

Pengujian Akurasi *Naïve Bayes* untuk Klasifikasi Sentimen Isu Keaslian Ijazah menggunakan *Orange*

Accuracy Evaluation of the Naïve Bayes Classifier for Sentiment Classification of Diploma Authenticity Issues using Orange

¹Eunike Loise Laapen, ²Indrastanti R. Widiasari

^{1,2}Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya
Wacana

^{1,2}Jl. Dr. O. Notohamidjojo, Blotongan, Salatiga, Jawa Tengah 50711, Indonesia

*e-mail: 672022190@student.uksw.edu, indrastanti@uksw.edu

(received: 12 June 2026, revised: 16 June 2026, accepted: 20 June 2026)

Abstrak

Pertumbuhan pesat data teks digital telah meningkatkan kebutuhan akan metode yang efektif untuk mengekstraksi informasi yang bermakna, khususnya dalam memahami opini publik. Analisis sentimen banyak digunakan untuk mengklasifikasikan opini ke dalam kategori positif, negatif, dan netral. Namun, tantangan seperti ambiguitas bahasa, subjektivitas, dan *imbalanced dataset* seringkali menurunkan performa klasifikasi. Penelitian ini bertujuan untuk mengevaluasi kinerja algoritma *Naïve Bayes* dalam klasifikasi sentimen pada isu keaslian ijazah menggunakan *dataset* publik, serta menganalisis pengaruh distribusi data terhadap performa model. Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen pada 1.014 data asli dan 1,767 data hasil *oversampling*, yang diproses melalui tahap *preprocessing*, transformasi *Bag of Words*, serta klasifikasi menggunakan *Orange Data Mining* dengan *10-fold cross validation*. Hasil penelitian menunjukkan bahwa model menghasilkan akurasi sebesar 37,2% dengan nilai AUC sebesar 0,704 pada *dataset* asli, yang menunjukkan performa klasifikasi masih rendah. Setelah dilakukan *oversampling* untuk menyeimbangkan data, akurasi meningkat signifikan menjadi 82,1% dengan nilai AUC sebesar 0,970. Hasil ini menunjukkan bahwa distribusi data memiliki pengaruh yang signifikan terhadap performa algoritma *Naïve Bayes* dalam klasifikasi sentimen.

Kata kunci: ketidakseimbangan data, *naïve bayes*, *oversampling*, analisis sentimen, *text mining*

Abstract

The rapid growth of digital text data has increased the demand for effective methods to extract meaningful information, particularly for understanding public opinion. Sentiment analysis is widely used to classify opinions into positive, negative, and neutral categories. However, challenges such as linguistic ambiguity, subjectivity, and class imbalance often degrade classification performance. This study aims to evaluate the performance of the *Naïve Bayes* algorithm for sentiment classification on the issue of diploma authenticity using a publicly available dataset, while examining the impact of data distribution on model performance. A quantitative experimental approach was employed using an original dataset of 1,014 instances and an oversampled dataset of 1,767 instances. The data were processed through preprocessing, Bag-of-Words feature extraction, and sentiment classification using *Orange Data Mining* with *10-fold cross-validation*. Model performance was evaluated using accuracy and the Area Under the Receiver Operating Characteristic Curve (AUC). The results indicate that the *Naïve Bayes* model achieved an accuracy of 37.2% and an AUC of 0.704 on the original imbalanced dataset, reflecting relatively poor classification performance. After applying oversampling to balance the class distribution, the model's accuracy increased substantially to 82.1%, while the AUC improved to 0.970. These findings demonstrate that class distribution has a significant impact on the performance of the *Naïve Bayes* algorithm in sentiment classification and highlight the importance of addressing class imbalance to achieve more reliable classification results.

Keywords: *imbalanced dataset*, *naïve bayes*, *oversampling*, sentiment analysis, *text mining*

1 Pendahuluan

Perkembangan teknologi informasi dan komunikasi di era digital telah mendorong pertumbuhan data teks dalam jumlah yang sangat besar, terutama dari berbagai platform digital dan sumber data terbuka. Fenomena ini menjadikan data teks sebagai sumber informasi penting dalam memahami opini publik terhadap isu-isu yang berkembang secara cepat dan luas. Dalam beberapa waktu terakhir, isu keaslian ijazah menjadi salah satu topik yang banyak diperbincangkan di ruang publik digital dan memicu beragam opini masyarakat yang bersifat pro, kontra, maupun netral. Kondisi ini menunjukkan pentingnya pemanfaatan pendekatan komputasional untuk mengelola dan menganalisis opini publik secara sistematis.

Salah satu pendekatan yang banyak digunakan adalah *text mining*, yang memungkinkan proses ekstraksi informasi dari data tidak terstruktur menjadi informasi yang bermakna. Dalam konteks ini, analisis sentimen berperan penting dalam mengklasifikasikan opini publik ke dalam kategori tertentu sehingga dapat memberikan gambaran yang lebih objektif terhadap suatu fenomena. Namun, dalam implementasinya, analisis sentimen menghadapi berbagai tantangan, seperti ambiguitas bahasa, variasi penggunaan kata, serta subjektivitas opini. Selain itu, karakteristik *imbalanced dataset* juga menjadi permasalahan utama yang memengaruhi performa model, di mana distribusi data yang tidak seimbang menyebabkan model cenderung bias terhadap kelas mayoritas dan mengurangi kemampuan dalam mengenali kelas minoritas.

