

ANALISIS SENTIMEN PADA NAMA DAN LOGO WHOOSH DI MEDIA SOSIAL

Shuhaib Yahya Banamah¹⁾, Dika Jatnika²⁾

Bachelor of Digital Business Study Program, Universitas Padjadjaran / Management and Business, Universitas Padjadjaran
shuhaib20001@mail.unpad.ac.id

Abstract (English)

The name and logo of "Whoosh" are part of a branding effort for a newly launched high-speed train service in Indonesia. With increasing public attention towards this visual identity, this study aims to analyze public sentiment regarding the Whoosh name and logo on social media using sentiment classification algorithms. The algorithms tested include Support Vector Machine (SVM) and K-Nearest Neighbor (KNN). The evaluation results show that SVM performs better than KNN, with F1-scores of 0.92 for the negative class, 0.83 for the neutral class, and 0.40 for the positive class, while KNN achieved F1-scores of 0.80, 0.55, and 0.33 for the same classes. Sentiment analysis on X (Twitter) reveals a predominance of negative sentiment, with words like "jelek" and "logo" frequently appearing in comments. In contrast, on Instagram, neutral sentiment is more dominant, with comments often mentioning words like "admin" and "tiket", while negative sentiment is also present but less prevalent compared to Twitter. This study recommends the use of additional data augmentation methods, integration of data from various sources, and exploration of other classification models to enhance sentiment classification accuracy in the future.

Article History

Submitted: 18 Juli 2024

Accepted: 24 Juli 2024

Published: 25 Juli 2024

Key Words

Sentiment Analysis, Support Vector Machine (SVM), K-Nearest-Neighbor (KNN), X (Twitter), Instagram

Abstrak (Indonesia)

Nama dan logo "Whoosh" adalah bagian dari upaya branding untuk layanan kereta api cepat yang baru diluncurkan di Indonesia. Dengan semakin meningkatnya perhatian publik terhadap identitas visual ini, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap nama dan logo Whoosh di media sosial menggunakan algoritma klasifikasi sentimen. Algoritma yang diuji meliputi Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN). Hasil evaluasi menunjukkan bahwa SVM memiliki kinerja yang lebih baik dibandingkan KNN, dengan F1-score masing-masing sebesar 0.92 untuk kelas negatif, 0.83 untuk kelas netral, dan 0.40 untuk kelas positif, sedangkan KNN memperoleh F1-score 0.80, 0.55, dan 0.33 untuk kelas yang sama. Analisis sentimen di X (Twitter) menunjukkan dominasi sentimen negatif, dengan kata-kata seperti "jelek" dan "logo" sering muncul dalam komentar. Sebaliknya, di Instagram, sentimen netral lebih dominan, dengan komentar yang sering menyebut kata-kata seperti "admin" dan "tiket", sementara sentimen negatif juga terdapat meskipun jumlahnya lebih sedikit dibandingkan di Twitter.

Sejarah Artikel

Submitted: 18 Juli 2024

Accepted: 24 Juli 2024

Published: 25 Juli 2024

Kata Kunci

Analisis Sentimen, Support Vector Machine (SVM), K-Nearest-Neighbor (KNN), X (Twitter), Instagram

Pendahuluan

Transportasi memegang peranan penting dalam kehidupan sehari-hari, pemerintahan, dan masyarakat (Putra & Adeswastoto, 2018). Di Indonesia, transportasi umum meliputi berbagai moda seperti taksi, ojek, kereta, pesawat, dan kapal. Perkembangan terbaru mencakup pemesanan moda transportasi yang kini dapat dilakukan melalui aplikasi, serta peningkatan jumlah

penumpang kereta api sebesar 163.20% pada Agustus 2023 dibandingkan Januari 2021 (Badan Pusat Statistik, 2023).

Salah satu proyek transportasi yang signifikan adalah Kereta Cepat Jakarta-Bandung. Pada 16 September 2023, PT KCIC mengumumkan masa uji coba yang berlangsung hingga 15 Oktober 2023, menarik lebih dari 15.000 penumpang (Intan, 2023). Pada 22 September 2023, pemerintah mengumumkan nama resmi kereta cepat ini sebagai "Whoosh," akronim dari 'Waktu Hemat, Operasi Optimal, Sistem Handal' (Biro Komunikasi dan Informasi Publik, 2023). Nama dan logo "Whoosh" dirancang untuk mencerminkan perubahan dan pertumbuhan yang dibawa oleh kehadiran kereta cepat dalam berbagai aspek kehidupan masyarakat.

