

ANALISIS SENTIMEN MASYARAKAT TERHADAP TIGA BANDARA TERSIBUK DI INDONESIA BERBASIS *TRANSFORMER*

Nurmahendra Harahap¹, Donna N. M. Sirait², Iin Verawati², Andre Togi Sinurat²

¹ Universitas Tjut Nyak Dhien

² Politeknik Penerbangan Medan

E-mail correspondence : nurmahendrasahap@gmail.com

Abstrak

Media sosial seperti X, pada umumnya sering digunakan oleh masyarakat untuk menyampaikan opini di berbagai hal, seperti kondisi lingkungan, pelayanan dan akses suatu bandara yang ada di Indonesia. Penelitian ini bertujuan untuk mengetahui sentimen masyarakat ditiga bandara tersibuk di Indonesia, yaitu Bandara Internasional Soekarno-Hatta, Bandara Internasional Ngurah Rai, dan Bandara Internasional Juanda. Data dikumpulkan dari platform X melalui metode *crawling data* untuk tweet pada periode May 2023 sampai April 2024. Terdapat 2809 tweet yang terkumpul dari periaode tersebut. Selanjutnya data yang diperoleh diberi label positif, negatif maupun netral secara manual. Kemudian dilakukan pembersihan data dari elemen-elemen yang tidak termasuk teks untuk keperluan analisis data. Digunakan perbandingan dengan dua buah model *large language Models* berbasis Transformers antara lain "akahana/indonesia-sentiment-roberta" dan "cardiffnlp/xlm-roberta-base-sentiment-multilingual," untuk melaksanakan analisis sentimen terhadap tiga bandara tersebut. Evaluasi dari performa penelitian ini menggunakan metrik presisi, *recall*, *F1-score*, dan akurasi. Hasilnya, pada model pertama tercapai akurasi hingga 51% sedangkan akurasi pada model kedua mencapai 63%. Dengan kata lain, dalam penelitian ini akurasi model kedua lebih baik dari model pertama. Walaupun demikian, masing-masing dari model masih memiliki keterbatasan dalam mengklasifikasi sentimen positif pada data yang ada.

Kata Kunci : Sentimen analisis, RoBERTa, Bandara, X (Twitter)

Abstract

Social media platforms like X are commonly used by the public to express opinions on various topics, such as environmental conditions, services, and access to airports in Indonesia. This study aims to analyze public sentiment toward the three busiest airports in Indonesia: Soekarno-Hatta International Airport, Ngurah Rai International Airport, and Juanda International Airport. Data was collected from the X platform using a data crawling method for tweets posted between May 2023 and April 2024. A total of 2,809 tweets were gathered during this period. The collected data was manually labeled as positive, negative, or neutral. Subsequently, data cleaning was performed to remove non-text elements for analysis purposes. A sentiment analysis was conducted on the three airports using a comparison between two Transformer-based large language models: "akahana/indonesia-sentiment-roberta" and "cardiffnlp/xlm-roberta-base-sentiment-multilingual." The performance of this study was evaluated using precision, recall, F1-score, and accuracy metrics. The first model achieved an accuracy of 51%, while the second model reached an accuracy of 63%. In other words, the second model outperformed the first in terms of accuracy. However, both models exhibited limitations in classifying positive sentiment within the dataset.

Keywords: Sentiment Analysis, RoBERTa, Airport, X (Twitter)

PENDAHULUAN

Perkembangan dunia digital pada saat ini mengalami peningkatan yang sangat cepat, hal ini dapat dilihat dari maraknya masyarakat menggunakan media social dalam kehidupan sehari-hari. Media sosial seperti X, pada umumnya sering digunakan masyarakat untuk menyampaikan opini tentang berbagai hal termasuk pelayanan, akses, meupun kondisi suatu bandar udara. Di Indonesia terdapat 3 bandara tersibuk yaitu: Bandara Internasional Soekarno-Hatta, Bandara Internasional Ngurah Rai dan Bandara Internasional Yogyakarta. Analisa sentimen dapat digunakan untuk mengevaluasi kepuasan masyarakat terhadap pelayanan disuatu bandara. Penelitian ini bertujuan untuk mengetahui pandangan masyarakat tentang beberapa bandara tersibuk di Indonesia pada platform X.

Terdapat beberapa penelitian yang berkaitan dengan penelitian ini seperti pandangan penumpang tentang kualitas beberapa bandara terkemuka didunia menggunakan pendekatan yang lebih manual dan deskriptif daripada Teknik analisis menggunakan pembelajaran mesin (Bae and Chi).

