

Analisis Sentimen Media Sosial *Youtube* Kereta Cepat (*Whoosh*) Menggunakan Algoritma *Bidirectional-LSTM*

Joy Lawa Rizky^{1*}, Windu Gata²

Ilmu Komputer, Universitas Nusa Mandiri, Depok, Indonesia

*e-mail *Corresponding Author*: joylawarizky7@gmail.com

Abstract

This study analyzes social media sentiment on YouTube regarding the high-speed train (Whoosh) using the Bidirectional-LSTM algorithm. The issue raised is the need for a deeper understanding of public perception of the high-speed train project, which can affect its acceptance and sustainability. The purpose of this paper is to evaluate the performance of the Bidirectional-LSTM algorithm in sentiment analysis compared to other algorithms. The method used involves collecting YouTube comment data, text preprocessing, and applying the Bidirectional-LSTM algorithm for sentiment classification. The parameters analyzed include accuracy, precision, and resilience to data variations. The research results show that the Bidirectional-LSTM algorithm achieves an accuracy of (0.86), which is significantly higher compared to the Multinomial Naïve Bayes algorithm (0.80), USE-Transfer learning (Tensorflow) (0.80), and Text Vectorization and Embedding (Tensorflow) (0.80). The conclusion of this study is that Bidirectional-LSTM is more effective and reliable in analyzing social media sentiment towards the high-speed train (Whoosh).

Keywords: *Sentiment Analysis; YouTube Fast Train (Whoosh); Bidirectional-LSTM.*

Abstrak

Penelitian ini menganalisis sentimen media sosial *YouTube* terhadap kereta cepat (*Whoosh*) menggunakan algoritma *Bidirectional-LSTM*. Masalah yang diangkat adalah perlunya pemahaman yang lebih mendalam tentang persepsi publik terhadap proyek kereta cepat, yang dapat mempengaruhi penerimaan dan keberlanjutannya. Tujuan penulisan ini adalah untuk mengevaluasi performa algoritma *Bidirectional-LSTM* dalam menganalisis sentimen dibandingkan dengan algoritma lain. Metode yang digunakan melibatkan pengumpulan data komentar *YouTube*, *preprocessing* teks, dan penerapan algoritma *Bidirectional-LSTM* untuk klasifikasi sentimen. Parameter-parameter yang dianalisis meliputi akurasi, presisi, dan ketahanan terhadap variasi data. Hasil penelitian menunjukkan bahwa algoritma *Bidirectional-LSTM* mencapai akurasi (0.86) yang secara signifikan lebih tinggi dibandingkan dengan algoritma *Multinomial Naïve Bayes* (0.80), *USE-Transfer learning* (0.80), dan *Text Vectorization and Embedding* (0.80). Simpulan penelitian ini adalah bahwa *Bidirectional-LSTM* lebih efektif dan andal dalam menganalisis sentimen media sosial *YouTube* kereta cepat (*Whoosh*).

Kata kunci: *Analisis Sentimen; Youtube Kereta Cepat (Whoosh); Bidirectional-LSTM;*

1. Pendahuluan

Kereta berkecepatan tinggi telah lama menjadi simbol kemajuan teknologi di bidang transportasi, membawa visi perjalanan yang cepat, efisien, dan nyaman bagi masyarakat di berbagai belahan dunia. Namun, di era ini, dengan meningkatnya tuntutan mobilitas dan kebutuhan akan solusi transportasi berkelanjutan yang lebih efisien, lahirlah sebuah inovasi yang menjanjikan revolusi baru dalam dunia kereta cepat: Kereta Cepat (*Whoosh*). Nama (*Whoosh*) bukan sekadar nama, tapi juga menggambarkan sensasi perjalanan cepat seolah-olah kita sedang bergerak di udara dengan kecepatan yang mengesankan. Dengan teknologi mutakhir yang memadukan prinsip levitasi magnetik (*maglev*) dan desain aerodinamis canggih, kereta cepat (*Whoosh*) memperkenalkan paradigma baru dalam transportasi berkecepatan tinggi.

Meskipun perkembangan teknologi kereta cepat seperti (*Whoosh*) membawa banyak manfaat, penerimaan publik terhadap proyek ini masih menjadi tantangan besar. Sentimen masyarakat terhadap kereta cepat ini sangat bervariasi, dengan adanya pandangan positif maupun negatif yang muncul di berbagai platform media sosial, termasuk *YouTube*. Hal ini

menunjukkan adanya gap antara kondisi ideal yang diharapkan dengan kondisi realitas yang ada. Kondisi idealnya adalah proyek ini diterima secara positif oleh seluruh lapisan masyarakat, mendukung keberlanjutannya dan memaksimalkan manfaat ekonomi dan sosial. Namun, kenyataannya, banyak komentar negatif yang menyoroti berbagai masalah, seperti biaya yang tinggi dan dampak lingkungan, yang dapat menghambat penerimaan dan kesuksesan proyek ini.

Untuk mengatasi masalah tersebut, penelitian ini mengusulkan penggunaan algoritma Bidirectional-LSTM dalam analisis sentimen media sosial *YouTube* terkait proyek kereta cepat (*Whoosh*). Algoritma *Bidirectional-LSTM* dipilih karena kemampuannya dalam menangkap konteks dari kedua arah, sehingga mampu memberikan pemahaman yang lebih baik terhadap struktur dan nuansa kalimat dibandingkan dengan model-model sebelumnya seperti *Multinomial Naive Bayes* dan *USE-Transfer learning (Tensorflow)*. Penelitian terdahulu menunjukkan bahwa Bidirectional-LSTM memiliki performa yang lebih baik dalam hal akurasi, presisi, recall, dan F1-score dalam berbagai aplikasi analisis sentimen [1]. Dengan mengadopsi model ini, diharapkan mampu memberikan gambaran yang lebih akurat tentang sentimen publik dan mendukung pengambilan keputusan yang lebih baik dalam pengelolaan proyek kereta cepat (*Whoosh*).