Pengumuman mengenai nama dan logo "Whoosh" dipublikasikan melalui media sosial, seperti akun X (Twitter) dan Instagram @KeretaCepatID (PT KCIC, 2023b) (PT KCIC, 2023a). Nama ini diambil untuk mewakili suara lesatan kereta cepat, dengan identitas visual yang dirancang agar mudah dipahami publik. Meskipun konsep tersebut bertujuan untuk mencerminkan dampak positif dari kereta cepat, tanggapan masyarakat di media sosial menunjukkan beragam reaksi, termasuk banyak komentar negatif terkait nama dan desain logonya. Hal ini menekankan pentingnya strategi komunikasi dan branding dalam mempengaruhi persepsi publik. Oleh karena itu, dibutuhkan dilakukannya analisis sentimen pada topik ini untuk memahami lebih dalam reaksi masyarakat.

Selain itu, respons negatif yang diterima di media sosial menunjukkan bahwa meskipun sebuah produk memiliki manfaat yang jelas, penerimaan masyarakat terhadap elemen branding seperti nama dan logo sangat krusial. Reaksi ini memberikan pelajaran penting bagi perusahaan dan pemerintah dalam mengelola peluncuran produk dan layanan publik, terutama dalam era digital dimana opini publik dapat dengan cepat tersebar luas. Evaluasi lebih mendalam terhadap tanggapan masyarakat bisa menjadi bahan pertimbangan untuk strategi komunikasi yang lebih efektif di masa mendatang.

Dalam konteks ini, analisis sentimen menjadi relevan. Analisis sentimen merupakan proses ekstraksi dan analisa terhadap pandangan, perasaan, sikap, persepsi, dan sebagainya dari individu terhadap berbagai hal seperti topik, produk, serta layanan (Birjali et al., 2021). Tugas analisis sentimen dapat dilihat sebagai suatu permasalahan klasifikasi teks, di mana langkah-langkah yang terlibat pada akhirnya memutuskan apakah teks yang diberikan mencerminkan sentimen positif, negatif, atau netral. Namun, meskipun mungkin tampak sederhana, analisis sentimen sesungguhnya melibatkan banyak subtugas dalam pemrosesan bahasa alami, termasuk deteksi sindiran dan subjektivitas. Selain itu, teks tidak selalu tersusun dengan rapi seperti buku atau surat kabar, dan seringkali berisi kesalahan ejaan, ungkapan idiomatik, atau singkatan (Birjali et al., 2021). Salah satu pendekatan yang digunakan dalam analisis sentimen adalah pendekatan machine learning.

Machine learning adalah suatu bentuk model komputasi yang menggunakan data yang telah dipelajari untuk meningkatkan kinerja atau membuat prediksi. Konsep machine learning melibatkan penggunaan algoritma prediksi yang akurat dan efisien (Mohri et al., 2018). Dalam

konteks analisis sentimen teks, algoritma machine learning digunakan untuk membangun model statistik berdasarkan data yang telah dipelajari, sehingga dapat memprediksi sentimen dalam teks (van Atteveldt et al., 2021). Dua algoritma machine learning yang umum digunakan untuk melakukan analisis sentimen yaitu Support Vector Machine dan K-Nearest Neighbor.

Support Vector Machine atau yang biasa disingkat SVM merupakan salah satu algoritma yang biasa digunakan dalam analisis sentimen. SVM berusaha untuk membuat hyperplane terbaik yang dapat memisahkan dua kelas. Hyperplane sendiri adalah batas keputusan yang memaksimalkan margin antara dua kelas. Salah satu keuntungan penggunaan SVM adalah tahan terhadap overfitting dengan menggunakan margin maksimum tersebut (James et al., 2017).

K-Nearest Neighbor adalah salah satu algoritma yang umum digunakan untuk klasifikasi, secara khususnya pada analisis sentimen. Algoritma yang umumnya disingkat menjadi KNN ini efektif untuk pengelompokan data. Dalam proses implementasi KNN, menentukan nilai k diperlukan. Nilai k terbaik untuk algoritma ini tergantung pada data yang dimiliki. Dan salah satu cara penentuan nilai k yang baik adalah dengan cross validation (Wisnu et al., 2020). Untuk membuat prediksi untuk data baru, algoritma akan mengidentifikasi titik-titik data yang paling dekat (Müller & Guido, 2016).

