Seminar Nasional Informatika (SENATIKA)
Prosiding Senatika 2025 **QRSBN:62-0151-02942-8**

Analisis Sentimen Publik Pada Platform X Menggunakan Algoritma Naïve Bayes: Isu Ijazah Presiden Jokowi

Wilson Christhoper[1], Deny Jollyta[2], Manuel Alexandro[3]
[1,2,3]Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Institut Bisnis dan Teknologi Pelita Indonesia
e-mail: [1]wilson.christhoper@student.pelitaindonesia.ac.id,
[2]deny.jollyta@lecturer.pelitaindonesia.ac.id, [3]manuelalexandro8@gmail.com

Abstrak

Perkembangan media sosial telah membuka ruang luas bagi masyarakat untuk mengekspresikan pendapat terhadap berbagai isu politik di Indonesia. Salah satu topik yang menimbulkan perhatian publik adalah dugaan ijazah palsu Presiden Joko Widodo (Jokowi) yang menimbulkan beragam tanggapan di aplikasi X. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap isu keaslian ijazah tersebut menggunakan algoritma Naive Bayes Classifier. Sebanyak 1.004 tweet berbahasa Indonesia dikumpulkan melalui proses scraping dan dibagi menjadi 80% data latih serta 20% data uji. Tahapan analisis meliputi preprocessing data dan klasifikasi menggunakan algoritma Naive Bayes. Hasil pengujian menunjukkan akurasi sebesar 76,6%, precision 56,8%, recall 40,4%, dan F1-score 47,2%. Berdasarkan confusion matrix, model lebih baik dalam mengidentifikasi sentimen negatif, sehingga algoritma ini cukup efektif untuk analisis sentimen politik pada platform X.

Kata kunci: *Analisis Sentimen, Naive Bayes, Platform X, Opini Publik, Isu Politik*

Abstract

The rapid development of social media has created a broad platform for the public to voice opinions on various political issues in Indonesia. One of the most prominent discussions is the alleged fake diploma of President Joko Widodo (Jokowi), which has sparked diverse public reactions on X application. This study aims to examine public sentiment toward the issue using the Naive Bayes Classifier algorithm. A total of 1,004 Indonesian-language tweets were collected through web scraping and divided into 80% training data and 20% testing data. The analysis involved several stages, including data preprocessing and sentiment classification using Naive Bayes. The results indicate that the model achieved an accuracy of 76.6%, precision of 56.8%, recall of 40.4%, and an F1-score of 47.2%. Based on the confusion matrix, the model performs better in classifying negative sentiments, suggesting that Naive Bayes is effective for political sentiment analysis on social media.

Keywords: *Sentiment Analysis, Naive Bayes, X Platform, Public Opinion, Political Issue*

---

1. Pendahuluan

Di era digital saat ini, media sosial tidak lagi sekadar digunakan untuk berbagi aktivitas atau informasi pribadi, melainkan telah berevolusi menjadi sarana ekspresi politik sekaligus ruang diskusi publik yang aktif dan dinamis. Beragam *platform* seperti aplikasi X, Facebook, dan Instagram, kini berperan penting sebagai media bagi masyarakat untuk menyampaikan pandangan terhadap isu-isu politik, kebijakan pemerintah, maupun figur publik. Salah satu isu yang ramai diperbincangkan dalam beberapa tahun terakhir adalah isu ijazah Presiden Joko Widodo (Jokowi). Isu ini memicu gelombang opini pro dan kontra di berbagai *platform* sosial, memunculkan banyak persepsi negatif, spekulasi, bahkan hoaks yang menyebar dengan cepat. Oleh karena itu, analisis

terhadap pola sentimen masyarakat mengenai isu tersebut menjadi penting untuk memahami bagaimana opini publik terbentuk dan menyebar di dunia maya.