Penelitian ini bertujuan untuk mengevaluasi kinerja algoritma *Bidirectional-LSTM* dalam menganalisis sentimen dibandingkan dengan algoritma lain, serta untuk memahami persepsi publik terhadap proyek kereta cepat (*Whoosh*) melalui analisis komentar di *YouTube*. Penelitian ini tidak hanya meningkatkan akurasi analisis sentimen, tetapi juga memberikan wawasan yang lebih mendalam tentang pandangan masyarakat, yang dapat digunakan sebagai dasar untuk strategi komunikasi dan kebijakan yang lebih efektif. Dengan demikian, penelitian ini berkontribusi pada pengembangan teknologi transportasi yang lebih baik dan berkelanjutan, serta meningkatkan penerimaan publik terhadap inovasi tersebut.

2. Tinjauan Pustaka

Kereta cepat, juga dikenal sebagai *high-speed rail (HSR)*, telah menjadi fokus utama dalam upaya meningkatkan konektivitas transportasi di berbagai negara. Studi-studi sebelumnya telah menyoroti keuntungan ekonomi, sosial, dan lingkungan dari pengembangan jaringan kereta cepat. Dan terkait analisis sentimen, di mana para peneliti telah menggunakan algoritma yang beragam untuk menangani permasalahan ini, berikut merupakan rangkuman dari penelitian tersebut :

- 1) Penelitian yang dilakukan oleh Johannes Florensus Sianipar, Yudhi Raymond Ramadhan, dan Irsan Jaelani ini bertujuan sentimen analisis pengguna *Twitter* terhadap pembangunan kereta cepat menggunakan metode *Naive Bayes*. Dari 2.390 data tentang kereta cepat Jakarta-Bandung yang dikumpulkan, setelah proses pembersihan data, tersisa 2.007 data yang dianalisis. Hasil analisis menunjukkan bahwa terdapat 673 tweet dengan komentar negatif, 668 tweet dengan komentar positif, dan 665 tweet dengan sentimen netral, dengan akurasi 71%, presisi 73%, dan recall 89%. Hasil penelitian ini menunjukkan bahwa sentimen publik terhadap pembangunan kereta cepat Jakarta-Bandung cenderung negatif. [1]
- 2) Penelitian ini dilakukan oleh Alam Rizki Fitriansyah dan Yuliant Sibaroni, yang menggunakan data berupa tweet dari kata kunci yang telah ditentukan serta menerapkan *GloVe word embedding* dan metode klasifikasi *Support Vector Machine*. penelitian ini, kombinasi parameter *GloVe* terbaik dengan nilai 200 untuk *no_of_component*, 0.001 untuk *learning_rate*, dan fitur TOP 1 menghasilkan peningkatan akurasi klasifikasi *SVM* dari 72.63% menjadi 77.72%, dibandingkan dengan *SVM* yang tidak menggunakan fitur ekspansi *GloVe*. [2]
- 3) Penelitian ini dilakukan oleh Tifani Agustiranti, Aulia Khalfani Izzati Kurdiana, Bilal Al Ghiffari, Elza Dwi Juniar, dan Diki Gita Purnama, menganalisis sentimen masyarakat penggunaan kereta cepat dengan memanfaatkan data ulasan yang diambil dari media sorhadap sial *Twitter*. Perbandingan akurasi antara kedua metode menunjukkan bahwa metode berbasis leksikon menghasilkan akurasi sebesar 67%, sementara metode *Naive Bayes Classifier* mencapai 78%. [3]
- 4) Penelitian yang dilakukan oleh Citra Pricylia Ananda Mulya, Pramudya Nugraha, dan Imam Santoso ini bertujuan menganalisis kelayakan kebijakan Transportasi Kereta Cepat dalam menciptakan transportasi yang ramah lingkungan. Hasil penelitian menunjukkan bahwa algoritma KNN menghasilkan akurasi dan AUC dengan tingkat keakuratan 82,70%, yang dikategorikan sebagai klasifikasi yang sangat baik. [4]
- 5) Penelitian ini dilakukan oleh Siti Nurwahyuni metode penelitian ini yaitu yang banyak digunakan dalam mengolah data dalam bentuk *text mining* dengan tingkat akurasi yang cukup

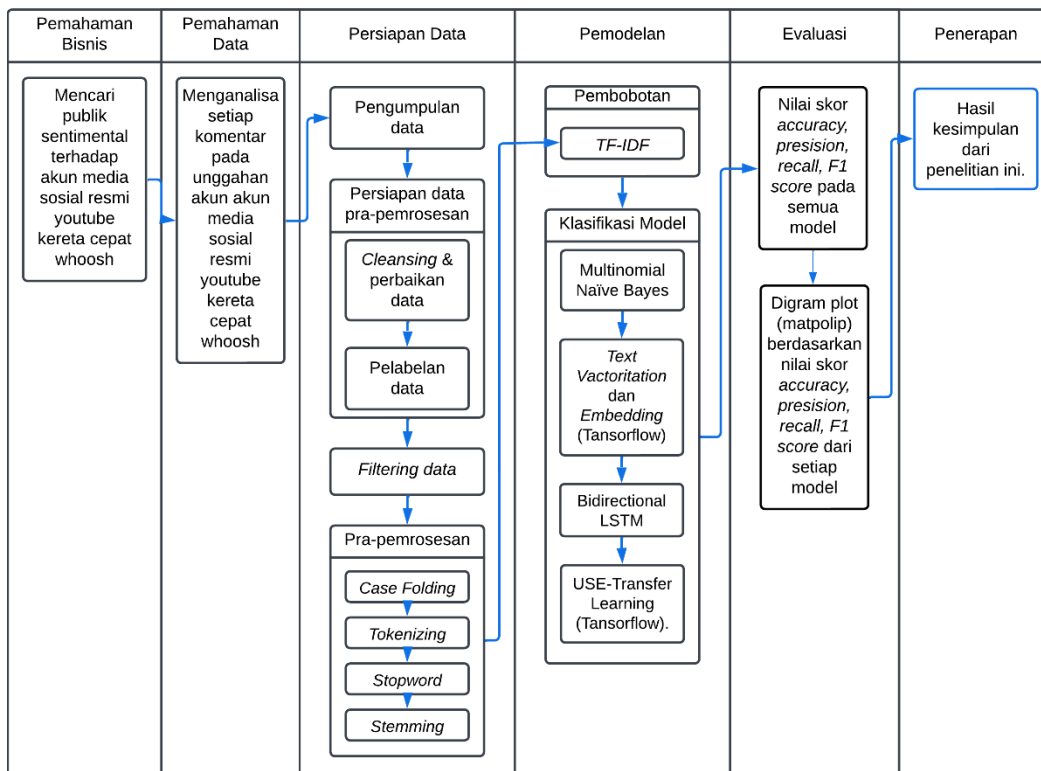
tinggi adalah algoritma *Naive Bayes Classifier (NBC)*. Menggunakan algoritma *Naive Bayes Classifier (NBC)* dengan hasil sebesar 84.00. [5]