SVM dan KNN beberapa kali telah digunakan untuk pembuatan model analisis sentimen dan memberikan hasil evaluasi yang cukup baik. Penelitian yang menunjukkan baiknya performa SVM dilakukan oleh (Muhammadi et al., 2022), (Styawati et al., 2021), dan (Ramasamy et al., 2021). Dan penelitian yang menunjukkan baiknya performa KNN dalam melakukan klasifikasi analisis sentimen dilakukan oleh (Wisnu et al., 2020), (Damarta et al., 2021), (Isnain et al., 2021).

Dalam penelitian ini, kinerja klasifikasi analisis sentimen dari algoritma Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN) akan dievaluasi dan dibandingkan untuk memilih algoritma yang paling tepat dalam analisis sentimen ini. Selain itu, penelitian ini juga bertujuan untuk mengidentifikasi sentimen masyarakat terhadap nama dan logo "Whoosh" di media sosial, khususnya di X (Twitter) dan Instagram.

Metode Penelitian

1. Pengumpulan Data

Data akan dikumpulkan dari balasan (reply) pada unggahan tweet akun resmi PT KCIC dengan username @KeretaCepatID terkait pengumuman nama "Whoosh" menggunakan Tweet Harvest. Alat ini mengekstraksi tweet berdasarkan kata kunci tertentu atau reply dari tweet spesifik (Satria, 2023). Selain itu, komentar dari unggahan Instagram terkait pengumuman yang sama pada akun @keretacepat_id akan dikumpulkan secara manual dengan menyalin setiap komentar yang relevan.

2. Pelabelan Data

Data yang telah dikumpulkan akan diberi label secara manual menjadi tiga kategori: Negatif, Netral, atau Positif. Proses pelabelan ini akan difokuskan pada data dari X (Twitter).

3. Data Preprocessing

- ◆ Data yang telah dikumpulkan akan melalui beberapa langkah preprocessing agar siap diolah. Langkah-langkah ini meliputi:
- ◆ *Data Cleansing*: Menghapus karakter-karakter tidak diperlukan seperti tag, link, tanda baca, dan emoji.
 - Case Folding*: Mengubah huruf kapital menjadi huruf kecil untuk menyamakan format penulisan.
 - Normalisasi: Menyeragamkan penulisan kata-kata yang memiliki makna sama.
 - Stopword Removal*: Menghapus kata-kata yang tidak memberikan informasi penting.
4. Pembagian Data dan Pembobotan Kata
- Data akan dibagi menjadi 70% untuk pelatihan dan 30% untuk pengujian. Pembobotan kata dilakukan menggunakan TF-IDF. Proses ini melibatkan evaluasi nilai TF (Term-Frequency) dan IDF (Inverse Document Frequency) untuk memberikan bobot pada kata-kata yang digunakan dalam proses pelatihan dengan algoritma SVM dan KNN (Muhammadi et al., 2022).
5. *Data Augmentation*
- Untuk mengatasi ketidakseimbangan data, metode Synthetic Minority Over-sampling Technique (SMOTE) akan digunakan. SMOTE meningkatkan jumlah contoh dari kelas minoritas dengan mensintesis data baru, membantu model belajar dari pola yang lebih bervariasi dan kompleks dalam kelas minoritas (Brandt & Lanzén, 2020).
6. Penerapan Support Vector Machine dan K-Nearest Neighbor
- Algoritma Support Vector Machine dan K-Nearest Neighbor akan diterapkan menggunakan library scikit-learn. Library ini menyediakan berbagai algoritma machine learning dan fungsi untuk evaluasi model.
7. Evaluasi
- Model akan dievaluasi menggunakan confusion matrix, yang membandingkan hasil prediksi model dengan hasil prediksi sebenarnya. Dari confusion matrix, precision, recall, dan f-1 score untuk masing-masing sentimen (Negatif, Positif, Netral) akan dihitung untuk menilai kinerja model.
8. Prediksi Sentimen Data Instagram
- Setelah model terbaik dipilih berdasarkan hasil evaluasi, model tersebut akan digunakan untuk memprediksi sentimen data dari Instagram. Data komentar dari Instagram akan melalui tahapan preprocessing dan pembobotan kata sebelum diprediksi.

