

SKRIPSI

**ANALISIS SENTIMEN MASYARAKAT TERHADAP KERETA CEPAT
JAKARTA-BANDUNG PADA MEDIA SOSIAL X MENGGUNAKAN
METODE *K-NEAREST NEIGHBOR***



ANDITYA RAHMAT TRIANTARA

NIM: 205610039

PROGRAM STUDI SISTEM INFORMASI

PROGRAM SARJANA

FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA

YOGYAKARTA

2024

SKRIPSI

**ANALISIS SENTIMEN MASYARAKAT TERHADAP KERETA CEPAT
JAKARTA-BANDUNG PADA MEDIA SOSIAL X MENGGUNAKAN
METODE *K-NEAREST NEIGHBOR***

Diajukan sebagai salah satu syarat untuk menyelesaikan studi

Program Sarjana

Program Studi Sistem Informasi

Fakultas Teknologi Informasi

Universitas Teknologi Digital Indonesia

Yogyakarta

Disusun Oleh

ANDITYA RAHMAT TRIANTARA

NIM: 205610039

PROGRAM STUDI SISTEM INFORMASI

PROGRAM SARJANA

FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA

2024

HALAMAN PERSETUJUAN

Judul : Analisis Sentimen Masyarakat Terhadap
Kereta Cepat Jakarta-Bandung Pada
Media Sosial X Menggunakan Metode *K-Nearest Neighbor*

Nama : Anditya Rahmat Triantara

NIM : 205610039

Program Studi : Sistem Informasi

Program : Sarjana

Semester : Ganjil

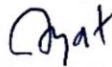
Tahun Akademik : 2023 / 2024

Telah diperiksa dan disetujui untuk diujikan di depan hadapan Dewan Penguji

Skripsi

Yogyakarta, 20 Februari....., 2024

Dosen Pembimbing,



Pulut Suryati, S.Kom., M.Cs.

NIDN: 0015037802

HALAMAN PENGESAHAN

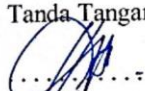
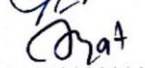
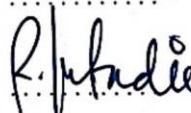
SKRIPSI

ANALISIS SENTIMEN MASYARAKAT TERHADAP KERETA CEPAT
JAKARTA-BANDUNG PADA MEDIA SOSIAL X MENGGUNAKAN
METODE *K-NEAREST NEIGHBOR*

Telah dipertahankan didepan Dewan Penguji Skripsi dan dinyatakan
diterima sebagai salah satu persyaratan untuk memperoleh Gelar Sarjana

Komputer
Program Studi Sistem Informasi
Fakultas Teknologi Informasi
Universitas Teknologi Digital Indonesia
Yogyakarta

Yogyakarta, 25 Maret....., 2024

Dewan Penguji	NIDN	Tanda Tangan
1. Cosmas Haryawan, S.TP., S. kom., M.Cs. (Ketua)	0519067401	
2. Pulut Suryati, S.Kom., M.Cs. (Sekretaris)	0015037802	
3. Rikie Kartadie, S.T., M.Kom. (Anggota)	0701037604	

Mengetahui

Ketua Program Studi Sistem Informasi


Deborah Kurniawati, S.Kom., M.Cs.

NPP: 051149

26 MAR 2024

HALAMAN PERNYATAAN

Dengan ini saya menyatakan bahwa Laporan Skripsi ini tidak terdapat karya yang penuh diajukan untuk memperoleh gelar Ahli Madya/keserjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya tidak terdapat karya atau pendapat yang penuh ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar Pustaka.