Analisis sentimen merupakan bagian dari cabang ilmu *Natural Language Processing* (NLP) yang berfokus pada pengenalan serta interpretasi emosi, opini, dan sikap seseorang yang tertuang dalam sebuah teks [1]. Tujuan utamanya adalah mengklasifikasikan suatu teks ke dalam kategori sentimen positif, negatif, atau netral. Hasil klasifikasi ini dapat dimanfaatkan sebagai dasar dalam proses pengambilan keputusan strategis di berbagai bidang, seperti politik, bisnis, maupun kebijakan publik [2]. Dalam ranah politik, analisis sentimen berperan penting sebagai sarana untuk menilai pandangan masyarakat terhadap figur politik maupun kebijakan pemerintah [3].

Salah satu algoritma yang paling umum digunakan dalam analisis sentimen adalah *Naive Bayes* (NB). Algoritma ini termasuk dalam pendekatan *supervised learning* dan bekerja dengan menerapkan prinsip teorema *Bayes*, disertai asumsi bahwa setiap fitur bersifat saling independen [4]. Penelitian yang saat ini dilakukan mempunyai tujuan untuk mengimplementasikan algoritma NB dalam menganalisis sentimen mengenai topik ijazah Jokowi dengan menggunakan data dari aplikasi X. Keunggulan utama NB terletak pada kesederhanaan konsepnya, kecepatan proses pelatihan model, serta kemampuannya dalam memberikan hasil yang baik meskipun diterapkan pada data teks berukuran besar dan memiliki variasi tinggi [5]. Selain itu, algoritma ini dikenal efisien dalam menangani klasifikasi teks berbahasa alami, termasuk bahasa Indonesia yang memiliki struktur morfologis kompleks.

Beberapa penelitian telah membuktikan efektivitas algoritma NB dalam konteks analisis sentimen politik di Indonesia. Gunawan (2024) melakukan penelitian mengenai opini masyarakat terhadap kandidat presiden 2024 di media sosial X menggunakan algoritma NB dan memperoleh akurasi sebesar 78%, dengan hasil menunjukkan bahwa mayoritas opini publik memiliki kecenderungan positif terhadap salah satu kandidat [6]. Sementara itu, Riyantoro dan Fauziah (2025) membandingkan performa algoritma NB dengan Support Vector Machine (SVM) dalam analisis opini publik terhadap kabinet Indonesia. Hasil penelitian menunjukkan bahwa NB memiliki keunggulan dari sisi efisiensi komputasi dan waktu pelatihan data, meskipun SVM memberikan hasil akurasi yang sedikit lebih tinggi [7].

Penelitian lain oleh Buntoro, Arifin, dan Syaifuddiin (2021) yang meneliti opini publik terhadap Pemilihan Presiden Indonesia tahun 2019, menemukan bahwa algoritma NB mampu mengklasifikasikan data sentimen dengan akurasi mencapai 83% dan menunjukkan bahwa algoritma ini dapat mengidentifikasi kecenderungan opini publik terhadap masing-masing calon presiden [8]. Selain itu, Zulfikar dan Atmadja (2023) juga menerapkan *Multinomial Naive Bayes* untuk menganalisis reaksi publik terhadap kebijakan pemerintah di media sosial, dan hasilnya menunjukkan efektivitas algoritma ini dalam memisahkan opini positif dan negatif secara signifikan, terutama ketika diterapkan pada data berbahasa Indonesia [9].

