

TESIS

**Analisis Sentimen Di Media Sosial Tentang Kualitas
Layanan KAI Setelah Pandemi Menggunakan Machine
Learning**



Stephanus Wisnu Setyawan

No. Mahasiswa: 215311587

MAGISTER INFORMATIKA

PROGRAM PASCASARJANA

UNIVERSITAS ATMA JAYA YOGYAKARTA

2024

HALAMAN PENGESAHAN

Tugas Akhir Berjudul

ANALISIS SENTIMEN DI MEDIA SOSIAL TENTANG KUALITAS LAYANAN KAI SETELAH PANDEMI
MENGUNAKAN MACHINE LEARNING

yang disusun oleh

Stephanus Wisnu Setyawan

215311587

dinyatakan telah memenuhi syarat pada tanggal 27 Februari 2024

Dosen Pembimbing 1	: Paulus Mudjihartono, S.T., M.T., Ph.D	Keterangan Telah Menyetujui
Dosen Pembimbing 2	: Prof. Dr. Pranowo, S.T., M.T.	Telah Menyetujui
Tim Penguji		
Penguji 1	: Paulus Mudjihartono, S.T., M.T., Ph.D	Telah Menyetujui
Penguji 2	: Prof. Dr. Ir. Alb. Joko Santoso, M.T.	Telah Menyetujui

Yogyakarta, 27 Februari 2024

Universitas Atma Jaya Yogyakarta

Teknologi Industri

Dekan

ttd.

Dr. Ir. Parama Kartika Dewa SP., S.T., M.T.

Dokumen ini merupakan dokumen resmi UAJY yang tidak memerlukan tanda tangan karena dihasilkan secara elektronik oleh Sistem Bimbingan UAJY. UAJY bertanggung jawab penuh atas informasi yang tertera di dalam dokumen ini

HALAMAN PERNYATAAN

Saya yang bertanda di bawah ini:

Nama : Stephanus Wisnu Setyawan
Nomor Mahasiswa : 215311587
Konsentrasi : *Intelligent Informatics*
Judul Tesis : Analisis Sentimen Di Media Sosial Tentang Kualitas
Layanan KAI Setelah Pandemi Menggunakan
Machine Learning.

Dengan ini menyatakan bahwa penelitian ini merupakan hasil karya penulis sendiri dan bukan merupakan hasil duplikasi dari penelitian sebelumnya. Penelitian terdahulu hanya menjadi referensi bagi penulis dan sudah disertakan acuan dan terlampir dalam daftar pustaka.

Yogyakarta, Februari 2024

Penulis

Stephanus Wisnu Setyawan

215311587

INTISARI

Media sosial adalah sebuah media *online*, di mana para penggunanya bisa dengan mudah berpartisipasi, berbagi dan menciptakan sebuah postingan. Jejaring sosial seperti Twitter merupakan bentuk media sosial yang paling umum digunakan oleh masyarakat di seluruh dunia. Salah satu postingan-postingan pada media sosial Twitter adalah tweet tentang Kereta Api Indonesia (KAI). Tweet-tweet tersebut biasanya akan menghasilkan sentimen, di mana sentimen-sentimen tersebut dapat diolah menjadi sebuah analisis sentimen.

Penelitian ini akan bertujuan untuk mengetahui analisis sentimen di media sosial Twitter tentang kualitas layanan KAI dimasa setelah masa pandemi COVID-19 berakhir dan mengetahui model *machine learning* yang terbaik dalam melakukan analisis sentimen tersebut. Data yang akan digunakan adalah tweet tentang @KAI121 pada rentang waktu 22 Juni 2023 hingga 31 Desember 2023. Penelitian ini akan menggunakan model *TF-IDF SVM*, *TF-IDF RF*, *TF-IDF KNN*, *CV SVM*, *CV RF* dan *CV KNN* yang akan digunakan untuk membandingkan mana model *machine learning* yang terbaik.

Hasil sentimen setiap dimensi kualitas layanan KAI dimasa berakhirnya pandemi COVID-19, menunjukkan bahwa penumpang KAI cenderung menunjukkan respon yang cukup netral. Selain itu *SVM* menunjukkan performa terbaik, baik menggunakan *feature extraction CV* ataupun *TF-IDF*, di mana selalu memberikan akurasi yang tinggi yaitu lebih dari 80% pada setiap dimensi.