Kebaruan penelitian ini mengusulkan penggunaan algoritma *Bidirectional-LSTM* dalam analisis sentimen media sosial *YouTube* terkait proyek kereta cepat Jakarta-Bandung. Algoritma *Bidirectional-LSTM* ini terbukti memberikan hasil yang lebih baik atau lebih akurat dibandingkan dengan metode lain seperti *Multinomial Naive Bayes* dan *USE-Transfer learning (Tensorflow)*. Hasil penelitian menunjukkan bahwa *Bidirectional-LSTM* memiliki performa terbaik dengan akurasi 0.86, presisi 0.85, recall 0.99, dan F1-score 0.92, yang mengungguli model-model sebelumnya yang telah digunakan dalam penelitian analisis sentimen di konteks serupa. Penelitian ini tidak hanya meningkatkan akurasi dari analisis sentimen, tetapi juga memperkenalkan penggunaan metode transfer learning dalam pengolahan data teks, yang belum banyak diterapkan dalam penelitian terdahulu

Penelitian ini menawarkan peningkatan signifikan dalam analisis sentimen dengan mengadopsi model *Bidirectional-LSTM* yang mengungguli model-model sebelumnya seperti *Multinomial Naive Bayes* dan *USE-Transfer learning (Tensorflow)*. Keunggulan *Bidirectional-LSTM* terletak pada kemampuannya menangkap konteks dari kedua arah, yang menghasilkan pemahaman yang lebih baik terhadap struktur dan nuansa kalimat. Berbeda dengan penelitian terdahulu yang menggunakan model statistik sederhana atau embedding kata seperti *GloVe* dan metode klasifikasi dasar seperti *SVM* dan *KNN*, penelitian ini memanfaatkan kemampuan transfer learning dan embedding yang lebih kompleks untuk meningkatkan performa analisis sentimen. Penggunaan parameter-parameter canggih dan teknik pemrosesan data yang lebih mutakhir menempatkan penelitian ini pada garis depan dalam bidang analisis sentimen media sosial, khususnya dalam konteks proyek infrastruktur besar seperti kereta cepat Jakarta-Bandung

3. Metodologi

Bab ini menjelaskan langkah-langkah yang diambil selama penelitian berlangsung. Untuk memudahkan pemahaman mengenai tahapan-tahapan yang digunakan, langkah-langkah tersebut akan digambarkan dalam bentuk diagram di Gambar 1.



Gambar 1. Metodologi

3.1 Pemahaman Bisnis

Kereta cepat *whoosh* merupakan proyek besar di industri transportasi yang menawarkan layanan Kereta cepat berkecepatan tinggi untuk menghubungkan berbagai tujuan. Bisnis kereta cepat *whoosh* dapat dirancang dan dijalankan secara efektif untuk memberikan layanan transportasi yang andal, efisien dan nyaman bagi penumpangnya.

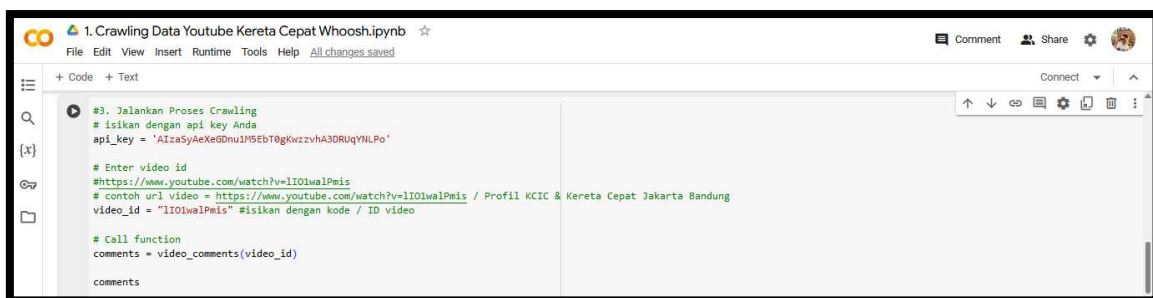
3.2. Pemahaman Data

Untuk memahami data sebagai bahan penelitian, peneliti melakukan observasi dengan cara melihat komentar dari postingan yang ada di akun media sosial resmi youtube kereta cepat *whoosh* dan menunjukkan beberapa komentar negatif dan positif terhadap pelayanan kereta cepat *whoosh*. Oleh karena itu, dilakukan penelitian mendalam untuk menganalisis sentimen publik terhadap kereta cepat *whoosh* dengan fokus pada akun media sosial resmi *youtube* kereta cepat *whoosh*.

3.3 Persiapan Data

3.3.1 Pengumpulan Data

dilakukan pada tanggal 22 Maret 2024. Sumber data berasal dari akun media sosial resmi youtube kereta cepat *whoosh*. Sumber data diambil berdasarkan komentar postingan awal pada youtube kereta cepat *whoosh*. Peneliti hanya mengambil 1148 komentar di awal postingan youtube kereta cepat *whoosh*. Pengambilan data komentar *youtube* menggunakan *crawling* dan menggunakan bahasa pemrograman python dengan menggunakan *API Key Google Colab* kemudian data yang dikumpulkan akan disimpan ke pada file ekstensi *.csv* proses *crawling* dapat dilihat di gambar 2



Gambar 2. Crawling Data

Proses ini merupakan mengubah data tidak terstruktur menjadi terstruktur, sehingga dapat disimpan dalam database. Tujuan utama dari *crawling* adalah mengumpulkan, menyimpan, dan memvalidasi data agar data tersebut dapat didefinisikan menjadi informasi yang bermanfaat.[2]

3.3.2 Persiapan Data Pra-pemrosesan

Langkah pertama dalam mempersiapkan data pra-pemrosesan yaitu melakukan perbaikan kata. Perbaikan kata ini dilakukan dengan cara menerjemahkan komentar yang berbahasa inggris menjadi bahasa indonesia menggunakan library *Googletrans* pada program python, dan juga memperbaiki kalimat kata-kata singkatan, memperbaiki kata-kata yang mengandung kesalahan penulisan (*typo*).