Terdapat pula *Study* (Haya Altajel and Alotaibi, 2021) mengenai Analisa sentiment dengan cara klasterisasi pada situs Skytrax dan Twitter tentang tiga bandara tersibuk di Arab Saudi. Penelitian (Putu et al.) melakukan analisis sentiment di Bandara Kertajati menggunakan *Naïve Bayes* yang hasilnya cenderung negatif. Artikel (Meddeb, Lavandier and Kotzinos,2020) menganalisis tentang sentimen kebisingan bandara menggunakan SVM dengan pendekatan berbasis leksikom. Terdapat pula sentiment analisis di domain keuangan yang mengusulkan dataset analisis berbahasa china yang baru. Pada *Study* (Lan et al.), LLMs dievaluasi dengan dataset dari penelitian ini.

Selanjutnya terdapat penelitian yang menyelidiki kemampuan dari LLMs yang dibandingkan dengan SLMs dalam berbagai tugas analisis sentiment (Zhang et al.). Referensi (Seong and Song) membahas penerapan model bahasa Korean

Funnel Transformer dalam menganalisis sentimen respons online terhadap seni pertunjukan. Penelitian (Li, Yang, and Huang) mengeksplorasi Large language models BERT dan tiga turunannya menggunakan dataset tweet sentiment pada maskapai penerbangan.

Studi (Fatouros et al.) menggali potensi ChatGPT 3.5 dalam menganalisis sentimen keuangan untuk pasar valuta asing yang selanjutnya dibandingkan dengan FinBERT. Pada (Assiri et al.) digunakan model hibrida yang menggabungkan Model DeBERTa dan GRU dalam melakukan analisis sentimen. Terdapat juga penggunaan Model BERT terhadap aksi boikot Israel di platform X (Sriyanti, Kartika, and Najaf).

Penelitian (Iriananda et al.) melakukan analisis sentiment menggunakan SVM, Grid Search dan kombinasi N-Gram pada komentar penggunaan *game* bergerak. Artikel (Vidya Chandradev, I Made Agus Dwi Suarjaya, and I Putu Agung Bayupati) menerapkan NLP BERT untuk menganalisa *review* dari pengunjung hotel di Indonesia. Terdapat pula penerapan model Bahasa berbasis jaringan saraf dalam menganalisis dan mengklasifikasikan teks menjadi emosi positif, negatif atau netral (Kumawat et al.). (Bello, Ng, and Leung) Mengkombinasikan BERT dengan CNN, RNN, dan BiLSTM dalam melakukan Analisa sentiment pada Twitter.

Pada artikel ini, kami melakukan analisis sentimen di tiga bandara tersibuk di Indonesia menggunakan *Large Language Models* berbasis *Transformer*. Disini kami menggunakan model yang di implementasikan dari RoBERTa dalam melakukan sentimen analisisnya.

METODE

Pada penelitian ini, kami mengumpulkan data dari tweet dan komentar-komentar yang terdapat pada *platform X* dengan metode *Crawling data*, dengan kata kunci "Soekarno-Hatta" dan "Soetta" untuk Bandara Internasional Soekarno-Hatta,

“Ngurah Rai” untuk Bandara Internasional Ngurah Rai dan “Bandara Juanda” untuk Bandara Internasional Juanda. Tweet dan komentar yang diperoleh, diambil mulai dari Mei 2023 sampai April 2024 dengan total 2809 tweet. Data yang diperoleh kemudian diberi label secara manual dengan ketentuan 100 tweet untuk setiap bandara, label yang diberikan bernilai positif, negatif dan netral.

Data yang telah diberi label kemudian masuk kedalam tahapan *pre-processing* untuk meningkatkan kualitas dari data yang akan dianalisis. Pada tahapan ini, dilakukan pembersihan teks dari elemen-elemen yang bukan teks seperti @, https, tanda baca dan lain lain. Setelah itu, digunakan dua model analisis berbasis *Transformer* yaitu *akahana/indonesia-sentiment-roberta* dan *cardiffnlp/xlm-roberta-base-sentiment-multilingual* dari Hugging Face untuk mengklasifikasikan sentimen positif, negatif dan netral pada dataset.