Algoritma *Naïve Bayes* merupakan salah satu metode yang banyak digunakan dalam klasifikasi sentimen karena kesederhanaan dan efisiensinya. Penelitian oleh Purnama dan Sipayung [1] menunjukkan bahwa algoritma *Naïve Bayes* mampu mencapai akurasi sebesar 91,2% dalam klasifikasi sentimen pada data layanan publik. Meskipun demikian, performa algoritma ini tidak selalu konsisten dan sangat bergantung pada karakteristik data yang digunakan. Dalam studi terkait isu politik, akurasi *Naïve Bayes* dapat berada di bawah metode lain seperti LSTM atau *Random Forest*, yang menunjukkan adanya keterbatasan dalam menangkap kompleksitas konteks bahasa [2]. Selain itu, perbedaan antara akurasi data pelatihan dan data uji juga mengindikasikan adanya potensi *overfitting* dan keterbatasan generalisasi model [3].

Di sisi lain, sebagian besar penelitian analisis sentimen masih bergantung pada data hasil *scraping* dari media sosial, yang seringkali menghadapi kendala teknis, keterbatasan akses, serta isu etika dan privasi. Oleh karena itu, penggunaan *dataset* publik menjadi alternatif yang lebih efisien dan terstruktur dalam penelitian berbasis data. Selain itu, penggunaan *tools* berbasis visual seperti *Orange Data Mining* masih relatif terbatas dalam penelitian analisis sentimen, padahal platform ini memiliki keunggulan dalam hal kemudahan implementasi dan visualisasi proses analisis.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menganalisis kinerja algoritma *Naïve Bayes* dalam klasifikasi sentimen pada isu keaslian ijazah menggunakan dataset publik, serta mengkaji pengaruh distribusi data terhadap performa model melalui perbandingan antara dataset asli dan *dataset* hasil *oversampling*. Dengan demikian, penelitian ini tidak hanya berfokus pada pengukuran akurasi, tetapi juga memberikan pemahaman yang lebih mendalam mengenai faktor-faktor yang mempengaruhi performa model klasifikasi sentimen.

2 Tinjauan Literatur

Berbagai penelitian telah dilakukan dalam bidang analisis sentimen menggunakan algoritma *Naïve Bayes*. Penelitian oleh Hanan et al. [4] menunjukkan bahwa *Naïve Bayes* mampu mencapai akurasi sebesar 90,28% dan 84,51% dalam klasifikasi sentimen opini masyarakat terhadap calon gubernur berdasarkan data media sosial, yang telah melalui tahapan *preprocessing* dan pelabelan berbasis leksikon. Hasil tersebut menunjukkan bahwa *Naïve Bayes* efektif dalam mengklasifikasikan data teks dan menangkap kecenderungan opini publik.

Penelitian lain oleh Siddiq et al. [5] menggabungkan *Naïve Bayes* dengan pembobotan TF-IDF dan menghasilkan akurasi sebesar 77% dengan F1-score tertinggi pada kelas netral sebesar 0,85. Hal ini menunjukkan bahwa representasi fitur memiliki pengaruh terhadap performa model. Sementara itu, Yudistira dan Isnain, (2025) menemukan bahwa meskipun *Naïve Bayes* mampu mencapai akurasi sebesar 82,02%, performa model tidak merata pada setiap kelas akibat adanya *imbalanced dataset*.

Kualitas data juga menjadi faktor penting dalam analisis sentimen. Astirani et al. [7] menyatakan bahwa validasi manual terhadap pelabelan data dapat meningkatkan akurasi model serta mengurangi

<http://sistemasi.ftik.unisi.ac.id>

bias klasifikasi. Selain itu, Dwi et al.[8] menunjukkan bahwa penggunaan Teknik *Random Over Sampling* (ROS) mampu meningkatkan performa model secara signifikan dengan menyeimbangkan distribusi data. Di sisi lain, penelitian oleh Riswan et al. [9] menegaskan bahwa analisis sentimen pada isu politik memiliki kompleksitas tinggi akibat ambiguitas bahasa dan variasi opini, sehingga memerlukan proses *preprocessing* yang lebih optimal.

Berdasarkan penelitian terdahulu tersebut, dapat disimpulkan bahwa performa *Naïve Bayes* sangat dipengaruhi oleh kualitas data, representasi fitur, serta distribusi *dataset*. Namun, masih terbatas penelitian yang secara khusus mengkaji pengaruh distribusi data menggunakan *dataset* publik serta implementasi berbasis *tools* visual seperti *Orange Data Mining*. Oleh karena itu, penelitian ini berfokus pada analisis perbandingan performa model pada *dataset* asli dan *dataset* hasil *oversampling*.

Algoritma *Naïve Bayes* merupakan metode klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes sebagai dasar perhitungan. Algoritma ini mengasumsikan bahwa setiap fitur bersifat *independent*, sehingga proses komputasi menjadi lebih sederhana dan efisien. Dalam klasifikasi teks, *Naïve Bayes* digunakan untuk menghitung probabilitas suatu dokumen termasuk ke dalam kelas tertentu berdasarkan distribusi kata yang terdapat di dalamnya [10].

Text Mining merupakan teknik untuk mengekstraksi informasi dari data teks tidak terstruktur melalui tahapan seperti *preprocessing*, transformasi data, dan analisis. Proses ini menjadi dasar dalam analisis sentimen karena memungkinkan data teks diubah menjadi bentuk yang dapat diproses oleh model [11]. Dalam penelitian ini, tahapan *preprocessing* meliputi *case folding*, tokenisasi, *stopword removal*, dan *stemming*.

