

ANALISA SENTIMEN INDONESIA MENGENAI KRL COMMUTER LINE MENGGUNAKAN METODE NAÏVE BAYES DAN SUPPORT VECTOR MACHINE

Adrian Nathanael Kalalo, Een Juhriah, Akhmad Aris Tantowi

Program Studi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer Universitas Indraprasta PGRI
adriankalalo654@gmail.com, eenzuhriah29@gmail.com, bgrjkt1973@gmail.com

Submitted March 25, 2025; Revised March 28, 2025; Accepted April 4, 2025

Abstrak

Penelitian ini menganalisis sentimen masyarakat Indonesia terhadap layanan KRL Commuter Line menggunakan metode Naive Bayes dan Support Vector Machine (SVM). Data diperoleh dari Twitter melalui API dan diproses melalui serangkaian tahapan pra-pemrosesan, termasuk pembersihan karakter khusus, URL, dan informasi tidak relevan lainnya. Labeling data dilakukan secara lexicon-based untuk mengidentifikasi sentimen positif, negatif, atau netral dari setiap tweet. Model Naive Bayes dan SVM kemudian dilatih dan diuji dengan data yang telah diberi label. Hasil penelitian menunjukkan bahwa kedua metode tersebut efektif dalam mengklasifikasikan sentimen, namun SVM menunjukkan kinerja yang lebih unggul dalam hal akurasi dibandingkan dengan Naive Bayes. Temuan ini menyoroti potensi besar kedua metode dalam analisis sentimen, dengan SVM sebagai pilihan yang lebih tepat untuk kasus ini. Penelitian ini diharapkan dapat berkontribusi pada peningkatan kualitas layanan KRL Commuter Line dengan menyediakan wawasan mengenai sentimen pengguna yang dapat digunakan sebagai dasar perbaikan layanan

Kata Kunci: Analisis Sentimen, KRL Commuter Line, Naive Bayes, Support Vector Machine (SVM)

Abstract

This study analyzes the sentiment of Indonesian society towards KRL Commuter Line services using Naive Bayes and Support Vector Machine (SVM) methods. Data was collected from Twitter through an API and processed through a series of preprocessing steps, including the removal of special characters, URLs, and other irrelevant information. Lexicon-based labeling was performed to identify the sentiment as positive, negative, or neutral for each tweet. The Naive Bayes and SVM models were then trained and tested on the labeled data. The results indicate that both methods are effective in classifying sentiment; however, SVM outperforms Naive Bayes in terms of accuracy. These findings highlight the significant potential of both methods in sentiment analysis, with SVM being the more suitable choice for this particular case. This study is expected to contribute to the improvement of KRL Commuter Line services by providing insights into user sentiment, which can be used as a basis for service enhancement.

Key Words : Sentiment Analysis, KRL Commuter Line, Naive Bayes, Support Vector Machine (SVM)

1. PENDAHULUAN

Dalam era digital saat ini, media sosial seperti Twitter berperan penting dalam menyampaikan opini publik, termasuk tentang layanan transportasi publik seperti KRL Commuter Line di Indonesia. Mengingat perannya yang vital, penting untuk memahami sentimen masyarakat terhadap layanan ini untuk meningkatkan kualitasnya.

Analisis sentimen merupakan cabang dari pemrosesan bahasa alami (Natural Language Processing/NLP) yang bertujuan untuk mengidentifikasi dan mengategorikan opini yang diekspresikan dalam teks, terutama untuk menentukan apakah sikap penulis terhadap topik tertentu positif, negatif, atau netral [1]. Dua metode yang sering digunakan dalam analisis sentimen adalah Naive Bayes dan

