

ANALISIS SENTIMEN MASYARAKAT TERHADAP PEMBANGUNAN KERETA CEPAT JAKARTA – BANDUNG MENGUNAKAN ALGORITMA K-NEAREST NEIGHBORS (KNN)

Citra Pricylia Ananda Mulya¹, Pramudya Nugraha², Imam Santoso³

¹Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika ²Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika ³Dosen Universitas Teknologi Muhammadiyah Jakarta
Jl Raden Inten II No. 8 Rt 05/ Rw 14 Duren Sawit Kecamatan Duren Sawit Kota Jakarta Timur
13440 Indonesia
E-mail : citra@stikomcki.ac.id¹, Pram@stikomcki.ac.id², imam.santoso@utmj.ac.id³

ABSTRAK

Proyek kereta cepat Jakarta-Bandung menjadi ramai dibicarakan di media social, karena pada pembangunannya terdapat banyak pihak yang merasa dirugikan, namun ada juga pihak yang merasa diuntungkan. Pada penelitian ini dilakukan analisis sentiment masyarakat terhadap pembangunan kereta cepat Jakarta – Bandung. Tujuan dari penelitian ini adalah untuk menganalisis kelayakan kebijakan Transportasi Kereta Cepat Jakarta-Bandung dalam mewujudkan transportasi ramah lingkungan. Penelitian ini menggunakan metode deskriptif, karena bermaksud untuk mendapatkan fakta empiris, dan makna yang mendalam, mengamati, menangkap realitas dan memeriksa perilaku individu dan kelompok objek penelitian. Instrumen dengan wawancara, focus group discussion, observasi, dan triangulasi. Penentuan Informan dengan teknik purposive.

Kata kunci : Kereta Cepat, Jakarta-Bandung, Proyek, Penelitian, Transpostasi, Media Sosial

ABSTRACT

The Jakarta-Bandung high-speed train project has become a topic of discussion on social media, as there are multiple parties who feel disadvantaged by its construction, but there are also those who feel benefited. This research aims to analyze the sentiment of the community towards the construction of the Jakarta-Bandung high-speed train. The objective of this study is to analyze the feasibility of the Jakarta-Bandung High-Speed Train Transportation policy in achieving environmentally-friendly transportation. This research adopts a descriptive method, as it intends to obtain empirical facts and profound meanings by observing, capturing reality, and examining the behaviors of individuals and groups as research subjects. The instruments used include interviews, focus group discussions, observations, and triangulation. The selection of informants was conducted using purposive techniques.

Keyword : Social Media, Jakarta-Bandung, Project, Train, High-speed train, sentiment analys

1. PENDAHULUAN

Media Sosial pada masa kini memberikan peran dan pengaruh yang sangat besar dalam perkembangan teknologi komunikasi. Media Sosial juga di jadikan sebagai media untuk mengekspresikan diri

