

**PERBANDINGAN PERFORMA ALGORITMA K-NEAREST NEIGHBOR
DAN SUPPORT VECTOR MACHINE DALAM ANALISIS SENTIMEN
TWITTER PADA PENGGUNA LAYANAN SELULER**

SKRIPSI



disusun oleh

Muhammad Ammar Farhan Putra

16.11.0627

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2020**

**PERBANDINGAN PERFORMA ALGORITMA K-NEAREST NEIGHBOR
DAN SUPPORT VECTOR MACHINE DALAM ANALISIS SENTIMEN
TWITTER PADA PENGGUNA LAYANAN SELULER**

SKRIPSI

Untuk memenuhi sebagian persyaratan
Mencapai gelar sarjana
Pada Program Studi Informatika



disusun Oleh :

Muhammad Ammar Farhan Putra

16.11.0627

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2020**

PERSETUJUAN

SKRIPSI

**PERBANDINGAN PERFORMA ALGORITMA K-NEAREST NEIGHBOR
DAN SUPPORT VECTOR MACHINE DALAM ANALISIS SENTIMEN
TWITTER PADA PENGGUNA LAYANAN SELULER**

yang dipersiapkan dan disusun oleh

Muhammad Ammar Farhan Putra

16.11.0627

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 18 November 2020

Dosen Pembimbing,

Lilis Dwi Farida, S.Kom., M.Eng.

NIK. 190302288

PENGESAHAN

SKRIPSI

**PERBANDINGAN PERFORMA ALGORITMA K-NEAREST NEIGHBOR
DAN SUPPORT VECTOR MACHINE DALAM ANALISIS SENTIMEN
TWITTER PADA PENGGUNA LAYANAN SELULER**

yang dipersiapkan dan disusun oleh

Muhammad Ammar Farhan Putra

16.11.0627

telah dipertahankan di depan Dewan Penguji
pada tanggal 18 November 2020

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Hartatik, S.T., M.Cs.

NIK. 190302232

Anna Balta, M. Kom.

NIK. 190302290

Lilis Dwi Farida, S.Kom., M.Eng

NIK. 190302288

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 Desember 2020

DEKAN FAKULTAS ILMU KOMPUTER

Krisnawati, S.Si, M.T.

NIK. 190302038

PERNYATAAN

Saya yang bertandatangan di bawah 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, 10 Oktober 2020