Support Vector Machine (SVM). Naive Bayes adalah metode probabilistik yang sederhana namun efektif dalam klasifikasi teks, sedangkan SVM dikenal karena kemampuannya dalam menghasilkan klasifikasi yang akurat dengan memaksimalkan margin antara kelas-kelas yang berbeda [2]. Penelitian terdahulu menunjukkan bahwa analisis sentimen telah berkembang pesat, dengan berbagai pendekatan yang diterapkan dalam berbagai konteks. Penelitian [3] mengkaji penerapan deep learning dalam analisis sentimen, sementara [4] melakukan survei terhadap berbagai pendekatan yang digunakan dalam opinion mining dan analisis sentimen. Beberapa penelitian juga menyoroti pentingnya pemilihan fitur dan teknik pra-pemrosesan dalam meningkatkan akurasi model analisis sentimen [5]. Selain itu, pendekatan lexicon-based yang digunakan dalam labeling data juga memiliki peran penting dalam menentukan hasil akhir analisis sentimen. Menurut [6] penggunaan kamus sentimen yang tepat dapat meningkatkan efektivitas labeling dan membantu model dalam menangkap nuansa sentimen yang lebih kompleks.

Di Indonesia, beberapa studi lokal juga telah mengeksplorasi analisis sentimen dalam konteks media sosial. Contohnya, [7] mengkaji penerapan analisis sentimen untuk menilai opini publik terhadap layanan publik di media sosial. Penelitian oleh [8] Diah Indarwati & Februriyanti juga menunjukkan penerapan teknik analisis sentimen untuk memahami opini masyarakat terhadap layanan transportasi publik di Indonesia. Dengan menggabungkan pendekatan tradisional seperti Naive Bayes dan SVM dengan teknik modern dan pra-pemrosesan yang cermat, penelitian ini bertujuan untuk memberikan wawasan yang komprehensif mengenai sentimen masyarakat terhadap layanan KRL Commuter Line, serta

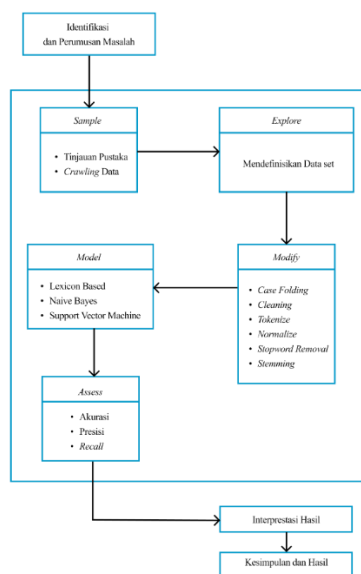
memberikan rekomendasi yang dapat digunakan untuk meningkatkan kualitas layanan tersebut. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat Indonesia terhadap layanan KRL Commuter Line melalui data yang diperoleh dari Twitter. Tujuan spesifik dari penelitian ini adalah:

1. Mengumpulkan dan memproses data tweet terkait KRL Commuter Line.
2. Melakukan labeling sentimen pada data menggunakan pendekatan lexicon-based.
3. Melatih dan menguji model Naive Bayes dan SVM untuk klasifikasi sentimen.
4. Membandingkan kinerja kedua metode tersebut dalam hal akurasi dan efektivitas klasifikasi.

Dalam penelitian di bidang analisis sentimen, beberapa studi sebelumnya telah menerapkan berbagai metode dan memperoleh hasil yang bervariasi. Penelitian [9] mengimplementasikan metode Naive Bayes dan Support Vector Machine (SVM) dengan pendekatan berbasis leksikon untuk analisis sentimen di Twitter, khususnya menggunakan data terkait COVID-19. Hasil penelitian menunjukkan akurasi sebesar 74% untuk metode Naive Bayes dan 85% untuk metode SVM. [10] fokus pada prediksi kinerja mahasiswa dengan menggunakan metode SVM dan software Matlab, mencapai akurasi 41,49% dengan kernel linear sebagai fungsi kernel paling optimal dan parameter C sebesar 5. Penelitian [11] menggunakan metode Naive Bayes, SVM, dan k-NN dengan software RapidMiner untuk analisis sentimen pada data Twitter, dan melaporkan akurasi Naive Bayes sebesar 75,58%, SVM 63,99%, serta k-NN 73,34%. Penelitian [12] menerapkan algoritma SVM untuk analisis kepuasan pengguna layanan data seluler, dengan hasil presisi 92,45%, recall 93,90%, dan akurasi 99,01% untuk kernel

Hasil dari penelitian ini diharapkan dapat memberikan wawasan yang berharga bagi pengelola KRL Commuter Line dalam memahami sentimen pengguna. Informasi ini dapat digunakan sebagai dasar dalam perencanaan dan implementasi perbaikan layanan, dengan tujuan akhir untuk meningkatkan kepuasan dan pengalaman penumpang.