dalam menanggapi suatu kejadian atau hal-hal yang terjadi di lingkungannya, salah satu kejadian yang ditanggapi oleh pengguna media social adalah proyek kereta cepat Jakarta – Bandung yang sudah dibangun sejak awal tahun 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. Berdasarkan catatan dari Direktorat Jenderal Perhubungan Darat (DJPD), tercatat kerugian akibat kemacetan di DKI Jakarta. Menurut Rini menteri BUMN, keuntungan dibangunnya kereta cepat Jakarta – Bandung diantaranya akan meningkatkan perekonomian, mengangkat sector pariwisata, dan membuka lapangan pekerjaan yang baru. Selain itu dalam pembangunan proyek kereta cepat Jakarta-Bandung 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 sentiment untuk mengetahui bagaimana sentiment yang ada pada media social mengenai pembangunan proyek kereta cepat Jakarta – Bandung. Analisis Sentimen adalah suatu cara untuk mendapatkan sebuah informasi sentiment pada suatu media social untuk memahami suatu opini atau preferensi dari pengguna media social. Informasi yang didapat bisa berupa sentiment yang bernilai positif atau negative. Salah satu metode yang digunakan untuk analisis sentiment adalah metode yang terdapat pada Machine Learning seperti KNN. Algoritma KNN merupakan algoritma klasifikasi yang bekerja dengan mengambil sejumlah K data terdekat sebagai acuan untuk menentukan kelas dari data baru. Algoritma ini mengklasifikasi data berdasarkan similarity atau kemiripan atau kedekatannya terhadap data lainnya. Algoritma K-Nearest Neighbor (K-NN) bertujuan untuk mengklasifikasikan objek berdasarkan atribut dan data training. Dari hasil penelitian sebelumnya yang berjudul Analisis Sentimen Terhadap Pembangunan Kereta Cepat Jakarta-Bandung pada media social dan GloVe Word Embedding didapatkan akurasi sebesar 77.72%. Pada penelitian ini dengan judul Analisis Sentimen Masyarakat Terhadap Pembangunan Kereta Cepat Jakarta-Bandung Algoritma K-Nearest Neighbor bertujuan untuk mengetahui akurasi yang dihasilkan. Berdasarkan pernyataan sebelumnya, penulis ingin meneliti bagaimana tanggapan pembangunan kereta

cepat Jakarta Bandung dengan menganalisis opini masyarakat di media social. Analisis sentimen pada media social menjadi pilihan penulis untuk melihat bagaimana respon masyarakat terhadap pembangunan kereta cepat Jakarta Bandung. Pada Penelitian ini akan diklasifikasikan respon masyarakat terhadap pembangunan kereta cepat Jakarta Bandung ke dalam dua sentimen, yaitu sentimen positif dan sentimen negative.

2. METODOLOGI

Data set yang digunakan pada penelitian ini merupakan data set yang diambil dari media sosial Twitter dimulai tanggal tweet 1 Januari 2020 menggunakan beberapa keyword seperti #keretacepat, #dukungkeretacepat dan #keretacepatjakartabandung[2].

Data Mining

Data mining adalah proses yang mempekerjakan satu atau lebih teknik machine learning untuk menganalisis dan mengekstrasi pengetahuan secara otomatis. Data Mining merupakan tahapan untuk menemukan pola atau informasi dalam kumpulan data dengan menggunakan teknik dan algoritma tertentu.

Analisis Sentimen

Analisis sentimen merupakan cara mengumpulkan pendapat khalayak umum menggunakan jejaring sosial yang didalamnya terdapat mengandung pelayanan umum serta isu terkini. Analisis Sentimen adalah mengekstrak opini public tentang topik tertentu, produk atau jasa yang di dalam nya terkandung teks-teks yang tidak terstruktur. Tujuan dari analisis sentimen merupakan untuk mengetahui opini seseorang ke arah positif atau negatif yang nantinya dapat digunakan untuk pengambilan keputusan.

Algoritma K-Nearest Neighbors

Algoritma K-Nearest Neighbor (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak,

dimana masing – masing dimensi merepresentasikan fitur dari data [3] Algoritma K-Nearest Neighbors memiliki prinsip sederhana, bekerja berdasarkan jarak terpendek dari sampel uji ke sampel latih. Algoritma k-nearest neighbors memiliki tahapan:

- a) Menentukan parameter k (jumlah tetangga paling dekat.
- b) Tentukan bobot untuk setiap term dengan menggunakan Term Weighting TF-IDF.
- c) Hitung kemiripan antar dokumen dengan menggunakan cosine similarity

$$\cos(\theta_{ij}) = \frac{\sum k(dik)j(k)}{\sqrt{\sum k dik} \sqrt{\sum k}}$$

- d) Urutkan hasil perhitungan *cosine similarity* dari besar ke kecil
- e) Ambil sebanyak K yang paling tinggi kemiripannya dengan dokumen yang diklasifikasikan, tentukan kelasnya.