Analisis sentimen adalah proses untuk mengidentifikasi dan mengklasifikasikan opini dalam teks ke dalam kategori positif, negatif, dan netral. Namun, proses ini sering menghadapi tantangan seperti ambiguitas bahasa dan subjektivitas opini, yang dapat mempengaruhi hasil klasifikasi [12]. Oleh karena itu, diperlukan metode yang tepat serta kualitas data yang baik.

Klasifikasi merupakan teknik dalam *data mining* yang digunakan untuk memprediksi label kelas dari data berdasarkan pola yang telah dipelajari. Keberhasilan proses klasifikasi dipengaruhi oleh kualitas data, metode yang digunakan, serta distribusi *dataset* [13]. Dalam konteks ini, *imbalanced dataset* menjadi salah satu permasalahan utama yang dapat menurunkan performa model.

Penggunaan *dataset* publik memberikan kemudahan dalam memperoleh data tanpa harus melakukan proses *scraping*, namun seringkali memiliki permasalahan distribusi data yang tidak seimbang. Oleh karena itu, teknik *oversampling* digunakan untuk meningkatkan jumlah data pada kelas minoritas sehingga distribusi menjadi lebih seimbang.

Selain itu, *Orange Data Mining* merupakan perangkat lunak berbasis visual yang mendukung proses analisis data secara interaktif tanpa memerlukan pemrograman. *Tools* ini menyediakan berbagai komponen seperti *preprocessing*, *feature extraction*, klasifikasi, dan evaluasi model, sehingga memudahkan proses implementasi analisis sentimen secara sistematis [14].

Meskipun berbagai penelitian menunjukkan performa yang cukup baik, sebagian besar penelitian tersebut belum secara khusus menganalisis pengaruh distribusi data terhadap hasil klasifikasi. Selain itu penggunaan *dataset* publik dan *tools* berbasis visual seperti *Orange Data Mining* juga masih terbatas. Oleh karena itu, penelitian ini berfokus pada analisis perbandingan performa model berdasarkan kondisi distribusi data yang berbeda.

3 Metode Penelitian

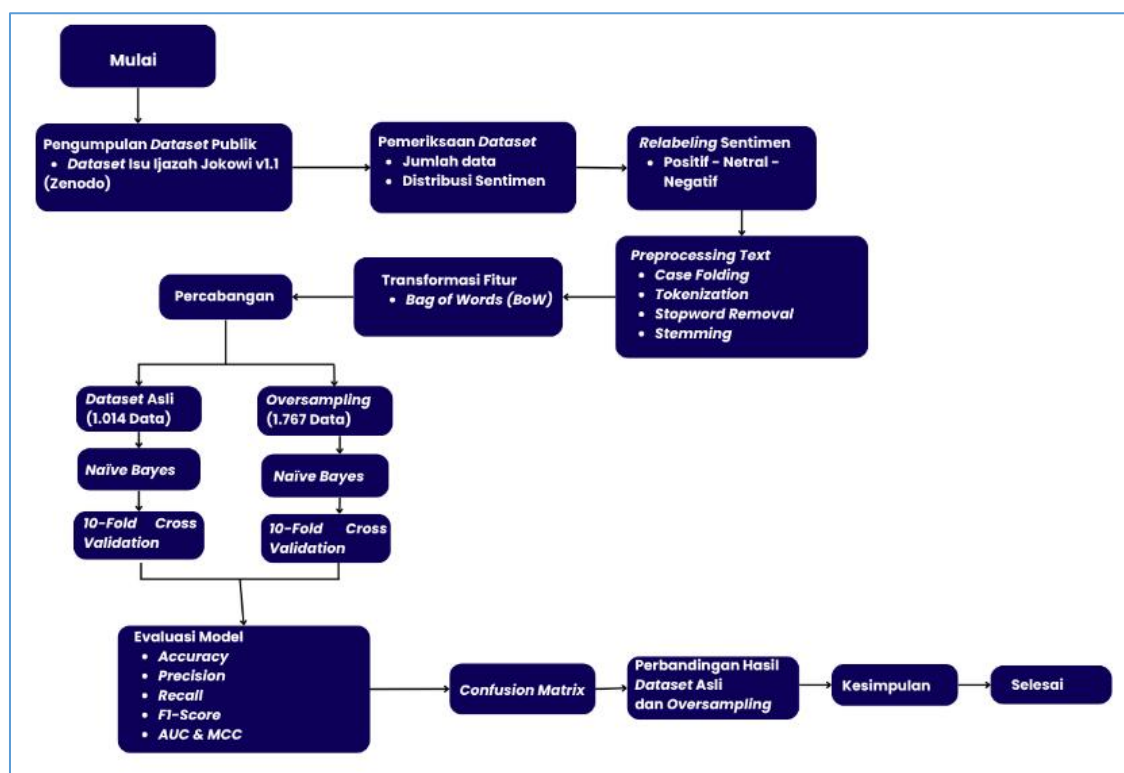
Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimen untuk menguji performa algoritma *Naïve Bayes* dalam klasifikasi sentimen terhadap isu keaslian ijazah. Pendekatan ini berfokus pada pengukuran kinerja model secara objektif menggunakan indikator seperti *accuracy*, *precision*, *recall*, *F1-score*, *Area Under Curve* (AUC), dan *Matthews Correlation Coefficient* (MCC). Metode eksperimen dilakukan dengan membangun model klasifikasi berbasis teks, kemudian mengevaluasi performanya menggunakan teknik validasi tertentu.