Hasil dan Pembahasan

1. Pengumpulan Data

Proses pengumpulan data dimulai dengan memperoleh data dari platform X (Twitter) dan Instagram. Data dari Twitter diambil menggunakan alat Tweet Harvest, yang mampu menarik antara 100 hingga 150 balasan dalam satu kali penarikan. Dengan melakukan 40 penarikan data, diperoleh 40 file CSV yang kemudian digabungkan menjadi satu dataset. Setelah proses penggabungan, dilakukan pembersihan data untuk menghapus duplikasi

serta data yang mengandung sarkasme. Hasil akhir dari data Twitter adalah 228 balasan yang bersih dan siap untuk analisis lebih lanjut. Di sisi lain, data dari Instagram dikumpulkan secara manual. Selama pengumpulan, data yang mengandung sarkasme disaring dan dihapus. Proses ini menghasilkan 423 data yang akan digunakan dalam penelitian.

2. Pelabelan Data

Pelabelan data dilakukan secara manual untuk data Twitter, menghasilkan 137 sentimen negatif, 79 netral, dan 11 positif. Data Instagram tidak dilabeli secara manual, melainkan akan prediksi menggunakan model yang dikembangkan.

3. Data Preprocessing

- Data Cleansing**, pembersihan data dilakukan menggunakan Google Colab dengan bantuan library regex. Tahap pertama adalah menghapus tag dan tautan yang terdapat dalam data, yang sering kali mengganggu proses analisis. Selanjutnya, emoji dan tanda baca juga dihapus menggunakan regex. Proses ini memastikan bahwa data yang tersisa hanya berisi teks yang relevan, memudahkan analisis lebih lanjut.
- Case Folding**, untuk menghindari perbedaan yang disebabkan oleh variasi kapitalisasi huruf, semua huruf dalam teks diubah menjadi huruf kecil menggunakan metode `.lower()` dalam Python. Ini penting untuk memastikan konsistensi dalam analisis.
- Normalisasi**, normalisasi dilakukan dengan menggunakan kamus yang dibuat secara manual, berisi 488 entri yang mencakup variasi penulisan tidak standar dalam teks. Proses ini bertujuan untuk menggantikan variasi penulisan dengan bentuk yang lebih konsisten dan standar, meningkatkan kualitas data untuk analisis.
- Stopword Removal**, penghapusan kata-kata umum yang tidak memberikan makna signifikan dilakukan dengan menggunakan kamus stopwords dari GitHub repository `irfandythalib/python-indonesia-stopwords-remover` yang diambil dari *Stoplist* dari penelitian (Tala, 2003). Penambahan beberapa kata khusus juga diperlukan untuk melengkapi kamus tersebut. Proses ini dimulai dengan memuat kamus stopwords kemudian kata-kata yang terdaftar dalam kamus stopwords dihapus untuk memastikan teks yang tersisa berisi kata-kata yang penting.

Tabel 1. Contoh Data Sebelum dan Setelah Data Preprocessing

Sebelum Data Preprocessing	Setelah Data Preprocessing
@KeretaCepatID mau tiket uji coba min	tiket uji coba admin
@KeretaCepatID Logonya lebih cocok buat raket sama sepatu badminton min	logonya cocok raket sepatu badminton admin
@KeretaCepatID kek logo sepatu 50kan	logo sepatu
@KeretaCepatID kerennn kerennn nama 'Whoosh' dan logo nya menggambarkan semangatt kemajuann mantap	keren keren nama whoosh logo menggambarkan semangat kemajuan keren
@KeretaCepatID logonyaa sukaaa bangettt minnn ditambah namanya mudah diingatt	logonya suka banget admin ditambah namanya mudah

4. Pembagian Data dan Pembobotan Kata

- a. Pembagian data dilakukan sebelum proses pembobotan kata seperti TF-IDF untuk mencegah kebocoran data. Teknik stratified sampling diterapkan dengan proporsi 30% untuk data testing dan 70% untuk data training, memastikan distribusi kategori dalam data testing dan training sesuai dengan distribusi keseluruhan dataset.
- b. Teknik TF-IDF digunakan untuk membobot kata dengan bantuan library scikit-learn. TF-IDF menghitung seberapa penting setiap kata dalam teks data training dan mengubahnya menjadi matriks fitur. Proses ini dilakukan pada data training untuk mendapatkan matriks fitur, yang kemudian diterapkan pada data testing tanpa menghitung ulang bobot, melainkan menggunakan bobot yang sudah dihitung dari data training.

