



**SENTIMEN ANALISIS KOMENTAR VIDEO *YOUTUBE* MENGENAI PENERAPAN  
JALAN BERBAYAR ELEKTRONIK (*ELECTRONIC ROAD PRICING*) DI DKI  
JAKARTA DENGAN METODE *SUPPORT VECTOR MACHINE* DAN ASOSIASI KATA**

**SKRIPSI**

**NATASYA FEBRIYANTI**

**2010512025**

**SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA  
2024**



**SENTIMEN ANALISIS KOMENTAR VIDEO *YOUTUBE* MENGENAI PENERAPAN  
JALAN BERBAYAR ELEKTRONIK (*ELECTRONIC ROAD PRICING*) DI DKI  
JAKARTA DENGAN METODE *SUPPORT VECTOR MACHINE* DAN ASOSIASI KATA**

**SKRIPSI**

**Diajukan Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Komputer**

**NATASYA FEBRIYANTI**

**2010512025**

**SISTEM INFORMASI**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA**

**2024**

## PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya saya sendiri, dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Natasya Febriyanti

NIM : 2010512025

Tanggal : 8 Maret 2024

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia dituntut dan diproses dengan ketentuan berlaku

Jakarta, 8 Maret 2024

Yang Menyatakan,

  
Natasya Febriyanti

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN  
AKADEMIS**

Sebagai civitas akademik Universitas Pembangunan Nasional "Veteran" Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Natasya Febriyanti

NIM : 2010512025

Fakultas : Ilmu Komputer

Program Studi : SI – Sistem Informasi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional "Veteran" Jakarta Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

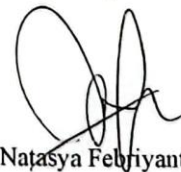
**Sentimen Analisis Komentar Video Youtube Mengenai Penerapan Jalan Berbayar Elektronik (*Electronic Road Pricing*) di DKI JAKARTA dengan Metode *Support Vector Machine* dan Asosiasi Kata**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional "Veteran" Jakarta berhak menyimpan, mengalih data/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenar-benarnya.

Dibuat di : Jakarta

Pada tanggal : 21 Mei 2024

Yang Menyatakan,



Natasya Febriyanti

## LEMBAR PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Natasya Febriyanti

NIM : 2010512025

Program Studi : S1 – Sistem Informasi

Judul Tugas Akhir : Sentimen Analisis Komentar Video Youtube Mengenai Penerapan Jalan Berbayar Elektronik (Electronic Road Pricing) di DKI JAKARTA dengan Metode Support Vector Machine dan Asosiasi Kata

Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian dari persyaratan yang diperlukan untuk memperoleh gelar Sarjana Ilmu Komputer pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jakarta.



Iin Ernawati S.Kom., M.Si.

Penguji 1



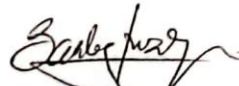
Ika Nurlaili, S.Kom., M.Sc.

Penguji 2



Nur Hafifah Matondang, S.Kom., M.M., M.T.I.

Pembimbing 1



Bambang Tri Wahyono, S.Kom., M.Si.

Pembimbing 2



Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM

Dekan



Anita Muliawati, S.Kom., M.TI

Ketua Program Studi S1 Sistem Informasi

Ditetapkan di : Jakarta

Tanggal Ujian : 3 April 2024

**SENTIMEN ANALISIS KOMENTAR VIDEO *YOUTUBE* MENGENAI PENERAPAN  
JALAN BERBAYAR ELEKTRONIK (*ELECTRONIC ROAD PRICING*) DI DKI  
JAKARTA DENGAN METODE SUPPORT VECTOR MACHINE DAN ASOSIASI  
KATA**

