

Analisis Sentimen Ulasan Co-Pilot Google Play dengan SVM, Neural Network, dan Decision Tree

Imam Kharits Najibulloh ^{1,*}, Imam tahyudin ¹, Dhanar Intan Surya Saputra ¹

¹ Program Studi Ilmu Komputer, Universitas Amikom Purwokerto, Indonesia

* Correspondence: kharits9@gmail.com

Copyright: © 2025 by the authors

Received: 1 Maret 2025 | Revised: 13 Maret 2025 | Accepted: 1 April 2025 | Published: 17 April 2025

Abstrak

Analisis sentimen adalah teknik yang digunakan untuk memahami opini pengguna melalui ulasan produk atau layanan. Tujuan penelitian ini adalah untuk membandingkan tiga metode klasifikasi, yaitu *Support Vector Machine* (SVM), *Neural Network* (NN), dan *Decision Tree* (DT) dalam menganalisis sentimen pengguna aplikasi *Microsoft Co-Pilot* berbahasa Indonesia yang diambil dari *Google Play Store*. Dataset terdiri dari 20.000 ulasan, yang terlebih dahulu melalui preprocessing seperti normalisasi, tokenisasi, penghapusan stopwords, dan stemming. Ketiga metode yang kami gunakan dalam penelitian ini, memiliki arsitektur *Multilayer Perceptron* (MLP) dengan tiga lapisan tersembunyi dan fungsi aktivasi ReLU, serta teknik regularisasi *dropout* untuk menghindari *overfitting*. Evaluasi model dilakukan menggunakan akurasi, presisi, recall, dan *F1-score*, dengan hasil menunjukkan bahwa NN mencapai akurasi tertinggi 95,5%, diikuti oleh SVM dengan 95,4% dan DT dengan 92,1%. Keunggulan metode NN terletak pada kemampuannya dalam mengenali pola yang lebih kompleks dalam bahasa Indonesia, terutama dalam menangani teks informal dan campuran bahasa (*code-mixing*). Penelitian ini berkontribusi pada pengembangan aplikasi berbasis *Artificial Intelligence* (AI) dengan memberikan wawasan tentang efektivitas metode klasifikasi dalam analisis sentimen bahasa Indonesia, yang penting untuk meningkatkan kualitas layanan dan pengembangan teknologi NLP di Indonesia. Implikasi praktis dari penelitian ini dapat digunakan dalam pengembangan aplikasi berbasis AI yang lebih responsif terhadap sentimen pengguna.

Kata kunci: *decision tree; machine learning; microsoft co-pilot; neural network; support vector machine*

Abstract

Sentiment analysis is a technique used to understand user opinions through product or service reviews. The purpose of this research is to compare three classification methods, namely *Support Vector Machine* (SVM), *Neural Network* (NN), and *Decision Tree* (DT) in analyzing the sentiment of users of the Indonesian-language *Microsoft Co-Pilot* application taken from the *Google Play Store*. The dataset consists of 20,000 reviews, which first went through preprocessing such as normalization, tokenization, stopwords removal, and stemming. The three methods we used in this study have a *Multilayer Perceptron* (MLP) architecture with three hidden layers and a ReLU activation function, as well as a dropout regularization technique to avoid overfitting. Model evaluation was conducted using accuracy, precision, recall, and *F1-score*, with the results showing that NN achieved the highest accuracy of 95.5%, followed by SVM with 95.4% and DT with 92.1%. The advantage of the NN method lies in its ability to recognize more complex patterns in Indonesian, especially in handling informal text and code-mixing. This research contributes to the development of *Artificial Intelligence* (AI)-based applications by providing insights into the effectiveness of classification methods in Indonesian sentiment analysis, which is important for improving service quality and the development of NLP technology in Indonesia. The practical implications of this research can be used in the development of AI-based applications that are more responsive to user sentiment.



Keywords: *decision tree; machine learning; microsoft co-pilot; neural network; support vector machine*

PENDAHULUAN

Perkembangan kecerdasan buatan telah membawa perubahan signifikan dalam banyak bidang, termasuk aplikasi produktivitas. Salah satu inovasi kecerdasan buatan atau *Artificial Intelligence* (AI) yang paling terkenal adalah *Microsoft Co-Pilot*, yang terintegrasi dalam produk *Microsoft 365* seperti *word* dan *excel*. *Co-Pilot* dirancang untuk membantu pengguna dalam menyelesaikan tugas-tugas seperti mengedit dokumen, menyusun laporan, dan mengelola data secara otomatis, yang membuatnya sangat berguna dalam meningkatkan efisiensi (Junianto et al., 2024). Namun, meskipun aplikasi ini menawarkan banyak keuntungan, *feedback* atau ulasan dari pengguna di platform seperti *Google Play Store* sering kali beragam, mencakup sentimen positif, negatif, dan netral (Budianto et al., 2024). Oleh karena itu, memahami persepsi pengguna melalui analisis sentimen sangat penting untuk memperbaiki kualitas layanan dan pengembangan aplikasi berbasis AI.