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa *dataset* publik yang diperoleh dari repositori ilmiah Zenodo dengan nama *Dataset Isu Ijazah Jokowi v1.1* yang dipublikasikan pada 10 juni 2025 [15]. *Dataset* ini dipilih karena memiliki keterkaitan langsung dengan topik penelitian, yaitu opini masyarakat mengenai isu keaslian ijazah yang berkembang di ruang digital. Penggunaan *dataset* publik dipilih karena lebih efisien dibandingkan proses

<http://sistemasi.ftik.unisi.ac.id>

pengambilan data secara mandiri (*scraping*) dari media sosial, serta mendukung aspek transparansi dan reproduksibilitas penelitian karena sumber data dapat diverifikasi dan digunakan kembali oleh peneliti lain. Jumlah data awal yang digunakan sebanyak 1.014 data teks yang telah dilabeli ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Meskipun *dataset* telah memiliki label sentimen, penelitian ini tetap melakukan proses *relabeling* sentimen melalui pemeriksaan ulang isi teks untuk memastikan kesesuaian label dengan konteks opini pada data. Proses *relabeling* dilakukan dengan menyesuaikan kategori sentimen ke dalam tiga kelas, yaitu positif, negatif, dan netral, sehingga kualitas label yang digunakan dalam proses klasifikasi menjadi lebih konsisten.

Dataset yang digunakan memiliki distribusi data yang tidak seimbang (*imbalanced dataset*), di mana kelas tertentu lebih dominan dibandingkan kelas lainnya sehingga berpotensi menyebabkan bias model dalam mengenali kelas minoritas. Kondisi ini sejalan dengan penelitian Dwi et al. [8] yang menyatakan bahwa ketidakseimbangan data dapat menurunkan kemampuan model dalam mengenali kelas minoritas. Untuk mengatasi permasalahan tersebut, dilakukan teknik *oversampling* dengan menduplikasi data pada kelas minoritas hingga distribusi menjadi seimbang, sehingga total *dataset* meningkat menjadi 1.767 data. Penggunaan *oversampling* terbukti mampu meningkatkan performa model, khususnya pada nilai *recall* dan akurasi [8]. Keaslian data dalam penelitian ini tetap dipertahankan karena data diperoleh langsung dari repositori Zenodo tanpa melakukan perubahan terhadap isi teks asli pada *dataset*. Peneliti hanya melakukan tahapan pengolahan data berupa *preprocessing*, *relabeling* sentimen, transformasi fitur, serta penyeimbangan distribusi data untuk kebutuhan eksperimen klasifikasi. Proses penelitian dilakukan secara sistematis menggunakan *Orange Data Mining* sebagai *tools* berbasis visual, yang mencakup tahapan pemuatan data, *preprocessing text*, transformasi fitur, pembangunan model, hingga evaluasi performa.



Gambar 1 Tahapan penelitian

Berdasarkan Gambar 1, penelitian dilakukan melalui beberapa tahapan yang disusun secara sistematis untuk menguji performa *Naive Bayes* dalam klasifikasi sentimen terhadap isu keaslian ijazah. Tahap awal penelitian dimulai dengan pengumpulan *dataset* publik yang diperoleh dari repositori Zenodo berupa *Dataset Isu Ijazah Jokowi v1.1*. Setelah data diperoleh, dilakukan pemeriksaan awal terhadap jumlah data, distribusi kelas sentimen, serta proses *relabeling* sentimen guna memastikan kesesuaian label dengan konteks opini pada *dataset*.

Pada tahap *preprocessing*, dilakukan beberapa proses penting untuk meningkatkan kualitas data, yaitu *case folding*, *tokenization*, *stopword removal*, dan *stemming*. *Case folding* dilakukan untuk menyeragamkan seluruh huruf menjadi huruf kecil, *tokenization* bertujuan memecah teks menjadi token kata, *stopword removal* digunakan untuk menghapus kata-kata umum yang tidak memiliki kontribusi signifikan terhadap makna sentimen, sedangkan *stemming* dilakukan untuk mengubah kata menjadi bentuk dasarnya. Tahapan ini bertujuan untuk mengubah data teks yang tidak terstruktur menjadi bentuk yang lebih terstandarisasi sehingga dapat diproses oleh model klasifikasi. Proses *preprocessing* terbukti berperan penting dalam meningkatkan akurasi model klasifikasi sentimen [16].

Selanjutnya, data ditransformasikan ke dalam bentuk numerik menggunakan metode *Bag of Words* (BoW), yang merepresentasikan frekuensi kemunculan kata dalam setiap dokumen tanpa mempertimbangkan urutan kata. Representasi ini memungkinkan algoritma *Naïve Bayes* untuk menghitung probabilitas setiap kelas berdasarkan fitur yang tersedia. Kombinasi antara BoW dan teknik menyeimbangkan data terbukti dapat meningkatkan performa model secara signifikan [17].

Penelitian ini dilakukan melalui dua skenario pengujian, yaitu menggunakan *dataset* asli sebanyak 1.014 data dan *dataset* hasil *oversampling* sebanyak 1.767 data. Teknik *oversampling* diterapkan pada kelas minoritas untuk mengurangi ketidakseimbangan distribusi data (*imbalanced dataset*) sehingga model dapat mempelajari pola dari setiap kelas sentimen secara lebih optimal. Model klasifikasi dibangun menggunakan algoritma *Naïve Bayes*, yaitu metode probabilistik berbasis Teorema Bayes yang digunakan untuk menghitung probabilitas posterior suatu kelas berdasarkan data yang diamati. Perhitungan probabilitas tersebut ditunjukkan pada Persamaan (1) [18].

$$P(m|n) = \frac{P(n|m) \times P(m)}{P(n)} \quad (1)$$

Pada Persamaan (1), m merepresentasikan kelas atau hipotesis, sedangkan n menunjukkan data yang akan diklasifikasikan [18].