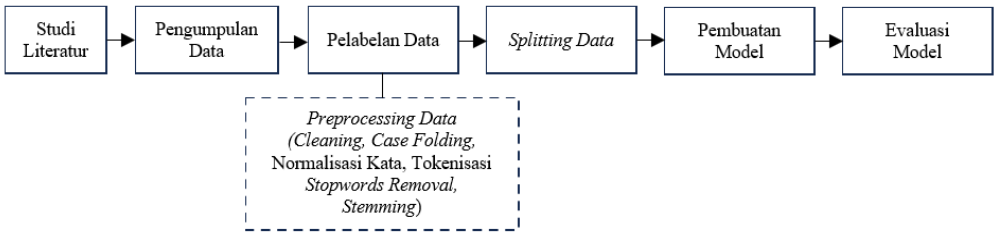
Namun demikian, penelitian yang secara khusus menyoroti isu sensitif seperti dugaan ijazah palsu Presiden Jokowi masih sangat terbatas. Padahal, isu ini menjadi salah satu fenomena yang menggambarkan polarisasi politik dan persepsi publik terhadap keabsahan figur pemimpin nasional. Dalam konteks ini, penerapan algoritma NB untuk menganalisis data sentimen dari media sosial diharapkan dapat memberikan gambaran yang objektif mengenai persepsi publik terhadap isu tersebut, sekaligus berkontribusi pada upaya literasi digital dan mitigasi penyebaran hoaks. Hasil dari penelitian ini diharapkan dapat menjadi acuan dalam memahami pola penyebaran opini publik, sekaligus memberikan kontribusi terhadap pengembangan kajian analisis sentimen politik di Indonesia menggunakan metode pembelajaran mesin.

2. Metode Penelitian

2.1. Metode Penelitian

Untuk mencapai tujuan penelitian yang diharapkan, telah disusun sejumlah tahapan penelitian yang sistematis seperti ditampilkan Gambar 1. Pada penelitian yang sedang dilakukan, tahapannya dibagi ke dalam tujuh kegiatan yakni studi literatur, pengumpulan data, *preprocesing data*, pelabelan data, *splitting data*, pelatihan dan pengujian model, serta evaluasi model.

Penjelasan setiap tahapan, disampaikan pada sub bab yang merupakan bagian dari sub Metode Penelitian.



Gambar 1. Tahapan Penelitian

2.2. Studi Literatur dan Pengumpulan Data

Tahap pertama dilakukan dengan mengumpulkan berbagai referensi dari jurnal, buku, maupun penelitian terdahulu yang berkaitan dengan analisis sentimen, pemrosesan bahasa alami (NLP), serta penerapan algoritma *Naïve Bayes* [10]. Selanjutnya dilakukan pengumpulan data memanfaatkan *harvest tweet* yang dapat dijalankan pada *google colab*, berdasarkan kata kunci seperti "ijazah jokowi", "ijazah presiden", dan variasi ejaan atau tagar terkait. Proses pengambilan data dilakukan menggunakan bahasa pemrograman Python dengan bantuan *Twitter Crawler*. Data yang dikumpulkan berupa komentar *tweet* berbahasa Indonesia sebanyak 1004, yang diambil dalam rentang waktu 30 Juli 2025 hingga 31 Juli 2025. Selanjutnya data disaring agar didapatkan *tweet* berbahasa Indonesia (lang:id) dan berada pada rentang waktu *tweet* yang diambil, dan bebas duplikasi berdasarkan *conversation_id*. Untuk *username* sendiri tidak dapat dilakukan *scraping* karena kebijakan dari *twitter* untuk melindungi privasi penggunanya. Setelah selesai dilakukan pengumpulan data maka data akan disimpan dalam bentuk file csv.

2.3. Preprocessing Data

Proses ini bertujuan untuk membersihkan serta menstandarkan data sehingga siap digunakan [11]. Adapun tahapan preprocessing yang dilakukan meliputi beberapa langkah yakni *cleaning, case folding*, normalisasi kata, tokenisasi, *stopword removal*, dan *stemming*. Pada Tabel 1 dijelaskan *preprocessing* dengan contoh yang diambil dari data *tweet*, yakni "@ethadisaputra @CakKhum Kasus ijazah palsu emangnye udah disidang pengadilan? Ape vonisnye? Mau kapan abolisi dikasihnye?". Maka hasil preprocessing ditampilkan pada Tabel 1.