Meskipun analisis sentimen dalam bahasa Indonesia telah banyak dilakukan, tantangan besar muncul akibat penggunaan bahasa yang tidak baku, singkatan, serta campuran bahasa (*code-mixing*) dalam ulasan pengguna (Soemedhy & others, 2023). Hal ini menyebabkan metode analisis tradisional, yang banyak digunakan untuk bahasa Inggris, tidak selalu efektif dalam memproses teks berbahasa Indonesia (Saputra & Noor Hasan, 2023). Keterbatasan ini memperburuk kesulitan dalam mencapai akurasi tinggi dalam analisis sentiment (Manueke et al., 2023), terutama dalam konteks aplikasi berbasis AI yang menggunakan bahasa alami. Selain itu, kebanyakan penelitian sebelumnya (Budianto et al., 2024) masih terbatas pada analisis produk digital umum atau dalam bahasa Inggris, sehingga sedikit yang meneliti analisis sentimen untuk produk AI berbahasa Indonesia (Sabrani et al., 2020).

Penelitian ini membandingkan tiga algoritma klasifikasi sentimen yaitu SVM, NN, dan DT. SVM adalah algoritma yang dipilih karena kemampuannya dalam menangani dimensi data yang tinggi dan memisahkan kelas sentimen dengan margin yang maksimal. Hal ini menjadikannya sangat efektif dalam mengklasifikasikan data yang kompleks dan memiliki dimensi tinggi (Dewi, 2021). NN merupakan algoritma yang memiliki kemampuan untuk mengenali pola yang lebih kompleks, khususnya dalam menangani hubungan non-linear dalam teks yang lebih bervariasi. Algoritma ini efektif untuk menganalisis data yang memiliki variabilitas tinggi (Pamungkas et al., 2024; Ramdhana & Pratiwi, 2023; Triginandri & Subhiyakto, 2024). Sementara DT adalah algoritma yang dipilih karena kelebihan interpretasinya, yang memungkinkan pemahaman yang lebih jelas terhadap faktor-faktor yang mempengaruhi klasifikasi sentimen dalam ulasan pengguna (Pang & Lee, 2008).

Penelitian sebelumnya menggunakan algoritma seperti *naïve bayes*, *logistic regression*, dan SVM dalam analisis sentimen pada berbagai aplikasi (Yolanda et al., 2022). Namun, sangat sedikit yang fokus pada analisis sentimen dalam bahasa Indonesia, terutama yang melibatkan aplikasi berbasis AI seperti *Co-Pilot*. Beberapa penelitian seperti yang dilakukan oleh (Ramadani et al., 2024) telah membahas analisis sentimen menggunakan SVM dan DT, tetapi mereka lebih banyak berfokus pada produk digital umum atau kebijakan sosial. Penelitian ini menganalisis ulasan *Microsoft Co-Pilot* dalam bahasa Indonesia, yang memiliki karakteristik bahasa yang berbeda dan lebih kompleks (Zamsuri et al., 2023).

Penelitian ini dilakukan selain untuk membantu *Microsoft* dalam mengevaluasi aplikasi *Microsoft Co-Pilot*, juga untuk memperkaya penelitian tentang analisis sentimen. Tujuan utama pada penelitian ini adalah untuk mengetahui persepsi pengguna *Co-Pilot* yang bernilai positif dan negatif yang berkaitan dengan fitur yang dimiliki *Co-Pilot* saat ini, yang selanjutnya dapat digunakan sebagai referensi dalam pengembangan aplikasi *Microsoft Co-Pilot* sesuai dengan kebutuhan pengguna. Hal tersebut yang belum dibahas pada penelitian sebelumnya.

Dengan adanya penelitian ini diharapkan dapat membantu *Microsoft* untuk meningkatkan kualitas aplikasi *Microsoft Co-Pilot*.

METODE

Penelitian ini menggunakan pendekatan kuantitatif untuk menganalisis sentimen ulasan pengguna aplikasi *Microsoft Co-Pilot* dalam bahasa Indonesia, dengan tujuan untuk membandingkan efektivitas tiga algoritma klasifikasi sentimen, yaitu SVM, NN, dan DT. *Microsoft Co-Pilot* adalah fitur berbasis kecerdasan buatan yang terintegrasi dalam produk *Microsoft 365* seperti *Word* dan *Excel*, yang bertujuan untuk membantu pengguna dalam menyelesaikan tugas dengan lebih efisien. Ulasan yang dianalisis dalam penelitian ini diambil dari platform *Google Play Store* (Nurian, 2023), dengan dataset yang terdiri dari 20.000 ulasan pengguna terbaru yang dikumpulkan hingga 28 Februari 2025 menggunakan teknik *web scraping*. Untuk pengumpulan data ini, digunakan pustaka *BeautifulSoup* dan *Scrapy*, yang memungkinkan pengambilan informasi dari halaman ulasan secara otomatis, serta mematuhi kebijakan pengumpulan data yang ada.

