Nama dan Tahun Penelitian	Ringkasan	Research Gap
		klasifikasi ujaran kebencian. untuk mencari nilai akurasi tertinggi	
7	Pemanfaatan Teknologi <i>Big Data Analytics</i> Untuk Mendukung Pertahanan Negara (Hesty FMT et al., 2021)	Menerapkan <i>Big Data Analytics</i> (dengan metode yang telah dicantumkan) dalam menganalisis prediksi konflik untuk mendukung pertahanan negara serta mengidentifikasi dan menganalisis prediksi konflik dari dampak pertumbuhan penduduk untuk mendukung pertahanan negara.	• Hanya menggunakan metode NBC • Memprediksi konflik dari dampak pertumbuhan penduduk
8	<i>Polarity of News Headlines in Social Media and Extent of Anti-Rohingya Rhetoric: Evidence from Online News Media in Bangladesh</i> (Volterrani & Meloni, 2020)	Penelitian yang menganalisis sentimen dan subjektivitas judul berita dan komentar yang diposting di dua halaman Facebook agensi berita populer di Bangladesh terkait krisis Rohingya menggunakan <i>Natural Language Toolkit</i> (NLTK) dalam bahasa pemrograman Python	• Hanya menggunakan metode <i>Natural Language Toolkit</i> (NLTK) • Mengambil data dari platform media sosial Facebook

No	Nama dan Tahun Penelitian	Ringkasan	Research Gap
9	Perbandingan Metode <i>Naïve Bayes</i> dan <i>Support Vector Machine</i> pada Analisis Sentimen Twitter (Fikri et al., 2020)	Klasifikasi analisis sentimen terhadap dataset Twitter yang berkaitan dengan UMM menggunakan metode NBC dan SVM dengan menggunakan tolak ukur akurasi, presisi, <i>recall</i> , dan F1-Score	• Tidak menggunakan metode pembandingan hasil akurasi masing masing algoritma • Objek yang diteliti terkait kampus UMM
10	Perbandingan Analisis Sentimen Mengenai BPJS pada Media Sosial Twitter Menggunakan <i>Naïve Bayes Classifier</i> (NBC) dan <i>Support Vector Machine</i> (SVM) (Nada et al., 2023)	Membandingkan hasil ketepatan klasifikasi pada data Twitter mengenai kinerja BPJS selama pandemi menggunakan metode <i>Naïve Bayes Classifier</i> dan <i>Support Vector Machine</i> .	• Tidak menggunakan metode pembandingan hasil akurasi masing masing algoritma • Objek yang diteliti terkait kinerja BPJS
11	<i>Assessing Sentiment in Conflict Zones Through Social Media</i> (Baker, 2016)	Bagaimana menganalisis sentimen publik dalam zona konflik melalui media sosial dengan menggunakan analisis media sosial untuk mengumpulkan dan menganalisis data dengan fokus pada penggunaan Twitter di Yaman.	• Tidak menggunakan <i>data mining</i>

No	Nama dan Tahun Penelitian	Ringkasan	Research Gap
12	<i>Naives Bayes Algorithm for Twitter Sentiment Analysis</i> (Samsir et al., 2021)	Penelitian ini bertujuan untuk menganalisis opini Masyarakat Indonesia melalui analisis jaringan sosial Twitter dengan metode NBC. Crawling data menggunakan kata kunci "jokowi" dan "coronavirus". Hasil penelitian menunjukkan bahwa opini negatif lebih dominan 46%, maka kesimpulannya Masyarakat Indonesia memiliki perspektif negatif terhadap kebijakan pemerintah pada awal masa pandemi COVID-19.	- Hanya menggunakan metode NBC - Hasil penelitian langsung mengambil dari jumlah data negatif, positif, dan netral menurut <i>matrix</i>.
13	<i>Comparison Performance of Naive Bayes Classifier and Support Vector Machine Algorithm for Twitter's Classification of Tokopedia Services</i> (Kusumawati et al., 2019)	Membandingkan kerja algoritma SVM dan NBC untuk mengklasifikasikan tweet mengenai layanan Tokopedia. Penelitian ini menggunakan akurasi, presisi, dan recall untuk membandingkan unjuk kerja kedua algoritma. Hasil evaluasi menunjukkan bahwa teknik SVM dengan	- Tema permasalahan yang diambil - Proses wordcloud langsung dilakukan sebelum mengklasifikasi data untuk mencari representasi kata pada kelas positif dan negatif