2. METODE PENELITIAN



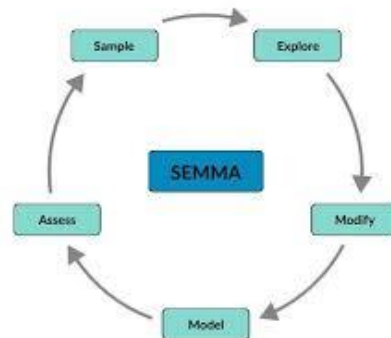
Sumber : Dokumen Pribadi

Gambar 1. Diagram Alir Penelitian

Tahapan penelitian ini berawal dari identifikasi dan perumusan masalah, dilanjutkan mengolah data menggunakan metode SEMMA yang memiliki 5 tahap yang terdiri atas sample, explore, modify, model, dan asses. Tahap sample terdiri dari atas tinjauan pustaka dan crawling data. Tahap explore dimana mendefinisikan data hasil crawling dan kolom. Selanjutnya adalah tahap modify terdiri atas case folding, cleaning, tokenizing, normalize, stopword removal, dan stemming, dilakukan agar data set lebih terstruktur. Tahap keempat adalah model dimana data set diberikan label kelas dengan metode Lexicon Based lalu diolah dengan metode

Naive Bayes dan Support Vector Machine. Tahap terakhir adalah assess atau evaluasi kinerja model yang terdiri atas akurasi, presisi, dan recall. Tahap terakhir adalah kesimpulan dan saran dari penelitian.

SEMMA model



Sumber : www.datascience-pm.com

Gambar 2. Metode SEMMA

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimental yang mengikuti metode SEMMA (Sample, Explore, Modify, Model, Assess) untuk analisis sentimen terhadap layanan KRL Commuter Line. Data dikumpulkan dari Twitter menggunakan API, dengan fokus pada tweet terkait KRL Commuter Line. Setelah pengumpulan, data dianalisis untuk memahami karakteristiknya, termasuk distribusi sentimen dan frekuensi kata, serta divisualisasikan menggunakan Wordcloud dan grafik. Data kemudian dimodifikasi melalui pembersihan dari elemen yang tidak relevan, penghapusan duplikat, dan pelabelan sentimen menggunakan metode lexicon-based. Selanjutnya, model analisis sentimen dibangun menggunakan Naive Bayes dan Support Vector Machine (SVM), dan kinerja kedua model ini dievaluasi berdasarkan akurasi, precision, recall, dan F1-score. Hasil evaluasi digunakan untuk menentukan model yang paling efektif dalam mengklasifikasikan sentimen terhadap KRL Commuter Line.

Penelitian ini bertujuan untuk memberikan wawasan yang mendalam tentang sentimen pengguna terhadap layanan KRL

Commuter Line dan untuk menentukan metode analisis yang paling efektif dalam konteks penelitian ini. Temuan dari penelitian ini diharapkan dapat digunakan untuk meningkatkan kualitas layanan berdasarkan feedback dari pengguna.

3. HASIL DAN PEMBAHASAN

Sentimen analisis dari dataset tweet yang didapatkan dengan kata kunci KRL Commuter Line kemudian membandingkan hasil klasifikasi antara metode Naïve Bayes dan Support Vector Machine. Penelitian ini menggunakan data set tweets yang berjumlah 1001 tweets.