Sebelum menerapkan model klasifikasi, dilakukan *preprocessing* yang mencakup beberapa tahapan penting, seperti *case folding*, yang mengubah seluruh huruf dalam teks menjadi huruf kecil untuk memastikan konsistensi, serta tokenisasi yang memecah teks menjadi kata-kata individu. Langkah berikutnya adalah *stopword removal*, di mana kata-kata yang umum digunakan namun tidak memberikan informasi penting (seperti "dan", "atau", "dengan") dihapus menggunakan daftar *stopwords* yang dikembangkan dari sumber *IDNstopwords* (Hendrawan et al., 2022). Setelah itu, dilakukan *stemming*, yaitu untuk mengubah kata menjadi bentuk dasarnya menggunakan pustaka *Sastrawi* versi 1.0.1. Selain itu, dilakukan juga langkah normalisasi untuk menangani kata-kata tidak baku, singkatan, dan variasi bahasa lain yang sering ditemukan dalam ulasan pengguna bahasa Indonesia.

Setelah proses *preprocessing* selesai, teks ulasan diubah menjadi representasi numerik menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF adalah teknik yang mengukur seberapa penting sebuah kata dalam dokumen relatif terhadap keseluruhan dataset, memberikan bobot lebih pada kata-kata yang lebih spesifik dan jarang muncul di dokumen lain. Penggunaan *n-gram* (*unigram* dan *bigram*) juga diterapkan pada parameter TF-IDF untuk menangkap konteks kata yang lebih luas dalam kalimat.

Selanjutnya, dataset dibagi menjadi dua bagian: 80% untuk data latih dan 20% untuk data uji. Pembagian ini menggunakan teknik *stratified sampling*, yang memastikan bahwa distribusi sentimen tetap seimbang dalam setiap subset data, menghindari adanya ketidakseimbangan yang dapat mempengaruhi hasil pelatihan model. Setelah data terbagi, tiga model klasifikasi yang diuji adalah SVM dengan kernel RBF, NN menggunakan arsitektur *Multilayer Perceptron* (MLP) dengan tiga lapisan tersembunyi yang masing-masing memiliki 128 neuron dan menggunakan fungsi aktivasi ReLU, serta DT menggunakan metode *Classification and Regression Tree* (CART) dengan kedalaman maksimum 10 dan teknik *pruning* untuk menghindari pohon keputusan yang terlalu dalam dan kompleks.

Pencarian *hyperparameter* optimal dilakukan menggunakan *grid search* dengan *cross-validation* untuk memilih kombinasi parameter terbaik bagi setiap model. Proses ini memastikan bahwa model yang dihasilkan memiliki performa terbaik berdasarkan data yang tersedia. Misalnya, dalam model NN, jumlah lapisan dan neuron pada setiap lapisan diuji dalam berbagai kombinasi untuk mencari struktur terbaik. Sedangkan pada model DT diuji untuk mendapatkan model yang seimbang antara akurasi dan interpretabilitas.

Evaluasi model dilakukan menggunakan metrik akurasi, *precision*, *recall*, dan *f1-score*. Metrik-metrik ini digunakan untuk menilai seberapa baik model dapat memprediksi kelas sentimen (positif, negatif, atau netral) dengan benar. Untuk memberikan gambaran yang lebih

lengkap mengenai kinerja model, juga digunakan *confusion matrix* untuk menganalisis distribusi kesalahan dan kesuksesan prediksi.

HASIL DAN PEMBAHASAN

Hasil

Penelitian ini berhasil menganalisis sentimen ulasan pengguna *Microsoft Co-Pilot* dalam bahasa Indonesia menggunakan tiga algoritma klasifikasi sentimen: NN, SVM, dan DT. Pada gambar 1 adalah dataset yang digunakan terdiri dari 20.000 ulasan yang diambil melalui teknik web scraping menggunakan pustaka BeautifulSoup dan Scrapy, yang mencakup ulasan terbaru hingga 28 Februari 2025. Sementara itu, gambar 2 merupakan ringkasan data dikumpulkan 28 Februari 2025. Pengumpulan data dilakukan dengan mematuhi etika pengumpulan data yang ada dan memastikan kepatuhan terhadap kebijakan *Google Play Store*.