Tabel 1. Tahapan *Preprocessing*

Preprocessing	Hasil
***Cleaning*:** menghapus elemen-elemen yang tidak relevan seperti URL, *mention, retweet* (RT), tanda baca, angka yang tidak memiliki makna penting, emoji, serta duplikasi data	Kasus ijazah palsu emangnye udah disidang pengadilan Ape vonisnye Mau kapan abolisi dikasihnye
***Case Folding*:** mengubah semua huruf pada teks diubah menjadi huruf kecil untuk menjaga konsistensi penulisan dan mempermudah proses analisis	kasus ijazah palsu emangnye udah disidang pengadilan ape vonisnye mau kapan abolisi dikasihnye
Normalisasi Kata: mengganti kata tidak baku, singkatan, atau istilah tidak standar diganti ke dalam bentuk kata baku sesuai dengan kamus besar bahasa Indonesia	kasus ijazah palsu emangnya sudah disidang pengadilan apa vonisnya mau kapan abolisi dikasihnya
Tokenisasi: pemisahan teks menjadi satuan kata (token) agar dapat dianalisis secara individual	['kasus', 'ijazah', 'palsu', 'emangnya', 'sudah', 'disidang', 'pengadilan', 'apa', 'vonisnya', 'mau', 'kapan', 'abolisi', 'dikasihnya']

***Stopword Removal*:** menghapus kata-kata umum seperti "yang", "dan", atau "di" yang tidak memberikan kontribusi bermakna terhadap konteks sentimen	ijazah palsu emangnya disidang pengadilan vonisnya abolisi dikasihnya
***Stemming*:** mengubah setiap kata ke bentuk dasarnya menggunakan algoritma *Sastrawi Stemmer* untuk mengurangi variasi bentuk kata yang memiliki makna serupa	ijazah palsu emangnya sidang adil vonisnya abolisi dikasihnya

2.4. Pelabelan dan *Splitting* Data

Data yang telah dilakukan melalui tahap *preprocessing* selanjutnya diberi label sentimen yang dikategorikan ke dalam dua kelas, yaitu positif dan negatif. Pelabelan ini dilakukan dengan kamus leksikon yang dijalankan pada Python, seperti yang diperlihatkan Gambar 2.

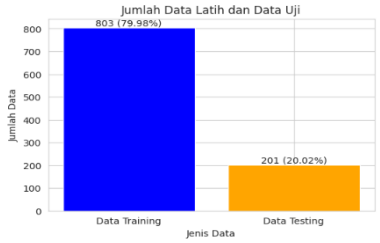
```
def determine_sentiment(text):
    if isinstance(text, str):
        positive_count = sum(1 for word in text.split() if word in positive_lexicon)
        negative_count = sum(1 for word in text.split() if word in negative_lexicon)
        sentiment_score = positive_count - negative_count
        sentiment = "Negatif" if sentiment_score <= 0 else "Positif"
        return sentiment_score, sentiment
    return 0, "Negatif"

data[['Score', 'Sentiment']] = data['stopword removal'].apply(lambda x: pd.Series(determine_sentiment(x)))
data.head(5)
```

	stopword removal	Score	Sentiment
0	tanda kuat deal bu mega amp prabowo bersatu pe...	-5	Negatif
1	ya iyalaaahhhijazah palsu cari pakai akal apap...	0	Negatif
2	wkwkwksdhlah jkw termul alihkan bayar ijazah a...	1	Positif
3	orang trtuduh kau lempar isu ijazah palsu back...	-2	Negatif
4	ijazah palsu emangnye disidang pengadilan voni...	-1	Negatif

Gambar 2. Proses dan Hasil Pelabelan Data

Proses pembagian data (*split data*) dilakukan terhadap total 1.004 data yang telah dikumpulkan. Dataset tersebut dibagi menjadi dua bagian, yaitu data latih sebesar 80% dan data uji sebesar 20%, dengan tujuan agar proses pelatihan dan pengujian model dapat berlangsung secara seimbang, representatif, serta mampu menggambarkan kinerja model secara objektif [12]. Gambar 3 merupakan bentuk pemisahan data yang dimaksud.