6000

Muhammad Ammar Farhan Putra

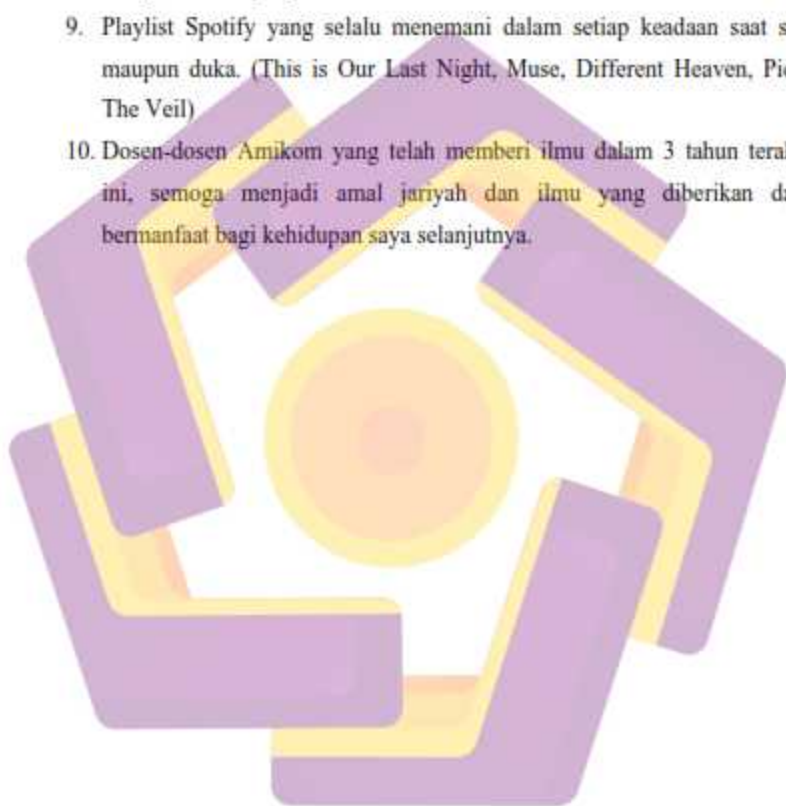
NIM. 16.11.0627

PERSEMBAHAN

Puji Syukur kepada Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya. Shalawat kepada Nabi Muhammad SAW yang telah memberikan contoh Akhlakul Kharimah bagi seluruh umat muslim di penjuru dunia. Didalam pengerjaan skripsi ini telah melibatkan banyak pihak yang sangat membantu. Oleh sebab itu, disini penulis sampaikan rasa terimakasih sedalam-dalamnya kepada:

1. Keluarga penulis, Orang Tua saya Ibu (Tri Suhartini, Dra.) dan Bapak (Jana Susila, Drs.), yang selalu mendoakan untuk kesuksesan anaknya, Kedua saudara saya Amalia Putri Nurlita Sari dan Alifia Putri Nur Ramadhani yang selalu memberikan semangat untuk menyelesaikan skripsi, dan semua dukungan yang telah diberikan sehingga saya bisa menyelesaikan studi dari Universitas Amikom Yogyakarta dengan baik dan lancar.
2. Dosen Pembimbing Ibu Lilis Dwi Farida, terimakasih atas bimbingan, ilmu, dan waktu yang telah diberikan selama pengerjaan skripsi sehingga penulis dapat menyelesaikan skripsi ini.
3. Salma Hayyu Nur Husna yang selalu mengingatkan untuk mengerjakan skripsi, selalu memberi semangat dan motivasi untuk segera menyelesaikan skripsi dan selalu memberikan semua bantuan hingga saat ini.
4. Jefri Setiawan dan Novi Nurkhaeni yang telah membantu saya dalam pengerjaan data skripsi.
5. Teman-teman TheWolfpack yang memberikan semangat dan hiburan dikala penulis merasa bosan dalam pengerjaan skripsi (Rafid Bambang Sumiyatno, Putra Jatmiko, Arif Rahmat Suifullah).
6. Keluarga besar Purna OSIS SMA N 1 Mertoyudan 14/15 yang selalu menemani dan selalu menghibur dalam keadaan senang dan susah (Ahmad Ghozali, Yogi Adi Saputra, Susilo Hadi Prabowo, Fahmi Hidayat, Annisa Dhiandini, Dheri Sukma Sayekti, Dwi Kristianti, Ovelina Ulfah, M. Firmansyah Arya K., Tahya Cahya Wiguna, M. Hasanuddin Pella, Marcellino F.A.).

7. Keluarga besar 16-IF-10 yang tidak bisa disebutkan satu persatu, kalian menakjubkan.
8. Masih banyak lagi teman-teman yang selalu membantu selama perkuliahan dan pengerjaan skripsi ini. Terimakasih banyak untuk waktu luang dan ilmu yang diberikan.
9. Playlist Spotify yang selalu menemani dalam setiap keadaan saat suka maupun duka. (This is Our Last Night, Muse, Different Heaven, Pierce The Veil)
10. Dosen-dosen Amikom yang telah memberi ilmu dalam 3 tahun terakhir ini, semoga menjadi amal jariyah dan ilmu yang diberikan dapat bermanfaat bagi kehidupan saya selanjutnya.



KATA PENGANTAR

Puji syukur saya ucapkan kehadirat Tuhan Yang Maha Kuasa atas limpahan berkah-Nya sehingga penulis dapat menyelesaikan laporan Penelitian Skripsi berjudul “Perbandingan Performa Algoritma K-Nearest Neighbor dan Support Vector Machine dalam Analisis Sentimen Twitter pada Pengguna Layanan Seluler”.

