

Analisa Sentimen Terhadap Kereta Cepat Jakarta – Bandung Menggunakan Algoritma Naïve Bayes Dan K-Nearest Neighbor

Zaky Rahman Hakim¹, Sugiyono²

^{1,2}Sistem Informasi, STIKOM Cipta Karya Informatika, Jakarta Timur, Indonesia

Email Penulis Korespondensi: zakyrhmanhakim@gmail.com

Abstrak– Kereta cepat Jakarta – Bandung merupakan salah satu proyek besar di Indonesia, dengan transportasi pertama di Indonesia ini kereta cepat Jakarta – Bandung banyak menuai banyak opini masyarakat Indonesia tanpa terkecuali media sosial seperti twitter, di jejaring sosial ini banyak menuai pro kontra mengenai proyek kereta cepat Jakarta – Bandung karena banyak bagian yang diuntungkan dan tidak sedikit juga yang dirugikan. Oleh karena itu pada penelitian ini, akan mengambil tweets dari twitter dengan kata kunci pencarian kereta cepat Jakarta – Bandung untuk diolah dan mengklasifikasikan teks dengan menggunakan metode analisa sentimen. Tujuan dari penelitian ini adalah untuk menganalisis kelayakan kebijakan Transportasi Kereta Cepat Jakarta-Bandung apakah sudah sesuai dengan apa yang masyarakat Indonesia sekarang butuhkan atau malah sebaliknya. Penelitian ini menggunakan metode *Naïve Bayes* dan metode *K-Nearest Neighbor*. Untuk prosesnya analisis sentiment ini dilakukan dengan klasifikasi teks dibagi menjadi dua kelas yaitu kelas sentimen positif dan kelas sentimen negatif. Data yang digunakan berjumlah 2000 tweets yang terdiri dari 1000 data latih dan 1000 data uji.

Kata Kunci: Data Mining, Twitter, Kereta Cepat Jakarta - Bandung, *Naïve Bayes*, *K-Nearest Neighbor*

Abstract– The Jakarta – Bandung fast train is one of the big projects in Indonesia, with the first transportation in Indonesia, the Jakarta-Bandung fast train reaps a lot of opinions from the Indonesian people without exception on social media such as Twitter, in this social network there are many pros and cons regarding the Jakarta fast train project. – Bandung because there are many parties who benefit and not a few who are harmed. Therefore, in this research, we will take tweets from Twitter with the keyword search for the Jakarta-Bandung fast train to be processed and classify the text using the sentiment analysis method. The purpose of this study is to analyze the feasibility of the Jakarta-Bandung Fast Train Transportation policy whether it is in accordance with what Indonesian people now need or vice versa. This study uses the *Naïve Bayes* method and the *K-Nearest Neighbor* method. For the sentiment analysis process, the text classification is divided into two classes, namely the positive sentiment class and the negative sentiment class. The data used amounted to 2000 tweets consisting of 1000 training data and 1000 test data.

Keywords: Data Mining, Twitter, The Jakarta-Bandung fast train, *Naïve Bayes*, *K-Nearest Neighbor*

I. PENDAHULUAN

Pada saat ini teknologi komunikasi berkembang pesat, dan peran media sosial pada zaman sekarang sangatlah penting termasuk penggunaan media sosial seperti twitter yang banyak digemari masyarakat sekarang ini dengan penggunaanya yang mudah dan mampu menjadi media dengan salah satu pengaruh yang besar dengan berbagai opini tentang berbagai topik dan membahas isu-isu yang terjadi saat ini.

Jumlah pengguna aktif dalam komunikasi online memiliki jumlah data yang sangat banyak sehingga memunculkan data besar. Munculnya teknologi big data yang merupakan himpunan data dalam jumlah yang sangat besar, rumit, dan tidak terstruktur menjadi salah satu sumber daya besar yang dapat diolah untuk memperoleh hasil sentimen yang lebih akurat[1].

Berdasarkan penelitian Semiocast, lembaga riset media sosial yang berpusat di Paris, Prancis, mengatakan bahwa jumlah pemilik akun Twitter di Indonesia merupakan yang terbesar kelima didunia, dan berada pada posisi ketiga negara yang paling aktif mengirim Tweet perhari[2], dengan begitu media sosial twitter dapat dijadikan sebagai acuan untuk mengetahui sentiment dan menentukan kecenderungan opini masyarakat terhadap

suatu kejadian yang terjadi disekitar masyarakat. Salah satu topik yang masih hangat sampai sekarang yaitu proyek kereta cepat Jakarta – Bandung yang sudah dibangun sejak 2016. Kereta Cepat Jakarta – Bandung merupakan transportasi massal masa depan yang menggunakan energi listrik atau energi terbarukan nonfossil, dan menjadi isu strategis semua bangsa untuk mengurangi kerugian yang sangat besar yang membebani energi fosil (BBM) yang bersubsidi.