Yogyakarta, ~~28~~ Februari....., 2024



Anditya Rahmat Triantara

NIM: 205610039

HALAMAN PERSEMBAHAN

Puji syukur Alhamdulillah saya panjatkan kepada Allah SWT, yang telah melimpahkan rahmat dan karunia-Nya sehingga saya dapat menyelesaikan skripsi ini yang saya persembahkan kepada :

1. Kedua orang tua, Bapak Ir.Trijaya Gane Putra, M.Si dan Ibu Marbubah, S.Pd., yang telah membesarkan dan memberikan sebuah dukungan serta memberikan kepercayaannya.
2. Kedua Saudara/i saya Renny Cahyaningtyas dan Ryan Akhmal Hidayat memberikan support dan bimbingan dan ilmunya terkait dengan penulisan skripsi ini serta dukungan yang tiada henti.
3. Ibu Pulut Suryati, S.Kom., M.Cs. selaku dosen pembimbing skripsi yang memberikan arahan dan dorongan yang telah diberikan sehingga skripsi ini dapat selesai dengan baik.
4. Teman-teman dan sahabat-sahabat yang namanya tidak saya sebutkan satu persatu yang selalu memberikan dukungan dan motivasi kepada saya hingga saat ini.

HALAMAN MOTTO

“Tidak sedang bersaing dengan siapapun. Silahkan mendahului saya dari sisi dan arah manapun. Jangan bandingkan apapun dengan saya. Kalian punya cerita sendiri. Begitu juga dengan saya. Saya bersyukur dan menikmati hidup saya sendiri apapun ceritanya” – (Penulis).

KATA PENGANTAR

Segala Puji syukur penulis panjatkan kehadirat Allah SWT, yang telah melimpahkan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan mengangkat judul “**Analisis Sentimen Masyarakat Terhadap Kereta Cepat Jakarta-Bandung Pada Media Sosial X Menggunakan Metode *K-Nearest Neighbor***” dapat terselesaikan dengan baik. Penulis menyadari bahwa dalam penyusunan Tugas Akhir Skripsi ini tidak lepas dari bantuan orang lain. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada yang terhormat:

1. Kedua orang tua Bapak Ir. Trijaya Gane, M. Si Putra dan Ibu Marbubah, S.Pd., yang telah senantiasa memberi dukungan, semangat, nasehat dan doa sehingga penulis dapat menyelesaikan skripsi ini.
2. Ibu Sri Redjeki, S.Si., M.Kom., Ph.D. Selaku Rektor Universitas Teknologi Digital Indonesia Yogyakarta.
3. Deborah Kurniawati, S.Kom., M.Cs, selaku Ketua Program Studi Sistem Informasi.
4. Ibu Pulut Suryati, S.Kom., M.Cs, dan Pembimbing skripsi yang telah banyak memberikan arahan bimbingan dan motivasi pada saat penulisan skripsi ini.
5. Terima kasih saya ucapkan kepada Bapak Cosmas Haryawan, S.TP., S. Kom., M.Cs. selaku penguji 1 serta Bapak Rikie Kartadie, S.T., M.Kom. selaku penguji 2 yang telah menguji dan memberi saran untuk menyempurnakan skripsi ini.
6. Terima kasih untuk rumah hijau HimaSI yang telah berproses bersama.

7. Terima kasih teman-teman kelas Sistem Informasi dan Sahabat-sahabat yang selalu membantu dalam suka maupun duka.

Penulis menyadari sepenuhnya bahwa di dalam pembuatan laporan ini masih banyak terdapat kekurangan. Untuk itu saran dan kritik yang membangun dari semua pihak sangat diharapkan oleh penulis. Dan harapan dari penulis adalah bahwa semoga laporan ini dapat memberi manfaat kepada pembaca pada umumnya, serta pihak-pihak lain yang terkait dan dapat bermanfaat bagi penulis khususnya. Dan kepada semua pihak penulis ucapkan terima kasih.