Gambar 3. Hasil Splitting Data

2.4. Algoritma *Naïve Bayes* (NB)

Algoritma NB bekerja dengan menghitung nilai probabilitas menggunakan teorema Bayes, kemudian mengombinasikannya dengan frekuensi kemunculan data yang ada [13] [14]. Dalam konteks tugas klasifikasi, tujuan utama dari metode ini adalah memprediksi label kelas suatu data atau sampel berdasarkan sekumpulan fitur atau karakteristik yang dimilikinya [15]. Pada penelitian ini, algoritma NB yang digunakan adalah *Multinomial* karena data yang masuk dalam kategori dokumen [16]. Berikut adalah rumus NB yang digunakan:

$$P(H \mid X) = \frac{P(X|H)\,P(H)}{P(X)} \qquad (1)$$

Keterangan :
X : Data dengan kelas yang belum diketahui
H : Hipotesis data merupakan suatu kelas spesifik
P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)
P(H) : Probabilitas hipotesis H (prior probabilitas)
P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H
P(X) : Probabilitas X

2.5. Evaluasi Model

Pada tahap ini, model yang dihasilkan algoritma NB dievaluasi menggunakan *Confusion Matrix*. *Confusion Matrix* adalah alat analisis prediktif yang menampilkan dan membandingkan nilai aktual atau benar dengan hasil prediksi model yang dapat digunakan untuk menghasilkan metrik evaluasi seperti Akurasi, Presisi, Recall, dan Skor F1 atau Pengukuran F [17]. Rumus yang digunakan sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1\text{-}Score = \frac{2TP}{2TP+FP+FN} \quad (5)$$

3. Hasil dan Pembahasan

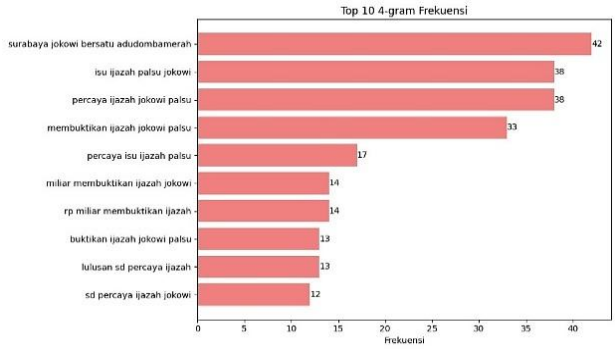
3.1. Implementasi Naïve Bayes dan Visualisasi

Implementasi dilakukan menggunakan algoritma NB yang mengacu pada rumus (1) dengan *source code* seperti pada Gambar 4. Model dibentuk dengan mengolah data *tweet* pelatihan menggunakan algoritma NB Multinomial yang umum digunakan untuk mengklasifikasikan kategori dokumen berdasarkan frekuensi kata-kata yang muncul di dalam dokumen tersebut.

```
# Initialize models
models = {
    "Naive Bayes": MultinomialNB(),
}

# Train models
results = {}
for model_name, model in models.items():
    model.fit(X_train_vec, y_train)
    y_pred = model.predict(X_test_vec)
    results[model_name] = {
        "accuracy": accuracy_score(y_test, y_pred),
        "classification_report": classification_report(y_test, y_pred, output_dict=True),
        "confusion_matrix": confusion_matrix(y_test, y_pred)
    }
```

Gambar 4. Implementasi Algoritma NB



Gambar 5. Visualisasi Model NB

Hasil pembentukan model oleh NB ditampilkan dalam bentuk grafik n-gram. Gambar 5 merupakan interpretasi model yang memperlihatkan bahwa kumpulan kata yang paling sering muncul dalam data teks berkaitan erat dengan isu "ijazah palsu Jokowi". Hampir seluruh n-gram yang menempati sepuluh besar mengandung kata "ijazah", "palsu", dan "Jokowi", yang menunjukkan bahwa topik utama pembahasan dalam korpus teks sangat terfokus pada isu