Langkah kedua adalah data *cleaning*, yaitu tahap untuk membersihkan data dan memperbaiki inkonsistensi dengan cara menghapus tanda baca, *hashtag*, *URL*, simbol, angka, dan atribut yang kosong. [1]

Langkah ketiga yaitu melakukan pelabelan *Sentiment analysis* atau *opinion mining* merujuk pada area luas dalam pengolahan bahasa alami, *komputasi linguistik*, dan *text mining* digunakan untuk menganalisis pendapat, sentimen, evaluasi, sikap, penilaian, dan emosi individu. Analisis sentimen dikategorikan ke dalam kategori positif dan negatif. Setelah proses pelabelan, data komentar akan dibersihkan secara manual dengan menghapus komentar yang duplikat menggunakan aplikasi *microsoft excel*. [2]

3.3.3 Filtering Data

Filtering Data merupakan proses penyeleksian komentar. Komentar yang sudah terfilter nantinya akan digunakan untuk proses pemodelan. Filtering data bertujuan untuk menyeleksi komentar yang berbobot, penghapusan duplikasi komentar, dan juga agar data yang disajikan jumlahnya mendekati seimbang antara komentar positif dan negatif.

3.3.4 Pra-pemrosesan

Proses *preprocessing* adalah langkah krusial untuk memperbaiki kualitas data mentah sebelum diproses lebih lanjut. Pada tahap ini, data komentar yang masih terdapat simbol, *hashtag*, dan kata tidak relevan dapat dibersihkan. Pra-pemrosesan dalam penelitian ini meliputi *cleansing*, *case folding*, *tokenisasi*, *stopword removal*, dan *stemming*. Proses ini mencakup beberapa kegiatan modifikasi data, yaitu [3]:

- 1) *Cleansing*, proses menghapus karakter, *mention*, *hashtag*, *URL*, dan karakter lainnya selain huruf.
- 2) *Case folding*, proses mengubah semua teks menjadi huruf kecil.
- 3) *Tokenizing*, proses memotong atau memisahkan kata yang dipisahkan oleh *whitespace* atau spasi.
- 4) *Stopword removal*, proses menghilangkan kata yang sering muncul dalam teks tetapi tidak memiliki arti penting.
- 5) *Stemming*, proses mengubah semua kata dalam teks menjadi bentuk dasar dengan menghilangkan awalan, sisipan, dan akhiran.

3.4 Pemodelan

3.4.1 Pembobotan

TF-IDF, atau *Term Frequency-Inverse Document Frequency*, yaitu nilai numerik yang digunakan untuk pemrosesan bahasa alami untuk menunjukkan seberapa penting suatu kata dalam satu dokumen dibandingkan dengan sekumpulan dokumen lainnya. [6] Ada beberapa tahapan klasifikasi yang menggunakan TF-IDF, termasuk menjalankan klasifikasi dengan *Naive Bayes*. TF-IDF bertujuan menghitung bobot setiap kata yang digunakan untuk fitur. Semakin banyak dokumen yang diproses, semakin banyak fitur yang dikumpulkan[3]

3.4.2 Klasifikasi Model

Semua proses pemodelan dengan menggunakan program python. Sebelum melakukan mengklasifikasikan komentar menggunakan model *Multinomial Naive Bayes*, *Text Vectorization and Embedding (Tensorflow)*, *Bidirectional-LSTM*, *USE-Transfer learning (Tensorflow)* harus dilakukan *training data* TDF IF. Pada tahap ini bertujuan untuk mencari *accuracy*, *precision*, *recall*, *F1-score* pada setiap model.[4]

3.5 Evaluasi

Klasifikasi mencakup tahap evaluasi yang melibatkan *confusion matrix* yaitu tabel yang membantu menentukan seberapa akurat proses klasifikasi yang telah dilakukan. *Visualisasi dashboard* digunakan untuk menampilkan data yang telah diproses dalam bentuk grafik. [1] Tahap ini mencakup menampilkan dan membandingkan nilai *Accuracy*, *Precision*, *Recall*, dan *F1-score* pada model-model seperti *Multinomial Naive Bayes*, *Text Vectorization and Embedding (TensorFlow)*, *Bidirectional-LSTM*, dan *USE-Transfer learning (TensorFlow)*, serta memvisualisasikan hasil perbandingan tersebut dengan diagram.

3.6 Penerapan

Penerapan adalah hasil kesimpulan dari proses penelitian. Pada tahap ini juga memberikan kesimpulan dari perbandingan nilai akurasi, presisi, *recall*, dan *F1-score* dari model *Multinomial Naive Bayes*, *Text Vectorization and Embedding (Tensorflow)*, *Bidirectional-LSTM*, *USE-Transfer learning (Tensorflow)*. Penulis juga memberikan saran berdasarkan hasil penelitian ini. [5]

4. Hasil dan Pembahasan

Dalam tahap perbaikan kata, dilakukan perbaikan kata-kata singkatan, kata-kata yang mengandung kesalahan penulisan dan juga menerjemahkan kalimat yang berbahasa Inggris ke

bahasa Indonesia. Untuk tahap cleansing, penghapusan atribut tidak jelas dan penghapusan *mention* nama orang. [7] Berikut komentar yang telah diperbaiki dan cleansing dari akun sosial media resmi Youtube Kereta Cepat Whoosh yang akan dijabarkan pada tabel 1