Evaluasi kedua model dilakukan dengan menggunakan beberapa metrik, yaitu presisi, *recall*, F1-Score dan akurasi. Persamaan (1) merupakan formulasi untuk menghitung presisi dimana TP adalah jumlah prediksi positif yang benar, dan FP adalah jumlah prediksi positif yang salah.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

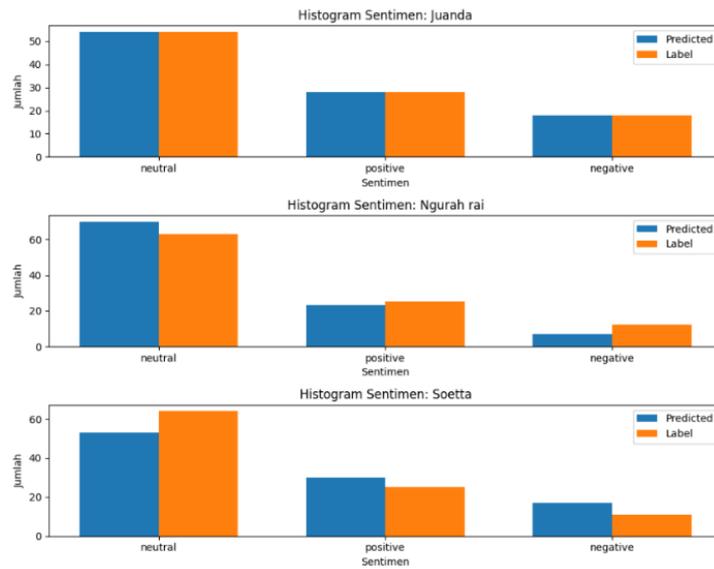
Persamaan (2) menunjukkan formula untuk *recall* dimana FN adalah jumlah data positif yang salah diklasifikasikan sebagai negatif. *F1-Score* (F_1) ditunjukkan pada persamaan (3) yang menggabungkan presisi dan recall menjadi satu nilai yang seimbang.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision+Recall} \quad (3)$$

Sementara persamaan (4) merupakan formula untuk mencari nilai dari akurasi dimana TN adalah jumlah prediksi negatif yang benar.

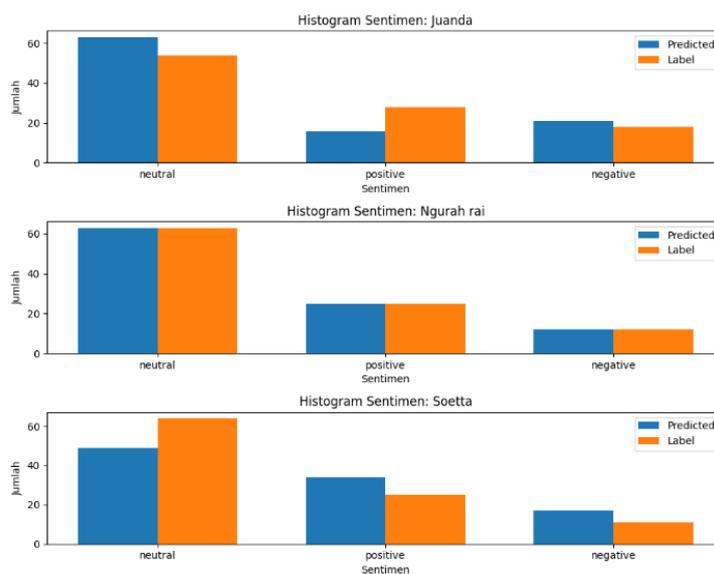
$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$



Gambar 3. Hasil Perbandingan antara Pelabelan Manual dengan Prediksi dari Model Pertama.

Gambar 3 menampilkan prediksi sentimen antara pelabelan manual dengan prediksi menggunakan model pertama. Data menunjukkan bahwa label manual dan prediksi memiliki nilai yang sangat mendekati untuk sentiment pada Bandara Juanda.

Sementara Gambar 4 menampilkan sentiment pada bandara Soekarno-Hatta yang menggunakan model kedua dan hasil dari prediksi dan pelabelan yang mendekati nilai yang sama.



Gambar 4. Hasil Perbandingan antara Pelabelan Manual dengan Prediksi dari Model Pertama

Tabel 1. Hasil Klasifikasi dari Model Pertama

<i>Classification Report:</i>				
	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
<i>Negative</i>	0.55	0.56	0.55	41
<i>Neutral</i>	0.61	0.60	0.60	181
<i>Positive</i>	0.27	0.28	0.28	78
<i>Accuracy</i>			0.51	300
<i>Macro Avg</i>	0.48	0.48	0.48	300
<i>Weighted Avg</i>	0.51	0.51	0.51	300

Tabel 1 menampilkan hasil klasifikasi dari model pertama. Berdasarkan laporan klasifikasi yang diperoleh, model yang digunakan berhasil mengklasifikasikan data dengan akurasi keseluruhan sebesar 51%. Artinya, model dapat memprediksi kelas yang benar untuk 51% dari seluruh data yang diuji.