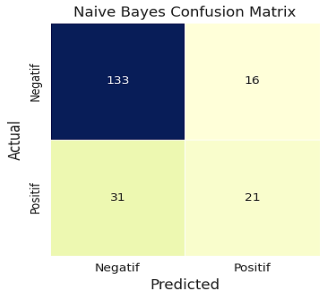
keaslian ijazah Presiden Jokowi. Beberapa frasa seperti “isu ijazah palsu jokowi”, “percaya ijazah jokowi palsu”, dan “membuktikan ijazah jokowi palsu” menggambarkan adanya pola narasi yang berulang seputar klaim dan opini publik mengenai keaslian ijazah tersebut. Sementara itu, frasa seperti “surabaya jokowi bersatu adudombamerah” tampaknya berasal dari nama akun atau *tag* tertentu, yang kemungkinan tidak berkaitan langsung dengan isi diskusi dan dapat dikategorikan sebagai *noise* dalam analisis teks. Secara keseluruhan, hasil analisis n-gram ini menunjukkan bahwa percakapan di media sosial atau teks yang dianalisis didominasi oleh wacana mengenai isu ijazah Jokowi.

Selanjutnya model klasifikasi juga ditampilkan dalam bentuk visualisasi *wordcloud*, seperti yang dapat dilihat pada Gambar 6. Gambar 6 menampilkan visualisasi frekuensi kemunculan kata dalam kumpulan data *tweet* terkait isu ijazah Presiden Jokowi. Ukuran kata yang lebih besar menunjukkan tingkat kemunculan yang lebih tinggi dalam korpus data. Berdasarkan hasil tersebut, terlihat bahwa kata “ijazah”, “palsu”, dan “Jokowi” mendominasi tampilan *wordcloud.*

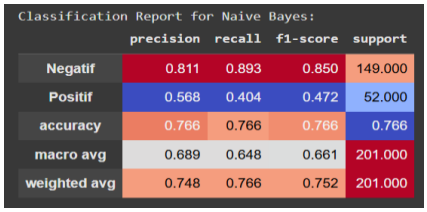
Gambar 6. Hasil *Wordcloud*

3.2. Evaluasi Model

Langkah selanjutnya adalah pengujian model. Proses perhitungan *Confusion Matrix* yang mengacu pada rumus (2), (3), (4) dan (5), dilakukan menggunakan *library scikit-learn*, yaitu pustaka Python yang juga digunakan pada tahap klasifikasi dengan algoritma NB. Hasil dari proses pelatihan (training model) kemudian diuji menggunakan data testing, yang menghasilkan matriks berukuran 2x2 sebagai representasi dari kelas aktual dan kelas prediksi, sehingga dapat menggambarkan performa model dalam mengklasifikasikan data secara akurat. Pengujian menghasilkan data klasifikasi yang menunjukkan jumlah sentimen positif dan negatif berdasarkan hasil evaluasi.