Tabel 1. Distribusi Sentimen

Tweets Sentimen	Jumlah Tweets	Presentase
Tweets Positive	812	81%
Tweets Neutral	37	3,7%
Tweets Negative	153	15,3%

Sumber : Dokumen Pribadi

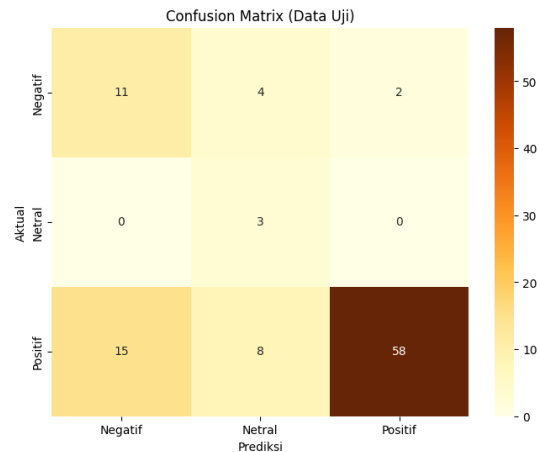
Pada Tabel 1 merupakan distribusi sentimen yang berisikan *tweets positive* berjumlah 812 dengan *presentase* 81%, lalu *tweets neutral* berjumlah 37 dan *tweet negative* berjumlah 153 dan 15,3%. Dari hasil tersebut, dapat dikatakan bahwa opini masyarakat terhadap KRL Commuter Line Indonesia didominasi oleh opini positif. Meskipun opini negatif banyak dikarenakan beberapa terlambatnya kereta tiba atau tidak sesuai jadwal serta fasilitas KRL Commuter Line yang tidak memuaskan.

Tabel 2. Hasil Akurasi, Presisi, Recall Naïve Bayes

Rasio Split	Akurasi	Presisi	Recall
90% : 10%	71,29%	53%	79%
80% : 20%	67,16%	65%	90%
70% : 30%	70,10%	66%	91%

Sumber : Dokumen Pribadi

Berdasarkan Tabel 2 menampilkan rasio split pada model Naïve Bayes dengan nilai akurasi, presisi, dan recall masing-masing. Hasil terbaik yaitu pada rasio split 90%:10% dengan akurasi 71,29%, presisi 53%, dan recall 79%.



Sumber : Dokumen Pribadi

Gambar 3. Confusion Matrix Naïve Bayes

Selanjutnya dengan rasio *split* terbaik yaitu 90%:10% mendapatkan *confusion matrix* Naïve Bayes pada Gambar 2 dengan didapat melakukan perhitungan nilai akurasi, presisi, dan *recall* setiap kelasnya.

$$\text{Akurasi} = \frac{\text{total prediksi benar}}{\text{total keseluruhan data}}$$

$$= \frac{72}{101} \times 100\% = 71,29\%$$

Kemudian menghitung presisi kelas positif, netral, dan negatif:

$$\begin{aligned} \text{Presisi negatif} &= \frac{\text{true negatif}}{\text{total prediksi negatif}} \\ &= \frac{11}{11 + 0 + 15} = \frac{11}{26} \times 100 = 42,31\% \end{aligned}$$

$$\begin{aligned} \text{Presisi netral} &= \frac{\text{true netral}}{\text{total prediksi netral}} \\ &= \frac{3}{4 + 3 + 8} \\ &= \frac{3}{15} \times 100\% = 20\% \end{aligned}$$

$$\begin{aligned} \text{Presisi positif} &= \frac{\text{true positif}}{\text{total prediksi positif}} \\ &= \frac{58}{2 + 0 + 58} \\ &= \frac{58}{60} \times 100\% = 96,67\% \end{aligned}$$

Selanjutnya menghitung *recall* kelas positif, netral, dan negatif:

$$\begin{aligned} \text{Recall negatif} &= \frac{\text{true negatif}}{\text{total recall negatif}} \\ &= \frac{11}{11 + 4 + 2} \\ &= \frac{11}{17} \times 100\% = 64,71\% \end{aligned}$$

$$\begin{aligned} \text{Recall netral} &= \frac{\text{true netral}}{\text{total recall netral}} \\ &= \frac{3}{0 + 3 + 0} \\ &= \frac{3}{3} \times 100\% = 100\% \end{aligned}$$

$$\begin{aligned} \text{Recall positif} &= \frac{\text{true positif}}{\text{total recall positif}} \\ &= \frac{58}{15 + 8 + 58} \\ &= \frac{58}{81} \times 100\% = 71,60\% \end{aligned}$$