Pengajuan skripsi ini ditujukan sebagai pemenuhan kelulusan pada jenjang perkuliahan Strata I Universitas Amikom Yogyakarta. Melewati penyusunan skripsi ini tentunya tidak terlepas dari hambatan, tantangan serta kesulitan, namun karena binaan dan dukungan dari semua pihak, akhirnya semua hambatan tersebut dapat teratasi.

Dalam penulisan skripsi ini tentunya penulis sadar akan banyak ditemukan kekurangan pada laporan ini. Baik itu dari segi kualitas maupun dari segi kuantitas bahan observasi yang penulis tampilkan. Selanjutnya penulis mengucapkan terimakasih yang sebanyak-banyaknya kepada segenap pihak yang telah memberikan dukungan, baik itu berupa bantuan, doa maupun dorongan dan beragam pengalaman selama proses penyelesaian penulisan skripsi ini.

Akhir kata, semoga semua bantuan yang telah diberikan oleh segenap pihak dapat menjadi ladang kebaikan. Dan semoga skripsi ini dapat memberikan manfaat.

Yogyakarta, 10 Oktober 2020

Penulis



Muhammad Ammar Farhan Putra

DAFTAR ISI

PERSETUJUAN	iii
PENGESAHAN	iv
PERNYATAAN.....	v
PERSEMBAHAN	vi
KATA PENGANTAR	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
INTISARI.....	14
ABSTRACT.....	15
BAB I PENDAHULUAN.....	16
1.1 Latar Belakang	16
1.2 Rumusan Masalah	19
1.3 Batasan Masalah.....	19
1.4 Tujuan Penelitian.....	20
1.5 Metode Penelitian.....	20
1.5.1 Studi Pustaka.....	20
1.5.2 Pengumpulan Data	21
1.5.3 Analisis dan Perancangan	21
1.5.4 Implementasi	22
1.5.5 Pengujian.....	22
1.5.6 Penulisan Laporan.....	22
1.6 Sistematika Penulisan.....	22
BAB II LANDASAN TEORI.....	24
2.1 Tinjauan Pustaka	24
2.2 Dasar Teori.....	26
2.2.1 Text Mining.....	26
2.2.2 Analisis Sentimen	27
2.2.3 Pre-Processing.....	27
2.2.4 Term Frequency – Inverse Document Frequency (TF-IDF).....	29
2.2.5 Klasifikasi	30

2.2.6	K-Nearest Neighbor	30
2.2.7	Support Vector Machine	32
2.2.8	Confusion Matrix	40
2.2.9	K-fold <i>cross-validation</i>	42
BAB III METODE PENELITIAN.....		44
3.1	Alat dan Bahan	44
3.1.1	Alat Penelitian	44
3.1.2	Bahan Penelitian.....	45
3.2	Alur Penelitian	45
3.2.1	Pengumpulan Data	47
3.2.2	<i>Pre-Processing</i> Data	48
3.2.3	<i>Term Weighting</i> (TF-IDF).....	52
3.2.4	Klasifikasi	57
3.2.5	Evaluasi	69
3.2.6	Validasi	69
BAB IV IMPLEMENTASI DAN PEMBAHASAN		70
4.1	Implementasi	70
4.1.1	Crawling Data	70
4.1.2	Pre-processing Data	74
4.1.3	Pembobotan.....	78
4.1.4	Pelatihan Data	80
4.1.5	Pengujian Data	81
4.1.6	Evaluasi Model.....	83
4.1.7	Validasi	83
4.2	Hasil dan Pembahasan.....	85
4.2.1	Hasil Evaluasi.....	85
4.2.2	Hasil Pengujian Sentimen	89
4.2.3	Hasil Validasi.....	90
BAB V KESIMPULAN DAN SARAN.....		94
5.1	Kesimpulan.....	94
5.2	Saran.....	95
DAFTAR PUSTAKA		96