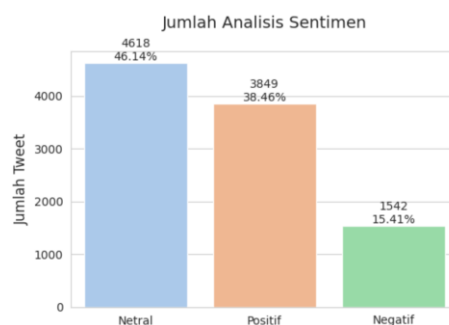
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20000 entries, 0 to 19999
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   Review ID       20000 non-null  object
1   Username        20000 non-null  object
2   Rating          20000 non-null  int64
3   Review Text     20000 non-null  object
4   Date            20000 non-null  object
dtypes: int64(1), object(4)
memory usage: 781.4+ KB
```

Gambar 1. Data scrapping

	Review ID	Username	Rating	Review Text	Date
0	79d46af8-3b11-4bac-8aec-40e88750e88f	Pengguna Google	4	good	2025-02-22 02:24:30
1	ce5d2428-76e3-4356-b0c6-c49795caf711	Pengguna Google	5	nice jobs is oke	2025-02-22 01:35:15
2	364eab47-36a6-4de8-a686-358beb84387b	Pengguna Google	5	bagus	2025-02-21 13:53:13
3	0997300a-f14a-4e91-8ddf-7905c849c458	Pengguna Google	5	mantab	2025-02-21 12:00:58
4	c594f686-e432-4a15-8f77-7f19bfb32b1a	Pengguna Google	5	Bagus	2025-02-21 10:04:14

Gambar 2. Distribusi sentimen

Setelah pengumpulan data, dilakukan *preprocessing* yang mencakup *case folding*, tokenisasi, *stopword removal*, *stemming*, dan normalisasi untuk menangani kata tidak baku dan singkatan. Pada tahap ini, jumlah ulasan berkurang menjadi 18.500 ulasan dari 20.000 ulasan awal. Pengurangan ini terjadi karena ulasan yang tidak relevan (seperti yang hanya berisi simbol atau tidak ada isi) dan duplikat (ulasan yang diulang oleh pengguna yang sama) dihapus. Kriteria penghapusan ini memastikan bahwa data yang digunakan untuk pelatihan model adalah data yang valid dan relevan, yang dapat dipertanggungjawabkan.



Gambar 3. Distribusi sentimen ulasan pengguna *microsoft co-pilot*

Distribusi sentimen dalam dataset menunjukkan bahwa 46,14% ulasan bersifat netral, 38,46% positif, dan 15,41% negatif. Hasil pada gambar 4 adalah distribusi ulasan Sentimen

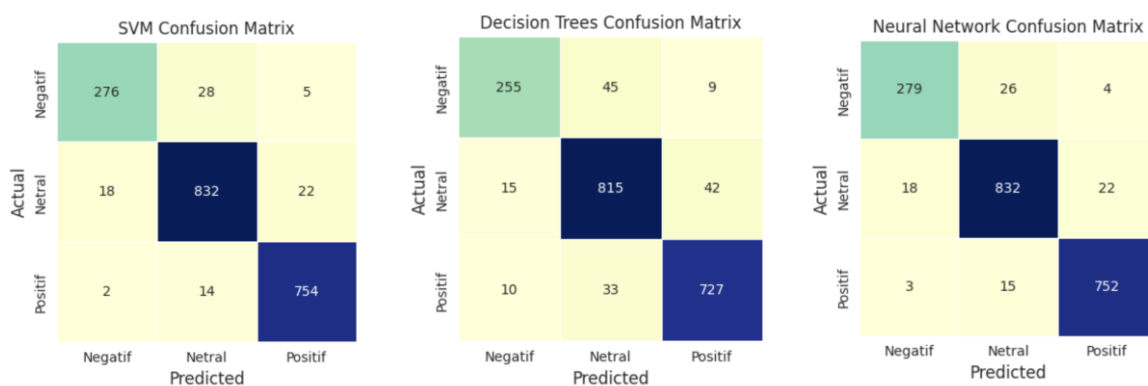
yang menunjukkan bahwa mayoritas ulasan berisi sentimen netral, yang mencerminkan pendapat pengguna yang tidak terlalu ekstrem dalam menilai aplikasi. Penjelasan yang lebih rinci mengenai sentimen netral adalah bahwa banyak ulasan yang menggambarkan pengalaman pengguna yang lebih objektif atau deskriptif tanpa adanya evaluasi positif atau negatif yang jelas.

Tabel 1. Perbandingan kinerja model dalam klasifikasi sentimen

Model	Akurasi	Precision	Recall	F1-Score
NN	95,50%	96,10%	94,80%	95,40%
SVM	95,40%	95,70%	94,20%	95,00%
DT	92,10%	91,80%	90,50%	91,10%

Tabel 1 menunjukan bahwa kinerja model. NN memberikan hasil terbaik dengan akurasi 95,5%, *precision* 96,1%, *recall* 94,8%, dan *f1-score* 95,4%. SVM sedikit lebih rendah dengan akurasi 95,4%, namun tetap sangat baik. Sementara itu, DT memiliki kinerja yang lebih rendah dengan akurasi 92,1%, *precision* 91,8%, *recall* 90,5%, dan *f1-score* 91,1%. Secara keseluruhan, model NN unggul, diikuti oleh SVM, sementara DT lebih cocok untuk aplikasi yang membutuhkan interpretabilitas yang lebih sederhana.

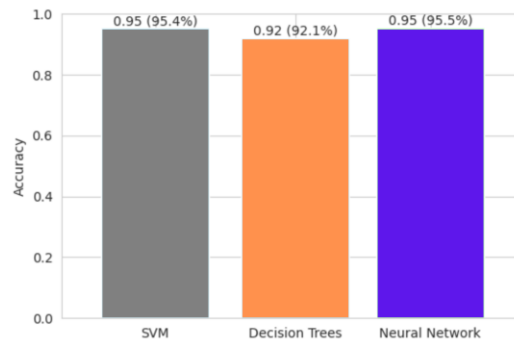
Fitur yang digunakan dalam pelatihan model dikodekan menggunakan TF-IDF untuk mengubah teks menjadi representasi numerik, dengan pengaturan n-gram (*unigram* dan *bigram*) untuk menangkap konteks kalimat. Setelah pembagian data menggunakan *stratified sampling*, tiga model klasifikasi dilatih menggunakan data latih, dan hasil evaluasi menunjukkan bahwa algoritma NN memberikan hasil terbaik dengan akurasi 95,5%, diikuti oleh SVM dengan 95,4% dan DT dengan 92,1%.



Gambar 4. Confusion matrix setiap model