Tabel 1. Perbaikan Kata dan *Cleansing* Youtube

No.	Sebelum	Perbaikan Kata & <i>Cleansing</i>
1.	Keren menuju Indonesia maju Mantap 🤝👍👍	Keren menuju Indonesia maju Mantap
2.	#Semoga berkah buat masyarakat Indonesia 🙏❤️🙏	Semoga berkah buat masyarakat Indonesia
—	—	—
1162.	KEMAHALAN TIKETNYA !!! 😞	KEMAHALAN TIKETNYA
1148.	Mana mungkin @ Proyek ini memajukan ekonomi masyarakat !	Mana mungkin Proyek ini memajukan ekonomi masyarakat

Langkah selanjutnya adalah melakukan pelabelan sesuai dengan sifat dari komentar tersebut dan dikategorikan menjadi komentar positif dan negatif. Namun ada beberapa komentar tidak bisa diidentifikasi sehingga dihapus, sehingga dataset untuk Youtube Kereta Cepat Whoosh menjadi. [8] Berikut komentar yang telah diberi label dari akun sosial media resmi Youtube Kereta Cepat Whoosh yang akan dijabarkan pada tabel 2.

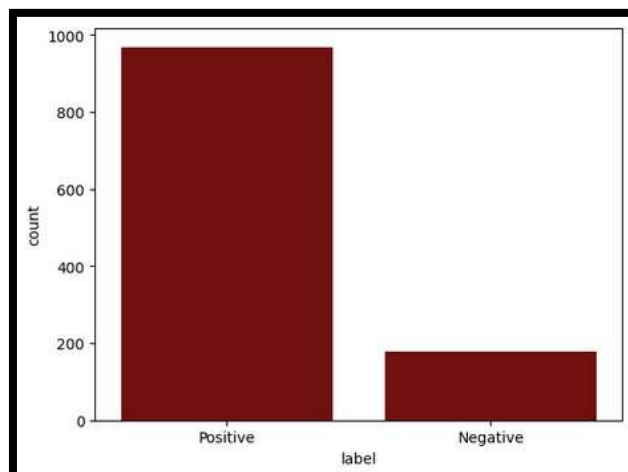
Tabel 2. Perbaikan dan *Cleansing* Komentar Youtube Kereta Cepat Whoosh

No.	Perbaikan Kata & <i>Cleansing</i>	Label
1.	Keren menuju Indonesia maju Mantap	Positif
2.	Semoga berkah buat masyarakat Indonesia	Positif
—	—	—
1162.	KEMAHALAN TIKETNYA	Negatif
1148.	Mana mungkin Proyek ini memajukan ekonomi masyarakat	Negatif

Jumlah komentar berdasarkan setiap label akan dijabarkan di tabel 3.

Tabel 3. Jumlah Komentar Berdasarkan Label

Sosial Media	Komentar Negatif	Komentar Positif	Total
Youtube	179 (15.59%)	969 (84.41%)	1148 (100%)



Gambar 3. Hasil Labelling

Gambar diatas adalah sebuah bar chart yang menunjukkan distribusi label pada dataset. Sumbu X (*Label*): Menunjukkan dua kategori label *Positive* (Positif): Menunjukkan jumlah sampel yang diberi label positif. *Negative* (Negatif) Menunjukkan jumlah sampel yang diberi label negatif. Sumbu Y (*Count*) Menunjukkan jumlah atau frekuensi dari setiap label dalam dataset. Interpretasi Bar Positif Jumlah sampel dengan label positif adalah sekitar 1000. Bar Negatif Jumlah sampel dengan label negatif adalah sekitar 200. Dataset ini menunjukkan jumlah sampel positif jauh lebih banyak dibandingkan dengan jumlah sampel negatif. Ini berarti ada banyak lebih banyak contoh kelas positif dibandingkan dengan kelas negatif. Implikasi untuk Machine Learning Ketidakseimbangan data seperti ini dapat menyebabkan model machine learning menjadi bias terhadap kelas mayoritas (kelas positif), dan kurang mampu mengenali atau memprediksi kelas minoritas (kelas negatif). Perlu adanya teknik penanganan data tidak seimbang seperti oversampling kelas minoritas, undersampling kelas mayoritas, atau menggunakan algoritma yang dirancang untuk menangani data tidak seimbang.

4.3 Filtering Data

Sebelum melakukan pra-pemrosesan, pembobotan dan klasifikasi model, dilakukan *filtering data* yang akan menjadi sumber data, karena komentar positif dan negatif perbedaanya terlalu jauh dan juga komentar yang seleksi adalah komentar yang berbobot, penghapusan duplikasi komentar. [9]Data yang digunakan dalam proses pembobotan dan klasifikasi model akan dijelaskan dalam tabel 4.

Tabel 4. *Filtering Data* Komentar

Sosial Media	Komentar Negatif	Komentar Positif	Total
Youtube	191	957	1148

4.4 Pra-pemrosesan

4.4.1 Case Folding

Proses mengubah *data teks* menjadi huruf kecil. [10] Berikut komentar yang telah dilakukan proses *case folding* dari akun sosial media resmi Youtube Kereta Cepat Whoosh yang akan dijabarkan pada tabel 5.

Tabel 5. *Case Folding* Youtube Kereta Cepat Whoosh

No.	Perbaikan Kata & <i>Cleansing</i>	<i>Case Folding</i>
1.	Keren menuju indonesia maju Mantap	keren menuju indonesia maju mantap
2.	Semoga berkah buat masyarakat Indonesia	semoga berkah buat masyarakat indonesia
—	—	—
1162.	KEMAHALAN TIKETNYA	kemahalan tiketnya
1148.	Mana mungkin Proyek ini memajukan ekonomi masyarakat	mana mungkin proyek ini memajukan ekonomi masyarakat

4.4.2 Tokenizing

Metode membagi teks menjadi komponen lebih kecil, seperti kata, kalimat, atau *bigram*. [11] Berikut komentar yang telah dilakukan proses *tokenizing* dari akun sosial media resmi Youtube Kereta Cepat Whoosh yang akan ditampilkan pada tabel 6.