Gambar 7. *Confusion Matrix* NB



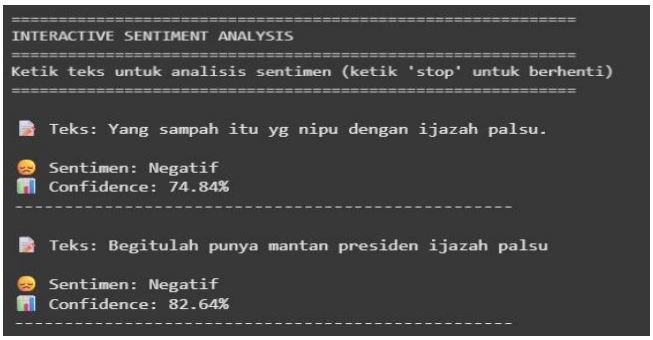
Gambar 8. *Clasification Report*

Gambar 7 memperlihatkan sebanyak 133 data dengan sentimen negatif berhasil diprediksi secara benar sebagai negatif *(True Negative)*, yang menunjukkan bahwa model mampu mengenali sebagian besar data berlabel negatif dengan akurat. Sementara itu, terdapat 16 data negatif yang salah terklasifikasi sebagai positif *(False Positive)*, menandakan adanya kesalahan prediksi di mana model mengidentifikasi sentimen positif pada teks yang seharusnya bersifat negatif. Secara keseluruhan, dari total 201 data uji, algoritma NB menunjukkan kinerja yang cukup baik, khususnya pada kelas negatif yang memiliki tingkat ketepatan prediksi lebih tinggi dibandingkan kelas positif.

Selanjutnya performa model juga didasarkan pada hasil *classification report* pada Gambar 8 dari pengujian model NB, untuk kelas negatif, model memperoleh nilai *precision* sebesar 0.811, *recall* sebesar 0.893, dan *F1-score* sebesar 0.850 dari total 149 data uji. Nilai tersebut mengindikasikan bahwa model mampu mengenali sentimen negatif dengan tingkat ketepatan yang tinggi, dimana sebagian besar data berlabel negatif berhasil diidentifikasi dan diprediksi secara akurat. Untuk kelas positif diperoleh tingkat akurasi sebesar 76.6%, dengan nilai *precision* untuk kelas positif sebesar 56.8%, *recall* sebesar 40.4%, dan *F1-score* sebesar 47.2% dari total 52 data uji. Hasil ini menunjukkan bahwa model masih menghadapi tantangan dalam mendeteksi data dengan sentimen positif. Beberapa kemungkinan dapat menyebabkan hal ini, seperti jumlah data, keseimbangan data, dan pembagian data.

3.3. *Deployment*

Bentuk *deploy* dari model penelitian ini terdapat pada Gambar 9. Berdasarkan *deploy* yang dibuat, pengguna dapat memasukan komentar terkait ijazah Jokowi dengan gaya bahasa yang diinginkan. Pada contoh pertama yang ditampilkan pada Gambar 9, model mengklasifkasi komentar ke dalam kelas negatif dengan keyakinan ketepatan mencapai 74.84%. Untuk komentar kedua, model memberikan keyakinan sebesar 82.64%. Jika dibandingkan dengan akurasi model pada kelas positif sebesar 76.6%, akurasi komentar baru yang diujikan ke model mendekati bahkan melampaui. Hal ini memperlihatkan bahwa model klasifikasi NB telah baik dalam membaca komentar terkait ijazah Jokowi.



Gambar 9. *Deployment* Model NB

4. Kesimpulan

Secara keseluruhan, dapat disimpulkan bahwa algoritma NB merupakan algoritma yang efektif untuk digunakan dalam analisis sentimen isu politik di media sosial, karena mampu memberikan akurasi yang stabil dengan komputasi yang efisien. Penelitian ini membuktikan bahwa akurasi komentar saat diimplementasikan dalam beberapa kali percobaan, mampu melewati akurasi model. Artinya model NB mampu mengklasifikasi komentar dengan baik. Meski demikian, peningkatan performa model masih dimungkinkan melalui beberapa langkah, seperti penyeimbangan distribusi data *(data balancing)*, optimasi fitur teks, serta eksperimen lanjutan dengan algoritma pembanding seperti SVM, Random Forest, atau model berbasis *deep learning* seperti Long Short-Term Memory (LSTM) pada penelitian berikutnya.

Daftar Pustaka

[1] D. Nurcahyono, W. P. Putra, and A. Najib, “Analysis sentiment in social media against election using the method naive Bayes,” in *Journal of Physics: Conference Series*, 2020, p. 012003. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1511/1/012003

[2] D. Sugiarto, E. Utami, and A. Yaqin, “Perbandingan Kinerja Model TF-IDF dan BOW untuk Klasifikasi Opini Publik Tentang Kebijakan BLT Minyak Goreng,” 2022.