Hasil analisis *confusion matrix* pada gambar 4 menunjukkan bahwa model NN memiliki tingkat kesalahan klasifikasi yang lebih rendah dibandingkan dengan model SVM dan DT. Namun, terdapat beberapa kesalahan yang signifikan, terutama dalam membedakan antara ulasan netral dan negatif. Model NN sering salah mengklasifikasikan ulasan netral sebagai ulasan negatif, yang disebabkan oleh kata-kata seperti "tidak berfungsi" atau "kurang membantu", yang meskipun bersifat netral dalam konteks kalimat, cenderung dipahami sebagai negatif oleh model. Kesalahan klasifikasi ini menunjukkan tantangan dalam klasifikasi sentimen bahasa Indonesia, yang memerlukan pemahaman lebih mendalam terhadap nuansa bahasa, seperti penggunaan kata-kata yang ambigu. Hasil ini juga memberikan Gambaran bagaimana model mengklasifikasikan ulasan sebagai positif, negatif, atau netral. Hasil ini memberikan wawasan yang lebih dalam mengenai bagaimana setiap model menangani kesalahan klasifikasi. Selain itu, gambar 4 mengilustrasikan bahwa model NN berhasil meminimalisir kesalahan pada kategori negatif dan netral, meskipun model ini masih sering

kesulitan dalam mengklasifikasikan ulasan yang memiliki ambiguitas atau ketidakjelasan dalam konteks.



Gambar 5. Hasil perbandingan algoritma

Selanjutnya hasil pada gambar 5 menunjukkan diagram perbandingan antara ketiga algoritma berdasarkan kinerja mereka. model NN unggul dengan skor akurasi tertinggi, diikuti oleh model SVM yang juga menunjukkan hasil yang sangat baik. Sementara itu, model DT menunjukkan kinerja yang sedikit lebih rendah, tetapi tetap memberikan kontribusi dalam klasifikasi sentimen dalam ulasan, dan memberikan gambaran yang jelas mengenai perbandingan kinerja model dalam hal akurasi dan *f1-score*, serta memberikan wawasan lebih jauh mengenai kekuatan dan kelemahan masing-masing model dalam menghadapi dataset ulasan pengguna.

Classification Report for SVM:					Classification Report for Decision Trees:				
	precision	recall	f1-score	support		precision	recall	f1-score	support
Negatif	0.932	0.893	0.912	309.000	Negatif	0.911	0.825	0.866	309.000
Netral	0.952	0.954	0.953	872.000	Netral	0.913	0.935	0.924	872.000
Positif	0.965	0.979	0.972	770.000	Positif	0.934	0.944	0.939	770.000
accuracy	0.954	0.954	0.954	0.954	accuracy	0.921	0.921	0.921	0.921
macro avg	0.950	0.942	0.946	1951.000	macro avg	0.919	0.901	0.910	1951.000
weighted avg	0.954	0.954	0.954	1951.000	weighted avg	0.921	0.921	0.921	1951.000

Classification Report for Neural Network:				
	precision	recall	f1-score	support
Negatif	0.930	0.903	0.916	309.000
Netral	0.953	0.954	0.954	872.000
Positif	0.967	0.977	0.972	770.000
accuracy	0.955	0.955	0.955	0.955
macro avg	0.950	0.945	0.947	1951.000
weighted avg	0.955	0.955	0.955	1951.000

Gambar 6. Matrix klasifikasi

Hasil matrix klasifikasi pada gambar 6 menunjukkan hasil evaluasi klasifikasi sentimen menggunakan SVM, DT, dan NN. Model NN memiliki performa terbaik secara keseluruhan dengan nilai *precision*, *recall*, dan *f1-score* tertinggi, terutama pada kelas positif (*f1-score* 0,972) dan akurasi keseluruhan 95,5%. Model SVM menyusul dengan hasil yang sangat mendekati, menunjukkan *f1-score* 0,972 untuk kelas positif dan akurasi total 95,4%. Sementara itu, DT menunjukkan performa paling rendah, khususnya pada kelas negatif dengan *f1-score* hanya 0,866 dan akurasi keseluruhan 92,1%.

