ABSTRAK

PERBANDINGAN ALGORITMA SUPPORT VECTOR MACHINE DAN NAÏVE BAYES PADA ANALISIS SENTIMEN KEANGGOTAAN INDONESIA DI BOARD OF PEACE MENGGUNAKAN PENDEKATAN SEMMA

Nama : Dendy Pratama

NPM : 2255201198

Pembimbing : Muntahanah, S.Kom., M.Kom.

Perkembangan media sosial telah mendorong munculnya berbagai opini publik terhadap isu politik yang berkembang di ruang digital. Salah satu isu yang memicu beragam respons masyarakat adalah keanggotaan Indonesia di *Board of Peace*. Penelitian ini bertujuan untuk menganalisis dan mengklasifikasikan sentimen publik terhadap isu tersebut menggunakan algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes* dengan menerapkan pendekatan kerangka kerja SEMMA (*Sample, Explore, Modify, Model, Assess*). Data yang digunakan dalam penelitian ini berjumlah 3.962 data unik. Tahapan penelitian meliputi pengumpulan data, pembersihan data (*cleaning*), pelabelan data manual, *case folding*, normalisasi, tokenisasi, *stopword removal, stemming*, pembagian dataset, serta pelatihan dan pengujian model. Evaluasi model dilakukan menggunakan *confusion matrix* serta metrik akurasi untuk mengukur performa kedua algoritma tersebut. Hasil penelitian menunjukkan bahwa model *Support Vector Machine* (SVM) mampu mengklasifikasikan sentimen dengan tingkat akurasi sebesar 74,15%, sedangkan *Naïve Bayes* mencapai 64,44%. Nilai tersebut menunjukkan bahwa model SVM memiliki kemampuan yang lebih baik dalam memahami pola sentimen pada teks berbahasa Indonesia yang bersifat tidak terstruktur dan mengandung bahasa informal. Penelitian ini membuktikan bahwa pendekatan SEMMA efektif digunakan dalam analisis sentimen pada isu politik serta mampu memberikan gambaran mengenai kecenderungan opini publik di ruang digital.

Kata Kunci: Analisis Sentimen, SEMMA, *Support Vector Machine, Naïve Bayes*, Klasifikasi Teks, Board of Peace.

ABSTRACT

Performance Comparison of Support Vector Machine and Naïve Bayes Algorithms in Public Sentiment Analysis Regarding Indonesia's Membership in the Board of Peace Using the SEMMA Approach

Nama: Dendy Pratama

NPM: 2255201198

Pembimbing: Muntahanah, S.Kom., M.Kom.

The development of social media has encouraged the emergence of various public opinions regarding political issues discussed in the digital sphere. One issue that has generated diverse public responses is Indonesia's membership in the Board of Peace. This study aims to analyze and classify public sentiment toward this issue using the Support Vector Machine (SVM) and Naïve Bayes algorithms by applying the SEMMA framework (Sample, Explore, Modify, Model, Assess). The dataset used in this study consisted of 3,962 unique data points. The research stages included data collection, data cleaning, manual data labeling, case folding. normalization, tokenization, stopword removal, stemming, dataset splitting, and model training and testing. Model evaluation was conducted using a confusion matrix and accuracy metrics to measure the performance of the two algorithms. The results showed that the Support Vector Machine (SVM) model achieved an accuracy of 74.15% in classifying sentiment, while the Naïve Bayes model achieved an accuracy of 64.44%. These results indicate that the SVM model performed better in identifying sentiment patterns in unstructured Indonesian-language text containing informal language. This study demonstrates that the SEMMA approach is effective for sentiment analysis of political issues and can provide an overview of public opinion trends in the digital sphere.