Proyek kereta cepat Jakarta – Bandung telah dimulai sejak tanggal 21 januari 2016 dengan dilakukannya groundbreaking oleh Jokowi di Perkebunan Mandalawangi Maswati, Cikalong Wetan, Bandung Barat, Jawa Barat. Menurut Rini mantan menteri BUMN, keuntungan dibangunnya kereta cepat Jakarta – Bandung diantaranya akan meningkatkan perekonomian, mengangkat sektor pariwisata, dan membuka lapangan pekerjaan yang baru. Namun, dalam pembangunannya terdapat beberapa masalah, seperti banjir yang terjadi di Bekasi dan menyebabkan kemacetan dan mengganggu kelancaran logistik. Permasalahan yang lain dalam proyek kereta cepat ini kurang memperhatikan kelancaran akses keluar – masuk jalan tol, pembiaran penumpukan material yang mengganggu fungsi drainase, pembangunan pilar LRT

tanpa izin, sampai persoalan keselamatan dan Kesehatan kerja (K3). Oleh karena itu, diperlukan analisis sentimen untuk mengetahui bagaimana sentimen yang ada pada media sosial mengenai pembangunan proyek kereta cepat Jakarta – Bandung.[3]

Dalam melakukan aktivitas utang luar negeri, Indonesia lebih banyak melakukan pinjaman kepada negara memiliki kemudahan prosedur dalam pengajuan utang serta tidak menerapkan syarat peminjaman ke Negara lain yang begitu menyulitkan, salah satunya ialah Negara tirai bambu Tiongkok, dengan salah satu contoh platform terbesar bagi Tiongkok untuk hadir dalam aktivitas pemberian pinjaman luar negerinya di Indonesia terletak pada pembiayaan proyek Kereta Cepat Jakarta – Bandung (KCBJ) di tahun 2015 sebesar Rp 13 Triliun dari keseluruhan struktur pembiayaan ini, 63% pinjaman dalam dollar AS bersumber dari China Development Bank (CDB), bank milik Tiongkok. CDB memberikan jangka waktu pengembalian utang itu hingga 40 tahun dengan tenor pelunasan 10 tahun.[4]

KAI Access dirilis sejak tahun 2014 dengan fitur yang terus berkembang hingga sekarang, dimana saat ini pemanfaatan teknologi pada pelayanan publik sudah menjadi hal yang umum. Hal ini menjadikan KAI Access sebagai layanan teknologi digital utama dalam pelayanan PT Kereta Api Indonesia[5].

Analisis sentimen adalah proses ekstraksi, Secara otomatis memproses dan memahami data dalam bentuk teks tidak terstruktur untuk mengambil informasi sentimen yang terkandung dalam kalimat opini atau opini[6].

Sentiment analisis adalah cabang dari data mining. Data, Mining, merupakan sebuah, proses, yang dapat mengekstrak informasi, sehingga menghasilkan informasi yang sangat berharga[7]. Tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, apakah pendapat yang dikemukakan dalam dokumen bersifat positif, negatif atau netral[8].

Ada banyak metode yang bisa digunakan dalam melakukan analisis sentimen akan tetapi penulis menggunakan dua metode yaitu metode Naïve bayes dan metode KNN. Algoritma Naïve bayes adalah metode machine learning yang menggunakan teorema kuno warisan abad ke-18 yang ditemukan oleh Thomas Bayes[9]. Penggunaan algoritma Naïve bayes pada penelitian ini untuk melakukan labeling secara otomatis pada sentiment jadi memudahkan dalam proses penelitian. Untuk metode algoritma K-Nearest Neighbor (K-NN) bertujuan untuk mengklasifikasikan objek berdasarkan atribut dan data training[10].

Naive Bayes dikembangkan oleh Reverend Thomas Bayes pada abad ke 18. Klasifikasi dengan metode Naïve Bayes secara umum dilakukan dengan pendekatan peluang atau probabilitas. Algoritma ini memprediksi probabilitas di masa depan berdasarkan pengalaman yang sudah ada di masa lalu. Dari sudut pandang peluang[11].