Kemudian jika sudah mendapatkan presisi kelas positif, netral, dan negatif dan *recall* kelas positif, netral, dan negatif dapat menghitung Presisi dan Recall Totalnya.

$$\begin{aligned} \text{Presisi} &= \frac{\text{negatif} + \text{Positif} + \text{Netral}}{3} \\ &= \frac{42,31 + 20 + 96,67}{3} = \frac{158,98}{3} \times 100\% \\ &= 52,99\% \end{aligned}$$

$$\begin{aligned} \text{Recall} &= \frac{\text{Negatif} + \text{Positif} + \text{Netral}}{3} \\ &= \frac{64,71 + 100 + 71,60}{3} \\ &= \frac{236,31}{3} \times 100\% \\ &= 78,77\% \end{aligned}$$

Table 3 Hasil Grid Search Tiap Kernel Support Vector Machine

Kernel	Nilai Parameter	Akurasi	Presisi	Recall
Linear	C=1	85,07%	70%	58%
RBF	C=100, gamma=0,01	86,57%	75%	65%
Poly	C=10, degree=1	84,58%	70%	62%

Sumber : Dokumen Pribadi

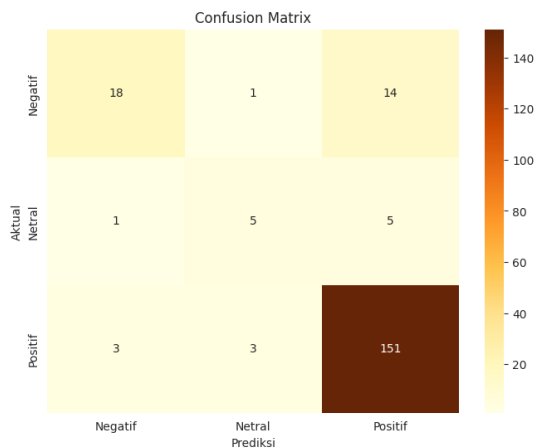
Tabel 3 merupakan hasil model *Support Vector Machine* dari masing-masing kernel yang digunakan untuk membandingkan mana kernel dan nilai parameter terbaik untuk dilanjutkan keproses selanjutnya. Jadi Kernel RBF dengan nilai parameter C=100, gamma=0,01 merupakan kernel terbaik dengan akurasi 86,57%, presisi 75%, dan *recall* 65%.

Tabel 4. Hasil Akurasi, Presisi, dan Recall Support Vector Machine

Rasio Split	Akurasi	Presisi	Recall
90% : 10%	82,18%	60%	56%
80% : 20%	86,57%	75%	65%
70% : 30%	86,38%	75%	64%

Sumber : Dokumen Pribadi

Pada Tabel 4 mendapatkan nilai hasil akurasi, presisi, dan *recall* dengan kernel terbaik pada Tabel 3. Rasio *split* 80%:20% memiliki nilai akurasi terbaik yaitu 86,57%. Setelah menentukan kernel terbaik dan rasio *split* terbaik selanjutnya menghitung nilai setiap sentimen (positif, negatif, dan netral) dengan menggunakan *confusion matrix* yang didapat pada Gambar 3.