**Natasya Febriyanti**

**ABSTRAK**

Dinas Perhubungan DKI Jakarta berencana akan menerapkan sistem jalan berbayar atau bisa disebut sebagai *Electronic Road Pricing* (ERP) kepada 25 jalan protokol yang ada di Jakarta. Banyak tanggapan dan respon masyarakat yang mereka sampaikan melalui video *Youtube* kanal CNBC Indonesia terhadap rencana kebijakan tersebut. Respon dalam bentuk komentar tersebut jika digabungkan, diproses dengan menerapkan pendekatan analisis sentimen akan menghasilkan klasifikasi berupa sentimen positif, negatif maupun netral. Penelitian ini akan menggunakan 1196 komentar untuk dapat dilakukan klasifikasi komentar masyarakat terhadap rencana kebijakan 25 Jalan Berbayar di DKI Jakarta menjadi sentimen positif, negatif dan netral dengan menggunakan metode *Support Vector Machine* (SVM). SVM sebagai salah satu metode klasifikasi memiliki ketidakmampuan memetakan data *non-linear* untuk memperoleh *hyperplane* yang optimal. Untuk itu diterapkan penambahan *kernel trick* bernama *Radial Basis Function* (RBF) yang memanfaatkan *Grid Search* untuk menentukan kombinasi parameter *Cost* dan *Gamma* untuk mendapatkan akurasi klasifikasi terbaik. Selain itu terdapat permasalahan lain, yaitu terjadinya ketidakseimbangan antara kelas mayor dan minor. Maka diperlukan metode SMOTE untuk mengatasi *imbalance* antar kelas tersebut. Pemodelan dengan kombinasi beberapa metode tersebut memperoleh nilai *accuracy* sebesar 90,6% atau dapat diketahui kombinasi metode akan menghasilkan *performance* yang dapat dikatakan cukup baik dalam melakukan prediksi dan klasifikasi data

**Kata Kunci:** ERP, Analisis Sentimen, Jalan Berbayar, Kemacetan, *Support Vector Machine*

**SENTIMEN ANALISIS KOMENTAR VIDEO *YOUTUBE* MENGENAI PENERAPAN  
JALAN BERBAYAR ELEKTRONIK (*ELECTRONIC ROAD PRICING*) DI DKI  
JAKARTA DENGAN METODE SUPPORT VECTOR MACHINE DAN ASOSIASI  
KATA**

**Natasya Febriyanti**

**ABSTRACT**

The DKI Jakarta Transportation Agency is planning to implement an *Electronic Road Pricing* (ERP) system on 25 protocol roads in Jakarta. There have been public responses to this policy plan, which have been conveyed through *youtube* videos on the CNBC Indonesia channel. If these comments are combined and processed using a sentiment analysis approach, they can be classified as having a positive, negative or neutral sentiment. This study aims to classify 1196 public comments on the policy plan for 25 Paid Roads in DKI Jakarta into positive, negative, and neutral sentiment using the Support Vector Machine (SVM) method. SVM is a classification method that cannot map non-linear data to obtain an optimal *hyperplane*. To achieve the best classification accuracy, an additional kernel trick called Radial Basis Function (RBF) is applied. This utilises Grid Search to determine the optimal combination of Cost and Gamma parameters. Additionally, there is an imbalance between major and minor classes that needs to be addressed. Therefore, the SMOTE method is used to overcome this imbalance. Using a combination of several methods, an accuracy value of 90.6% was achieved. This indicates that the combination of methods is effective in classification.

**Keywords:** ERP, *Sentiment Analysis*, *Paid Road*, *Congestion*, *Support Vector Machine*

## KATA PENGANTAR

Dengan memanjatkan puji dan syukur kehadirat Allah SWT yang telah melimpahkan rahmat, taufik, dan hidayah-Nya sehingga peneliti dapat menyelesaikan Proposal Tugas Akhir (Seminar Teknologi Informasi) dengan judul **“Sentimen Analisis Komentar Video Youtube Mengenai Penerapan Jalan Berbayar Elektronik (*Electronic Road Pricing*) di DKI JAKARTA dengan Metode *Support Vector Machine* dan Asosiasi Kata”** dengan baik. Proposal Tugas Akhir ini ditempuh karena merupakan prasyarat untuk pengambilan Tugas Akhir/Skripsi.

Dalam merampungkan proposal ini tidak lepas dari pertolongan banyak pihak yang telah memberikan saran, motivasi serta dukungan. Untuk itu peneliti mengucapkan banyak terima kasih kepada:

1. Allah SWT atas segala rahmat, ridha, dan karunia-Nya.
2. Kedua orang tua yang selalu memberikan motivasi dan semangat baik secara moral maupun material.
3. Bapak Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM, selaku Dekan Fakultas Ilmu Komputer UPNVJ
4. Ibu Anita Muliawati, S.Kom., M.T.I, selaku Ketua Program Studi Sarjana Jurusan Sistem Informasi.
5. Ibu Nur Hafifah Matondang, S.Kom., MM., M.T.I, selaku Dosen Pembimbing 1
6. Bapak Bambang Tri Wahyono, S.Kom., M.Si, selaku Dosen Pembimbing 2
7. Teman-teman semasa SMP terdiri dari Tiara Juniarti, Farah Garnisha, Fathan Mubiyna, Fauzan Akmal, dan Ryan Fathurrachman yang telah membantu memotivasi penulis hingga sampai dititik ini
8. Serta semua pihak yang tidak dapat peneliti sebutkan satu persatu tanpa mengurangi rasa hormat.

