

**PERBANDINGAN ANALISIS SENTIMEN TWITTER MENGGUNAKAN
METODE SUPPORT VECTOR MACHINE DAN K-NEAREST
NEIGHBOR PADA TRANSPORTASI ONLINE
DI INDONESIA**

SKRIPSI



disusun oleh

Dzaki Fauzi Mubarok

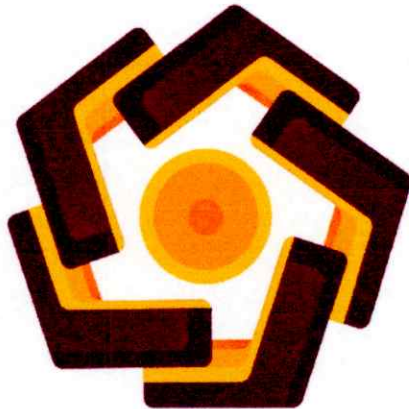
15.11.9324

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2019**

**PERBANDINGAN ANALISIS SENTIMEN TWITTER MENGGUNAKAN
METODE SUPPORT VECTOR MACHINE DAN K-NEAREST
NEIGHBOR PADA TRANSPORTASI ONLINE
DI INDONESIA**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Dzaki Fauzi Mubarok

15.11.9324

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2019**

PERSETUJUAN

SKRIPSI

PERBANDINGAN ANALISIS SENTIMEN TWITTER MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN K-NEAREST NEIGHBOR PADA TRANSPORTASI ONLINE DI INDONESIA

yang dipersiapkan dan disusun oleh

Dzaki Fauzi Mubarak

15.11.9324

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 16 Oktober 2018

Dosen Pembimbing,



Windha Mega Pradnya D, M.Kom.

NIK. 190302185

PENGESAHAN

SKRIPSI

PERBANDINGAN ANALISIS SENTIMEN TWITTER MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN K-NEAREST NEIGHBOR PADA TRANSPORTASI ONLINE

DI INDONESIA

yang dipersiapkan dan disusun oleh
Dzaki Fauzi Mubarak

15.11.9324

telah dipertahankan di depan Dewan Penguji
pada tanggal 15 Februari 2019

Susunan Dewan Penguji

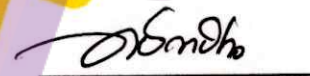
Nama Penguji

Heri Sismoro, M.Kom
NIK. 190302057

Bety Wulan Sari, M.Kom
NIK. 190302254

Windha Mega P.D., M.Kom
NIK. 190302185

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 28 Februari 2018

DEKAN FAKULTAS ILMU KOMPUTER




Krisnawati, S.Si., M.T.
NIK. 190302038

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 28 Februari 2019



Dzaki Fauzi Mubarak

NIM. 15.11.9324

PERSEMBAHAN

Alhamdulillah segala puji syukur atas kehadiran Allah SWT yang telah memberikan limpahan rahmat dan karunia-Nya sehingga penelitian ini dapat dilakukan dan diselesaikan dengan sebaik-baiknya. Saya juga ucapkan terimakasih untuk dukungan dan bantuan semua pihak yang membantu selesainya penelitian ini.

Skripsi ini saya persembahkan untuk kampus saya, Universitas AMIKOM Yogyakarta.

Untuk kedua orang tua saya yang telah susah payah mendukung saya tanpa kenal lelah.

Untuk ibu Windha Mega Pradnya, yang sudah dengan sabar membimbing saya. Saya ucapkan terimakasih sedalam-dalamnya.

Untuk para pembaca, saya ucapkan terima kasih sebesar-besarnya. Semoga menjadi amal jariyah atas ilmu yang saya bagikan.

Dan untuk semua pihak yang telah banyak membantu dalam penyusunan skripsi ini yang tidak dapat saya sebutkan satu persatu.