Sumber : Dokumen Pribadi
Gambar 4. Confusion Matix Support Vector Machine

Selanjutnya dengan rasio *split* terbaik yaitu 90%:10% mendapatkan *confusion matix Support Vector Machine* pada Gambar 3 dengan didapat melakukan perhitungan nilai akurasi, presisi, dan *recall* setiap kelasnya.

$$\begin{aligned} \text{Akurasi} &= \frac{\text{total prediksi benar}}{\text{total keseluruhan data}} \\ &= \frac{18 + 1 + 14 + 1 + 5 + 5 + 3 + 3 + 151}{18 + 1 + 14 + 1 + 5 + 5 + 3 + 3 + 151} \\ &= \frac{174}{201} \times 100\% = 86,57\% \end{aligned}$$

Kemudian menghitung presisi kelas positif, netral, dan negatif:

$$\begin{aligned} \text{Presisi negatif} &= \frac{\text{true negatif}}{\text{total prediksi negatif}} \\ &= \frac{18}{18 + 1 + 3} \\ &= \frac{18}{22} \times 100\% = 81,82\% \end{aligned}$$

$$\begin{aligned} \text{Presisi netral} &= \frac{\text{true netral}}{\text{total prediksi netral}} \\ &= \frac{5}{1 + 5 + 3} \\ &= \frac{5}{9} \times 100\% = 55,56\% \end{aligned}$$

$$\begin{aligned} \text{Presisi positif} &= \frac{\text{true positif}}{\text{total prediksi positif}} \\ &= \frac{151}{14 + 5 + 151} = \frac{151}{170} \times 100\% \\ &= 88,82\% \end{aligned}$$

Selanjutnya menghitung *recall* kelas positif, netral, dan negatif :

$$\begin{aligned} \text{Recall negatif} &= \frac{\text{true negatif}}{\text{total recall negatif}} \\ &= \frac{18}{18 + 1 + 14} \\ &= \frac{18}{33} \times 100\% = 54,55\% \end{aligned}$$

$$\begin{aligned} \text{Recall netral} &= \frac{\text{true netral}}{\text{total recall netral}} \\ &= \frac{5}{1 + 5 + 5} \\ &= \frac{5}{11} \times 100\% = 45,45\% \end{aligned}$$

$$\begin{aligned} \text{Recall positif} &= \frac{\text{true positif}}{\text{total recall positif}} \\ &= \frac{151}{3 + 3 + 151} \\ &= \frac{151}{157} \times 100\% = 96,11\% \end{aligned}$$

Kemudian jika sudah mendapatkan presisi kelas positif, netral, dan negatif dan *recall* kelas positif, netral, dan negatif dapat menghitung Presisi dan Recall Totalnya.

$$\begin{aligned} \text{Presisi} &= \frac{\text{negatif} + \text{Positif} + \text{Netral}}{3} \\ &= \frac{42,31 + 20 + 96,67}{3} = \frac{158,98}{3} \times 100\% \\ &= 52,99\% \end{aligned}$$

$$\begin{aligned} \text{Recall} &= \frac{\text{Negatif} + \text{Positif} + \text{Netral}}{3} \\ &= \frac{64,71 + 100 + 71,60}{3} \\ &= \frac{236,31}{15} \times 100\% \\ &= 78,77\% \end{aligned}$$

Selanjutnya melakukan Evaluasi Model atau Interpretasi Hasil, Hasil *Crawling* data sejumlah 1002 data mentah lalu dilakukan *text pre-processing* sehingga menjadi 1001

data bersih. Selanjutnya, dilakukan pelabelan 1001 data set dengan metode *Lexicon based* dimana pada Gambar 4 adalah hasil dari proses label. Dimana terdapat 81% data label positif kemudian 3,7% data label netral dan 15,3% data negatif. Berdasarkan proses pengujian klasifikasi metode Naive Bayes dan *Support Vector Machine* (SVM) pada data *tweet* didapatkan hasil akurasi dari kedua metode tersebut pada Tabel 4.33. Hasil dari penelitian menggunakan klasifikasi teks data *tweet* tentang KRL Commuter Line menggunakan metode Naive Bayes mendapatkan akurasi 71,29%. Sementara itu, akurasi yang didapatkan menggunakan metode *Support Vector Machine* (SVM) adalah 86,57%.