Peneliti menyadari bahwa dalam menyelesaikan Skripsi ini masih banyak kekurangan, baik dari materi maupun teknik penulisan, mengingat kurangnya pengetahuan dan pengalaman peneliti. Oleh karena itu, kritik dan saran yang membangun akan sangat bermakna bagi peneliti.

Jakarta, 8 Maret 2024

Peneliti



## DAFTAR ISI

PERNYATAAN ORISINALITAS .....	ii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS .....	iii
LEMBAR PENGESAHAN.....	iv
ABSTRAK .....	v
ABSTRACT .....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR .....	xi
DAFTAR TABEL .....	xii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	6
1.3 Tujuan Penelitian .....	6
1.4 Manfaat Penelitian .....	7
1.4.1 Bagi Pemerintah .....	7
1.4.2 Bagi Masyarakat.....	7
1.4.3 Bagi <i>Programmer</i> /Peneliti .....	7
1.5 Ruang Lingkup .....	8
1.6 Luaran Penelitian .....	8
1.7 Sistematika Penulisan .....	9
BAB II TINJAUAN PUSTAKA.....	10
2.1 Lalu Lintas .....	10
2.2 Jalan Raya.....	10
2.3 Protokol Berkendara .....	11
2.4 Pertumbuhan Volume Kendaraan Bermotor .....	13
2.5 Kemacetan .....	15
2.6 <i>Electronic Road Pricing (ERP)</i> .....	16
2.7 <i>Data mining</i> .....	17
2.8 <i>Text mining</i> .....	18
2.9 Machine Learning.....	18
2.10 Analisis Sentimen .....	19
2.11 <i>Youtube</i> .....	20
2.12 <i>Pre-processing Data</i> .....	21
2.12.1 <i>Case folding</i> .....	21
2.12.2 <i>Cleaning</i> .....	21

2.12.3	<i>Normalization</i> .....	22
2.12.4	<i>Tokenizing</i> .....	22
2.12.5	<i>Steeming</i> .....	22
2.12.6	<i>Stopword</i> .....	22
2.13	<i>Lexicon Based (InSet Lexicon)</i> .....	23
2.14	<i>Pembobotan TF-IDF</i> .....	23
2.15	<i>SMOTE</i> .....	25
2.16	<i>Support Vector Machine (SVM)</i> .....	25
2.16.1	<i>Kernel Radial Basis Function</i> .....	27
2.16.2	<i>Grid Search</i> .....	28
2.17	<i>Asosiasi Kata</i> .....	29
2.18	<i>Penelitian Terdahulu</i> .....	30
<b>BAB III METODOLOGI PENELITIAN</b> .....		34
3.1	<i>Tahapan Penelitian</i> .....	34
3.1.1	<i>Identifikasi Masalah</i> .....	34
3.1.2	<i>Pengumpulan Data</i> .....	35
3.1.3	<i>Pre-processing Data</i> .....	36
3.1.4	<i>Pelabelan Data</i> .....	36
3.1.5	<i>Pembobotan Kata</i> .....	37
3.1.6	<i>SMOTE (Synthetic Minority Oversampling Technique)</i> .....	37
3.1.7	<i>Pembagian Data</i> .....	38
3.1.8	<i>Uji Coba Model Support Vector Machine</i> .....	38
3.1.9	<i>Evaluasi Hasil Model SVM</i> .....	39
3.1.10	<i>Visualiasi &amp; Asosiasi Hasil Model</i> .....	40
3.1.11	<i>Asosiasi Kata</i> .....	41
3.2	<i>Alat Bantu Penelitian</i> .....	41
3.3	<i>Jadwal Penelitian</i> .....	42
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....		44
4.1	<i>Pengumpulan Data</i> .....	44
4.2	<i>Pre-processing Data</i> .....	45
4.2.1	<i>Case Folding</i> .....	45
4.2.2	<i>Cleaning Data</i> .....	46
4.2.3	<i>Normalization</i> .....	46
4.2.4	<i>Tokenizing</i> .....	47
4.2.5	<i>Steeming</i> .....	48
4.2.6	<i>Stopword</i> .....	48
4.3	<i>Pelabelan Data</i> .....	49
4.4	<i>Pembobotan Kata</i> .....	51