K-Nearest Neighbor (K-NN) ialah metode sederhana mudah diimplementasikan, data yang digunakan memiliki label sehingga memudahkan dalam proses

pengelompokkan ke dalam kelas yang paling sesuai dan memiliki keunggulan dapat mengklasifikasikan data dengan adanya data latih dan data uji serta memiliki kemudahan menerjemahkan hasil dan akurasi dari prediksi dengan secara akurat memilah terlebih dulu nilai k terdekat dengan tepat[12].

Tahapan – tahapan dasar yang dilakukan dalam menjalankan metode ini berupa crawling data, preprocessing, menambahkan label, menambahkan fitur, melakukan klasifikasi dan melakukan perhitungan sentimen.[3].

Pada penelitian ini bertujuan untuk dapat menerapkan konsep data mining menggunakan Algoritma K- Nearest Neighbors guna mengetahui presentase sentimen positif dan negatif tentang kereta cepat Jakarta – Bandung guna hasil tersebut akan menjadi pertimbangan untuk masyarakat agar mengetahui sentimen masyarakat.

II. METODE PENELITIAN

Pada penelitian ini metode yang digunakan adalah *Naïve Bayes* dan *K-Nearest Neighbors Algorithm Klasifikasi*. Data yang digunakan pada penelitian ini data tweet yang search menggunakan hashtag #keretacepatJakartaBandung.

A. Penerapan Metodologi

Data Mining adalah klasifikasi data, yaitu memetakan (mengklasifikasikan) data ke dalam satu kelas atau beberapa kelas yang sudah didefinisikan sebelumnya. Salah satu metoda dalam klasifikasi data adalah Naïve Bayes yaitu salah satu metode machine learning yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, cara kerja Naive Bayes yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya[13].

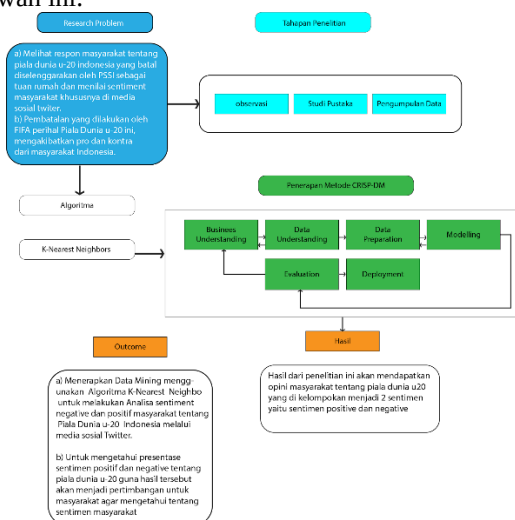
Penelitian ini menggunakan metode CRISP-DM dan rapidminer sebagai aplikasi yang digunakan untuk melakukan pengolahan datanya.

Rapidminer sendiri merupakan perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari Institute of Teknologi Blanchardstown dan Ralf Klinkenberg dari rapid-i.com dengan tampilan GUI (Graphical User Interface) sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini[14].

CRISP-DM (Cross-Industry Standard Process for Data Mining) yang di kembangkan tahun 1996 oleh analisis dari beberapa industry seperti Daimler Chrysler, SPSS dan NCR. CRISP-DM menyediakan standar proses data mining sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian. CRISP-DM adalah metode yang dapat diterapkan ke dalam strategi pemecahan masalah umum serta metodologi yang menyediakan standar baku untuk data mining[15].

Pada tahapan penelitian ini akan dilakukan pengumpulan data awal maka proses selanjutnya akan dilakukan pembersihan data, kemudian setelah itu dilakukan Cleansing Data, lalu pada tahap selanjutnya dilakukan Pemilihan Atribut, tahap berikutnya melakukan penerapan

metode CRISP-DM yang pada proses penelitian ini mengacu pada enam tahap CRISP-DM. Berikut tahapan metodologi penelitian yang dapat dilihat pada Gambar dibawah ini.



Gambar 1. Tahapan Penerapan Metodologi

Pada gambar diatas terdapat enam tahap dalam dalam CRISP-DM ini yakni dijelaskan sebagai berikut:

1) Pemahaman Bisnis (*Business Understanding*)

Pemahaman bisnis dilakukan dengan menganalisa objek penelitian yaitu analisa sentiment publik tentang kereta cepat Jakarta – Bandung. Pada tahap ini, tujuan bisnis dari penelitian ini adalah untuk menentukan sentiment public tentang kereta cepat Jakarta – Bandung. Untuk itu akan dilakukan pengambilan data dari twitter dengan menggunakan extention API twitter dari aplikasi rapid miner lalu setelah mendapatkan data tersebut akan melalui beberapa proses yaitu dengan menggunakan fitur replace RT , replace link, replace hastag, replace mention, replace symbol, trim, remove duplicates setelah melewati proses tersebut , maka akan menghasilkan data tweet yang sudah siap digunakan untuk melakukan analisa klasifikasi menggunakan algoritma K-Nearest Neighbors.

