

ANALISIS SENTIMEN MASYARAKAT TERHADAP LAYANAN KERETA CEPAT JAKARTA – BANDUNG MENGGUNAKAN METODE *MULTICLASS SUPPORT VECTOR MACHINE*

HERDIAN PARTAWIJAYA



**DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Analisis Sentimen Masyarakat Terhadap Layanan Kereta Cepat Jakarta – Bandung Menggunakan Metode *Multiclass Support Vector Machine*” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juli 2024

Herdian Partawijaya
G1401201022

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



ABSTRAK

HERDIAN PARTAWIJAYA. Analisis Sentimen Masyarakat Terhadap Layanan Kereta Cepat Jakarta – Bandung Menggunakan Metode *Multiclass Support Vector Machine*. Dibimbing oleh BAGUS SARTONO dan PIKA SILVIANTI.

Penambahan volume kendaraan bermotor terutama kendaraan pribadi dapat menimbulkan berbagai dampak negatif. Saat ini pemerintah menghadirkan fasilitas transportasi publik terbaru yaitu kereta cepat Jakarta – Bandung sebagai salah satu solusi untuk mengatasi hal tersebut. Penelitian ini bertujuan untuk mengklasifikasikan sentimen masyarakat terhadap layanan kereta cepat Jakarta – Bandung dan mengidentifikasi performa penggunaan metode *Multiclass Support Vector Machine*, serta mengidentifikasi topik dalam sentimen masyarakat terhadap layanan kereta cepat Jakarta – Bandung. Hanya sebanyak 40% data yang diberikan label sentimen secara manual yang digunakan dalam proses pelatihan model dan sisanya akan diprediksi menggunakan model terbaik. Hasil dari penelitian ini diperoleh model terbaik adalah algoritma *One-Against-All* kernel *sigmoid* dengan *hyperparameter* $C = 1000$, dan $\gamma = 0,0001$. Performa yang dihasilkan yaitu akurasi sebesar 74,8% dan *macro average f1-score* sebesar 74,6%. Pelabelan menggunakan model terbaik terhadap 14.255 data diperoleh sentimen positif sebanyak 3.715, netral sebanyak 5.288, dan negatif sebanyak 5.252. Hasil *clustering* sentimen positif menunjukkan berbagai pujian dari masyarakat, serta rasa bangga karena merupakan kereta cepat pertama di Asia Tenggara. Masyarakat juga merasa kagum dengan kecepatan dan efisiensi waktu perjalanan kereta cepat. Sedangkan sentimen negatif menunjukkan kekecewaan biaya kereta cepat yang menggunakan APBN dan kritik bahwa kereta cepat tidak terlalu penting karena tidak ada urgensinya apalagi ibu kota Indonesia akan dipindahkan. Masyarakat juga mengkritik biaya kereta cepat yang terus membengkak hingga ratusan triliun dan stasiunnya yang hanya sampai di Padalarang.

Kata kunci: analisis sentimen, kereta cepat, k-means clustering, multiclass support vector machine, twitter

ABSTRACT

HERDIAN PARTAWIJAYA. Sentiment Analysis of the Public Towards the Jakarta-Bandung High-Speed Train Service Using the Multiclass Support Vector Machine Method. Supervised by BAGUS SARTONO and PIKA SILVIANTI.

The increase in the volume of motorized vehicles, especially private vehicles, can lead to various negative impacts. Currently, the government has introduced a new public transportation facility, the Jakarta-Bandung high-speed train, as a solution to address these issues. This research aims to classify public sentiment towards the Jakarta - Bandung high-speed train service and identify the performance of the Multiclass Support Vector Machine method, as well as identifying topics in public sentiment towards the Jakarta - Bandung high-speed train service. Only 40% of the data manually labeled with sentiment was used in the model training process, with the remainder being predicted using the best model. The results of this research show that the best model is the One-Against-All algorithm with a sigmoid kernel, using hyperparameters $C = 1000$ and $\gamma = 0.0001$. The performance achieved was an accuracy of 74.8% and a macro average F1-score of 74.6%. Labeling the remaining 14,255 data points using the best model resulted in 3,715 positive sentiments, 5,288 neutral sentiments, and 5,252 negative sentiments. Clustering of the positive sentiments revealed various praises from the public, including pride in having the first high-speed train in Southeast Asia and admiration for its speed and travel time efficiency. On the other hand, negative sentiments highlighted disappointment over the high-speed train's use of state budget funds, criticisms about its perceived lack of necessity given the upcoming relocation of Indonesia's capital, and concerns over the escalating costs exceeding hundreds of trillions, with the train only reaching Padalarang station.