Hasil penelitian menunjukkan bahwa NN adalah model paling efektif dalam menganalisis sentimen ulasan aplikasi *Microsoft Co-Pilot* dalam Bahasa Indonesia, dengan akurasi, *precision*, dan *f1-score* yang lebih tinggi dibandingkan model SVM dan DT. Keunggulan NN terletak pada kemampuannya untuk menangkap pola kompleks dalam teks bahasa Indonesia, yang sering mengandung *code-mixing* dan kata-kata tidak baku. Meskipun SVM menunjukkan kinerja hampir setara dengan NN, *recall*-nya sedikit lebih rendah. Meskipun lebih rendah dalam akurasi, DT tetap memiliki manfaat dalam klasifikasi yang lebih mudah dipahami, terutama dalam kasus yang lebih sederhana. Secara keseluruhan, penelitian ini memberikan wawasan penting bagi pengembang aplikasi untuk memahami sentimen pengguna. Model NN dapat digunakan untuk meningkatkan kualitas aplikasi berbasis AI dengan mengoptimalkan fitur berdasarkan *feedback* yang diperoleh dari analisis sentimen.

Pembahasan

Hasil penelitian ini menunjukkan bahwa tiga model klasifikasi sentiment NN, SVM, dan DT memiliki performa yang berbeda dalam mengklasifikasikan ulasan pengguna aplikasi *Microsoft Co-Pilot* berbahasa Indonesia. Dari ketiga model tersebut, NN menunjukkan kinerja terbaik dengan akurasi sebesar 95,5%, diikuti oleh SVM dengan 95,4%, dan DT dengan 92,1%. Model NN unggul secara konsisten pada seluruh kategori sentimen. Untuk kategori positif, model ini mencapai *precision* sebesar 0,967 dan *recall* sebesar 0,977, sementara untuk kategori negatif, *precision*-nya mencapai 0,930 dan *recall*-nya 0,903. Keunggulan ini menunjukkan bahwa NN mampu mengenali pola bahasa yang kompleks dan menangani variasi seperti *code-mixing* serta penggunaan kata-kata tidak baku yang lazim ditemukan dalam ulasan aplikasi.

Perbedaan performa ketiga model dapat dijelaskan melalui karakteristik masing-masing algoritma. SVM bekerja dengan mencari *hyperplane* terbaik untuk memisahkan kelas sentimen dan dikenal efektif dalam menangani data dengan dimensi tinggi (Sihombing & Yuliati, 2021). Namun, hasil penelitian ini menunjukkan bahwa NN memiliki keunggulan dibandingkan SVM, terutama dalam mengenali ulasan netral dan negatif. Hal ini disebabkan karena NN memiliki lebih banyak parameter yang dapat disesuaikan, sehingga lebih fleksibel dalam menangani hubungan *non-linear* antar kata dalam teks ulasan.

Sementara itu, DT memiliki akurasi yang lebih rendah dibandingkan dua model lainnya, terutama dalam mengenali ulasan negatif. Berdasarkan *confusion matrix*, model ini cenderung salah mengklasifikasikan sentimen negatif sebagai netral, yang menunjukkan bahwa DT kurang mampu menangkap pola sentimen dengan baik dalam teks yang lebih kompleks (Singgalen, 2022). Temuan ini sejalan dengan hasil penelitian sebelumnya yang menyatakan bahwa ulasan pengguna sering kali mencerminkan aspek pengalaman langsung mereka dengan aplikasi, baik dalam hal performa maupun kegunaan (Rahayu et al., 2022).

Penelitian sebelumnya telah memperlihatkan pergeseran performa antar algoritma. Penelitian Setiawan & Suryono (2024) telah menggunakan SVM dan *naïve bayes* dalam menganalisis sentimen terhadap kebijakan Ibu Kota Nusantara menemukan bahwa SVM unggul dengan akurasi 84%, sementara *Naïve Bayes* hanya 77%. Namun, dalam penelitian ini, SVM masih kalah dibandingkan NN, yang menunjukkan bahwa model berbasis jaringan saraf lebih mampu menangani keragaman pola bahasa pada ulasan aplikasi. Selain itu, penelitian Kurniawan & Waluyo (2022) mengenai ulasan aplikasi transportasi *online* juga menunjukkan bahwa DT memiliki interpretabilitas tinggi, namun kurang efektif dalam menangani bahasa tidak baku. Temuan tersebut sejalan dengan hasil penelitian ini, di mana DT kesulitan dalam mengklasifikasikan ulasan yang mengandung bahasa informal atau slang.