2) Pemahaman Data (*Data Understanding*)

Berdasarkan data yang telah didapat, tahap selanjutnya dalam metode CRISP-DM yaitu pemahaman terhadap kebutuhan data terkait dengan pencapaian tujuan dalam menentukan strategi penjualan yang efektif dan efisien. Data dari twitter yang telah di ambil menggunakan API twitter berjumlah 5000 tweet dengan 3 atribut yang akan dilakukan dengan cara pemilihan atribut menjadi tiga atribut yaitu tweet, positive dan negative. Data yang didapatkan dalam penelitian ini diperoleh tanggal 1 April sampai 7 April 2023. Setelah

data diperoleh, proses selanjutnya adalah melakukan replace dan cleansing data.

3) Data Preparation

Persiapan data mencakup semua proses untuk membangun dataset yang akan diterapkan ke dalam alat pemodelan, dari data mentah berupa data tweet publik dan selanjutnya akan dilakukan proses data mining. Fungsi utamanya yaitu khusus sebagai alat pemodelan classification. Persiapan data merupakan tahap yang padat dengan aktivitas pengolahan data.

4) Modeling

Tahap pemodelan akan menggunakan metode classification dengan algoritma Naïve bayes dan K-Nearest Neighbors. Dalam penerapan metode classification akan dibagi menjadi 3 atribut yang akan di kelompokkan menjadi tweet, positive dan negative. Alat pemodelan yang digunakan adalah Rapidminer.

5) Evaluation

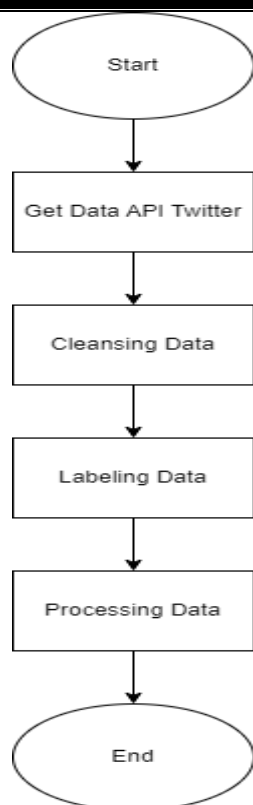
Pada tahap ini akan dilakukan analisa atau pengukuran ketepatan terhadap pemodelan yang telah dilakukan. Evaluasi ditujukan untuk mengetahui pemodelan yang dilakukan apakah sudah tepat dan sesuai diterapkan pada kasus penelitian ini serta sudah sesuai rencana awal penelitian. Selanjutnya dari hasil evaluasi tersebut adalah menentukan langkah berikutnya apakah bisa dilanjutkan atau diulang dari awal karena tidak sesuai dengan rencana awal. penelitian.

6) Deployment

Tahap selanjutnya yaitu tahap penyebaran hasil dari penelitian yang sudah dilakukan dijadikan sebagai laporan atau presentasi dari pengetahuan yang telah didapat berdasarkan pemodelan dan evaluasi pada proses data mining.

B. Proses Pengumpulan Data

Data tersebut diperoleh dari twitter dengan menggunakan API twitter, Tweet yang diambil berupa tweet berbahasa Indonesia dengan kata kunci kereta cepat Jakarta – Bandung dengan limit 10.000 dan berhasil di dapat berjumlah 5.726 data Tweet yang telah selesai melalui tahap *text preprocessing*.



Gambar 2. Flowchart Pengumpulan Data

Dalam gambar diatas bisa dijelaskan penelitian ini meliputi 4 proses yang di lakukan yaitu :

1) GET data API Twitter

Objek pada deskripsi data penelitian ini adalah opini pengguna twitter terhadap isu kereta cepat Jakarta – Bandung. Tweet yang diambil berupa tweet berbahasa Indonesia dengan kata kunci kereta cepat Jakarta – Bandung.

2) Cleansing data

Data tweet yang belum diolah melalui beberapa proses, yaitu: cleaning, tokenisasi, transform cases, stopword, dan filtering.