DAFTAR TABEL

Tabel 2.1 Tabel Confusion Matrix.....	41
Tabel 2.2 Ilustrasi <i>K-fold cross-validation</i> , $k=10$	43
Tabel 3.1 Spesifikasi Perangkat Keras yang digunakan.....	44
Tabel 3.2 Spesifikasi Perangkat Lunak yang digunakan.....	44
Tabel 3.3 Penghitungan nilai bobot Term Frequency.....	53
Tabel 3.4 Penghitungan nilai Inverse Document Frequency.....	55
Tabel 3.5 Penghitungan nilai TF-IDF.....	56
Tabel 3.6 Penentuan Data Latih dan Data Uji.....	60
Tabel 3.7 Nilai <i>distance</i> hasil perhitungan <i>euclidean distance</i> data uji terhadap data latih.....	61
Tabel 3.8 Data Latih berbentuk <i>Support Vector</i>	63
Tabel 3.9 Matriks hasil kernelisasi $\mathbf{x_iT_x}$	67
Tabel 3.10 Nilai Kernel Hasil Perkalian Data Uji dengan Data Latih.....	68
Tabel 4.1 Tabel Nilai <i>accuracy</i> , <i>recall</i> , <i>precision</i> , dan <i>F1-Score</i> (dalam %).	87
Tabel 4.1 Nilai akurasi, <i>F1-Score</i> , dan Waktu proses metode KNN dan SVM pada Seluruh <i>Dataset</i>	88
Tabel 4.2 Tabel Hasil Klasifikasi <i>Tweet</i> dengan Metode KNN.....	89
Tabel 4.3 Tabel Hasil Klasifikasi <i>Tweet</i> dengan Metode SVM.....	90
Tabel 4.4 Tabel Rata-rata <i>F1-Score</i> dan Waktu Proses.....	93

DAFTAR GAMBAR

Gambar 2.1 Visualisasi Algoritma K-Nearest Neighbor	31
Gambar 2.2 Ilustrasi Algoritma Support Vector Machine dengan <i>margin</i> dan <i>hyperplane</i> optimal.....	32
Gambar 3.1 Diagram alir tahapan penelitian	45
Gambar 3.2. Diagram alir tahap <i>pre-processing</i>	49
Gambar 3.3 Diagram Alir Tahap Klasifikasi <i>K-Nearest Neighbor</i>	59
Gambar 3.7 Diagram Alir Tahap Klasifikasi Support Vector Machine.....	62
Gambar 4.1 Keys dan token yang didapat dari registrasi Twitter API	71
Gambar 4.2 Implementasi kode program untuk crawling data dari Twitter.....	71
Gambar 4.3 Implementasi kode program untuk memasukkan data ke dalam file <i>*.csv</i>	72
Gambar 4.4 Hasil dari <i>crawling data</i> Twitter menggunakan <i>library tweepy</i>	73
Gambar 4.5 Hasil proses <i>balancing data</i> terhadap data latih.....	74
Gambar 4.6 Implementasi kode program proses <i>cleaning tweet</i>	75
Gambar 4.7 Hasil dari proses <i>cleaning tweet</i>	75
Gambar 4.8 Implementasi kode program untuk <i>replace slang words</i>	76
Gambar 4.9 Implementasi kode program untuk <i>stemming</i>	77
Gambar 4.10 Implementasi kode program untuk <i>stopword removing</i>	78
Gambar 4.11 Implementasi kode program untuk <i>tokenization</i>	78
Gambar 4.12 Implementasi kode program untuk pembobotan.....	79
Gambar 4.13 Implementasi kode program <i>classifier</i> KNN.....	81
Gambar 4.14 Implementasi kode program <i>classifier</i> SVM.....	81
Gambar 4.15 Implementasi kode program untuk prediksi dan menyimpan hasil klasifikasi data uji.....	82
Gambar 4.16 Implementasi kode program memberikan nilai <i>recall</i> , <i>precision</i> , dan <i>F1-Score</i>	83

Gambar 4.17 Implementasi kode program untuk <i>K-fold cross-validation</i> , dengan nilai $k=10$	84
Gambar 4.18 Grafik nilai akurasi tiap k pada KNN dan SVM terhadap <i>dataset</i> Indosat dan XL.....	86
Gambar 4.19 Grafik nilai <i>F1-Score</i> tiap k pada KNN dan SVM terhadap <i>dataset</i> Indosat dan XL.....	87
Gambar 4.20 Grafik nilai <i>F1-Score</i> model <i>classifier</i> KNN tiap iterasi <i>10-fold cross-validation</i>	91
Gambar 4.21 Grafik nilai <i>F1-Score</i> model <i>classifier</i> SVM tiap iterasi <i>10-fold cross-validation</i>	92