MOTTO

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

” فَإِنَّ مَعَ الْعُسْرِ يُسْرًا إِنَّ مَعَ الْعُسْرِ يُسْرًا ”

- Al-Qur'an Surah Al-Insyiraah (94: 5-6)

” Our main business is not to see what lies dimly as a distance, but to do what lies clearly at at hand. ”

- Sir William Osler book *How To Stop Worrying and Start Living*

” Hidup kita dibentuk oleh pemikiran kita sendiri ”

- Marcus Aurelius

” Rasa aman yang menyenangkan, ketentraman batin yang mendalam, serta kebahagiaan sejati akan menyegarkan seluruh syaraf manusia bila manusia asyik melaksanakan tugas pekerjaannya ”

- John Cowper Powys book *The Art of Forgetting the Unpleasant*

7. Makhluk-makhluk Kontrakan yang sudah menemani dan bersenang-senang bersama penulis selama perkuliahan.

Penulis menyadari dengan betul bahwa dalam pembuatan skripsi ini masih banyak sekali kekurangan dan kelemahan. Oleh karena itu penulis berharap untuk semua pihak yang telah membaca dan memahami penelitian ini untuk dapat menyampaikan kritik dan saran yang membangun sehingga dapat menambah kesempurnaan skripsi ini.

Akhir kata, semoga skripsi ini dapat bermanfaat bagi pihak terkait dan pembaca pada umumnya.

Kamis , 28 Februari 2019

Dzaki Fauzi Mubarak

15.11.9324

KATA PENGANTAR

Puji syukur penulis haturkan atas kehadiran Allah *Subhanahu wa Ta'ala* yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “Perbandingan Analisis Sentimen Twitter Menggunakan Support Vector Machine dan K-Nearest Neighbor Pada Transportasi Online di Indonesia” dengan sebaik-baiknya. Tidak lupa shalawat serta salam penulis haturkan kepada junjungan kita, nabi agung Nabi Muhammad *Shallallahu 'alaihi wa Sallam* yang telah membawa cahaya ilmu pengetahuan ke tengah-tengah kita.

Dengan selesainya skripsi ini, penulis mengucapkan banyak terima kasih kepada:

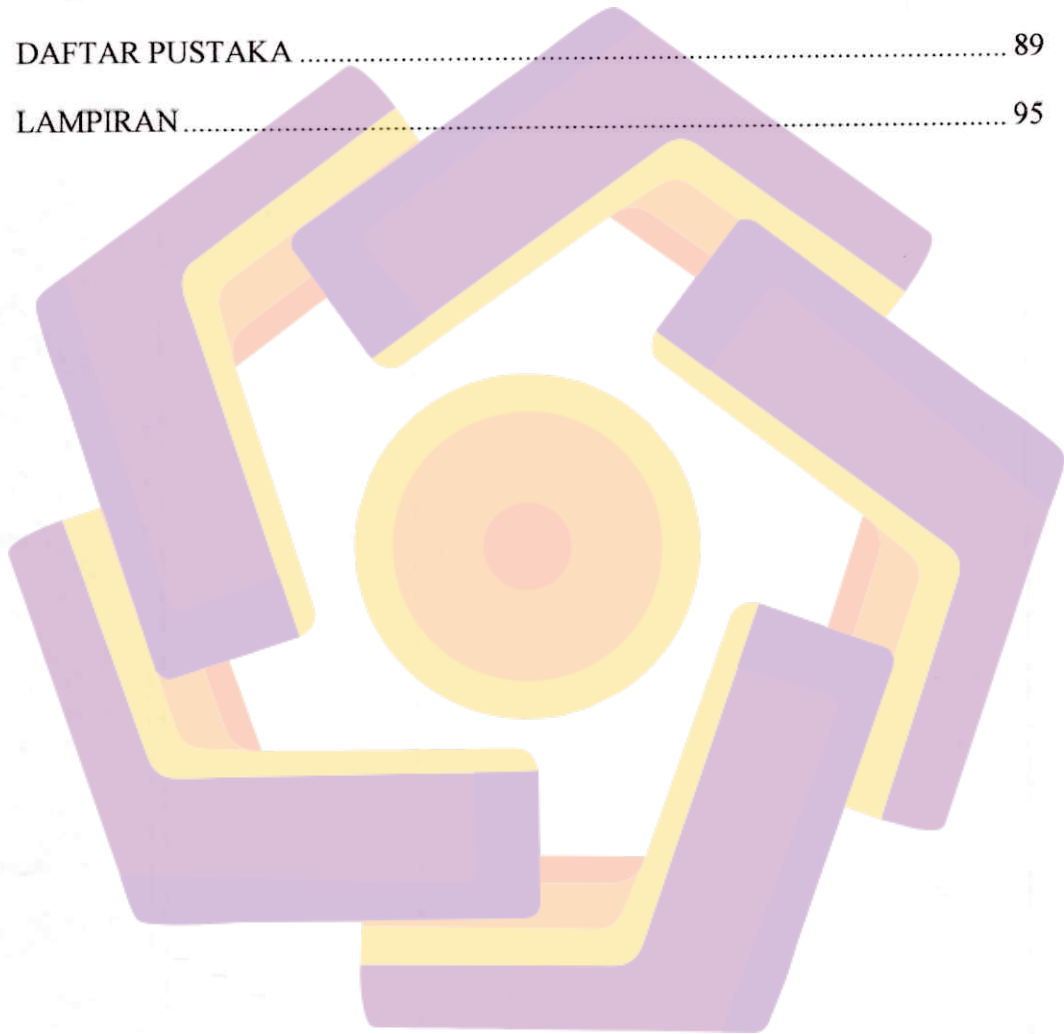
1. Bapak Prof. Dr. M. Suyanto selaku Rektor Universitas AMIKOM Yogyakarta
2. Ibu Krisnawati, S.Si., M.T. selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Tri Susanto, M.Kom selaku Dosen wali.
4. Ibu Windha Mega Pradnya D, M.Kom. selaku pembimbing yang telah dengan sabar membimbing dan banyak memberikan kontribusi bagi penulis dalam pembuatan skripsi ini.
5. Para Dosen dan Staff Universitas AMIKOM Yogyakarta yang telah banyak memberi bantuan informasi, pengalaman dan kontribusi lainnya selama penulis berkuliah di kampus ini hingga terselesaikannya skripsi.
6. Orang tua dan saudara penulis yang telah memberikan dukungan terbaiknya selama berkuliah.