Yogyakarta, ~~28 Februari~~.....,2024



Anditya Rahmat Triantara

Nim: 205610039

DAFTAR ISI

SKRIPSI.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN.....	v
HALAMAN PERSEMBAHAN.....	vi
HALAMAN MOTTO.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xii
DAFTAR TABEL.....	xiii
INTISARI.....	xiv
<i>ABSTRACT</i>	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Ruang Lingkup.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penelitian.....	4
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI.....	6
2.1 Tinjauan Pustaka.....	6
2.2 Landasan Teori.....	9
2.2.1 Analisis Sentimen.....	9
2.2.2 X (<i>Twitter</i>).....	10
2.2.3 <i>Python</i>	10
2.2.4 <i>Tweet Harvest</i>	11
2.2.5 Seleksi Fitur.....	11
2.2.6 <i>SMOTE Upsampling</i>	12
2.2.7 <i>RapidMiner</i>	12
2.2.8 <i>K-Nearest Neighbor</i>	13
2.2.9 Model Evaluasi.....	14
BAB III METODE PENELITIAN.....	17

3.1	Bahan.....	17
3.2	Peralatan	17
3.3	Prosedur dan Pengumpulan Data.....	18
3.3.1	Pengumpulan Data	19
3.3.2	Pelabelan Data.....	20
3.3.3	<i>Preprocessing</i>	21
3.3.4	Seleksi Fitur (<i>Term Frequency</i>).....	24
3.3.5	Klasifikasi <i>K-Nearest Neighbor</i>	24
3.3.6	Pengujian Evaluasi.....	25
BAB IV IMPLEMENTASI DAN PEMBAHASAN		27
4.1	Pengumpulan Data	27
4.2	Pelabelan Data.....	28
4.3	<i>Preprocessing</i>	31
4.4	Pembobotan Kata	39
4.5	Klasifikasi <i>K-Nearest Neighbor</i>	40
4.6	Model Evaluasi.....	47
BAB V KESIMPULAN DAN SARAN.....		55
5.1	Kesimpulan.....	55
5.2	Saran.....	55
DAFTAR PUSTAKA		57
LAMPIRAN.....		59

DAFTAR GAMBAR

Gambar 2. 1 Kurva Roc (Sumber: Sucipto,2012)	15
Gambar 3. 1 Tahapan Penelitian	19
Gambar 3. 2 Alur Pelabelan	20
Gambar 3. 3 Tahapan Text Preprocessing	21
Gambar 4. 1 Kode Program Crawling Data	27
Gambar 4. 2 Hasil Crawling	28
Gambar 4. 3 Kode Program Import Data Crawling	28
Gambar 4. 4 Hasil Import Data Crawling	28
Gambar 4. 5 Kode Program Translate.....	29
Gambar 4. 6 Hasil Translate	29
Gambar 4. 7 Kode Program Penentuan Kelas Sentimen.....	30
Gambar 4. 8 Hasil Penentuan Kelas Sentimen	30
Gambar 4. 9 Kode Program Jumlah Sentimen.....	30
Gambar 4. 10 Hasil jumlah Sentimen	30
Gambar 4. 11 Kode Program Cleaning Text.....	31
Gambar 4. 12 Hasil Cleaning Text.....	32
Gambar 4. 13 Kode Program Case Folding	32
Gambar 4. 14 Hasil Case Folding	33
Gambar 4. 15 Kode Program Tokenizing	33
Gambar 4. 16 Hasil Tokenizing	34
Gambar 4. 17 Kode Program Normalisasi	34
Gambar 4. 18 Hasil Normalisasi	35
Gambar 4. 19 Kode Program Stopword.....	35
Gambar 4. 20 Hasil Stopword.....	36
Gambar 4. 21 Kode Program Stemming	37
Gambar 4. 22 Hasil Stemming	38
Gambar 4. 23 Kode Program Hapus Duplikat	38
Gambar 4. 24 Hasil Jumlah Sentimen setelah hapus duplikat	38
Gambar 4. 25 Process Documents From Data	39
Gambar 4. 26 Hasil Process Documents From Data.....	40
Gambar 4. 27 Tools Rapid Miner	41
Gambar 4. 28 Hasil Visualisasi Skema 90:10.....	44
Gambar 4. 29 Visualisasi Skema 80:20	45
Gambar 4. 30 Kurva ROC Skema 70:30.....	46
Gambar 4. 31 Evaluasi Data 90:10	47
Gambar 4. 32 Evaluasi Data 80:20	49
Gambar 4. 33 Evaluasi Data 70:30	51
Gambar 4. 34 Visualisasi Akurasi Pembagian Data	53