Namun, terdapat ketidakseimbangan performa dalam mengklasifikasikan setiap kelas. Kelas "netral" merupakan kelas yang paling mudah dikenali oleh model, dengan nilai presisi, recall, dan F1-score yang cukup baik. Sebaliknya, kelas "positif" adalah kelas yang paling sulit dikenali, ditandai dengan nilai presisi, recall, dan F1-score yang rendah. Hal ini menunjukkan bahwa model cenderung lebih baik dalam mengidentifikasi data yang termasuk dalam kelas mayoritas (dalam hal ini, kelas "netral") dibandingkan dengan kelas minoritas (kelas "positif").

Sementara pada model kedua yang ditampilkan pada Tabel 2 menunjukkan akurasi sebesar 63% dalam mengklasifikasikan data. performa terbaik dicapai pada kelas "neutral" dengan nilai presisi 71%, recall 69%, dan F1-score 70%. Sementara

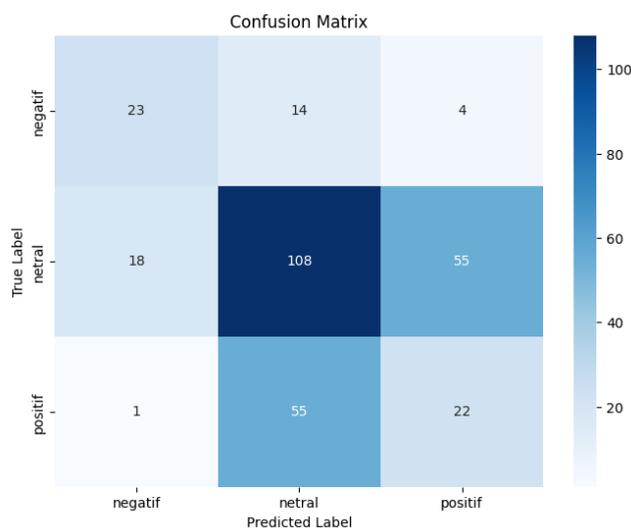
Tabel 2. Hasil Klasifikasi dari Model Kedua

<i>Classification Report:</i>				
	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
<i>Negative</i>	0.52	0.63	0.57	41
<i>Neutral</i>	0.71	0.69	0.70	181
<i>Positive</i>	0.51	0.49	0.50	78
<i>Accuracy</i>			0.63	300
<i>Macro Avg</i>	0.58	0.60	0.59	300
<i>Weighted Avg</i>	0.63	0.63	0.63	300

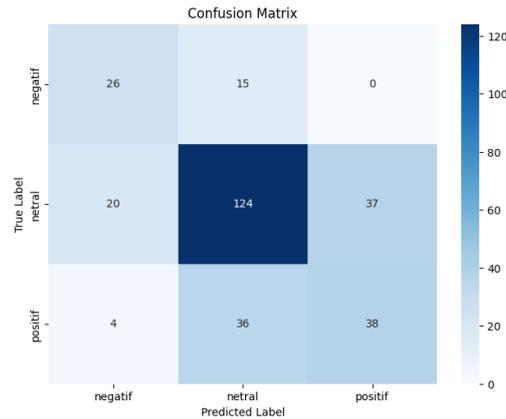
itu, performa terendah ada pada kelas "positive" dengan nilai presisi 51%, recall 49%, dan F1-score 50%. Hal ini menunjukkan bahwa model lebih mudah dalam mengidentifikasi data "neutral" dibandingkan dengan data "positive".

Gambar 5 dan 6 menunjukkan *confusion matrix* untuk mengamati secara visual seberapa baik jaringan membuat prediksi menggunakan model pertama dengan kedua. Gambar 5 menjelaskan bahwa jumlah tweet pada prediksi netral yang sesuai dengan pelabelan netral sebanyak 108 tweet, sementara pelabelan netral yang diprediksi negatif dan positif sebanyak 18 dan 55 tweet. Untuk pelabelan negatif yang diprediksi negatif, netral dan positif adalah sebanyak 23, 14 dan 4. Sementara pelabelan positif yang diprediksi negatif, netral dan positif adalah sebanyak 1, 55 dan 22.

Gambar 6 menjelaskan bahwa jumlah tweet pada prediksi netral yang sesuai dengan pelabelan netral sebanyak 124 tweet, sementara pelabelan netral yang diprediksi negatif dan positif sebanyak 20 dan 37 tweet. Untuk pelabelan negatif yang diprediksi negatif, netral dan positif adalah sebanyak 26, 15 dan 0. Sementara pelabelan positif yang diprediksi negatif, netral dan positif adalah sebanyak 4, 36 dan 38.