[3] T. Santoso and A. A. Yana, "Sentiment analysis of Facebook comments on Indonesian presidential candidates using the Naïve Bayes method," in *Journal of Physics: Conference Series*, 2020, p. 012012. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1641/1/012012

[4] A. Mustolih, P. Arsi, and P. Subarkah, "Sentiment Analysis Motorku X Using Applications Naive Bayes Classifier Method," *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 6, no. 2, p. 231, Aug. 2023, doi: 10.24014/ijaidm.v6i2.24864.

[5] C. Setianingsih and I. Ramadhan, "Deciphering Political Discourse on X: A Deep Dive into SVM vs. Naïve Bayes Approaches," in *IEEE International Conference on Data Science and Engineering*, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10475056/

[6] Y. Gunawan, "Sentiment Analysis of Twitter Towards the 2024 Indonesian Presidential Candidates Using the Naive Bayes Algorithms," 2024. [Online]. Available: http://repository.ulb.ac.id/id/eprint/944

[7] R. Riyantoro and F. Fauziah, "Sentiment Analysis of Twitter Data on Indonesia's Cabinet Using Naïve Bayes and Support Vector Machine Algorithms," *Tekno Journal*, vol. 30, no. 2, 2025, [Online]. Available: https://ejournal.gunadarma.ac.id/index.php/tekno/article/view/13953

[8] G. A. Buntoro, R. Arifin, and G. N. Syaifuddiin, "The Implementation of the Machine Learning Algorithm for the Sentiment Analysis of Indonesia's 2019 Presidential Election," *IIUM Engineering Journal*, vol. 22, no. 1, pp. 45–53, 2021, [Online]. Available: https://journals.iium.edu.my/ejournal/index.php/iiumej/article/view/1532

[9] W. B. Zulfikar and A. R. Atmadja, "Sentiment analysis on social media against public policy using multinomial naive bayes," *Scientific Journal of Informatics*, 2023, [Online]. Available: http://shura.shu.ac.uk/31645

[10] M. N. F. Putri, H. W. Herwanto, and E. R. Wardhani, "Tinjauan Literatur Analisis Sentimen Produk E- Commerce: Dataset, Pendekatan, Metode, Dan Performa," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 3, pp. 2282–2289, Aug. 2025, doi: 10.29100/jipi.v10i3.8026.

[11] M. Siino, I. Tinnirello, and M. La Cascia, "Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers," *Inf Syst*, vol. 121, Mar. 2024, doi: 10.1016/j.is.2023.102342.

[12] V. R. Joseph, "Optimal ratio for data splitting," *Stat Anal Data Min*, vol. 15, no. 4, pp. 531–538, Aug. 2022, doi: 10.1002/sam.11583.

[13] S. N. Sofyan and M. Iqbal, "Analisis Sentimen Terhadap Dampak Inflasi Menggunakan Naive Bayes," *Bulletin of Information Technology (BIT)*, vol. 6, no. 1, pp. 1–8, 2025, doi: 10.47065/bit.v5i2.1796.

[14] A. H. Tandiano and D. Jollyta, "Classification Of Fake News In Indonesian Language Using Support Vector Machine Method," *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, vol. 10, no. 2, pp. 315–322, Mar. 2024, doi: 10.33330/jurteksi.v10i2.2895.

[15] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *Jurnal KomtekInfo*, pp. 1–7, Jan. 2023, doi: 10.35134/komtekinfo.v10i1.330.

[16] R. Rinaldi and R. Goejantoro, "Penerapan Metode Klasifikasi Multinomial Naive Bayes (Studi Kasus: PT Prudential Life Samarinda Tahun 2019) Application of Naive Bayes Multinomial Classification Method (Case Study: PT Prudential Life Samarinda in 2019)."

[17] J. H. Joloudari, A. Marefat, M. A. Nematollahi, S. S. Oyelere, and S. Hussain, "Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks," *Applied Sciences (Switzerland)*, vol. 13, no. 6, Mar. 2023, doi: 10.3390/app13064006.