Tabel 2. Contoh Hasil Pembobotan Kata

argo
0
0
0
0,298798
0
0
0,202184

5. *Data Augmentation*

Dalam tahap ini, teknik SMOTE (Synthetic Minority Over-sampling Technique) diterapkan untuk mengatasi ketidakseimbangan kelas dalam data training. Sebelum SMOTE, distribusi data menunjukkan ketidakseimbangan yang signifikan dengan 95 data untuk kelas negatif, 55 untuk kelas netral, dan 8 untuk kelas positif. SMOTE digunakan untuk menambah jumlah data pada kelas yang kurang terwakili dengan menciptakan data sintesis baru berdasarkan pola yang ada dalam data. Setelah penerapan SMOTE, distribusi data menjadi seimbang dengan 95 data untuk setiap kelas, membantu model dalam belajar dari data yang lebih seimbang

6. Penerapan Support Vector Machine dan K-Nearest Neighbor

- a. Support Vector Machine (SVM). Implementasi SVM dimulai dengan pencarian hyperparameter terbaik menggunakan GridSearchCV dari library scikit-learn. GridSearchCV mencari hyperparameter terbaik seperti C, gamma, dan kernel untuk mendapatkan performa optimal. Hyperparameter C mengontrol trade-off antara margin dan kesalahan klasifikasi, gamma mengatur pengaruh satu data point terhadap data point lainnya, dan kernel menentukan cara data diubah ke dalam ruang fitur yang lebih tinggi. Setelah menemukan hyperparameter terbaik (C=1, kernel='linear'), model SVM dilatih ulang dengan data training yang telah ditingkatkan dengan SMOTE.

- b. K-Nearest Neighbor (KNN). Implementasi KNN juga dimulai dengan pencarian hyperparameter optimal menggunakan GridSearchCV. Parameter yang diuji meliputi jumlah tetangga (`n_neighbors`), jenis bobot (`weights`), dan metrik jarak (`metric`). Hasil terbaik ditemukan dengan hyperparameter (`metric='euclidean'`, `n_neighbors=2`, `weights='distance'`). Data training kemudian diresampling menggunakan SMOTE, dan model KNN dilatih kembali dengan hyperparameter terbaik untuk memastikan kinerja optimal pada data testing.
7. Evaluasi
- a. Evaluasi SVM, model SVM menunjukkan performa yang bervariasi untuk setiap kelas. Untuk kelas negatif, F1-score mencapai 0.92 dengan precision 0.91 dan recall 0.93. Kelas netral memiliki F1-score 0.83 dengan precision dan recall masing-masing 0.83. Namun, untuk kelas positif, F1-score hanya 0.40 dengan precision dan recall yang lebih rendah. Confusion matrix mengonfirmasi bahwa SVM lebih akurat dalam mengklasifikasikan kelas negatif dan netral dibandingkan dengan KNN.

Tabel 3. Precision, Recall, dan F-1 Score dari Model SVM

Kelas	Precision	Recall	F-1 Score
Negatif	0.91	0.93	0.92
Netral	0.83	0.83	0.83
Positif	0.50	0.33	0.40

- b. Evaluasi KNN, kinerja KNN menunjukkan F1-score 0.80 untuk kelas negatif, dengan precision 0.74 dan recall 0.88. Untuk kelas netral, F1-score mencapai 0.55 dengan precision 0.69 dan recall 0.46. Kelas positif menunjukkan F1-score terendah yaitu 0.33 dengan precision dan recall masing-masing 0.33. Confusion matrix menunjukkan bahwa KNN memiliki kesulitan yang lebih besar dalam mengklasifikasikan kelas positif dan netral dibandingkan SVM.

Tabel 4. Precision, Recall, dan F-1 Score dari Model KNN

Kelas	Precision	Recall	F-1 Score
Negatif	0.74	0.88	0.80
Netral	0.69	0.46	0.55
Positif	0.33	0.33	0.33

Berdasarkan hasil evaluasi, SVM adalah model yang lebih baik dibandingkan KNN untuk klasifikasi data. SVM memiliki performa yang lebih tinggi dalam mengklasifikasikan kelas negatif dan netral, serta sedikit lebih baik dalam mendeteksi kelas positif. Oleh karena itu, SVM dipilih sebagai model terbaik untuk memprediksi sentimen dari data Instagram.