No	Nama dan Tahun Penelitian	Ringkasan	Research Gap
		kernel linier memiliki akurasi yang lebih baik dibandingkan NBC, yaitu 83,34%.	
14	<i>Sentiment Analysis For Extracting Student Opinion Data On Higher Education Services Using The Naive Bayes Classifier And Support Vector Machine Methods (Case Study Akprind Institute Of Science And Technology Yogyakarta)</i> (Lestari et al., 2022)	Penelitian ini bertujuan untuk membuat sistem pemrosesan data opini mahasiswa menggunakan metode Naive Bayes Classifier (NBC) dan Support Vector Machine (SVM) yang dioptimalkan dengan pemilihan fitur Information Gain (IG). Metode SVM menghasilkan akurasi yang lebih tinggi daripada NBC mencapai 99,09%, sedangkan NBC 96,56%. Penerapan IG tidak secara signifikan mempengaruhi akurasi analisis, tetapi sangat memengaruhi durasi analisis SVM, yang dapat memperpendek waktu hingga 195,71%.	• Metode SVM dan NBC yang digunakan ditambah dengan fitur IG • Hasil penelitian digunakan untuk evaluasi pengembangan kampus
15	<i>Comparison of SVM & Naïve Bayes Algorithm for Sentiment Analysis Toward West Java</i>	Penelitian ini membahas tentang perbandingan SVM dan NBC untuk melakukan analisis sentimen terhadap	• Tujuan penelitian lebih fokus untuk mencari sentimen positif