DAFTAR ISI

COVER.....	i
PERSETUJUAN.....	iii
PENGESAHAN.....	iv
PERNYATAAN.....	v
PERSEMBAHAN.....	vi
MOTTO.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
INTISARI.....	xvi
<i>ABSTRACT</i>	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Maksud dan Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.6 Metode Penelitian.....	4
1.7 Sistematika Penulisan.....	7
BAB II LANDASAN TEORI.....	8
2.1 Tinjauan Pustaka.....	8
2.2 Teori.....	10
2.2.1 Text Mining.....	10

2.2.2	Analisis Sentimen	11
2.2.3	Pre-processing	11
2.2.4	Klasifikasi	12
2.2.5	Term Frequency – Inverse Document Frequency	13
2.2.6	K-Nearest Neighbor	15
2.2.7	Support Vector Machine	17
2.2.8	Evaluasi	24
2.2.9	Validasi	25
BAB III METODE PENELITIAN.....		27
3.1	Alat dan Bahan	27
3.1.1	Alat Penelitian	27
3.1.2	Bahan Penelitian.....	28
3.2	Alur Penelitian	28
3.2.1	Pengumpulan Data	30
3.2.2	Pre-Processing Data	31
3.2.3	TF-IDF	33
3.2.4	Klasifikasi	40
3.2.5	Evaluasi	55
3.2.6	Validasi	56
BAB IV HASIL DAN PEMBAHASAN		57
4.1	Eksperimen.....	57
4.1.1	Pengumpulan Data	57
4.1.2	Pre-processing Data	61
4.1.3	Pembobotan	68
4.1.4	Pelatihan Data	70
4.1.5	Pengujian Data	71
4.1.6	Evaluasi Model.....	73
4.1.7	Validasi	74
4.2	Hasil dan Pembahasan.....	76
4.2.1	Hasil Evaluasi.....	76

4.2.2	Pengujian Sentimen.....	81
4.2.3	Hasil Validasi.....	82
BAB V KESIMPULAN DAN SARAN.....		87
5.1	Kesimpulan.....	87
5.2	Saran.....	88
DAFTAR PUSTAKA.....		89
LAMPIRAN.....		95