3) Labeling data

Data yang sudah melalui proses cleansing kemudian melakukan pelabelan secara manual. Pelabelan secara manual dilakukan pada seluruh tweet.

4) Processing data

Pada penelitian ini akan dilakukan pengujian menggunakan confusion matrix untuk melihat hasil pengujian data yang diperoleh dari tahapan modelling dengan menggunakan algoritma K-NN untuk menilai akurasi dari algoritma tersebut.

III. HASIL DAN PEMBAHASAN

Pada penelitian ini metode yang digunakan adalah K-Nearest Neighbors Algorithm Classification. Data yang digunakan ialah tweet yang diambil dari aplikasi media sosial twitter menggunakan API Twitter dengan cara

menggunakan ekstensi dari Rapidminer yaitu search twitter yang di ambil dari 1 April sampai 7 april 2023 setelah data tersebut di dapatkan lalu masuk ke proses penerapan CRISP-DM yaitu pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi dan penyebaran (Deployment). Data tersebut akan diproses dalam pengujian algoritma K-Nearest Neighbors menggunakan tools Rapidminer supaya dapat nilai akurasi klasifikasi yang baik dan optimal.

Pengujian ini dapat dilakukan dengan menggunakan metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang memiliki 6 tahapan yaitu :

A. Pemahaman Bisnis (*Business Understanding*)

Kereta cepat merupakan transportasi rel yang menggunakan kereta api yang jauh lebih cepat daripada kereta api tradisional, menggunakan sistem terintegrasi dari bakal pelanting khusus dan jalur khusus. Kereta cepat Jakarta – Bandung merupakan proyek kereta cepat pertama di Indonesia, banyak yang masyarakat yang menantikan kereta cepat ini dan hal ini membuat banyak masyarakat membicarakan perihal kereta cepat ini selain banyak keuntungan yang didapat dari adanya kereta cepat Jakarta – Bandung, akan tetapi ada juga pihak yang dirugikan dengan adanya kereta cepat ini. Oleh karena itu akan dianalisa sentiment mengenai kereta cepat Jakarta – Bandung.

B. Pemahaman Data (*Data Understanding*)

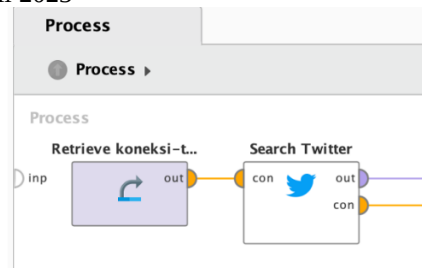
Pada tahapan ini maka peneliti mencoba untuk memahami data yang akan digunakan Data dari twitter yang telah di ambil menggunakan API twitter berjumlah 5000 tweet dengan 3 atribut yang akan dilakukan dengan cara pemilihan atribut menjadi tiga atribut yaitu tweet, positive dan negative. Data yang didapatkan dalam penelitian ini diperoleh tanggal 1 April sampai 7 April 2023.

C. Data Preparation

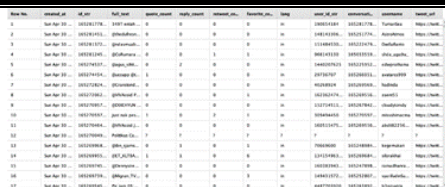
Pada tahap data preparation dilakukan beberapa tahapan antara lain sebagai berikut :

1) Pengumpulan data

Pengumpulan data awal yaitu melakukan pengumpulan data yang dibutuhkan untuk mendukung dalam melakukan pemahaman data. Adapun sumber data utama yang digunakan dalam penelitian ini adalah data tweet Kereta Cepat Jakarta-Bandung pada 1 April sampai 7 April 2023