Untuk memberikan gambaran yang lebih rinci mengenai mekanisme kerja algoritma *Naïve Bayes* pada penelitian ini, digunakan representasi alur proses dalam bentuk *pseudocode*. Tahapan klasifikasi sentimen mulai dari proses pemuatan data, *preprocessing text*, transformasi fitur, pembangunan model, hingga evaluasi hasil klasifikasi ditunjukkan pada Tabel 1.

Tabel 1 Tahapan kerja *naïve bayes* pada penelitian

Input : *Dataset* sentimen isu keaslian ijazah

Output : Hasil klasifikasi sentimen (positif, negatif, netral)

1. Memuat *dataset* penelitian
 - a. *Dataset* asli (1.014 data)
 - b. *Dataset* hasil *oversampling* (1.767 data)
 2. Melakukan relabeling sentimen
 - a. Positif
 - b. Negatif
 - c. Netral
 3. Melakukan *preprocessing text*
 - a. *Case folding*
 - b. *Tokenization*
 - c. *Stopword removal*
 - d. *Stemming*
 4. Melakukan transformasi fitur
Menggunakan *Bag of Words* (BoW)
 5. Membangun model *Naïve Bayes*
 - a. Menghitung probabilitas prior P(Kelas)
 - b. Menghitung probabilitas fitur P(Fitur|Kelas)
 6. Menghitung probabilitas posterior
P(Kelas|Dokumen)
 7. Menentukan kelas sentimen
-

-
- Berdasarkan probabilitas terbesar
8. Melakukan evaluasi model
Menggunakan *10-fold cross validation*
 9. Mengukur performa model :
Accuracy, precision, recall, F1-score, AUC, MCC.
 10. Menganalisis confusion matrix
 11. Membandingkan hasil *dataset* asli dan *oversampling*
-

Pendekatan ini banyak digunakan dalam klasifikasi teks karena efisiensi komputasinya, meskipun memiliki keterbatasan dalam menangkap hubungan antar fitur [17]. Implementasi model dilakukan menggunakan *Orange Data Mining* sebagai *tools* berbasis visual untuk mempermudah proses pengolahan data, pembangunan model, serta evaluasi hasil klasifikasi. Evaluasi model dilakukan menggunakan metode *10-fold cross validation*, di mana *dataset* dibagi menjadi sepuluh bagian dan diuji secara bergantian untuk memperoleh hasil yang lebih stabil. Metode ini terbukti efektif dalam menghasilkan evaluasi yang lebih akurat dan konsisten [19], serta banyak digunakan dalam pengujian model *Naïve Bayes* pada analisis sentiment [20].

Kinerja model diukur menggunakan beberapa parameter evaluasi, yaitu *accuracy, precision, recall, F1-score, AUC* dan *MCC*, untuk memberikan gambaran performa yang lebih komprehensif [16]. Selain itu, digunakan *confusion matrix* untuk menganalisis distribusi kesalahan klasifikasi secara lebih rinci. Analisis ini memungkinkan identifikasi pola kesalahan model dalam mengklasifikasikan data, terutama pada kondisi *dataset* yang tidak seimbang. Penelitian sebelumnya menunjukkan bahwa model *Naïve Bayes* cenderung mengalami kesalahan klasifikasi sebelum dilakukan penyeimbangan data, yang terlihat melalui *confusion matrix* [21]. Dengan tahapan penelitian yang dilakukan secara sistematis tersebut, penelitian ini tidak hanya mengukur performa model klasifikasi, tetapi juga menganalisis pengaruh distribusi data terhadap hasil klasifikasi sentimen melalui perbandingan performa antara *dataset* asli dan *dataset* hasil *oversampling*.

4 Hasil dan Pembahasan

Penelitian ini bertujuan untuk menguji kinerja algoritma *Naïve Bayes* dalam klasifikasi sentimen terhadap isu keaslian ijazah dengan membandingkan dua kondisi dataset, yaitu *dataset* asli sebanyak 1.014 data dan *dataset* hasil *oversampling* sebanyak 1.767 data. Pengujian dilakukan menggunakan *Orange Data Mining* dengan metode *10-fold cross validation* untuk memperoleh hasil evaluasi yang stabil dan representatif.

4.1. Hasil Pengujian *Dataset* Asli dan *Dataset* *Oversampling*

Berdasarkan hasil pengujian menggunakan *Test and Score*, performa model pada *dataset* asli menunjukkan hasil yang relatif rendah dengan nilai *accuracy* sebesar 37,2%, *AUC* sebesar 0,704, serta nilai *precision, recall*, dan *F1-score* yang masih rendah yang bisa dilihat pada Gambar 2. Hasil tersebut menunjukkan bahwa model belum mampu mengklasifikasikan data sentimen secara optimal. Rendahnya performa model dipengaruhi oleh kondisi distribusi data yang tidak seimbang (*imbalanced dataset*), sehingga model cenderung lebih dominan mengenali kelas mayoritas dibandingkan kelas minoritas.