8. Prediksi Sentimen Data Instagram

Setelah data Instagram dibaca, teks melalui proses pembersihan yang sama seperti pada data Twitter. Teks dibersihkan dari tag, tautan, dan karakter non-huruf, kemudian diubah

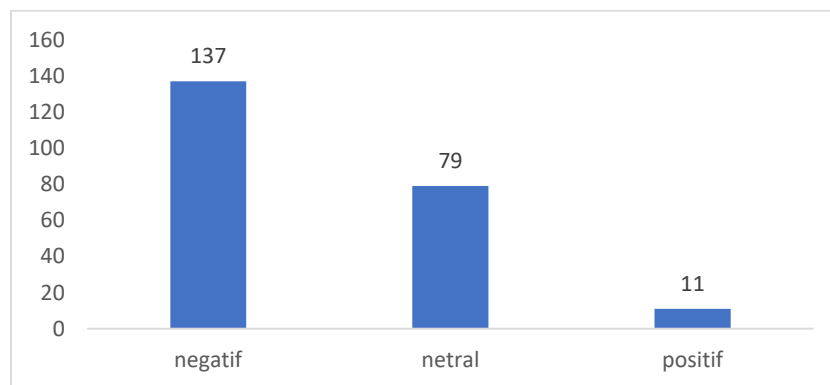
menjadi huruf kecil, dinormalisasi, dan dihapus stopwords-nya. TF-IDF diterapkan pada data yang telah dibersihkan, dan model SVM digunakan untuk memprediksi sentimen. Namun, sekitar 21.3% dari total 423 entri diprediksi salah oleh model, menunjukkan bahwa ada kebutuhan untuk peningkatan dalam prediksi sentimen.

Tabel 5. Contoh Prediksi Sentimen yang Salah

Text	Predicted Sentiment	Actual Sentiment
@ahasyari.inet setuju. Jaman skrg nama2 kereta bagus yg simple dan menyiratkan kecepatan. Whoosh = cepat	Negatif	Positif
@mandafi.kos hooh mentang2 cepet aja di namain whoosh 😏 apaan anying gaada nilai dan maknanya banget	Netral	Negatif
Anjirrrr beneraan whooshh, gantiiii weiiii gw ikut maluuuu 🤔	Netral	Negatif
kece min logo nya 🤔	Negatif	Positif

9. Sentimen Masyarakat di X (Twitter) dan di Instagram

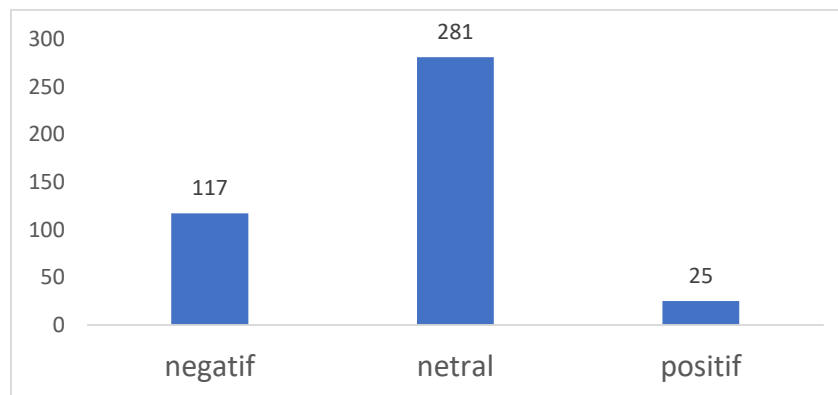
a. Sentimen di X (Twitter), dari 227 balasan, 137 memiliki sentimen negatif, 79 netral, dan 11 positif. Analisis kata-kata negatif mengungkapkan ketidakpuasan terhadap desain logo dan nama baru Whoosh, dengan kata-kata seperti "jelek", "logo", "norak", "merek", "argo", "sepatu", "badminton", "olahraga", dan "nama" sering muncul. Untuk sentimen netral, ketertarikan lebih pada informasi tiket dengan kata-kata seperti "admin", "tiket", "turun", "coba", "kereta", "bandung", "jakarta", dan "whoosh". Sementara itu, sentimen positif diwakili oleh kata-kata seperti "keren", "banget", "oke", "lucu", "modern", "logonya", dan "namanya".