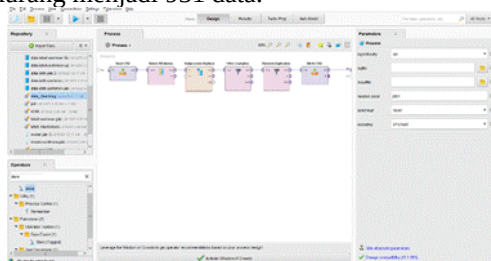
Gambar 3. Model Proses Pengambilan Data



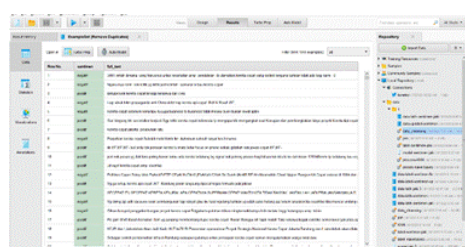
Gambar 4. Hasil Data Yang Di Dapatkan

2) Cleansing data

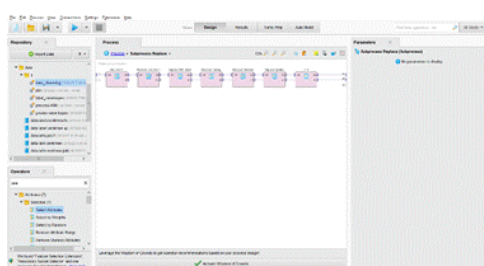
Pada tahapan ini peneliti akan melakukan proses Cleansing dengan menggunakan rapidminer sebelum melakukan pemberian label pada sentimen, agar dataset pada file csv terbaru terhindar dari duplikasi data dan tanda yang tidak diperlukan. Pada proses ini akan dilakukan pembersihan dari berbagai noise seperti menghilangkan link URL, username, retweet, digit angka dan karakter. Setelah selesai menjalankan proses Cleansing dengan menggunakan rapidminer. Awalnya terdapat 1.454 data tweet pada file csv, kemudian setelah melewati proses Cleansing maka data tweet berkurang menjadi 931 data.



Gambar 5. Model Cleansing



Gambar 6. Hasil Cleansing Data

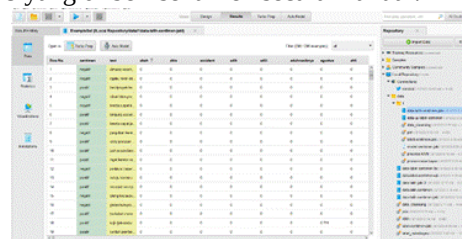


Gambar 7. Model Subproses Replace

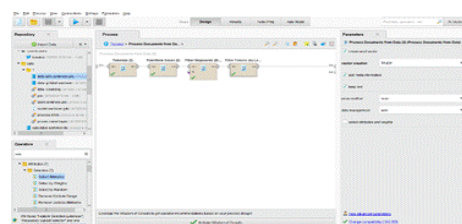
3) Labeling

Kemudian tahap selanjutnya adalah melakukan klasifikasi atau penentuan sentimen berdasarkan masing-masing tweet. Selama proses

penentuan sentimen berlangsung, disini peneliti melakukan pemberian sentimen secara *automatis* menggunakan algoritma *Naive Bayes*. Tahap pertama yaitu membuat data latih yang berjumlah 298 yang di beri sentimen secara manual.



Gambar 7. Hasil Data Latih Dan Tokenizing

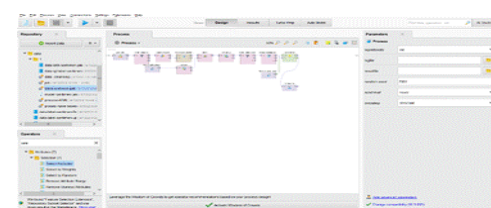


Gambar 8. Model Proses Tokenizing

Setelah itu *Naive Bayes* akan mempelajari data latih tersebut dan disimpan menggunakan attribute 'Store' pada Rapidminer.

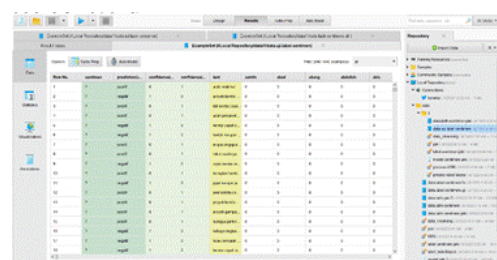


Gambar 9. Model Proses Pengujian Data Latih



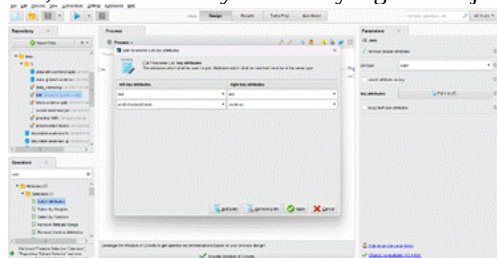
Gambar 10. Model Proses Labeling Automatis

Setelah di proses data, dengan data latih yang berjumlah 298 dan data uji berjumlah 646, maka hasil nya seperti pada gambar dibawah ini.