Tabel 6. *Tokenizing* Youtube Kereta Cepat Whoosh

No.	<i>Case Folding</i>	<i>Tokenizing</i>
1.	keren menuju indonesia maju mantap	['keren', 'menuju', 'indonesia', 'maju', 'mantap']
2.	semoga berkah buat masyarakat indonesia	['semoga', 'berkah', 'buat', 'masyarakat', 'indonesia']
—	—	—
153.	kemahalan tiketnya	['kemahalan', 'tiketnya']
154.	mana mungkin proyek ini memajukan ekonomi masyarakat	['mana', 'mungkin', 'proyek', 'ini', 'memajukan', 'ekonomi', 'masyarakat']

4.4.3 Stopword

Proses penghapusan kata yang tidak diperlukan dalam pengolahan data, atau kata yang tidak memiliki makna signifikan [12] *Stopword* tahap mengambil kata penting dari hasil membuang kata kurang penting dari proses *tokenizing*. Berikut komentar yang telah dilakukan proses *stopword* dari akun sosial media resmi Youtube Kereta Cepat Whoosh yang akan dijabarkan pada tabel 7.

Tabel 7. *Stopword* Youtube Kereta Cepat Whoosh

No.	<i>Tokenizing</i>	<i>Stopword</i>
1.	['keren', 'menuju', 'indonesia', 'maju', 'mantap']	['keren', 'menuju', 'indonesia', 'maju', 'mantap']
2.	['semoga', 'berkah', 'buat', 'masyarakat', 'indonesia']	['semoga', 'berkah', 'buat', 'masyarakat', 'indonesia']

No.	Tokenizing	Stopword
—	—	—
153.	['kemahalan', 'tiketnya']	['kemahalan', 'tiketnya']
154.	['mana', 'mungkin', 'proyek', 'ini', 'memajukan', 'ekonomi', 'masyarakat']	['mana', 'mungkin', 'proyek', 'ini', 'memajukan', 'ekonomi', 'masyarakat']

4.4.4 Stemming

Proses mengubah kata dalam *teks* menjadi bentuk dasar dengan menghilangkan awalan, sisipan, dan akhiran. [13] *Stemming* merupakan proses menghilangkan imbuhan dalam kata. Berikut komentar yang telah dilakukan proses *stemming* dari akun sosial media resmi Youtube Kereta Cepat Whoosh yang akan ditampilkan pada tabel 8.

Tabel 8. *Stemming* Youtube Kereta Cepat Whoosh

No.	Stopword	Stemming
1.	['keren', 'menuju', 'indonesia', 'maju', 'mantap']	['keren', 'tuju', 'indonesia', 'maju', 'mantap']
2.	['semoga', 'berkah', 'buat', 'masyarakat', 'indonesia']	['moga', 'berkah', 'buat', 'masyarakat', 'indonesia']
—	—	—
153.	['kemahalan', 'tiketnya']	['mahal', 'tiket']
154.	['mana', 'mungkin', 'proyek', 'ini', 'memajukan', 'ekonomi', 'masyarakat']	['mana', 'mungkin', 'proyek', 'ini', 'maju', 'ekonomi', 'masyarakat']

4.5 Pemodelan

4.5.1 Pembobotan

Pembobotan menggunakan metode *TF-IDF*. Hasil dari pembobotan Youtube menggunakan metode *TF-IDF* akan diuraikan pada tabel 9.

Tabel 9. *TF-IDF* Youtube

(0, 515)	0.3389288520832195
(0, 431)	0.5052056022779143
—	—
(801, 86)	0.4123005261937698

4.5.2 Klasifikasi Model

Dalam penelitian ini, pengujian akurasi dilakukan menggunakan *Multinomial Naive Bayes*, *Text Vectorization* dan *Embedding (TensorFlow)*, *Bidirectional-LSTM*, serta *USE-Transfer Learning (TensorFlow)*. Pemodelan dilakukan dengan menggunakan *library MultinomialNB*, *Keras*, dan *TensorFlow*. [14] Hasil dari pemodelan akan dibahas secara rinci di Bab 4.6.

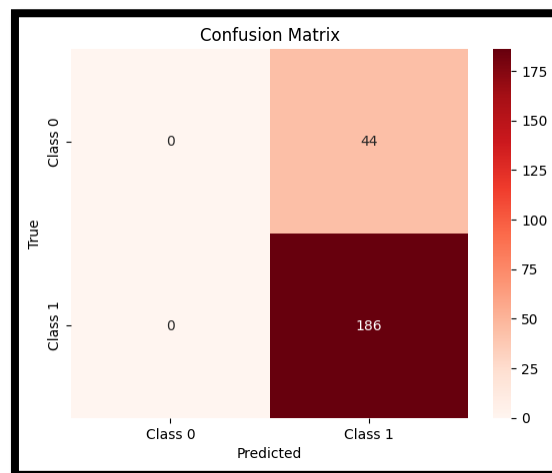
4.6 Evaluasi

Pada penelitian ini, dilakukan pengujian *Accuracy*, *Precision*, *Recall*, *F1-score* menggunakan model klasifikasi *Multinomial Naive Bayes*, *Text Vactoritation and Embedding (Tensorflow)*, *Bidirectional-LSTM*, *USE-Transfer learning (Tensorflow)* [15] pada data yang berasal dari *Youtube* yang ditampilkan pada gambar 3.

	accuracy	precision	recall	f1-score
Multinomial Naive Bayes (MultinomialNB)	0.808696	0.808696	1.000000	0.894231
Bidirectional-LSTM	0.860870	0.856481	0.994624	0.920398
USE-Transfer learning (Tensorflow)	0.804348	0.873016	0.887097	0.880000
Text Vactoritation and Embedding (Tensorflow)	0.808696	0.808696	1.000000	0.894231