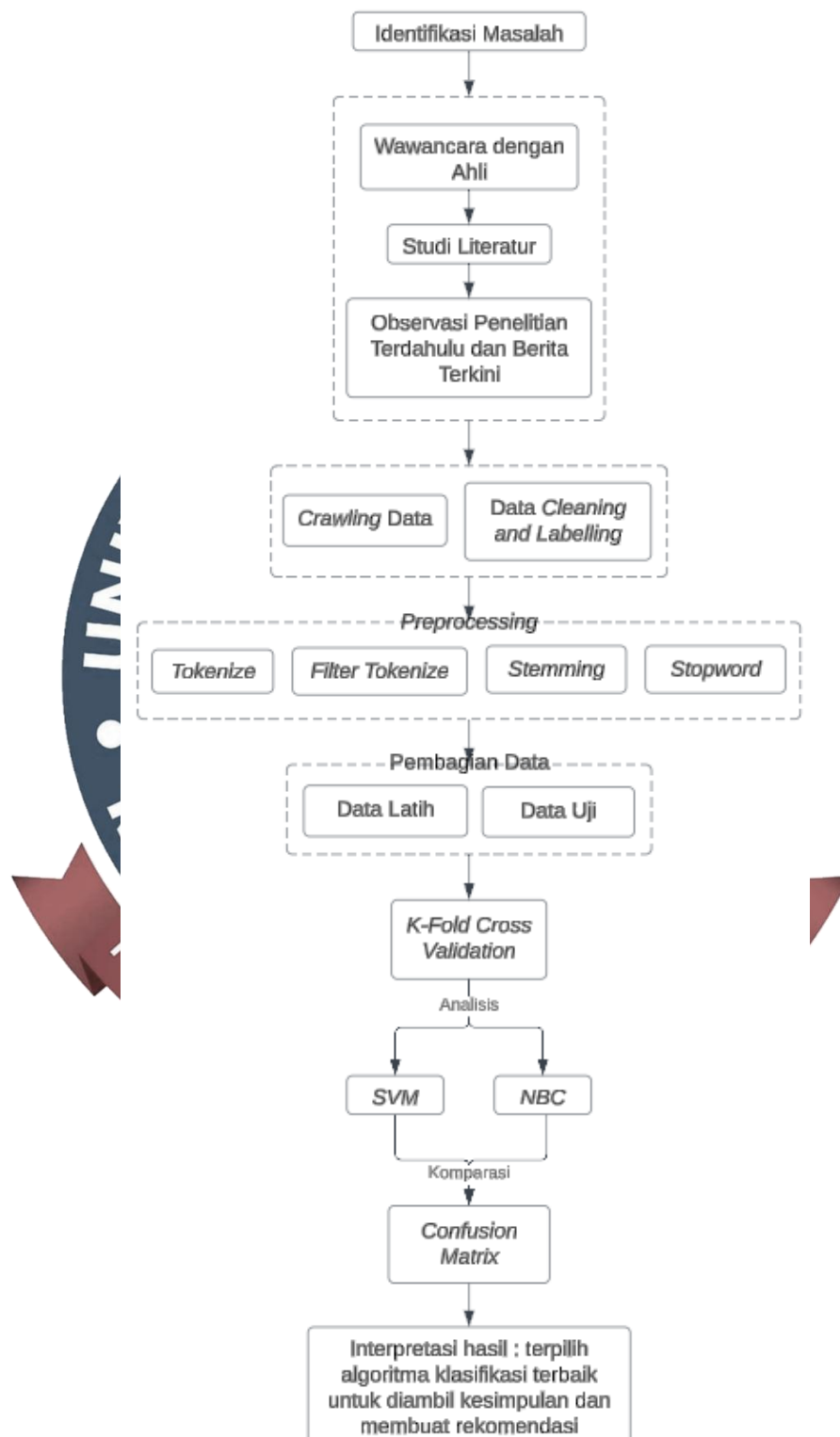
No	Nama dan Tahun Penelitian	Ringkasan	Research Gap
	<i>Governor Candidate Period 2018-2023 Based on Public Opinion on Twitter</i> (Kristiyanti et al., 2019)	opini publik di Twitter mengenai kandidat Gubernur Jawa Barat periode 2018-2023. Penulis menggunakan data tweet berbahasa Indonesia dengan kata kunci "Rindu", "Hasanah", "Asyik", dan "2DM" dengan jumlah dataset 800 tweet. Hasil penelitian menunjukkan bahwa algoritma NBS memiliki tingkat akurasi yang lebih tinggi daripada SVM, hingga mencapai 94% untuk pasangan kandidat Deddy Mizwar-Dedi Mulyadi. Tujuan dari penelitian ini adalah untuk mengklasifikasikan sentimen opini publik terhadap empat pasangan kandidat Gubernur Jawa Barat agar dapat diprediksi pasangan mana yang memiliki sentimen positif dan lebih mungkin terpilih menjadi Gubernur Jawa Barat periode 2018-2023.	atau negatif bagi masing masing kandidat calon gubernur Model <i>confusion matrix</i> digunakan pada masing masing kata kunci

Sumber: Diolah oleh peneliti

Merujuk dari tabel 2.1 yang berisi daftar penelitian terdahulu, rangkuman, dan *research gap* tersebut, sudah banyak penelitian tentang analisis sentimen opini masyarakat terkait suatu kasus menggunakan metode *data mining* dan hasilnya terbukti akurat. Analisis sentimen tersebut rata-rata diambil dari platform media sosial seperti Twitter, Youtube, Facebook, dan surat kabar. Kasus yang diteliti banyak mencerminkan tentang bencana aktual yang sedang terjadi dan membutuhkan opini pengguna media sosial untuk dinilai sentimennya. Seperti kasus yang sedang ramai sekarang yaitu konflik antara pihak keamanan dan pertahanan dengan masyarakat sipil tentang pro kontra kepada bantuan kemanusiaan untuk pengungsi Rohingya di Indonesia. Berkaca dari penelitian sebelumnya, penulis memilih untuk memberikan rekomendasi pengambilan keputusan bagi pimpinan pertahanan dengan memperhatikan sentimen pada masyarakat sipil terkait dengan konflik Rohingya di Indonesia dengan membandingkan metode SVM dan *Naïve Bayes*.