Pada penelitian ini dilakukan dengan pengujian dengan 1002 tweet tentang KRL Commuter Line. Tahap preprocessing dilakukan untuk membersihkan data set dari karakter yang tidak memiliki sentimen untuk selanjutnya dilakukan pelabelan ke dalam kelas positif, netral, dan negatif dengan metode *Lexicon Based*. Implementasi metode klasifikasi diterapkan pada data set dimana hasil akurasi metode Naive Bayes adalah 71,29% dan akurasi metode *Support Vector Machine* dengan kernel paling optimal dan terbaik adalah RBF (parameter $C=100$ dan $\gamma=0,01$) adalah 86,57%.

Rasio	Naïve Bayes			Support Vector Machine		
	Akurasi	Presisi	Recall	Akurasi	Presisi	Recall
90%:10%	71,29%	53%	79%	82,18%	60%	56%
80%:20%	67,16%	65%	90%	86,57%	75,10%	65%
70%:30%	70,10%	66%	91%	86,38%	75%	64%

Sumber : Dokumen Pribadi

Gambar 5. Kinerja Naïve Bayes dan Support Vector Machine

4. SIMPULAN

Penelitian ini telah mengimplementasikan model klasifikasi teks dari Twitter mengenai KRL Commuter Line di Indonesia dengan menggunakan metode *lexicon-based* untuk pemberian label kelas, visualisasi kata menggunakan *wordcloud*, serta evaluasi performa model menggunakan Naive Bayes dan *Support Vector Machine* (SVM). Berdasarkan hasil klasifikasi teks dari data *tweet*, beberapa kesimpulan penting dapat diambil.

Pertama, setelah melalui proses *text preprocessing*, data yang dihasilkan diberi label menggunakan metode *lexicon-based*. Dari hasil pelabelan tersebut, ditemukan bahwa terdapat 812 *tweet* yang termasuk dalam kelas positif, 153 *tweet* dalam kelas negatif, dan 37 *tweet* dalam kelas netral. Hal ini menunjukkan bahwa sentimen pengguna Twitter terhadap KRL Commuter Line di Indonesia cenderung positif. Kedua, metode Naive Bayes menghasilkan performa terbaik pada rasio data latih dan data uji 90%:10%, dengan akurasi sebesar 71,29%, presisi 53%, dan *recall* 79%. Di sisi lain, metode SVM diuji dengan tiga jenis kernel, yaitu linear, RBF (*Radial Basis Function*), dan poly, di mana kernel terbaik adalah RBF dengan parameter optimal $C=100$ dan $\gamma=0,01$. Model ini menghasilkan akurasi sebesar 86,57%, presisi 75%, dan *recall* 65%. Berdasarkan hasil ini, dapat disimpulkan bahwa metode SVM dengan kernel RBF lebih unggul dibandingkan Naive Bayes dalam konteks klasifikasi sentimen terhadap KRL Commuter Line di Indonesia.

Penelitian ini memberikan kontribusi yang signifikan dalam analisis sentimen terkait KRL Commuter Line dan diharapkan dapat menjadi referensi bagi penelitian selanjutnya di bidang *Natural Language Processing*, khususnya dalam klasifikasi teks menggunakan metode Naive Bayes dan SVM berbasis *lexicon*. Berdasarkan

hasil penelitian ini, ada beberapa saran untuk penelitian selanjutnya. Pertama, penelitian ini hanya mencakup tweet yang menggunakan bahasa Indonesia, sehingga penelitian di masa mendatang bisa mengeksplorasi sentimen tweet dalam berbagai bahasa dan lokasi pengguna Twitter yang berbeda. Ini akan memberikan wawasan yang lebih komprehensif tentang variasi sentimen berdasarkan konteks bahasa dan geografi. Kedua, dalam pemberian label pada dataset, penelitian ini menggunakan kamus lexicon InSet atau Indonesia Sentiment yang mencakup sekitar 10.250 kata dengan bobot antara -5 hingga +5. Namun, kamus ini masih terbatas dan belum mencakup semua kemungkinan sentimen dalam bahasa Indonesia. Oleh karena itu, penelitian selanjutnya bisa mempertimbangkan penggunaan metode lain seperti Vader atau TextBlob dengan menerjemahkan dataset terlebih dahulu, untuk memperkaya akurasi pelabelan sentimen. Penelitian ini menyoroti pentingnya pilihan model dan metode dalam analisis sentimen, serta menunjukkan bahwa pemilihan kernel yang tepat dalam SVM dapat secara signifikan meningkatkan performa klasifikasi. Adanya pemahaman yang lebih baik tentang sentimen pengguna dapat memberikan wawasan yang berharga bagi peningkatan layanan KRL Commuter Line di masa depan.