DAFTAR TABEL

Tabel 3. 1 Spesifikasi Perangkat keras dan Perangkat Lunak	27
Tabel 3. 2 Contoh proses pembobotan Term Frequency	34
Tabel 3. 3 Contoh proses pembobotan Inverse Document Frequency	36
Tabel 3. 4 Contoh hasil proses TF-IDF	38
Tabel 3. 5 Penentuan data uji dan data latih	41
Tabel 3. 6 Hasil perkalian scalar data uji dengan seluruh data latih.....	43
Tabel 3. 7 Perhitungan panjang setiap dokumen	45
Tabel 3. 8 Data latih dalam format Support Vector.....	49
Tabel 3. 9 Hasil kernelisasi matrix $x_i^T x$	53
Tabel 4. 1 Hasil perhitungan akurasi dan waktu proses dataset Gojek.....	77
Tabel 4. 2 Hasil perhitungan akurasi dan waktu proses dataset Grab	78
Tabel 4. 3 Hasil perhitungan Recall, precision, F1 Score.....	79
Tabel 4. 4 Hasil perhitungan Recall, precision, F1 Score.....	80
Tabel 4. 5 Hasil perhitungan Akurasi dan Proses	80
Tabel 4. 6 Hasil perhitungan Akurasi dan Proses K=3	81
Tabel 4. 7 Hasil perhitungan Akurasi dan Proses SVM	81
Tabel 4. 8 Hasil perhitungan akurasi K-NN (K=3) dengan validasi	82
Tabel 4. 9 Hasil perhitungan akurasi SVM dengan validasi.....	84
Tabel 4. 10 Hasil perhitungan akurasi dengan waktu proses K-Fold Cross Validation.....	85

DAFTAR GAMBAR

Gambar 2. 1 Alur kerja klasifikasi.....	13
Gambar 2. 2 Ilustrasi metode K-NN.....	16
Gambar 2. 3 Ilustrasi Metode SVM.....	18
Gambar 2. 4 Maximum hyperplane dengan support vector.....	19
Gambar 2. 5 <i>Confusion Matrix</i>	24
Gambar 2. 6 Ilustrasi k-Fold Cross Validation.....	26
Gambar 3. 1 Diagram alir tahapan eksperimen.....	30
Gambar 3. 2 Diagram alir tahapan pre-processing.....	32
Gambar 3. 3 Diagram alir tahapan K-NN.....	41
Gambar 3. 4 Diagram alir tahapan SVM.....	48
Gambar 4. 1 Registrasi ke Twitter API.....	58
Gambar 4. 2 Keys yang disediakan oleh Twitter.....	58
Gambar 4. 3 Keys dimasukkan ke dalam file Python.....	59
Gambar 4. 4 Listener ke database Twitter.....	60
Gambar 4. 5 Streaming ke database Twitter.....	60
Gambar 4. 6 Konversi dari file DB ke CSV.....	61
Gambar 4. 7 <i>Import</i> file CSV dengan <i>library</i> Pandas.....	61
Gambar 4. 8 Implementasi <i>case folding</i>	61
Gambar 4. 9 Cuplikan hasil <i>case folding</i>	62
Gambar 4. 10 Implementasi <i>cleaning</i>	63
Gambar 4. 11 Cuplikan hasil <i>cleaning</i>	63
Gambar 4. 12 Cuplikan hasil <i>cleaning</i>	64
Gambar 4. 13 Implementasi <i>Stemming</i>	64
Gambar 4. 14 Contoh hasil <i>Stemming</i>	65
Gambar 4. 15 Implementasi <i>stopwords removing</i>	65
Gambar 4. 16 Contoh hasil <i>stopwords removing</i>	66
Gambar 4. 17 Implementasi <i>filtering</i>	66
Gambar 4. 18 Contoh hasil <i>filtering</i>	67