DAFTAR TABEL

Tabel 2. 1 Perbandingan Dengan Penelitian Sebelumnya	8
Tabel 3. 1 Pengumpulan data Tweet Harvest.....	20
Tabel 3. 2 Contoh Pelabelan Data.....	21
Tabel 3. 3 Hasil Pre-Processing	23
Tabel 3. 4 Rencana Uji.....	25
Tabel 3. 5 Confussion Matrix.....	26
Tabel 4. 1 Jumlah Pembagian Data.....	43
Tabel 4. 2 Pembagian Data 90:10	43
Tabel 4. 3 Pembagian Data 80:20	45
Tabel 4. 4 Pembagian Data 70:30	46
Tabel 4. 5 Perhitungan Presisi 90:10	48
Tabel 4. 6 Perhitungan Recall 90:10	49
Tabel 4. 7 Perhitungan Presisi 80: 20	50
Tabel 4. 8 Perhitungan Recall 80:20	51
Tabel 4. 9 Perhitungan Presisi 70:30	52
Tabel 4. 10 Perhitungan Recall 70:30	52
Tabel 4. 11 Tabel Hasil Akurasi	53

INTISARI

Pemerintah Indonesia sedang berupaya meningkatkan pelayanan transportasi dan mendukung pembangunan di kawasan Jakarta-Bandung dengan membangun kereta cepat. Proyek kereta cepat Jakarta-Bandung yang dibangun atas kerja sama Indonesia dengan China. Dalam proyek kereta cepat Jakarta-Bandung, masyarakat sebagai pengguna kereta cepat memberikan pengalaman dan opini terkait dengan kereta cepat Jakarta Bandung.

Tahap pada proses penelitian ini yang pertama yaitu crawling data dengan menggunakan *library tweet harvest* dengan *keyword*, preprocessing, yaitu terdiri dari *case folding*, *tokenizing*, normalisasi, *stopword* dan *stemming*. Pada penelitian ini melakukan analisis sentimen terhadap pengguna kereta cepat Jakarta-Bandung pada media sosial X dengan menggunakan metode *K-Nearest Neighbor*.

Hasil dari pengklasifikasian menggunakan metode *K-Nearest Neighbor* terkait dengan analisis sentimen masyarakat terhadap pengguna kereta cepat Jakarta-Bandung pada media sosial X. Dengan penggunaan nilai k 1 sampai 10 angka ganjil, mendapatkan hasil akurasi paling tinggi pada nilai $k=3$ dengan skema pembagian 90% data latih dan 10% data uji, atau setara dengan 1.644 data latih dan 183 data uji mendapatkan jumlah akurasi sebesar 93,87%.

Kata Kunci: *Kereta Cepat Jakarta-Bandung, K-Nearest Neighbor, Media Sosial X.*

ABSTRACT

The Indonesian government is working to improve transportation services and support development in the Jakarta-Bandung area by building fast trains. The Jakarta-Bandung high-speed train project was built in collaboration with Indonesia and China. In the Jakarta-Bandung high-speed train project, the public as fast train users provide experiences and opinions regarding the Jakarta-Bandung high-speed train.

The first stage in this research process is crawling data using the Tweet Harvest Library with the keyword, preprocessing, which consists of case folding, tokenizing, normalization, stopword and stemming. In this research, sentiment analysis was carried out on Jakarta-Bandung fast train users on social media X using the K-Nearest Neighbor method.

The results of the classification using the K-Nearest Neighbor method are related to the analysis of public sentiment towards Jakarta-Bandung fast train users on social media 90% training data and 10% test data, or the equivalent of 1.644 training data and 183 test data to get an accuracy of 93.87%.

Keywords: Fast Train Jakarta-Bandung, K-Nearest Neighbor, Social Media X.