DAFTAR PUSTAKA

- [1] Prastyo, P. H., Ardiyanto, I., & Hidayat, R. (2020). A Review of Feature Selection Techniques in Sentiment Analysis Using Filter, Wrapper, or Hybrid Methods. 2020 6th International Conference on Science and Technology (ICST), 1, 1–6. <https://doi.org/10.1109/ICST50505.2020.9732885>
- [2] Ahmad, M., Aftab, S., Bashir, M. S., & Hameed, N. (2018). Sentiment Analysis using SVM: A Systematic Literature Review. *International Journal of Advanced Computer Science and Applications*, 9. <https://api.semanticscholar.org/CorpusID:3695786S>.
- [3] Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *WIREs Data Mining and Knowledge Discovery*, 8(4), e1253. <https://doi.org/https://doi.org/10.1002/widm.1253>
- [4] Liu, B., & Zhang, L. (2012). A Survey of Opinion Mining and Sentiment Analysis. In C. C. Aggarwal & C. Zhai (Eds.), *Mining Text Data* (pp. 415–463). Springer US. https://doi.org/10.1007/978-1-4614-3223-4_13
- [5] Sahoo, C., Wankhade, M., & Singh, B. K. (2023). Sentiment analysis using deep learning techniques: a comprehensive review. *International Journal of Multimedia Information Retrieval*, 12(2), 41. <https://doi.org/10.1007/s13735-023-00308-2>
- [6] Verma, B., & Thakur, R. S. (2018). Sentiment Analysis Using Lexicon and Machine Learning-Based Approaches: A Survey. In B. Tiwari, V. Tiwari, K. C. Das, D. K. Mishra, & J. C. Bansal (Eds.), *Proceedings of International Conference on Recent Advancement on Computer and Communication* (pp. 441–447). Springer Singapore.
- [7] Kamila Nadia L. (n.d.). *dibimbing.id - Memahami Opini Publik dengan Analisis Sentimen Media Sosial*. Retrieved August 27, 2023, from <https://dibimbing.id/blog/detail/memahami-opini-publik-dengan-analisis-sentimen-media-sosial>
- [8] Diah Indarwati, K., & Februariyanti, H. (n.d.). *Analisis Sentimen Terhadap*

- Kualitas Pelayanan Aplikasi Go-Jek Menggunakan Metode Naive Bayes Classifier.
- [9] Nuri, A. (2022). Implementasi naive bayes dan support vector machine dengan lexicon based untuk analisis sentimen pada twitter. Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta.
- [10] Hilmiyah, F. (2017). Prediksi Kinerja Mahasiswa Menggunakan Support Vector Machine untuk Pengelola Program Studi di Perguruan Tinggi (Studi Kasus: Program Studi Magister Statistika ITS). Departemen Manajemen Teknologi Bidang Keahlian Manajemen Teknologi Informasi Fakultas Bisnis Dan Manajemen Teknologi Institut Teknologi Sepuluh Nopember Surabaya, 1–99.
- [11] Wongkar, M., & Angdresey, A. (2019). Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter. 2019 Fourth International Conference on Informatics and Computing (ICIC), 1–5. <https://doi.org/10.1109/ICIC47613.2019.8985884>
- [12] Chory, R. N., Nasrun, M., & Setianingsih, C. (2018). Sentiment Analysis on User Satisfaction Level of Mobile Data Services Using Support Vector Machine (SVM) Algorithm. 2018 IEEE International Conference on Internet of Things and Intelligence System (IOTAIS), 194–200. <https://doi.org/10.1109/IOTAIS.2018.8600884>