Keywords: high speed rail, k-means clustering, multiclass support vector machine, sentiment analysis, twitter



@Hak cipta milik IPB University

IPB University

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2024¹
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

ANALISIS SENTIMEN MASYARAKAT TERHADAP LAYANAN KERETA CEPAT JAKARTA – BANDUNG MENGGUNAKAN METODE *MULTICLASS SUPPORT VECTOR MACHINE*

HERDIAN PARTAWIJAYA

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Statistika dan Sains Data

**DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**



@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Skripsi:
Prof. Dr. Ir. Anik Djuraidah, M. S.

Perpustakaan IPB University

Judul Skripsi : Analisis Sentimen Masyarakat Terhadap Layanan Kereta Cepat
Jakarta – Bandung Menggunakan Metode Multiclass Support
Vector Machine

Nama : Herdian Partawijaya
NIM : G1401201022

Disetujui oleh

Pembimbing 1:
Dr. Bagus Sartono, S.Si., M.Si

Pembimbing 2:
Pika Silvianti, S.Si., M.Si



Diketahui oleh

Ketua Departemen Statistika:
Dr. Bagus Sartono, S.Si., M.Si
NIP. 197804112005011002

Tanggal Ujian:
(18 Juli 2024)

Tanggal Lulus:



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

PRAKATA

Puji dan syukur kehadiran Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Januari 2024 sampai bulan Juni 2024 ini ialah analisis sentimen, dengan judul “Analisis Sentimen Masyarakat Terhadap Layanan Kereta Cepat Jakarta – Bandung Menggunakan Metode Multiclass Support Vector Machine”.

Terima kasih penulis ucapkan kepada pihak-pihak yang telah banyak membantu dan memberikan dukungan selama proses penulisan karya ilmiah ini dari awal sampai penulisan ini selesai, diantaranya:

1. Bapak Habi Bakrin, Ibu Zauriyati selaku orang tua penulis dan seluruh keluarga besar yang senantiasa memberikan doa dan dukungan.
2. Bapak Dr. Bagus Sartono, S.Si., M.Si dan Ibu Pika Silvianti, S.Si., M.Si selaku dosen pembimbing yang telah banyak memberikan saran dan membimbing dalam penulisan karya ilmiah ini.
3. Seluruh dosen dan tenaga pendidik Departemen Statistika IPB yang telah membantu segala proses penulisan karya ilmiah ini.
4. Nabil Izzany dan Naufal Syafiq yang telah membantu penulis dalam mengolah data penelitian.
5. Cahya Ireno, M. Zhillan, Farhan Zahid, Irsyad Robbani, Akmal Riza, Alwi Sidiq, M. Nachnoer dan Fetri Wahyuningsih yang telah memberikan bantuan, baik itu saran, kritik dan dukungan kepada penulis.
6. Seluruh mahasiswa/i Departemen Statistika IPB angkatan 57 yang telah kebersamai perjalanan penulis.
7. Seluruh pihak yang telah membantu penulisan karya ilmiah ini yang tidak bisa penulis sebutkan satu persatu.

Penulis sadar bahwa karya ilmiah ini masih terdapat kekurangan dan kekeliruan, oleh sebab itu penulis mengucapkan mohon maaf karena karya ilmiah ini masih belum sempurna. Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Juli 2024

Herdian Partawijaya



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

DAFTAR ISI

DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	3
1.4 Manfaat	3
1.5 Ruang Lingkup	3
II TINJAUAN PUSTAKA	4
2.1 Analisis Sentimen	4
2.2 <i>Support Vector Machine (SVM)</i>	4
2.3 <i>Multiclass Support Vector Machine</i>	7
2.4 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	8
2.5 <i>K-Means Clustering</i>	8
2.6 <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	9
2.7 <i>Hyperparameter Tuning</i>	10
2.8 Validasi Silang Lipat-k	12
III METODE	13
3.1 Bahan dan Data	13
3.2 Tahapan Penelitian	13
IV HASIL DAN PEMBAHASAN	21
4.1 Praproses Data Awal	21
4.2 Pengambilan Sampel Data dan Pelabelan Data Sebagian	21
4.3 Praproses Data Lanjutan	22
4.4 Eksplorasi Data	23
4.5 Ekstraksi Fitur TF-IDF	25
4.6 Pemodelan <i>Multiclass SVM</i>	25
4.7 Pelabelan Kategori Sentimen Pada Data Tak Berlabel	31
4.8 Eksplorasi Lanjutan	32
V SIMPULAN DAN SARAN	41
5.1 Simpulan	41
5.2 Saran	41
DAFTAR PUSTAKA	42
LAMPIRAN	45
RIWAYAT HIDUP	48