Keywords: *Sentiment Analysis, SEMMA, Support Vector Machine, Naïve Bayes, Text Classification, Board of Peace.*

BAB I
PENDAHULUAN

1.1 Latar Belakang

Board of Peace (BoP) merupakan sebuah forum internasional yang diinisiasi dengan tujuan utama mendorong resolusi konflik dan menjaga stabilitas perdamaian global. Kendati memiliki tujuan yang positif, wacana keikutsertaan Indonesia sebagai anggota dalam organisasi tersebut telah memicu perdebatan publik dan polarisasi opini yang masif, terutama di ruang diskusi digital seperti platform X (sebelumnya Twitter). Dinamika pro dan kontra yang berkembang berkaitan erat dengan identitas sosiopolitik Indonesia sebagai negara berpenduduk mayoritas Muslim, yang secara historis maupun konstitusional memiliki komitmen teguh dalam mendukung kemerdekaan Palestina. Dari perspektif publik yang menolak (sentimen negatif), sorotan utama ditujukan pada keterlibatan Israel di dalam struktur kelembagaan BoP, di mana partisipasi Indonesia dikhawatirkan dapat mencederai nilai-nilai solidaritas serta melemahkan posisi Indonesia terhadap perjuangan kemerdekaan rakyat Palestina. Sebaliknya, dari perspektif publik yang mendukung (sentimen positif), keanggotaan ini dipandang sebagai manuver diplomasi yang strategis agar pemerintah memiliki ruang diplomasi yang lebih inklusif dan proaktif (Suryadi & Pratama, 2024).

Memahami sentimen masyarakat apakah mendukung, menolak, atau netral sangat penting sebagai bentuk evaluasi terhadap respons publik atas kebijakan luar negeri. Namun, tingginya volume data opini ini tidak memungkinkan untuk dianalisis secara manual. Di sinilah pesatnya kemajuan teknologi

informasi, khususnya di bidang Kecerdasan Buatan (*Artificial Intelligence*) dan Pemrosesan Bahasa Alami (*Natural Language Processing*), memberikan solusi yang krusial. Perkembangan teknologi tersebut telah mendorong pemanfaatan metode komputasional yang mampu mengolah dan mengklasifikasikan sentimen publik secara otomatis, cepat, dan akurat (Maisyaroh et al., 2025). Analisis sentimen telah banyak dimanfaatkan untuk memahami persepsi masyarakat terhadap berbagai isu publik melalui media sosial karena mampu mengelompokkan opini menjadi sentimen positif, netral, dan negatif secara sistematis (Khairi Basri dkk., 2025).

Sebagai solusi untuk memecahkan masalah klasifikasi sentimen teks tersebut, penelitian ini menerapkan algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes*. Kedua algoritma tersebut merupakan metode klasifikasi yang banyak digunakan pada penelitian analisis sentimen berbasis TF-IDF karena memiliki kemampuan yang baik dalam mengolah data teks serta dievaluasi menggunakan metrik *accuracy, precision, recall, F1-score*, dan *confusion matrix* (Tito Dian Permana dkk., 2025). Pemilihan kedua algoritma ini didasarkan pada karakteristiknya yang sangat relevan untuk pemrosesan bahasa alami. SVM dipilih karena kemampuannya yang krusial dalam menangani data non-linear melalui *hyperplane* optimal di ruang fitur berdimensi tinggi, sehingga sangat unggul dalam menangkap pola kata yang kompleks pada teks berbahasa Indonesia. Sebaliknya, *Naïve Bayes* dipilih sebagai algoritma pembanding (baseline) karena efisiensinya yang tinggi

dalam klasifikasi probabilistik pada dataset besar dengan asumsi independensi fitur (Lestari & Hutagalung, 2025).

Perbandingan penelitian terdahulu pada domain ulasan produk membuktikan bahwa SVM seringkali mengungguli *Naïve Bayes* secara signifikan (Rahmadzani dkk., 2025; Rizki dkk., 2025; Yasir Alghifari dan M. Rudi., 2025), di mana keunggulan kinerjanya pada konteks teks opini dan kebijakan menjadi landasan utama pemilihan kedua model tersebut.