Kata kunci: Analisis Sentimen, SVM, KNN, RF.

ABSTRACT

Social media is an online medium, where users can easily participate, share and create posts. Social networks such as Twitter are the most common form of social media used by people throughout the world. One of the posts on Twitter social media is a tweet about the Kereta Api Indonesia (KAI). These tweets will usually produce sentiment, where these sentiments can be processed into sentiment analysis.

This research will aim to determine sentiment analysis on Twitter social media regarding the quality of KAI services in the period after the COVID-19 pandemic ends and find out the best machine learning model for carrying out this sentiment analysis. The data that will be used are tweets about @KAI121 in the period 22 June 2023 to 31 December 2023. This research will use the models TF-IDF SVM, TF-IDF RF, TF-IDF KNN, CV SVM, CV RF, and CV KNN will be used to compare which machine learning model is the best.

The sentiment results for each dimension of KAI service quality during the end of the COVID-19 pandemic show that KAI passengers tend to show a fairly neutral response. Apart from that, the SVM shows the best performance using either feature extraction CV or feature extraction TF-IDF, which always provides high accuracy, namely more than 80% in each dimension.

Keywords: Sentiment Analysis, SVM, KNN, RF.

KATA PENGANTAR

Puji dan syukur penulis ucapkan kepada Tuhan Yang Maha Esa karena berkat rahmat dan karunia-Nya penulis dapat menyelesaikan pembuatan tugas akhir “Analisis Sentimen Di Media Sosial Tentang Kualitas Layanan KAI Setelah Pandemi Menggunakan *Machine Learning*” ini dengan baik.

Penulisan tugas akhir ini bertujuan untuk memenuhi salah satu syarat untuk mencapai derajat magister Informatika dari Program Studi Magister Informatika, Program Pascasarjana di Universitas Atma Jaya Yogyakarta.

Penulis menyadari bahwa dalam pembuatan tugas akhir ini penulis telah mendapatkan bantuan, bimbingan, dan dorongan dari banyak pihak. Untuk itu, pada kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Tuhan Yesus Kristus, Bunda Maria, dan Allah Bapa yang selalu membimbing dalam iman-Nya, memberikan berkat-Nya, dan menyertai penulis selalu.
2. Program studi Magister Informatika yang menjadi tempat penulis melanjutkan pendidikan magister.
3. Bapak Yonathan Dri Handarkho S.T., M.Eng., Ph.D. selaku Kepala Program Studi Magister Informatika, Universitas Atma Jaya Yogyakarta.
4. Bapak Paulus Mudjihartono, S.T., M.T., Ph.D., selaku dosen pembimbing I yang telah membimbing dan memberikan masukan serta motivasi kepada penulis untuk menyelesaikan tugas akhir ini.

5. Bapak Prof. Dr. Pranowo, S.T., M.T., selaku dosen pembimbing II yang telah membimbing dan memberikan masukan serta motivasi kepada penulis untuk menyelesaikan tugas akhir ini.
6. Kedua orang tua, kakak dan saudara-saudara yang telah mendoakan dan memberikan dukungan baik moral maupun material kepada penulis untuk selalu berusaha mencapai hasil yang terbaik.
7. Teman-teman yang selalu mendoakan dan menyemangati penulis selama penulis berproses di jenjang kuliah ini.
8. Semua pihak yang tidak dapat disebutkan satu per satu yang selalu memberikan doa, dukungan dan semangat kepada penulis.

Demikian laporan tugas akhir ini dibuat, dan penulis mengucapkan terima kasih kepada semua pihak. Semoga laporan tesis ini dapat bermanfaat bagi pembaca.