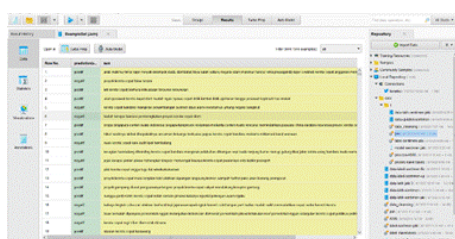


Gambar 11. Hasil Proses labeling Automatis

Setelah data sudah selesai di labeli secara otomatis, lalu peneliti melakukan penggabungan data, dari data latih dan data uji dengan menggunakan parameter 'join' yang ber tipe *outer*, lalu memilih *key attribute* yang mau di join.



Gambar 12. Proses Join Table

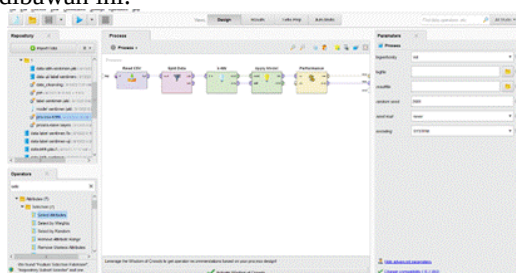


Gambar 13. Hasil Join Table

Data tersebut lalu di save dengan nama 'sentimen-kereta-jakarta-bdg.csv' Dan siap digunakan untuk proses selanjutnya yaitu melakukan klasifikasi menggunakan metode algoritma *K-Nearest Neighbor*

D. Modeling

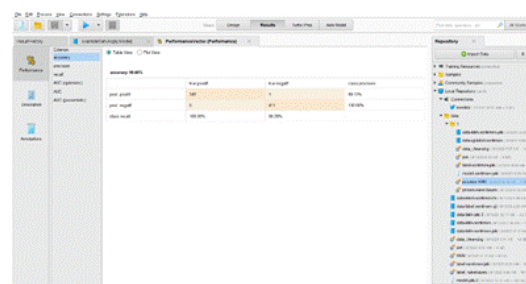
Pada tahapan ini peneliti akan melakukan preprocessing pada dataset hal ini ditunjukkan untuk menyiapkan data yang bersih dan data yang bebas dari noise. Pada proses ini juga untuk menghitung pembobotan kata untuk keperluan pada proses modelling nanti. Setelah melewati tahapan text preparation, maka jumlah data tweet didapatkan menjadi 944 data yang merupakan data bersih untuk dipakai pada tahap selanjutnya. Untuk tahapan modelling peneliti akan melakukan pengukuran performa klasifikasi dengan Pemodelling menggunakan split data. Pemodelan ini data akan dibagi menggunakan split menjadi 2 yaitu 0.8 dan 0.2. Pada tahapan ini akan dilakukan pengukuran performa klasifikasi dengan menggunakan split data. Berikut proses Modelling yang dilakukan di software RapidMiner yang ditunjukkan pada Gambar dibawah ini.



Gambar 13. Model Proses

E. Evaluasi

Pada penelitian ini akan dilakukan pengujian menggunakan confusion matrix untuk melihat hasil pengujian data yang diperoleh dari tahapan modelling dengan menggunakan algoritma K-NN. Total dataset yang dikumpulkan adalah 944 data. Berikut ini adalah confusion matrix hasil dari tahapan modelling yang dapat dilihat hasil dari perhitungan hasil RapidMiner dapat dilihat pada Gambar 4.14 untuk menunjukkan dan membuktikan hasil prediksi dari model K-Nearest Neighbor.



Gambar 14. Hasil Akurasi Confusion Matrix



Gambar 15. Grafik Area Under Curve (AUC)

Berdasarkan pada Gambar Diatas dapat dibuat kesimpulan bahwa precision menunjukkan tingkat ketepatan data yang diprediksi positif terhadap banyaknya data yang benar diprediksi positif yang menghasilkan presentase ketepatannya adalah 99.49%, sedangkan untuk data bersentimen negatif memiliki precision sebesar 100%. Untuk sentimen negatif memiliki recall (Specificity) adalah 99.53 % sehingga dapat disimpulkan bahwa model dapat menemukan kembali informasi atau data yang benar-benar negatif dengan sangat baik, sedangkan pada sentimen positif memiliki recall sebesar 100 %. Nilai accuracy yang dihasilkan menggunakan model K-Nearest Neighbor adalah 99.76 %, sehingga dapat disimpulkan bahwa algoritma K-Nearest Neighbor dapat mengklasifikasi sentimen dengan baik menggunakan data Kereta Cepat Jakarta-Bandung.