Cross Validation Confusion Matrix

Salah satu cara untuk menghitung akurasi adalah dengan menggunakan metode confusion matrix. Confusion matrix merupakan sebuah cara yang berguna untuk menganalisis seberapa baik classifier mengenali data. jika data set hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif. [4]

Tabel 1. Model Confusion Matrix [8]

		True class	
		Positive	Negative
Predicted Class	Positive	True positive (TP)	False negative (FN)
	Negative	False positive (FN)	True negative (TN)

3. LANDASAN TEORI

1.1 Text Mining

Text mining adalah proses ekstraksi pola (informasi dan pengetahuan yang berguna) dari sejumlah sumber data melalui identifikasi pola yang menarik. Pada kasus text mining, sumber data adalah bertujuan untuk menemukan informasi yang tidak diketahui, sesuatu yang belum diketahui dan belum dapat ditulis.

Analisis Sentimen

Analisis sentiment atau opinion mining merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentiment yang terkandung dalam suatu kalimat opini. Analisis sentiment dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negative atau positif.

Pre-Processing Data

Karena Bahasa di media sosial memiliki banyak item teks yang unik, maka item teks tersebut dihilangkan untuk mengurangi ruang fitur. Item tersebut perlu dihilangkan karena tidak memiliki nilai informasi dalam konteks klasifikasi sentiment. *Preprocessing* pertama dilakukan menggunakan gataframework, yaitu pengolahan data teks berbasis *web*, dengan menggunakan bahasa pemrograman *php*. Gataframework dapat diakses di alamat link www.gataframework.com/textmining. Peneliti menggunakan Gataframework karena memiliki keunggulan dalam melakukan proses *Stemming* bahasa Indonesia. Pengolahan data selanjutnya dilakukan menggunakan tools *Rapidminer*. Pre-processing yang akan dilakukan, Berikut tahapan *preprocessing* data, antara lain :

1. *Indonesian Stopword Removal.*

Yaitu proses pada sebuah kalimat jika mengandung kata-kata yang sering keluar dan dianggap tidak penting, seperti kata penghubung. contoh : dengan, dan, juga.

2. *Indonesian Stemming.*

Proses menemukan kata dasar (root word) dari kata berimbuhan dengan cara menghilangkan semua imbuhan yang terdiri dari awalan, sisipan, akhiran dan kombinasi awalan dan akhiran. Dalam text analytics, stemming merupakan

salah satu proses penting yang sangat mempengaruhi kualitas hasil analisis.

3. *Transformation Not Negative.*
Sebenarnya prosesnya tidaklah menghapus kata melainkan diambil untuk menilai bahwa kalimat yang diproses mengandung kalimat negative contoh: tidak, jangan, bukan.
4. *Normalization Indonesian Slank.*
Pada proses ini untuk mengkonversi kata-kata yang tidak baku (bahasa gaul) menjadi kata baku, contoh: ngeyel artinya keras kepala..
5. *Transforms Cases.*
Mengubah huruf kapital yang masih ada di dataset menjadi huruf - huruf kecil. Hal ini bertujuan agar terjadi keseragaman text pada model klasifikasi dan tidak terjadi kesalahan pada proses *tokenize*.
6. *Tokenize.*
Memecah sekumpulan karakter atau kalimat menjadi sebuah potongan karakter atau kata – kata sesuai dengan kebutuhan, biasa juga disebut tokenisasi.
7. *Filter Tokens (By Lenght).*
Menghilangkan kata – kata dengan panjang karakter tertentu, biasanya kata yang memiliki hanya 2 karakter tidak memiliki arti.
8. *Filter Stopword.*
Membuang kata – kata yang diabaikan pada sentimen analis, biasanya yang berupa kata sambung dan kata keterangan.

Sentiment Scoring

Pada sentiment scoring terdapat input kamus sentiment, boosterwords dan input kata negasi. Kamus sentiment 1 s.d. 5 (memiliki sentiment positif), dan -1 s.d. -5 (memiliki sentiment negative). Boosterwords adalah kata yang dapat meningkatkan atau mengurangi intensitas sentiment kata disebelahnya. Kata ini diberi bobot 1-2 untuk menambah atau mengurangi skor kata disampingnya. Kata negasi merupakan kata yang terdapat dalam suatu kalimat yang dapat mengubah orientasi dari suatu opini.