Gambar 4. 19 Implementasi tokenizing.....	67
Gambar 4. 20 Contoh hasil tokenizing	68
Gambar 4. 21 Implementasi TF-IDF	68
Gambar 4. 22 Contoh hasil TF-IDF	69
Gambar 4. 23 Implementasi K-NN.....	71
Gambar 4. 24 Implementasi SVM.....	71
Gambar 4. 25 Load model classifier pada file pickle	72
Gambar 4. 26 Transformasi dan prediksi sentimen data uji.....	72
Gambar 4. 27 Scoring model classifier.....	73
Gambar 4. 28 Perhitungan waktu model classifier	74
Gambar 4. 29 Implementasi K-Fold Cross Validation	75
Gambar 4. 30 Perhitungan waktu K-Fold Cross Validation.....	75
Gambar 4. 31 Grafik batang presentase akurasi setiap K pada K-NN.....	76
Gambar 4. 32 Grafik batang presentase akurasi setiap K pada K-NN.....	85
Gambar 4. 33 Grafik batang presentase waktu proses setiap K pada K-NN	86

INTISARI

Transportasi online merupakan salah satu jenis transportasi yang paling sering digunakan dalam beberapa tahun terakhir, dengan jumlah 15,37 juta pengguna android menggunakan aplikasi transportasi online menurut data comScore per Desember 2017. Dengan adanya transportasi berbasis online, pemanggilan akan lebih mudah dilakukan. Dalam hal ini, opini akan sangat dibutuhkan untuk dijadikan penilaian kualitas pelayanan transportasi online tersebut. Opini-opini tersebut akan mudah dijumpai di Twitter, salah satu platform media sosial yang paling terkenal saat ini. Dengan mengakses Twitter, kita akan dengan mudah menjumpai berbagai ulasan tentang transportasi online yang kita gunakan. Akan tetapi, analisis sentimen tetap dibutuhkan karena Twitter tidak bisa menyimpulkan kualitas Transportasi online tersebut.

Penelitian ini akan membuat model analisis sentimen twitter pada transportasi online di Indonesia yang paling populer digunakan yaitu Gojek dan Grab. Pada penelitian ini juga akan membandingkan model terbaik di antara kedua metode yang digunakan, yaitu Support Vector Machine dan K-Nearest Neighbor dan membandingkan metode mana yang memiliki performa yang tinggi dan baik. Evaluasi model menggunakan metode K-fold Cross Validation.

Hasil pada perhitungan akurasi menunjukkan bahwa metode Support Vector Machine lebih unggul dibanding KNN dengan 80% tanpa validasi dan menggunakan validasi 81.5% pada dataset gojek dan pada dataset grab SVM 88% tanpa validasi, menggunakan validasi 82.5%. sedang waktu proses KNN lebih cepat dibanding SVM dengan waktu 0.0790 tanpa validasi dan 1.040second menggunakan validasi pada dataset gojek dan pada dataset grab 0.0390 second tanpa validasi dan 1.305 second menggunakan validasi.

Kata Kunci: *machine learning*, sentimen, klasifikasi, support vector machine, k-nearest neighbor, k-fold cross validation

ABSTRACT

Online transportation is one of transportation types that often used in recent years, with 15,37 million android users are using online-based transportation, according to comScore per December 2017. With online-based transportations, users can call the services easily. In this case, opinion is very important to be the quality assessment of online-based transportation service. These opinions are easy to find in Twitter, one of most popular social media platform. By accessing Twitter, we will be easy to find reviews about the online transportation that we are going to use. But, sentiment analysis is still needed because Twitter cannot assess the quality of the online transportation.

This research will make Twitter sentiment analysis model on most popular online-based transportation in Indonesia: Go-jek and Grab. In this research will compare the best methods between two: Support Vector Machine and K-Nearest Neighbor, and find which method is in the highest and best perform. Model evaluation will use K-fold Cross Validation method.

The calculation of accuracies result that Support Vector Machine method is superior than KNN with 80% accuracy without validation and 81.5% accuracy with validation for Go-Jek dataset, while 88% accuracy without validation and 82.5% with validation for Grab dataset. The calculation of process times result that KNN is superior than Support Vector Machine with 0.0790s without validation and 1.040s with validation for Go-Jek dataset, while 0.0390s without validation and 1.305s with validation for Grab dataset.

Keyword: *machine learning, sentiment, classification, support vector machine, k-nearest neighbor, k-fold cross validation*