Sementara itu, penelitian kami memberikan kontribusi dalam pengembangan teknologi analisis sentimen. Perbedaan utama terletak pada penerapan model klasifikasi di bahasa Indonesia, yang memerlukan penanganan lebih lanjut terhadap variasi bahasa tidak baku dan

code-mixing, yang lebih sering ditemukan dalam ulasan Bahasa Indonesia dibandingkan dengan Bahasa Inggris.

Hasil penelitian kami ini membandingkan model yang lebih canggih dengan menerapkan metode NN dan SVM, serta mampu menangani variasi bahasa yang lebih kompleks, mengisi kekurangan pada penelitian sebelumnya yang lebih berfokus pada bahasa Inggris dan teks formal. Sebagian besar penelitian sebelumnya lebih banyak berfokus pada analisis dalam bahasa Inggris, sementara penelitian ini menunjukkan bahwa *machine learning* dapat diterapkan secara efektif untuk klasifikasi sentimen dalam Bahasa Indonesia, meskipun terdapat tantangan seperti variasi kata tidak baku dan campuran bahasa (*code-mixing*). Dengan meningkatnya penggunaan AI dalam interaksi manusia, penelitian ini dapat menjadi dasar bagi pengembangan model analisis sentimen yang lebih canggih untuk aplikasi lainnya, seperti asisten *virtual*, *chatbot* layanan pelanggan, dan sistem rekomendasi berbasis opini pengguna.

SIMPULAN

Penelitian ini membandingkan tiga model klasifikasi sentiment NN, SVM, dan DT untuk menganalisis sentimen ulasan pengguna aplikasi *Microsoft Co-Pilot* dalam bahasa Indonesia. NN menunjukkan akurasi terbaik 95,5%, meskipun memerlukan waktu komputasi lebih lama dibandingkan SVM (95,4%) dan DT (92,1%). Meskipun NN sering mengklasifikasikan ulasan netral sebagai negatif, model ini lebih efektif dalam mengurangi kesalahan klasifikasi dibandingkan yang lain. Kontribusi utama penelitian ini adalah penerapan analisis sentimen dalam bahasa Indonesia, yang penuh tantangan, memberikan panduan bagi pengembang aplikasi AI untuk lebih memahami umpan balik pengguna. Implikasi penelitian ini menunjukkan bahwa meskipun NN memberikan hasil terbaik, pemilihan model perlu mempertimbangkan waktu komputasi dan sumber daya.

REFERENSI

- Budianto, A. G., Rusilawati, R., Suryo, A. T. E., Cahyono, G. R., Zulkarnain, A. F., & Martunus, M. (2024). Perbandingan Performa Algoritma Support Vector Machine (SVM) dan Logistic Regression untuk Analisis Sentimen Pengguna Aplikasi Retail di Android. *Jurnal Sains Dan Informatika*, 10(2), 1-10. <https://doi.org/10.34128/jsi.v10i2.911>
- Dewi, F. K. S. (2021). Klasifikasi Berita Menggunakan Metode Multinomial Naïve Bayes. *Scan: Jurnal Teknologi Informasi dan Komunikasi*, 16(3), 1-8. <https://doi.org/10.33005/scan.v16i3.2870>
- Hendrawan, R., Pratama, D., & Sari, A. (2022). Preprocessing Data untuk Analisis Sentimen Bahasa Indonesia: Studi Kasus pada Media Sosial. *Jurnal Informatika*, 8(2), 99–112.
- Junianto, H., Saputro, R. E., Kusuma, B. A., & Saputra, D. I. S. (2024). Comparison Of Logistic Regression and Random Forest in Sentiment Analysis of Disdukcapil Application Reviews. *Jurnal Teknik Informatika (Jutif)*, 5(6), 1539-1547.
- Kurniawan, A., & Waluyo, R. (2022). Klasifikasi Sentimen Ulasan Aplikasi Transportasi Online Menggunakan Decision Tree dan Naïve Bayes. *Jurnal Teknologi Informasi Dan Komputer*, 10(3), 233–247.
- Manueke, E. G. Y., Lumando, L., Mangkey, L., & Manueke, G.W. (2023). Sentiment Analysis of Lyrics on Recycled and Recent Songs using Natural Language Processing Technology. *Proceedings of International Seminar on Innovation Challenges Multidisciplinary Research for Sustainable Development Goals*, 47-57. Maharashtra, India: Novateur Publication.