DAFTAR TABEL

1	Fungsi kernel pada SVM	7
2	Contoh data yang digunakan	13
3	Ilustrasi pengambilan contoh acak sistematis pada data	15
4	Pedoman Pelabelan Data Teks	16
5	Ilustrasi pembobotan TF-IDF	18
6	<i>Hyperparameter</i> yang dioptimalkan	19
7	<i>Confusion matrix</i>	19
8	Hasil tahapan praproses data	23
9	Kinerja model menggunakan <i>default hyperparameter</i>	27
10	Hasil <i>hyperparameter tuning</i>	28
11	Performa model terbaik menggunakan <i>10-fold cross validation</i>	29
12	Hasil validasi model terbaik terhadap data uji	30
13	Contoh <i>tweet</i> hasil prediksi model terbaik pada data uji	31

DAFTAR GAMBAR

1	Ilustrasi <i>hyperplane</i> terbaik yang memisahkan dua kelas data secara sempurna (Utami 2014)	5
2	Ilustrasi <i>hyperplane</i> terbaik yang memisahkan dua kelas data secara tidak sempurna (Utami 2014)	6
3	Ilustrasi algoritma metode SMOTE dalam membangkitkan data sintetik kelas minoritas (Graa dan Rekik 2019).	10
4	Ilustrasi pengaruh nilai parameter C terhadap lebar margin pada kurva <i>decision boundary</i> (Ben-Hur dan Weston 2010)	11
5	Ilustrasi pengaruh nilai parameter gamma terhadap bentuk kurva <i>decision boundary</i> (Ben-Hur dan Weston 2010)	11
6	Ilustrasi pengaruh nilai parameter C dan gamma secara bersamaan terhadap bentuk kurva <i>decision boundary</i> (Ben-Hur dan Weston 2010)	12
7	Ilustrasi validasi silang lipat-5 (Setiawan <i>et al.</i> 2022)	12
8	Diagram alir penelitian	14
9	Proporsi sentimen hasil pelabelan manual oleh peneliti (a) dan pelabel kedua (b) serta hasil pelabelan <i>final</i> (c)	22
10	Frekuensi <i>tweet</i> setiap kategori sentimen dari periode Januari 2022 – Januari 2024 menggunakan data sampel	24
11	<i>Word cloud</i> sentimen positif (a) dan netral (b) serta negatif (c)	24
12	Sepuluh kata yang memiliki skor TF-IDF tertinggi	25
13	Proporsi label pada data latih (a) dan data uji (b)	26
14	Evaluasi model terbaik menggunakan validasi silang lipat-10	29
15	<i>Confusion Matrix</i> data uji dengan model terbaik	30
16	Proporsi kategori sentimen dari data tak berlabel yang tidak digunakan dalam pemodelan setelah diprediksi menggunakan model terbaik	32
17	Frekuensi <i>tweet</i> setiap kategori sentimen dari periode Januari 2022 Januari 2024 menggunakan seluruh data	33
18	Persentase <i>tweet</i> dari periode Januari 2022 – Januari 2024	33
19	<i>Wordcloud</i> (a) dan diagram batang (b) kata dominan pada sentimen positif	34
20	<i>Wordcloud</i> (a) dan diagram batang (b) kata dominan pada sentimen negatif	35
21	<i>Wordcloud</i> (a) dan diagram batang (b) kata dominan pada sentimen netral	35
22	Nilai <i>silhouette coefficient</i> pada jumlah cluster 2 sampai 10 masing-masing kategori sentimen	36
23	<i>Word cloud cluster</i> satu (a), dua (b), tiga (c), empat (d), lima (e), enam (f), tujuh (g) dan delapan (h) pada data sentimen positif	37
24	<i>Word cloud cluster</i> satu (a), dua (b), tiga (c), empat (d), lima (e), enam (f), tujuh (g) dan delapan (h) pada data sentimen negatif	39
25	<i>Word cloud cluster</i> satu (a), dua (b), tiga (c) dan empat (d) pada data sentimen netral	40



DAFTAR LAMPIRAN

1	Lampiran 1 Sebagian daftar kata kamus normalisasi yang disusun peneliti	45
2	Lampiran 2 Sebagian daftar kata kamus normalisasi dari Ibrohim dan Budi (2019)	46
3	Lampiran 3 Sebagian daftar akun yang dieliminasi	47
4	Lampiran 4 Sebagian daftar <i>stopword</i> berdasarkan <i>library sastrawi</i>	47

@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.