INTISARI

Dalam menunjang kelancaran dan kemudahan komunikasi diperlukan penyedia layanan telekomunikasi seluler yang sesuai dengan kebutuhan. Dua penyedia layanan telekomunikasi seluler yang populer di Indonesia adalah Indosat dan XL. Jumlah pengikut di Twitter dari kedua penyedia layanan seluler tersebut juga cukup banyak dibandingkan dengan penyedia layanan seluler yang lain. Dari data yang diambil pada Oktober 2019, pengikut dari akun Twitter Indosat sebanyak 898,8 ribu dan XL memiliki pengikut sebanyak 1,34 juta, dapat disimpulkan bahwa kedua penyedia layanan seluler tersebut populer dan banyak pengguna yang mengakses dan berinteraksi dengan akun Twitter dari indosat maupun XL.

Analisis sentimen dapat digunakan untuk mengetahui opini para pengguna layanan seluler tersebut. Analisis sentimen digunakan untuk melihat pendapat atau kecenderungan opini seseorang terhadap suatu permasalahan atau objek, yang kemudian akan diketahui apakah opini tersebut bernilai positif atau negatif.

Untuk mengetahui kecenderungan opini pengguna layanan seluler digunakan algoritma *K-Nearest Neighbor* dan *Support Vector Machine* yang kemudian akan dibandingkan performa dan waktu proses antara kedua metode klasifikasi tersebut. Kedua algoritma dipilih karena merupakan 10 besar algoritma dalam *data mining*. Berdasarkan hasil pengujian tanpa *10-fold cross-validation* metode klasifikasi SVM memiliki nilai akurasi dan *F1-Score* lebih tinggi daripada KNN dengan nilai 95,1% dan 94%. Pada pengujian menggunakan *10-fold cross-validation* SVM mendapatkan nilai *F1-Score* sebesar 92,79%. Metode SVM memiliki waktu proses yang lebih cepat dibandingkan dengan metode KNN, dengan waktu yang didapatkan 0,0156 detik pada pengujian tanpa *10-fold cross-validation* dan rentang waktu 0,3497 detik sampai dengan 0,4375 detik pada pengujian dengan *10-fold cross-validation*.

Kata Kunci: *K-Nearest Neighbor, Support Vector Machine, Analisis Sentimen, TF-IDF, K-Fold Cross-validation*

ABSTRACT

To support the smoothness and ease of communication, telecommunication service providers are needed. Two popular telecommunication service providers in Indonesia are Indosat and XL. The number of followers on Twitter of the two cellular service providers is also quite large compared to the other cellular service providers. From the data taken in October 2019, the Indosat Twitter account has 898.8 thousand followers and XL had 1.34 million followers, it can be concluded that the two cellular service providers are popular and many users access and interact with both Twitter accounts from Indosat and XL.

Sentiment analysis can be used to find out the opinions of these cellular service users. Sentiment analysis is used to see the opinion or tendency of a person's opinion on a problem or object, which will then determine whether the opinion is positive or negative.

To determine the tendency of mobile service user opinions, the K-Nearest Neighbor and Support Vector Machine algorithms are used, which will then compare the performance and processing time between the two classification methods. Both algorithms were chosen because they are the top 10 algorithms in data mining. Based on the test results without 10-fold cross-validation, the SVM classification method has an accuracy value and F1-score higher than the KNN with a value of 95.1% and 94%. In testing using the 10-fold cross-validation, SVM got an F1-score of 92.79%. The SVM method has a faster processing time than the KNN method, with a time obtained of 0.0156 seconds on testing without 10-fold cross-validation and a time span of 0.3497 seconds to 0.4375 seconds on texts with 10-fold cross-validation.

Keywords: *K-Nearest Neighbor, Support Vector Machine, Sentiment Analysis, TF-IDF, K-Fold Cross-validation*