Gambar 1. Jumlah Data X (Twitter) per Sentimen

b. Sentimen di Instagram, dari 423 komentar, 117 memiliki sentimen negatif, 281 netral, dan 25 positif. Analisis kata-kata negatif di Instagram menunjukkan ketidakpuasan serupa dengan di Twitter, dengan kata-kata seperti "merek", "olahraga", "argo", "jelek", dan "sepatu" sering muncul. Untuk sentimen netral, kata-kata seperti "admin", "tiket",

"whoosh", "beli", dan "gratis" mendominasi, mencerminkan ketertarikan pada informasi praktis terkait perjalanan dengan Whoosh. Sentimen positif di Instagram menunjukkan bahwa beberapa pengguna menghargai desain dan nama baru dengan kata-kata seperti "keren".



Gambar 2. Jumlah Data Instagram per Sentimen (Hasil Prediksi)

Kesimpulan

Dari hasil analisis dan evaluasi yang telah dilakukan, dapat disimpulkan bahwa algoritma Support Vector Machine (SVM) menunjukkan performa yang lebih unggul dibandingkan dengan K-Nearest Neighbor (KNN) dalam klasifikasi sentimen. SVM mampu mengklasifikasikan data dengan akurasi yang tinggi, terutama untuk kelas negatif dan netral, dengan F1-score masing-masing 0.92 dan 0.83. Meskipun performanya pada kelas positif masih kurang memuaskan dengan F1-score 0.40, SVM tetap lebih efektif dibandingkan KNN, yang memiliki F1-score lebih rendah untuk semua kelas, terutama pada kelas positif dengan F1-score 0.33. Sementara itu, analisis sentimen masyarakat di Twitter menunjukkan dominasi sentimen negatif terhadap nama dan logo "Whoosh", sedangkan di Instagram, meskipun sentimen negatif masih mendominasi, terdapat distribusi sentimen yang lebih seimbang dengan beberapa komentar positif yang mencerminkan respons yang lebih beragam.

Untuk pengembangan penelitian selanjutnya, disarankan agar mencoba metode data augmentation yang berbeda, seperti *back and forth translation*, dipertimbangkan untuk mengatasi masalah ketidakseimbangan data. Metode ini dapat memperkaya dataset dan meningkatkan kemampuan model dalam menggeneralisasi sentimen dari teks yang bervariasi. Selain itu, integrasi data dari berbagai sumber seperti Instagram dan Twitter dalam satu dataset training dapat meningkatkan generalisasi model terhadap berbagai jenis data dan platform, mengurangi ketidakseimbangan hasil klasifikasi yang mungkin timbul dari penggunaan data yang terpisah.

Terakhir, eksplorasi terhadap model dan teknik klasifikasi lainnya seperti Random Forest, Gradient Boosting, atau teknik deep learning dapat dilakukan untuk menilai apakah model-model ini dapat memberikan hasil yang lebih baik dalam klasifikasi sentimen. Menggabungkan beberapa model dalam teknik ensemble juga bisa dipertimbangkan untuk meningkatkan performa klasifikasi secara keseluruhan. Dengan pendekatan ini, diharapkan dapat diperoleh model yang lebih robust dan akurat dalam menganalisis sentimen di berbagai platform sosial media