F. Deployment

Deployment merupakan tahap akhir dalam pembuatan laporan hasil kegiatan data mining. Laporan akhir yang berisi tentang pengetahuan yang diperoleh atau pengenalan pola dalam proses data mining.

IV. KESIMPULAN

Penelitian yang telah dilakukan mengenai klasifikasi opini masyarakat Indonesia terhadap kereta cepat Jakarta – Bandung dapat diambil kesimpulan yaitu hasil Sentimen masyarakat terhadap kereta cepat Jakarta – Bandung ditemukan 44,18% bersentimen positif dan 55,82% bersentimen negatif. Maka dari hasil tersebut dapat ditarik kesimpulan bahwa masyarakat Indonesia sebagian besar memiliki respon negatif terhadap kereta cepat Jakarta – Bandung dan berdasarkan klasifikasi model algoritma K-Nearest Neighbor, menggunakan split data perbandingan 0.8 : 0.2 dengan nilai $k = 3$ terhadap dataset kereta cepat Jakarta – Bandung, didapatkan nilai accuracy sebesar 99,76%. Sehingga dapat dikatakan bahwa algoritma K-KNN dapat mengklasifikasikan data secara baik dan benar.

UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini.

V. REFERENSI

- [1] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, p. 113, 2018, doi: 10.26418/jp.v4i2.27526.
- [2] P. Antinasari, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 12, pp. 1718–1724, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [3] A. R. FITRIANSYAH, "Analisis Sentimen Terhadap Pembangunan Kereta Cepat Jakarta - Bandung pada Media Sosial Twitter Menggunakan Metode SVM dan GloVe Word Embedding," *J. Tugas Akhir Fak. Inform.*, vol. 9, no. 5, pp. 6078–6083, 2022, [Online]. Available: <https://repository.telkomuniversity.ac.id/home/catalog/id/181602>
- [4] V. Ramadana Ovilia, P. Parameswari, and R. Muhaimin, "Analisis Dampak Kerja Sama Tiongkok-Indonesia melalui Aktivitas Utang Luar," vol. 2019, 2015.
- [5] N. N. AINII, "Klasifikasi Sentimen, Topik, Dan Detail Topik Dari Ulasan Kai Access Menggunakan Multilayer Perceptron (Mlp) Dan Bidirectional Long Short Term Memory (Bilstm)," 2022, [Online]. Available: <https://dspace.uui.ac.id/handle/123456789/40914%0Ahttps://dspace.uui.ac.id/bitstream/handle/123456789/40914/18523252.pdf?sequence=1&isAllowed=y>
- [6] D. Aryanti, "Analisis Sentimen Ibukota Negara Baru Menggunakan Metode Naive Bayes Classifier," *J. Inf. Syst. Res.*, vol. 3, no. 4, pp. 524–531, 2022, doi: 10.47065/josh.v3i4.1944.
- [7] E. N. Hamdana, "Pengembangan Sistem Analisis Sentimen Berbasis Java Pada Data Twitter Terhadap Omnibus Law Menggunakan Algoritma Naive Bayes dan K-Nearest Neighbor (K-NN)," *J. Inform. Polinema*, vol. 7, no. 2, pp. 79–84, 2021, doi: 10.33795/jip.v7i2.688.
- [8] F. Nurhuda, S. W. Sihwi, and A. Doewes, "Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier," *ITSmart J. Teknol. dan Inf.*, vol. 2, no. 2, pp. 35–42, 2013.
- [9] F. Sodik and I. Kharisudin, "Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter," *Prisma*, vol. 4, pp. 628–634, 2021.
- [10] C. Pricylia, A. Mulya, P. Nugraha, and I. Santoso, "ANALISIS SENTIMEN MASYARAKAT TERHADAP PEMBANGUNAN KERETA CEPAT JAKARTA – BANDUNG MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS," vol. 7, no. 2.
- [11] Evasaria M. Sipayung, Herastia Maharani, and Ivan Zefanya, "Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier," *J. Sist. Inf.*, vol. 8, no. 1, pp. 2355–4614, 2016, [Online]. Available: <http://ejournal.unsri.ac.id/index.php/jsi/index>
- [12] M. Furqan, S. Sriani, and S. M. Sari, "Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia," *Techno.Com*, vol. 21, no. 1, pp. 51–60, 2022, doi: 10.33633/tc.v21i1.5446.
- [13] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional," *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
- [14] S. Haryati, A. Sudarsono, and E. Suryana, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)," *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.
- [15] A. D. Adhi Putra, "Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.