Gambar 5. Hasil Klasifikasi Youtube

Model yang diuji dari gambar diatas adalah *Naive Bayes (MultinomialNB)*, *Bidirectional-LSTM*, *USE-Transfer learning (Tensorflow)*, *Text Vectorization and Embedding (Tensorflow)*. Metrik yang digunakan untuk evaluasi, *Accuracy* (Akurasi), Mengukur seberapa banyak prediksi yang benar dari total prediksi. *Precision* (Presisi), Mengukur proporsi prediksi positif yang benar dari keseluruhan prediksi positif disebut precision. *Recall* mengukur proporsi prediksi positif yang benar dari jumlah total data positif yang sebenarnya. *F1-Score* adalah rata-rata harmonis dari precision dan recall. Evaluasi dilakukan untuk setiap model, termasuk Multinomial Naive Bayes (MultinomialNB). *Accuracy* (0.80), *Precision* (0.80), *Recall* (1.00), *F1-Score* (0.89). *Bidirectional-LSTM Accuracy* (0.86), *Precision* (0.85), *Recall* (0.99), *F1-Score* (0.92). *USE-Transfer learning (Tensorflow Accuracy* (0.80), *Precision* (0.87), *Recall* (0.88), *F1-Score* (0.88). *Text Vectorization and Embedding (Tensorflow Accuracy* (0.80), *Precision* (0.80), *Recall* (1.00) *F1-Score* (0.89). Dari hasil ini, dapat dilihat bahwa model algoritma *Bidirectional-LSTM* memiliki performa terbaik dengan akurasi, presisi, dan recall yang tinggi, serta *F1-Score* yang paling tinggi di antara semua model yang diuji.

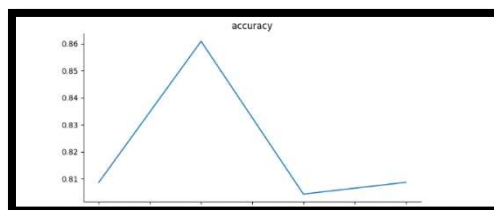


Gambar 6. Hasil Confusion Matrix

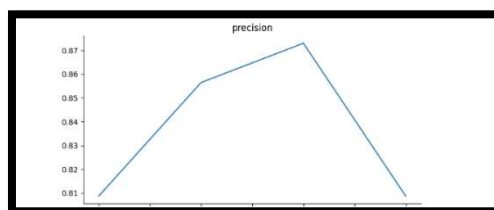
Confusion Matrix yang digunakan untuk mengevaluasi kinerja model klasifikasi. Sumbu Y (*True*): Menunjukkan kelas sebenarnya dari data. Class 0 kelas negatif. Class 1 kelas positif. Sumbu X (*Predicted*) Menunjukkan prediksi kelas yang dibuat oleh model. Class 0 Kelas negatif Class 1 Kelas positif. Matriks *True Positive* (TP) Kotak di baris *Class 1 (True)* dan kolom *Class 1 (Predicted)*, yaitu 186. Ini berarti model memprediksi 186 sampel sebagai positif (komentar) dan memang benar positif. *True Negative* (TN) Kotak di baris *Class 0 (True)* dan kolom *Class 0 (Predicted)*, yaitu 0. Ini berarti tidak ada sampel yang benar-benar negatif diprediksi sebagai

negatif oleh model. *False Positive* (FP) Kotak di baris *Class 0 (True)* dan kolom *Class 1 (Predicted)*, yaitu 44. Ini berarti model memprediksi 44 sampel sebagai positif (komentar) padahal sebenarnya negatif. *False Negative* (FN) Kotak di baris *Class 1 (True)* dan kolom *Class 0 (Predicted)*, yaitu 0. Ini berarti tidak ada sampel yang benar-benar positif (komentar) diprediksi sebagai negatif oleh model. Interpretasi Model ini memiliki *True Positive* yang sangat tinggi (186) dan *False Negative* yang sangat rendah (0), menunjukkan bahwa model sangat baik dalam mendeteksi sampel yang benar-benar positif (komentar). Namun, model memiliki *True Negative* yang sangat rendah (0) dan *False Positive* yang cukup tinggi (44), yang berarti model kurang baik dalam mendeteksi sampel yang benar-benar negatif. Warna merah yang lebih gelap menunjukkan jumlah yang lebih besar, sedangkan warna merah yang lebih terang menunjukkan jumlah yang lebih kecil. Secara keseluruhan, model ini sangat efektif dalam mendeteksi kasus positif, tetapi kurang efektif dalam mendeteksi kasus negatif.

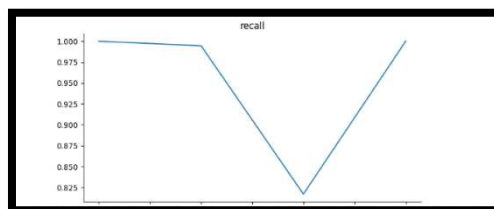
Hasil perbandingan klasifikasi pada tiap model berdasarkan *dataset* Youtube akan digambarkan pada gambar 7, 8, 9, 10



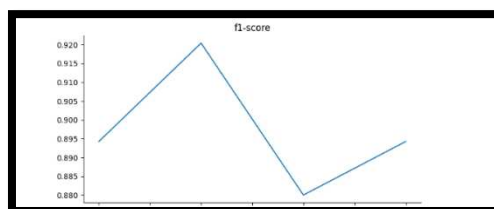
Gambar 7. Perbandingan *Accuracy* Tiap Model Pada Youtube



Gambar 8. Perbandingan *Precision* Tiap Model Pada Youtube



Gambar 9. Perbandingan *Recall* Tiap Model Pada Youtube



Gambar 10. Perbandingan *F1-Score* Tiap Model Pada Youtube

Berdasarkan grafik pada gambar 7, 8, 9, 10 Menunjukkan bahwa nilai tertinggi ada pada model Bidirectional-LSTM dengan *accuracy* (0.86), *precision* (0.85), *recall* (0.99), *F1-score* (0.92) sedangkan nilai terkecil ada pada model USE Transfer learning (Tensorflow) dengan *accuracy* (0.80), *precision* (0.87), *recall* (0.88), *F1-score* (0.88).

5. Simpulan

Kesimpulan Hasil penelitian menunjukkan bahwa algoritma berbasis *Bidirectional-LSTM* (0.86) secara signifikan mengungguli algoritma lainnya seperti *MultinomialNB* (0.80), *USE-Transfer learning (Tensorflow)* (0.80), dan *Text Vactoritation and Embedding (Tensorflow)* (0.80) dalam hal akurasi dan ketahanan terhadap variasi data, dari penelitian ini Ditemukan 179 (15.59%) komentar negatif dan 969 (84.41%) komentar positif pada *Youtube*. Nilai *accuracy* tertinggi *Youtube* terdapat pada model *Bidirectional-LSTM* sedangkan nilai akurasi terkecil ada pada model *USE Transfer learning (Tensorflow)*. Tingkat akurasi masih dibawah (0.90), Hal ini menunjukkan bahwa masih perlu adanya penelitian lebih lanjut agar mendapatkan nilai *accuracy* yang lebih baik lagi.