Pembobotan

Menyatakan bahwa pembobotan kata bertujuan untuk memberikan bobot pada fitur kata berdasarkan frekuensi kemunculan kata. Term Frequency (TF) merupakan jumlah

kemunculan atau frekuensi kata pada suatu dokumen. Sedangkan Inverse Document Frequency (IDF) bertujuan untuk mengetahui apakah term yang dicari cocok dengan kata kunci yang diinginkan, term yang sering muncul akan memberikan pengaruh yang kecil dalam menentukan keterkaitan kata kunci dokumen. TF-IDF dihitung dengan menggunakan persamaan seperti berikut:

$$W_{j,i} = n_{j,i} \sum k n_{k,i} . \log_2 D_{dj}$$

dimana:

$W_{j,i}$ = Pembobotan TF-IDF untuk term ke j pada dokumen ke i.

$n_{j,i}$ = Jumlah kemunculan term ke j pada dokumen ke i.

$\sum k n_{k,i}$ = Jumlah kemunculan seluruh term pada dokumen ke i.

D = Banyaknya dokumen yang dibangkitkan.

d_j = Banyaknya dokumen yang mengandung term ke j

4. HASIL DAN PEMBAHASAN

Hasil pengumpulan data yang didapat dari komentar postingan twitter mengenai pro dan kontra perkuliahan tatap muka dan daring, peneliti mengumpulkan 350 data opini pengguna twitter. Kemudian data tersebut dikumpulkan dalam format file xls, setelah itu data dikelompokkan menjadi 2 bagian yaitu data positif dan data negatif. Pengelompokan data ini dilakukan oleh beberapa koresponden, sehingga kemudian dihasilkan 265 opini Positif dan 80 opini negatif.

Berikut gambar proses pengolahan data pada *tools rapidminer*.

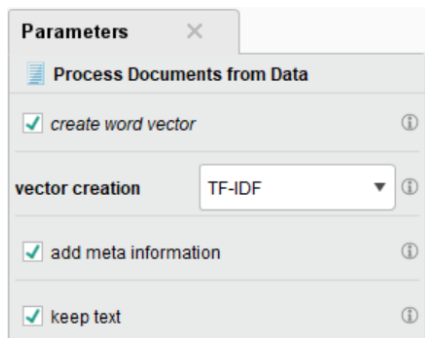


Gambar Proses Pengolahan Data Rapidminer.

Data hasil *preprocessing* yang berupa kata akan diubah ke dalam bentuk angka dengan dilakukan proses pembobotan kata yang bertujuan untuk menghitung bobot pada masing-masing kata yang akan digunakan sebagai fitur, semakin banyak dokumen yang akan diproses maka semakin banyak fitur. Pada tahapan ini terdapat dua bagian proses

yaitu *TF* (*Term Frequency*) dan *IDF* (*Inverse Document Frequency*),

TF-IDF adalah sebuah metode pembobotan kata yang digunakan untuk mengekstraksi ciri dari suatu teks, terdapat dua hal dalam perhitungan nilai bobot yaitu term frequency atau TF dan invers document frequency atau IDF. TF digunakan untuk mencari nilai dari kemunculan kata pada suatu dokumen yang mana semakin banyak suatu kata yang muncul maka makin tinggi nilai TF. Dan IDF adalah nilai kemunculan dari kata pada keseluruhan dokumen, nilai IDF berbanding terbalik dengan nilai TF, semakin banyak kata yang muncul maka nilai IDF akan semakin kecil.[5]



Gambar Proses *Tf Idf*.

Penelitian ini menggunakan algoritma KNN (K-Nearest Neighbor) untuk melakukan analisis sentimen. Kemudian dalam menentukan evaluasinya peneliti menggunakan Accuracy dan AUC (Area Under Curve).