Gambar 5 Confusion Matrix untuk Prediksi dan Pelabelan pada Model Pertama



Gambar 6. Confusion Matrix untuk Prediksi dan Pelabelan pada Model Kedua

PENUTUP

Kesimpulan

Penelitian ini menunjukkan bahwa sentiment netral lebih mendominasi dari sentiment positif atau negatif pada ketiga bandara (Soetta, Juanda dan Ngurah rai). Model pertama memiliki akurasi hingga 51% dan menunjukkan kinerja tertinggi pada kelas netral. Meskipun performa pada kelas yang lainnya rendah terutama pada kelas positif. Sementara pada model ke dua, akurasi yang ditampilkan lebih baik dengan persentase 63% dengan performa tertinggi pada kelas netral (presisi 71%, recall 69%, F1-score 70%) namun performa pada kelas positif masih rendah dengan presisi 51%, recall 49%, F1-score 50%. Kedua model ini menunjukkan kekurangan dalam menangani kelas terendah terutama pada sentimen positif.

DAFTAR PUSTAKA

- Assiri, Adel et al. "DeBERTa-GRU: Sentiment Analysis for Large Language Model." *Computers, Materials and Continua* 79.3 (2024): 4219–4236. Web.
- Bae, Wonmi, and Junwook Chi. "Content Analysis of Passengers' Perceptions of Airport Service Quality: The Case of Honolulu International Airport." *Journal of Risk and Financial Management* 15.1 (2022): n. pag. Web.
- Bello, Abayomi, Sin Chun Ng, and Man Fai Leung. "A BERT Framework to Sentiment Analysis of Tweets." *Sensors* 23.1 (2023): n. pag. Web.

- Fatouros, Georgios et al. "Transforming Sentiment Analysis in the Financial Domain with ChatGPT." *Machine Learning with Applications* 14 (2023): 100508. Web.
- Haya Altajel, Eng, and Faisal Alotaibi. *Measuring the Correlation of Service Expectations of Travelers in Selected International KSA Airports through Bilingual Sentiment and Clustering Analysis*. N.p. Print.
- Iriananda, Syahroni Wahyu et al. "Optimasi Klasifikasi Sentimen Komentar Pengguna Game Bergerak Menggunakan Svm, Grid Search Dan Kombinasi N-Gram." *Jurnal Teknologi Informasi dan Ilmu Komputer* 11.4 (2024): 743–752. Web.
- Kumawat, Spraha et al. "Sentiment Analysis Using Language Models: A Study." *Proceedings of the Confluence 2021: 11th International Conference on Cloud Computing, Data Science and Engineering* (2021): 984–988. Web.
- Lan, Yinyu et al. "Chinese Fine-Grained Financial Sentiment Analysis with Large Language Models." (2023): n. pag. Web.
- Li, Zehong, Chuyang Yang, and Chenyu Huang. "A Comparative Sentiment Analysis of Airline Customer Reviews Using Bidirectional Encoder Representations from Transformers (BERT) and Its Variants." *Mathematics* 12.1 (2024): n. pag. Web.
- Meddeb, Iheb, Catherine Lavandier, and Dimitris Kotzinos. *Using Twitter Streams for Opinion Mining: A Case Study on Airport Noise*. N.p. Print.
- Putu, Ni et al. "SENTIMENT ANALYSIS OF WEST JAVA INTERNATIONAL AIRPORT (BIJB) KERTAJATI ON TWITTER." *Jurnal Manajemen dan Bisnis* 4.2 (2020): n. pag. Print.
- Seong, Baekryun, and Kyungwoo Song. "Sentiment Analysis of Online Responses in the Performing Arts with Large Language Models." *Heliyon* 9.12 (2023): n. pag. Web.
- Sriyanti, Zilvi Azus, Dhian Satria Yudha Kartika, and Abdul Rezha Efrat Najaf. "Implementasi Model Bert Pada Analisis Sentimen Pengguna Twitter Terhadap Aksi Boikot Produk Israel." *Jurnal Informatika dan Teknik Elektro Terapan* 12.3 (2024): 2335–2342. Web.
- Vidya Chandradev, I Made Agus Dwi Suarjaya, and I Putu Agung Bayupati. "Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT." *Jurnal Buana Informatika* 14.02 (2023): 107–116. Web.
- Zhang, Wenxuan et al. "Sentiment Analysis in the Era of Large Language Models: A Reality Check." (2023): n. pag. Web.