Untuk memastikan tahapan pemodelan berjalan sistematis dan komprehensif, penelitian ini menggunakan pendekatan SEMMA (Sample, Explore, Modify, Model, Assess). Alasan utama pemilihan SEMMA dibandingkan kerangka kerja lain (seperti CRISP-DM) adalah karena pendekatannya yang berfokus langsung pada siklus teknis data mining yang iteratif. Tahapan ini sangat ideal untuk analisis sentimen karena memungkinkan efisiensi pengolahan data teks yang masif, mulai dari pengambilan sampel yang representatif, transformasi fitur teks (Modify), hingga pembangunan dan evaluasi keakuratan model SVM dan *Naïve Bayes* (Model dan Assess) secara terstruktur dan cepat (Agustin & Nabil Nur Afrizal, 2025).

Perbandingan penelitian terdahulu menunjukkan SVM secara konsisten mengungguli *Naïve Bayes* dalam analisis sentimen teks Indonesia. Penelitian Tito Dian Permana dkk. (2025) menunjukkan bahwa algoritma SVM menghasilkan akurasi 73,97%, sedikit lebih tinggi dibandingkan *Naïve Bayes* sebesar 73,85%, serta memiliki performa yang lebih konsisten dalam

mengklasifikasikan sentimen negatif dan netral. Demikian pula, pada review Female Daily skincare, SVM raih 87% akurasi versus 80% Naïve Bayes, berkat kemampuan SVM menangani pola rumit (Rahmadzani dkk., 2025). Studi TikTok Shop juga mengonfirmasi superioritas SVM (68,86%) atas Naïve Bayes (64,48%) pada dataset seimbang (Rizki A dkk., 2025).

Berdasarkan uraian permasalahan di atas, penelitian ini bertujuan untuk mengklasifikasikan sentimen publik dan membandingkan kinerja algoritma *Support Vector Machine* (SVM) serta Naïve Bayes terhadap wacana keanggotaan Indonesia di Board of Peace. Evaluasi performa model diukur menggunakan *metrik accuracy, precision, recall, dan F1-score* untuk menemukan model klasifikasi yang paling optimal dalam menangani data teks berbahasa Indonesia. Hasil penelitian ini diharapkan dapat memberikan kontribusi teknis mengenai algoritma machine learning terbaik untuk klasifikasi teks analitik, sekaligus menyajikan informasi berbasis data terkait kecenderungan opini masyarakat terhadap isu tersebut.

1.2 Pertanyaan Penelitian

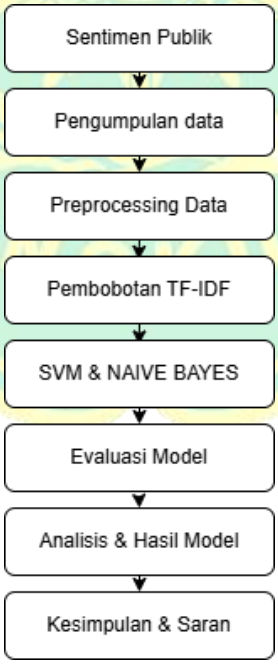
1. Bagaimana efektivitas dan kinerja komparasi algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes* dalam mengklasifikasikan sentimen publik terhadap keanggotaan Indonesia di Board of Peace menggunakan pendekatan SEMMA?
2. Bagaimana hasil evaluasi komparatif performa kedua model tersebut berdasarkan metrik *accuracy, precision, recall, dan F1-score*?

3. Bagaimana pola kecenderungan opini publik (positif, netral, dan negatif) yang dihasilkan dari klasifikasi sentimen tersebut?

1.3 Tujuan Penelitian

Tujuan penelitian ini adalah untuk mengidentifikasi dan mengklasifikasikan opini masyarakat menjadi sentimen positif, netral, dan negatif terkait keanggotaan Indonesia di *Board of Peace* dengan menerapkan komparasi algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes* menggunakan pendekatan SEMMA, guna mengetahui kecenderungan dominasi opini publik yang berkembang terhadap isu diplomasi tersebut, serta mengevaluasi kinerja model terbaik secara komprehensif menggunakan metrik akurasi, *precision, recall,* dan *F1-score* dalam mengolah data teks berbahasa Indonesia yang bersifat tidak terstruktur.

1.4 Kerangka Kerja Penelitian *(Research Framework)*



Gambar 1.1 Framework Penelitian