Sebaliknya, pada *dataset* hasil *oversampling*, performa model meningkat secara signifikan dengan nilai *accuracy* sebesar 82,1% dan *AUC* sebesar 0,970 seperti yang terdapat pada Gambar 3. Peningkatan ini menunjukkan bahwa distribusi data yang lebih seimbang memberikan pengaruh terhadap kemampuan model dalam mempelajari pola pada masing-masing kelas sentimen secara lebih optimal.

Model	AUC	CA	F1	Prec	Recall	MCC
Naive Bayes	0.704	0.372	0.325	0.682	0.372	0.233

Gambar 2 *Test and score dataset* asli

Model	AUC	CA	F1	Prec	Recall	MCC
Naive Bayes	0.970	0.821	0.814	0.861	0.821	0.754

Gambar 3 Test and score data oversampling

Untuk memberikan gambaran hasil pengujian secara lebih rinci, dilakukan evaluasi model menggunakan beberapa parameter, yaitu *accuracy*, *precision*, *recall*, *F1-score*, *AUC*, dan *MCC*. Hasil evaluasi performa model pada *dataset* asli dan *dataset* hasil *oversampling* dapat dilihat pada Tabel 2.

Tabel 2 Hasil evaluasi model *naïve bayes*

Parameter	<i>Dataset</i> Asli	<i>Dataset</i> <i>Oversampling</i>
<i>Accuracy</i>	37,2%	82,1%
<i>Precision</i>	0.682	0.861
<i>Recall</i>	0.372	0.821
<i>F1-score</i>	0.325	0.814
<i>AUC</i>	0.704	0.970
<i>MCC</i>	0.233	0.754

Berdasarkan Tabel 1, seluruh parameter evaluasi mengalami peningkatan setelah diterapkan teknik *oversampling*. Nilai *accuracy* meningkat sebesar 44,9% yaitu dari 37,2% menjadi 82,1% yang menunjukkan peningkatan kemampuan model dalam melakukan klasifikasi sentimen secara tepat. Selain itu, peningkatan pada nilai *precision*, *recall*, dan *F1-score*, menunjukkan bahwa model menjadi lebih baik dalam mengenali setiap kategori sentimen, khususnya pada kelas minoritas. Nilai *AUC* meningkat dari 0,704 menjadi 0,970, yang menunjukkan bahwa kemampuan model dalam membedakan antar kelas sentimen menjadi semakin baik setelah distribusi data diseimbangkan. Sementara itu, peningkatan nilai *MCC* menunjukkan bahwa kualitas klasifikasi model secara keseluruhan mengalami peningkatan, terutama pada kondisi *dataset* yang sebelumnya tidak seimbang.

4.2. Analisis *Confusion Matrix*

Analisis lebih lanjut dilakukan menggunakan *confusion matrix* untuk mengetahui pola kesalahan klasifikasi pada masing-masing sentimen. Pada *dataset* asli, model masih menunjukkan tingkat kesalahan klasifikasi yang tinggi, terutama pada kelas minoritas yang sering diprediksi sebagai kelas mayoritas. Kondisi ini menunjukkan bahwa model mengalami bias terhadap distribusi data yang tidak seimbang sehingga kemampuan model dalam mengenali seluruh kategori sentimen menjadi kurang optimal.

		Predicted			Σ
		NEGATIF	NETRAL	POSITIF	
Actual	NEGATIF	73	230	286	589
	NETRAL	4	206	76	286
	POSITIF	3	38	98	139
Σ		80	474	460	1014

Gambar 4 *Confusion matrix* *dataset* asli

		Predicted			Σ
		NEGATIF	NETRAL	POSITIF	
Actual	NEGATIF	332	221	36	589
	NETRAL	11	577	2	590
	POSITIF	1	46	541	588
Σ		344	844	579	1767

Gambar 5 Confusion matrix data oversampling

Berdasarkan Gambar 4, terlihat bahwa jumlah prediksi yang salah masih cukup tinggi pada beberapa kelas sentimen. Hal ini menunjukkan bahwa model belum mampu mengenali pola data secara merata. Sebaliknya, pada *dataset* hasil *oversampling* yang ditunjukkan pada Gambar 5, distribusi kesalahan klasifikasi mengalami penurunan dan prediksi antar kelas menjadi lebih seimbang. Kondisi tersebut menunjukkan bahwa teknik *oversampling* membantu meningkatkan representasi data pada kelas minoritas sehingga model mampu mengenali pola klasifikasi secara lebih baik. Hasil ini sejalan dengan penelitian Yudistira dan Isnain [6] yang menyatakan bahwa kondisi *imbalanced dataset* dapat menyebabkan model klasifikasi menjadi bias terhadap kelas mayoritas. Selain itu, penelitian Dwi et al.[8] juga menunjukkan bahwa teknik *oversampling* mampu meningkatkan performa model dengan memperbaiki distribusi data pelatihan.

4.3. Pembahasan Hasil Penelitian

Secara keseluruhan, hasil penelitian menunjukkan bahwa performa algoritma *Naïve Bayes* sangat dipengaruhi oleh karakteristik *dataset* yang digunakan. Pada *dataset* asli, rendahnya performa model dipengaruhi oleh ketidakseimbangan distribusi data, sehingga model cenderung mengalami kesulitan dalam mengenali kelas minoritas. Selain itu, karakteristik data teks yang mengandung ambiguitas, subjektivitas, serta variasi ekspresi bahasa menyebabkan asumsi independensi fitur pada algoritma *Naïve Bayes* menjadi kurang optimal dalam menangkap hubungan antar kata. Hal ini juga didukung oleh penelitian Panggabean dan Junika, (2024) yang menyatakan bahwa *Naïve Bayes* memiliki keterbatasan dalam menangkap konteks bahasa yang kompleks [2].