Dari tahapan - tahapan pengujian yang sudah dilakukan dengan menggunakan dataset sebanyak 350 komentar opini pengguna Tiktok mengenai video Penghapusan Jalur Mandiri untuk masuk PTN, maka hasil Akurasi Algoritma KNN (K-Nearest Neighbor) tanpa menggunakan Fitur Seleksi Particle Swarm Optimization yaitu sebesar 82,69%,

Berikut ini adalah Tabel Confusion Matrix Algoritma KNN ((K-Nearest Neighbor) tanpa fitur seleksi.

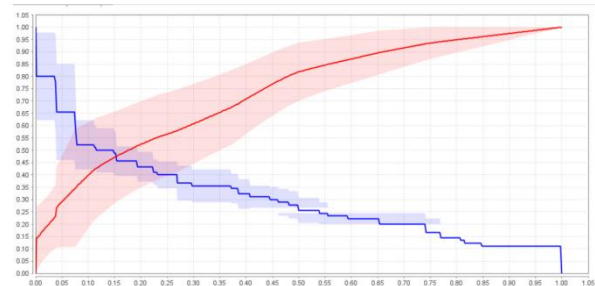
accuracy: 82.69% +/- 5.87% (micro average: 82.70%)

	true negative	true positive	class precision
pred. negative	38	11	77.55%
pred. positive	39	201	83.75%
class recall	49.35%	94.81%	

AUC: 0.735 +/- 0.106 (micro average: 0.735) (positive class: negative)
— ROC — ROC (Thresholds)

Gambar Accuracy

Dalam dataset yang diklasifikasikan negatif. False positif merupakan jumlah record negatif dalam dataset yang diklasifikasikan positif. False negatif (fn) merupakan jumlah record positif dalam dataset yang diklasifikasikan negatif.



Gambar Kurva ROC.

5. KESIMPULAN

Berdasarkan dari penelitian yang sudah dilakukan, dapat disimpulkan bahwa hasil akurasi dan AUC dengan algoritma KNN, hasil pengujian pengklasifikasian AUC nilai keakuratannya 82.70% dapat dikategorikan sebagai Excellent Classification. signifikan. Lalu juga di penelitian ini banyak sekali yang berkomentar atau beropini negative, ini membuktikan bahwa masyarakat yang tinggal di Bandung dan Jakarta lebih pro terhadap pembangunan kereta cepat Jakarta Bandung

Untuk mendukung penelitian ini, maka peneliti akan mengembangkan aplikasi opini dari komentar youtube dan twitter mengenai Kereta cepat Jakarta Bandung untuk mengklasifikasikan opini negatif dan positif menggunakan bahasa pemrograman php yang dilakukan oleh Rapid Miner.

DAFTAR PUSTAKA

- [1] N. Alfiah, “Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan Menggunakan Metode Naive Bayes,” *Respati*, vol. 16, no. 1, p. 32, 2021, doi: 10.35842/jtir.v16i1.386.
- [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,” vol. 10, no. 2, pp. 1713–1723, 2022.
- [3] F. Liantoni, “Klasifikasi Daun Dengan Perbaikan Fitur Citra Menggunakan Metode K-Nearest Neighbor,” *J. Ultim.*, vol. 7, no. 2, pp. 98–104, 2016, doi: 10.31937/ti.v7i2.356.
- [4] 2018: 41) James A.F Stoner, “Landasan Teori اندج,” *Dasar-Dasar Ilmu Polit.*, no. 2014, pp. 17–39, 1988.
- [5] T. W. D. Sari, “Penerapan Text Mining Dengan Menggunakan Algoritma TF-IF Untuk Klasifikasi Genre Novel,” *Pelita Inform. Inf. dan Inform.*, vol. 10, pp. 29–37, 2021, [Online]. Available: <http://www.stmik-budidarma.ac.id/ejurnal/index.php/pelita/article/view/3142>