Daftar Referensi

- [1] J. F. Sianipar, Y. R. Ramadhan, and I. Jaelani, "Analisis Sentimen Pembangunan Kereta Cepat Jakarta-Bandung di Media Sosial Twitter Menggunakan Metode Naive Bayes," *Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 1, pp. 360–367, 2023, doi: 10.30865/klik.v4i1.1033.
- [2] A. R. Fitriansyah and Y. Sibaroni, "Analisis Sentimen Terhadap Pembangunan Kereta Cepat Jakarta-Bandung Pada Media Sosial Twitter Menggunakan Metode SVM dan GloVe Word Embedding," *e-Proceeding of Engineering*, vol. 10, no. 2, pp. 1713–1723, Apr. 2023, Accessed: Jul. 17, 2024. [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/20000>
- [3] T. Agustiranti, A. Khalfani Izzati Kurdiana, B. Al Ghiffari, E. Dwi Juniar, and D. Gita Purnama, "Penerapan Naive Bayes Terhadap Sentimen Analisis Media Sosial Twitter Pengguna Kereta Cepat Jakarta-Bandung (Whoosh)," *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, vol. 7, no. 1, pp. 297–305, 2024.
- [4] C. Pricylia, A. Mulya, P. Nugraha, and I. Santoso, "Analisis Sentimen Masyarakat Terhadap Pembangunan Kereta Cepat Jakarta-Bandung Menggunakan Algoritma K-Nearest Neighbors (KNN)," Jakarta, Jul. 2023. Accessed: Jul. 17, 2024. [Online]. Available: www.gataframework.com/textmining.Peneliti
- [5] Y. A. Singgalen, "Pemilihan Metode dan Algoritma dalam Analisis Sentimen di Media Sosial : Sistematic Literature Review," *Journal of Information Systems and Informatics*, vol. 3, no. 2, pp. 56-64, 2021, [Online]. Available: <http://journal-isi.org/index.php/isi>
- [6] D. Susilo and H. Harliantara, "Framing Analysis of Indonesian Fast Train on detik.com," *Gema Kampus IISIP YAPIS Biak*, vol. 18, no. 2, pp. 148–162, Oct. 2023, doi: 10.52049/gemakampus.v18i2.327.
- [7] S. Nurwahyuni, "Analisis Sentimen Aplikasi Transportasi Online Krl Access Menggunakan Metode Naive Bayes," *JURNAL SWABUMI*, vol. 7, no. 1, pp. 31–38, Mar. 2019.
- [8] F. Amaliah and I. K. D. Nuryana, "Perbandingan Akurasi Metode Lexicon Based Dan Naive Bayes Classifier Pada Analisis Sentimen Pendapat Masyarakat Terhadap Aplikasi Investasi Pada Media Twitter," *Journal of Informatics and Computer Science*, vol. 03, no. 03, pp. 384–393, 2022.
- [9] P. A. Permatasari, L. Linawati, and L. Jasa, "Survei Tentang Analisis Sentimen Pada Media Sosial," *Majalah Ilmiah Teknologi Elektro*, vol. 20, no. 2, pp. 177-186, Dec. 2021, doi: 10.24843/rite.2021.v20i02.p01.
- [10] S. Wahyu, "Konferensi Nasional Ilmu Komputer (KONIK) 2023 Perbandingan Model Algoritma Klasifikasi Pada Analisis Sentimen Opini Masyarakat Terhadap Layanan Kereta Cepat Jakarta Bandung (The Whoosh)," *Konferensi Nasional Ilmu Komputer (KONIK) 2023*, p.8, 2023, Accessed: Jul. 17, 2024. [Online]. Available: <https://prosiding.konik.id/index.php/konik/article/view/214>
- [11] S. Chohan, A. Nugroho, A. Maezar Bayu Aji, and W. Gata, "Analisis Sentimen Aplikasi Duolingo Menggunakan Metode Naïve Bayes dan Synthetic Minority Over Sampling

- Technique,” *Paradigma – Jurnal Informatika dan Komputer*, vol. 22, no. 2, pp. 141-150, Sep. 2020, doi: 10.31294/p.v21i2.
- [12] D. Novira and A. Lattu, “Analisis Sentimen Masyarakat Di Instagram Mengenai Kereta Cepat Jakarta-Bandung Menggunakan Algoritma Naïve Bayes ClassifieR,” *SISMATIK (Seminar Nasional Sistem Informasi dan Manajemen Informatika)*, pp. 213-223, Aug. 2023, Accessed: Jul. 17, 2024. [Online]. Available: <https://sismatik.nusaputra.ac.id/index.php/sismatik/article/view/214>
- [13] L. Yusuf and S. Masripah, “Sentimen Analisis Chatgpt Dengan Algoritma Naïve Bayes Dan Optimasi PSO,” *INTI Nusa Mandiri*, vol. 18, no. 1, pp. 59–64, Aug. 2023, doi: 10.33480/inti.v18i1.4230.
- [14] C. P. Yanti, N. L. W. S. R. Ginantra, D. A. P. Wulandari, and N. P. A. I. Paramita, “Komparasi Metode Single Moving Average dan Double Exponential Smoothing untuk Peramalan Penjualan Produk Gerabah pada UD. Amerta Sedana,” *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 3, pp. 536-545, Jun. 2022, doi: 10.30865/jurikom.v9i3.4143.
- [15] N. T. Romadloni, I. Santoso, and S. Budilaksono, “Perbandingan Metode Naive Bayes, Knn Dan Decision Tree Terhadap Analisis Sentimen Transportasi Krl Commuter Line,” *Jurnal IKRA-ITH Informatika*, vol. 3, no. 2, pp. 1–9, Jul. 2019, Accessed: Jul. 17, 2024. [Online]. Available: <https://ojs.upi-yai.ac.id/index.php/ikraith-informatika/article/view/311>