2.3 Kerangka Berpikir



Gambar 2.3 Diagram Kerangka Berpikir

Sumber: Diolah oleh peneliti

2.3.1 Identifikasi Masalah

Pada gambar 2.3 tahap awal dalam melakukan suatu penelitian ilmiah adalah mengidentifikasi masalah yang akan dikaji. Salah satu isu penting yang perlu dikaji adalah dinamika hubungan sipil militer dalam kasus etnis Rohingya di Myanmar. Selama bertahun-tahun, telah terjadi konflik antara pemerintah Myanmar dan komunitas Rohingya yang mayoritas Muslim. Masalah ini bermula dari adanya diskriminasi, pengucilan, dan penindasan yang dilakukan oleh pihak militer dan aparat keamanan Myanmar terhadap warga Rohingya. Mereka dianggap sebagai imigran gelap dan tidak diakui sebagai warga negara, sehingga mengalami pembatasan akses ke layanan dasar, pendidikan, dan kesehatan. Pada beberapa kesempatan, telah terjadi aksi kekerasan berskala besar yang menyebabkan pembantaian, pemerkosaan, dan pengungsian paksa warga Rohingya. Hal ini menimbulkan pertanyaan mendasar tentang peran dan tanggung jawab militer dalam melindungi warga sipil, serta bagaimana membangun hubungan yang harmonis antara masyarakat dan aparat keamanan. Identifikasi akar permasalahan ini menjadi penting agar dapat ditemukan solusi yang komprehensif dan berkeadilan. Dengan adanya masalah tersebut muncul opini-opini Masyarakat terkait bantuan kemanusiaan oleh TNI kepada pengungsi Rohingya yang berkaitan dengan bagaimana pandangan publik terhadap masalah yang dikaji tersebut.

2.3.2 Observasi Penelitian

Peneliti melakukan observasi masalah dengan melaksanakan studi literatur seperti mencari penelitian terdahulu, latar belakang masalah, dan berita-berita terkait konflik. Setelah dirasa cukup, peneliti melaksanakan diskusi terkait konflik dengan ahli analisis pertahanan negara, Mohammad Nazar, S.Si., M.T.

2.3.3 Klasifikasi

Selanjutnya tahap pengumpulan data menggunakan *library* tweet-harvest dan mencari *tweet* yang mencakup periode waktu tertentu dengan *keyword* yang telah ditentukan. Setelah data yang dibutuhkan terkumpul selanjutnya melakukan tahap *preprocessing* untuk membersihkan dan mempersiapkan data untuk diberi label menjadi data latih dan data uji. Berikutnya data tersebut diuji menggunakan *k-fold cross validation* untuk mengevaluasi validitas model secara objektif. Dengan *k-fold* penulis dapat menghitung *matrix* evaluasi secara konsisten pada tiap *fold* sehingga meningkatkan reliabilitas hasil. Setelah model dinyatakan valid, penulis melakukan klasifikasi menggunakan metode SVM dan NBC yang akan menghasilkan prediksi sentimen dan akurasi metode mana yang lebih baik. Dari hasil tersebut peneliti mengolah *confusion matrix* untuk membandingkan keakuratan, presisi, *recall*, dan *f1-score*.

2.3.4 Interpretasi Hasil

Tahap terakhir yaitu menginterpretasikan hasil *data mining* menggunakan *word cloud* dan menarik kesimpulan berdasarkan temuan yang diperoleh dari algoritma klasifikasi terbaik.