4.5	SMOTE (Syntethic Minority Oversampling Technique) .....	58
4.6	Split Data .....	60
4.7	Klasifikasi dengan <i>Support Vector Machine</i> (SVM).....	60
4.8	Evaluasi .....	61
4.9	Visualisasi & Asosiasi Hasil Klasifikasi .....	64
4.9.1	Negatif.....	64
4.9.2	Positif .....	72
BAB V KESIMPULAN DAN SARAN.....		76
5.1	Kesimpulan.....	76
5.2	Saran .....	76
DAFTAR PUSTAKA .....		78
RIWAYAT HIDUP.....		83
LAMPIRAN.....		85

## DAFTAR GAMBAR

Gambar 1. 1 Jumlah Lapangan Kerja DKI Jakarta .....	1
Gambar 1. 2 Jumlah Pencari Kerja DKI Jakarta .....	2
Gambar 2. 1 Support Vector Machine Classification .....	25
Gambar 2. 2 Nilai Margin Support Vector Machine.....	26
Gambar 2. 3 Pemetaan Data ke Dimensi Lebih Tinggi.....	28
Gambar 3. 1 Tahapan Penelitian .....	34
Gambar 3. 2 Tahapan Preprocess Data .....	36
Gambar 3. 3 Alur Pelabelan dengan InSet Lexicon .....	37
Gambar 4. 1 Dataset Hasil Scrapping .....	45
Gambar 4. 2 Hasil Persebaran Data Tiap Kelas/Label.....	50
Gambar 4. 3 Syntax Melakukan Perhitungan IDF .....	52
Gambar 4. 4 Syntax Kalkulasi Bobot TF-IDF .....	58
Gambar 4. 5 Syntax SMOTE .....	59
Gambar 4. 6 Perbedaan Jumlah Data Train Sebelum dan Sesudah SMOTE.....	59
Gambar 4. 7 Syntax Pemodelan SVM dengan Kernel RBF dan GridSearch .....	61
Gambar 4. 8 Hasil Evaluasi Model SVM.....	62
Gambar 4. 9 Diagram Confussion Matrix untuk SVM .....	62
Gambar 4. 10 Top 10 Kata Paling Banyak Muncul Komentar Negatif .....	65
Gambar 4. 11 Wordcloud Sentimen Negatif Electronic Road Pricing .....	65
Gambar 4. 12 Top 10 Kata Paling Banyak Muncul Komentar Positif.....	72
Gambar 4. 13 Wordcloud Sentimen Postif Electronic Road Pricing .....	73

## DAFTAR TABEL

Tabel 2.1 Jumlah Kendaraan Bermotor Jenis Kendaraan DKI Jakarta.....	14
Tabel 2. 2 Klasifikasi <i>Polarity score</i> .....	23
Tabel 2. 3 Penelitian Terdahulu .....	30
Tabel 3. 1 Contoh Komentar video '25 Jalan yang akan berbayar' .....	36
Tabel 3. 2 Nilai Parameter Yang Akan Digunakan.....	38
Tabel 3. 3 Confusion Matrix .....	39
Tabel 3. 4 Jadwal Penelitian.....	42
Tabel 4. 1 Variabel dan Deskripsi .....	44
Tabel 4. 2 Hasil Case Folding .....	46
Tabel 4. 3 Hasil Cleaning Data .....	46
Tabel 4. 4 Hasil Normalization .....	47
Tabel 4. 5 Hasil Tokenizing .....	47
Tabel 4. 6 Hasil Stemming .....	48
Tabel 4. 7 Hasil Stopword.....	48
Tabel 4. 8 Hasil Pelabelan Data (InSet Lexicon).....	49
Tabel 4. 9 Dokumen Komnetar .....	52
Tabel 4. 10 Perhitungan Nilai TF.....	53
Tabel 4. 11 Kalkulasi Nilai DF .....	54
Tabel 4. 12 Perhitungan Nilai IDF .....	56
Tabel 4. 13 Kalkulasi Bobot TF-IDF .....	57
Tabel 4. 14 Perbandingan Data Latih dan Data Uji .....	60