- Nurian, A. (2023). Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naïve Bayes. *Jurnal Informatika Dan Teknik Elektro Terapan*, 11(3), 829-835. <https://doi.org/10.23960/jitet.v11i3s1.3348>

- Pamungkas, A. A., Alam, C. N., Atmadja, A. R., & Juliansyah, R. (2024). Integrasi Kamus Multibahasa pada Feed Forward Neural Network dan IndoBERT dalam Pengembangan Chatbot Mobile. *Edumatic: Jurnal Pendidikan Informatika*, 8(2), 635-644. <https://doi.org/10.29408/edumatic.v8i2.27886>
- Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135. <https://doi.org/10.1561/15000000011>
- Rahayu, N. A., Wibowo, H., & Nugroho, A. (2022). Analisis Sentimen Pengguna Marketplace dengan TF-IDF dan Support Vector Machine. *Jurnal Ilmu Komputer Dan Sistem Informasi*, 14(1), 87-102.
- Ramadani, N. C., Tahyudin, I., & Barkah, A. S. (2024). Perbandingan Algoritma Support Vector Machine, Decision Tree, dan Logistic Regression pada Analisis Sentimen Ulasan Aplikasi Netflix. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 10(2), 110-117. <https://doi.org/10.25077/TEKNOSI.v10i2.2024.110-117>
- Ramdhana, A. C., & Pratiwi, N. (2023). Perbandingan Kinerja Model Convolutional Neural Network pada Klasifikasi Kanker Kulit. *Edumatic: Jurnal Pendidikan Informatika*, 7(2), 197-206. <https://doi.org/10.29408/edumatic.v7i2.19823>
- Sabrani, A., Wedashwara W., I. G. W., & Bimantoro, F. (2020). Multinomial Naïve Bayes untuk Klasifikasi Artikel Online tentang Gempa di Indonesia. *Jurnal Teknologi Informasi, Komputer, Dan Aplikasinya (JTika)*, 2(1), 89-100. <https://doi.org/10.29303/jtika.v2i1.87>
- Saputra, F. A., & Noor Hasan, M. (2023). Implementasi Neural Network dalam Analisis Sentimen Twitter Berbahasa Indonesia. *Jurnal Kecerdasan Buatan*, 6(2), 123-137.
- Setiawan, A., & Suryono. (2024). Analisis Sentimen Terhadap Kebijakan Ibu Kota Nusantara Menggunakan Machine Learning. *Jurnal Sistem Informasi*, 12(1), 45-58.
- Sihombing, T., & Yuliati, N. (2021). Support Vector Machine untuk Klasifikasi Opini Publik terhadap Layanan Digital Banking. *Jurnal Data Mining Indonesia*, 9(1), 78-92.
- Singgalen, F. (2022). Penerapan Decision Tree dalam Analisis Sentimen Produk Digital. *Jurnal Teknologi Dan Inovasi Digital*, 5(2), 45-61. <https://doi.org/10.30865/json.v5i2.7081>
- Soemedhy, C. A. A., & others. (2023). Analisis Komparasi Algoritma Machine Learning untuk Sentiment Analysis (Studi Kasus: Komentar YouTube 'Kekerasan Seksual'). *Jurnal Teknologi Informasi Dan Komunikasi*, 11(2), 35-45. <https://doi.org/10.30591/jpit.v7i2.3547>
- Triginandri, R., & Subhiyakto, E. R. (2024). Deteksi Dini Cacar Monyet menggunakan Convolutional Neural Network (CNN) dalam Aplikasi Mobile. *Edumatic: Jurnal Pendidikan Informatika*, 8(2), 516-525. <https://doi.org/10.29408/edumatic.v8i2.27625>
- Yolanda, R., Firmansyah, A., & Widodo, S. (2022). Perbandingan Kinerja Support Vector Machine dan Naïve Bayes dalam Analisis Sentimen pada E-Commerce. *Jurnal Sains Data Dan Informatika*, 7(3), 89-104.
- Yuliana, D., Purwanto, & Supriyanto, C. (2019). Klasifikasi Teks Pengaduan Masyarakat Dengan Menggunakan Algoritma Neural Network. *Jurnal KomtekInfo*, 5(3), 92-116. <https://doi.org/10.35134/komtekinfo.v5i3.35>
- Zamsuri, R., Putra, A. F., & Wijaya, S. (2023). Decision Tree Berbasis SMOTE dalam Analisis Sentimen Masyarakat terhadap Vaksin COVID-19. *Jurnal Riset Manajemen Dan Informatika Komputer*, 10(1), 1-10.