Setelah diterapkan teknik *oversampling*, model menunjukkan peningkatan performa yang signifikan karena distribusi data menjadi lebih seimbang. Kondisi ini memungkinkan model mempelajari pola dari setiap kelas sentimen dengan lebih optimal, sehingga menghasilkan peningkatan pada nilai *accuracy*, *recall*, *precision*, *F1-score*, *AUC*, *MCC*. Hasil penelitian ini menunjukkan bahwa keberhasilan klasifikasi sentimen tidak hanya dipengaruhi oleh algoritma yang digunakan, tetapi juga sangat dipengaruhi oleh kualitas serta distribusi data.

Jika dibandingkan dengan penelitian terdahulu, hasil penelitian ini menunjukkan bahwa performa *Naïve Bayes* dapat bervariasi tergantung pada karakteristik data yang digunakan. Beberapa penelitian menunjukkan akurasi tinggi pada *dataset* yang lebih terstruktur dan seimbang [4][5], sementara pada data opini publik yang kompleks, performa model cenderung menurun [2][9]. Hal ini menunjukkan bahwa perbedaan hasil bukan disebabkan oleh kesalahan metode, melainkan oleh karakteristik data yang digunakan. Temuan ini memperkuat bahwa kualitas data dan distribusinya merupakan faktor kunci dalam keberhasilan model klasifikasi.

Dengan demikian, penelitian ini menunjukkan bahwa keberhasilan klasifikasi sentimen tidak hanya bergantung pada algoritma yang digunakan, tetapi juga sangat dipengaruhi oleh kualitas dan distribusi data. Penggunaan teknik *oversampling* terbukti efektif dalam meningkatkan performa model, sehingga dapat menjadi solusi dalam mengatasi permasalahan *imbalanced dataset*. Temuan penelitian menunjukkan bahwa teknik *oversampling* efektif dalam meningkatkan performa klasifikasi pada kondisi *dataset* yang tidak seimbang, sehingga dapat dipertimbangkan sebagai pendekatan penanganan *imbalanced dataset* pada analisis sentimen berbasis teks.

5 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma *Naïve Bayes* dapat diterapkan dalam klasifikasi sentimen pada data teks berbahasa Indonesia, khususnya pada isu keaslian ijazah. Hasil pengujian menunjukkan bahwa performa model sangat dipengaruhi oleh distribusi data yang digunakan. Pada *dataset* asli yang memiliki kondisi *imbalanced dataset*, model menghasilkan nilai *accuracy* sebesar 37,2% dengan AUC sebesar 0,704, yang menunjukkan bahwa kemampuan klasifikasi model masih rendah karena model cenderung bias terhadap kelas mayoritas. Setelah dilakukan penyeimbangan data menggunakan teknik *oversampling*, performa model meningkat secara signifikan dengan akurasi sebesar 82,1% dan AUC sebesar 0,970. Selain itu, nilai *precision*, *recall*, *F1-score*, dan MCC juga mengalami peningkatan, yang menunjukkan bahwa model menjadi lebih baik dalam mengenali setiap kategori sentimen secara lebih seimbang. Hasil tersebut membuktikan bahwa distribusi data memiliki pengaruh yang signifikan terhadap performa algoritma *Naïve Bayes* dalam klasifikasi sentimen. Hasil penelitian juga menunjukkan bahwa rendahnya performa pada *dataset* asli dipengaruhi oleh ketidakseimbangan distribusi data serta karakteristik data teks yang mengandung ambiguitas, subjektivitas, dan variasi ekspresi bahasa. Kondisi tersebut menyebabkan asumsi independensi fitur pada algoritma *Naïve Bayes* menjadi kurang optimal dalam menangkap hubungan antar kata. Secara keseluruhan, penelitian ini menunjukkan bahwa penggunaan teknik *oversampling* efektif dalam meningkatkan performa klasifikasi sentimen pada kondisi *dataset* yang tidak seimbang. Dengan demikian, keberhasilan model klasifikasi tidak hanya dipengaruhi oleh algoritma yang digunakan, tetapi juga sangat dipengaruhi oleh kualitas dan distribusi data yang digunakan pada proses pelatihan model.