Yogyakarta, Februari 2024

Stephanus Wisnu Setyawan

215311587

DAFTAR ISI

HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN	iii
INTISARI.....	iv
ABSTRACT.....	v
KATA PENGANTAR	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
BAB I.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Pertanyaan Penelitian	4
1.4 Batasan Masalah.....	4
1.5 Tujuan dan Manfaat Penelitian	5
1.5.1 Tujuan Penelitian	5
1.5.2 Manfaat Penelitian	5
1.6 Sistematis Penulisan.....	5

BAB II.....	7
2.1 Tinjauan Pustaka	7
BAB III	19
3.1 Analisis Sentimen	19
3.2 Kualitas Layanan.....	19
3.3 PT KAI.....	21
3.4 <i>Random Forest</i>	22
3.5 <i>K-Nearest Neighbors</i>	22
3.6 <i>Support Vector Machine</i>	23
3.7 <i>Feature Extraction</i>	24
3.7.1 <i>CountVectorizer</i>	24
3.7.2 <i>Term Frequency-Inverse Document Frequency</i>	24
3.8 <i>Confusion Matrix</i>	26
BAB IV	29
4.1 Studi Literatur	30
4.2 Pengumpulan Data	30
4.3 <i>Preprocessing Data</i>	30
4.3.1 <i>Clean Text</i>	31
4.3.2 <i>Steam Text</i>	31
4.3.3 <i>Remove Stopword</i>	31

4.3.4	<i>Spellcheck</i>	32
4.4	Pengelompokan Data Dimensi Kualitas Layanan.....	32
4.5	Proses Labeling	34
4.6	Proses Analisis Dan Model Evaluasi	43
4.7	Pelaporan.....	43
BAB V.....		45
5.1	Analisis Sentimen Untuk Mengetahui Nilai Kualitas Layanan	45
5.1.1	Dimensi <i>Tangibles</i>	46
5.1.2	Dimensi <i>Reliability</i>	47
5.1.3	Dimensi <i>Responsiveness</i>	48
5.1.4	Dimensi <i>Assurance</i>	49
5.1.5	Dimensi <i>Empathy</i>	50
5.2	Hasil Analisis dan Model Evaluasi Dengan <i>Machine Learning</i>	51
BAB VI.....		58
6.1	Kesimpulan	58
6.2	Saran.....	59
DAFTAR PUSTAKA		61

DAFTAR TABEL

Tabel 1. Penelitian Terdahulu.....	11
Tabel 2. Contoh Corpus Lima Dimensi Kualitas Layanan.....	32
Tabel 3. Hasil Pengelompokan Data Ke Dalam Dimensi Kualitas Layanan.	33
Tabel 4. Score Pada Sentimen.	34
Tabel 5. Contoh Kamus Kata Positif.....	35
Tabel 6. Contoh Kamus Kata Negatif.	36
Tabel 7. Hasil Labeling Dari Setiap Dimensi Kualitas Layanan.....	42
Tabel 8. Rangkuman Model Maching Learning Pada Tiap Dimensi.....	57

DAFTAR GAMBAR

Gambar 1. Confusion Matrix.....	27
Gambar 2. Diagram Metodologi Penelitian.....	29
Gambar 3. Kode Untuk Melakukan Import Kamus Kata.	36
Gambar 4. Kode Untuk Memberikan Tanda Positif Dan Negatif.	37
Gambar 5. Contoh Hasil Pemberian Tanda [+] dan Tanda [-].	38
Gambar 6. Contoh Hasil Skor Positif.	39
Gambar 7. Contoh Hasil Skor Negatif.	40
Gambar 8. Contoh Hasil Perhitungan Skor.	41
Gambar 9. Contoh Hasil Pemberian Label.	42
Gambar 10. Diagram Lima Dimensi Kualitas Layanan Dan Hasil Sentimen.	46
Gambar 11. Sentimen Pada Dimensi <i>Tangibles</i>	47
Gambar 12. Sentimen Pada Dimensi <i>Reliability</i>	48
Gambar 13. Sentimen Pada Dimensi <i>Responsiveness</i>	49
Gambar 14. Sentimen Pada Dimensi <i>Assurance</i>	50
Gambar 15. Sentimen Pada Dimensi <i>Empathy</i>	51
Gambar 16. Model Maching Learning Pada Dimensi <i>Tangibles</i>	52
Gambar 17. Model Maching Learning Pada Dimensi <i>Reliability</i>	53
Gambar 18. Model Maching Learning Pada Dimensi <i>Responsiveness</i>	54
Gambar 19. Model Maching Learning Pada Dimensi <i>Assurance</i>	55
Gambar 20. Model Maching Learning Pada Dimensi <i>Empathy</i>	56