Referensi

- Badan Pusat Statistik. (2023). *Badan Pusat Statistik*.
<https://www.bps.go.id/indicator/17/72/1/jumlah-penumpang-kereta-api.html>
- Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, 226, 107134. <https://doi.org/10.1016/j.knosys.2021.107134>
- Biro Komunikasi dan Informasi Publik. (2023). "Whoosh" Jadi Identitas Jenama Kereta Api Cepat Indonesia Kementerian Perhubungan Republik Indonesia. [https://dephub.go.id/post/read/"whoosh"-jadi-identitas-jenama-kereta-api-cepat-indonesia](https://dephub.go.id/post/read/)
- Brandt, J., & Lanzén, E. (2020). A Comparative Review of SMOTE and ADASYN in Imbalanced Data Classification. 202142, . <https://www.diva-portal.org/smash/record.jsf?pid=diva2:1519153>
- Damarta, R., Hidayat, A., & Abdullah, A. S. (2021). The application of k-nearest neighbors classifier for sentiment analysis of PT PLN (Persero) twitter account service quality. *Journal of Physics: Conference Series*, 1722(1). <https://doi.org/10.1088/1742-6596/1722/1/012002>
- Intan, G. (2023). *Menilik Antusiasme Masyarakat Menjajal Kereta Cepat Jakarta-Bandung*. <https://www.voaindonesia.com/a/menilik-antusiasme-masyarakat-menjajal-kereta-cepat-jakarta-bandung/7278369.html>
- Isnain, A. R., Supriyanto, J., & Kharisma, M. P. (2021). Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(2), 121. <https://doi.org/10.22146/ijccs.65176>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). An Introduction to Statistical Learning. In *Handbook of Quantile Regression*. <https://doi.org/10.1201/9781315120256>
- Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*. books.google.com. <https://books.google.com/books?hl=en&lr=&id=dWB9DwAAQBAJ&oi=fnd&pg=PR5&dq=machine+learning&ots=AytWTXp5o6&sig=GeNWDes13Lz7EHciDqsgzOunkAg>
- Muhammadi, R. H., Laksana, T. G., & Arifa, A. B. (2022). Combination of Support Vector Machine and Lexicon-Based Algorithm in Twitter Sentiment Analysis. *Khazanah Informatika: Jurnal Ilmu Komputer Dan Informatika*, 8(1), 59–71. <https://doi.org/10.23917/khif.v8i1.15213>
- Müller, A., & Guido, S. (2016). Introduction to Machine Learning with Python. In *Introduction to Machine Learning with Python*. <https://doi.org/10.2174/97898151244221230101>
- PT KCIC. (2023a). *Whoosh di Instagram: 🎵 Naik Kereta Cepat Whoosh Whoosh Whoosh 🎵 Siapa hendak turun? Ke Bandung dan Jakarta 🎵 Hari ini diresmikan jenama baru untuk Kereta tercepat se-Asia Tenggara dengan nama "Whoosh" Logo dan jenama ini terinspirasi dari efek riak dari*. https://www.instagram.com/p/Cx4QIHLppDM/?img_index=1

- PT KCIC. (2023b). *Whoosh di X: “🎵 Naik Kereta Api Whoosh Whoosh Whoosh 🎵🎵 Siapa hendak turun? Ke Bandung dan Jakarta 🎵 https://t.co/WB8r39EvDN” / X.* <https://twitter.com/KeretaCepatID/status/1708665234477740121>
- Putra, A. A., & Adeswastoto, H. (2018). Transportasi Publik Dan Aksesibilitas Masyarakat Perkotaan. *Jurnal Teknik Industri Terintegrasi*, 1(1), 55–60. <https://doi.org/10.31004/jutin.v1i1.312>
- Ramasamy, L. K., Kadry, S., Nam, Y., & Meqdad, M. N. (2021). Performance analysis of sentiments in Twitter dataset using SVM models. *International Journal of Electrical and Computer Engineering*, 11(3), 2275–2284. <https://doi.org/10.11591/ijece.v11i3.pp2275-2284>
- Satria, H. (2023). *Cara Crawl Data Twitter Thread atau Komentar.* <https://helmisatria.com/blog/cara-crawl-data-twitter-thread-komentar>
- Styawati, S., Isnain, A. R., Hendrastuty, N., & Andraini, L. (2021). Comparison of Support Vector Machine and Naïve Bayes on Twitter Data Sentiment Analysis. *Jurnal Informatika: Jurnal Pengembangan IT*, 6(1), 56–60. <https://doi.org/10.30591/jpit.v6i1.3245>
- Tala, F. Z. (2003). A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia. *M.Sc. Thesis, Appendix D*, pp, 39–46.
- van Atteveldt, W., van der Velden, M. A. C. G., & Boukes, M. (2021). The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms. *Communication Methods and Measures*, 15(2), 121–140. <https://doi.org/10.1080/19312458.2020.1869198>
- Wisnu, H., Afif, M., & Ruldevyani, Y. (2020). Sentiment analysis on customer satisfaction of digital payment in Indonesia: A comparative study using KNN and Naïve Bayes. *Journal of Physics: Conference Series*, 1444(1). <https://doi.org/10.1088/1742-6596/1444/1/012034>