Referensi

- [1] A. A. Purnama and Y. R. Sipayung, “Sentiment Analysis of Public Service using Naïve Bayes Classifier,” *J. Inf. Syst. Informatics*, Vol. 7, No. 3, pp. 2439–2457, 2025, DOI: 10.51519/journalisi.v7i3.1207.
- [2] S. Panggabean and A. Junika, “Sentiment Analysis on Public Opinions Regarding the 2024 Regional Elections using Long Short-Term Memory (LSTM), Random Forest , and Naive Bayes,” *JOISTECH: Journal Inf. Syst. Technol.*, Vol. 01, No. 02, pp. 67–75, 2024.
- [3] A. A. Hisyam and A. T. Ayunda, “Analisis Sentimen Persepsi Publik terhadap Bank DKI Pada Twitter menggunakan Metode Naive Bayes Classifier,” *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, Vol. 14, No. 1.
- [4] M. N. Hanan, H. M. Jumasa, and I. Y. Pasa, “Analisis Sentimen Calon Gubernur Jawa Tengah 2024 menggunakan Metode Naïve Bayes,” *J. FASILKOM*, Vol. 15, No. 3, pp. 491–499, 2025.
- [5] M. J. Siddiq *et al.*, “Analisis Sentimen Opini Masyarakat terhadap PILKADA 2024 di Media Sosial Twitter menggunakan Algoritma Naive Bayes,” *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, Vol. 13, No. 2, pp. 609–622, 2025.
- [6] F. Yudistira and A. R. Isnain, “Analisis Sentimen terhadap Seleksi CPNS Tahun 2024 berbasis Media Sosial X menggunakan Algoritma Naïve Bayes Program Studi Informatika , Fakultas Teknik dan Ilmu Komputer , Universitas Teknokrat Indonesia,” *J. Pendidik. dan Teknol. Indones.*, Vol. 5, No. 3, pp. 887–897, 2025.
- [7] W. Astriani, O. S. Bachri, and B. Irawan, “Classification of Sentiment of Emina Product Reviews using the Naive Bayes Algorithm,” *Bit-Tech (Binary Digit. - Technol.)*, Vol. 8, No. 2, 2025, DOI: 10.32877/bt.v8i2.3554.
- [8] N. Dwi, H. Sadikin, and S. Susanti, “Analisis Sentimen Publik terhadap Kampanye Pengurangan Sampah Plastik menggunakan Algoritma Naïve Bayes,” *J. FASILKOM*, Vol. 15, No. 2, pp. 202–212, 2025.
- [9] M. Riswan, A. Primajaya, A. Susilo, and Y. Irawan, “Analisis Sentimen terhadap Pemberitaan Hasil Rekapitulasi PEMILU Presiden 2024 pada Media Sosial Instagram menggunakan Naive Bayes,” *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, Vol. 13, No. 1, 2025.
- [10] I. F. Rozi, E. N. Hamdana, and Muhammad Balya Iqbal Alfahmi, “Pengembangan Aplikasi Analisis Sentimen Twitter menggunakan Metode Naive Bayes Classifier (Studi Kasus SAMSAT Kota Malang),” *J. Inform. Polinema*, Vol. 4, No. 2, p. 149, 2018, DOI: 10.33795/jip.v4i2.164.

- [11] I. Juventius, T. Gurning, P. P. Adikara, and R. S. Perdana, "Analisis Sentimen Dokumen Twitter menggunakan Metode *Naïve Bayes* dengan Seleksi Fitur GU Metric," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, Vol. 7, No. 5, pp. 2169–2177, 2023, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12665>
- [12] M. A. Kausar, A. Soosaimanickam, and M. Nasar, "Public Sentiment Analysis on Twitter Data during COVID-19 Outbreak," *Int. J. Adv. Comput. SCI. Appl.*, Vol. 12, No. 2, pp. 415–422, 2021, DOI: 10.14569/IJACSA.2021.0120252.
- [13] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru di Twitter menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, Vol. 14, No. 2, p. 115, 2020, DOI: 10.33365/jti.v14i2.679.
- [14] S. I. Nurhafida and F. Sembiring, "Analisis Text Clustering Masyarakat di Twiter mengenai Mcdonald'Sxbts menggunakan Orange Data Mining," *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.*, pp. 28–35, 2021.
- [15] T. Mufhimah, S. A. Yumnatusta, and N. A. Rakhmawati, "Dataset Isu Ijazah Jokowi." 2025. DOI: <https://doi.org/10.5281/zenodo.15636033>.
- [16] J. Saputra, L. Maryani, D. Wulandari, W. Eka, P. T. Informatika, and P. T. Komputer, "Analisis Performa *Naive Bayes* dan SVM terhadap Sentimen Teks Media Sosial dengan Word2Vec dan SMOTE," *J. INSTEK (Informatika Sains dan Teknol.*, Vol. 10, No. 1, pp. 143–155, 2025.
- [17] M. A. Hermawan, A. Faqih, G. Dwilestari, T. Informatika, and S. Informasi, "Implementasi Akurasi Model *Naive Bayes* menggunakan *Smote* dalam Analisis Sentimen Pengguna Aplikasi Brimo," *JITET (Jurnal Inform. dan Tek. Elektro Ter.*, Vol. 13, No. 1, 2025.
- [18] A. Rasool, R. Tao, K. Marjan, and T. Naveed, "Twitter Sentiment Analysis: A Case Study for Apparel Brands," *J. Phys. Conf. Ser.*, Vol. 1176, No. 2, 2019, DOI: 10.1088/1742-6596/1176/2/022015.
- [19] S. Khoerunnisa, D. F. Shiddiq, and D. Nurhayati, "Application of the Naive Bayes Algorithm with TF-IDF and Cross Validation Techniques for Sentiment Analysis Towards Starlink Penerapan Algoritma *Naive Bayes* dengan Teknik TF-IDF dan Cross Validation untuk Analisis Sentimen Terhadap Starlink," *MALCOM Indones. J. Mach. Learn. Comput. SCI.*, Vol. 5, No. April, pp. 566–577, 2025.
- [20] R. N. Irawan, K. M. Hindrayani, and M. Idhom, "Penerapan Cross Validation sebagai Analisis Sentimen Pelayanan Publik Kereta Api Lokal Daop 8 menggunakan Metode *Multinomial Naive Bayes*," *G-Tech J. Teknol. Terap.*, Vol. 8, No. 2, pp. 954–963, 2024.
- [21] M. F. Yulianto, F. M. Hana, and A. Prihandono, "Penerapan Algoritma *Naive Bayes* dalam Analisis Sentimen terhadap Mobil Listrik," *Sainteks*, Vol. 22, No. 1, pp. 109–115, 2025, DOI: 10.30595/sainteks.v22i1.26036.