

**SENTIMEN ANALISIS MENGGUNAKAN *SUPPORT VECTOR MACHINE*
(SVM) UNTUK PENGUKURAN PELAYANAN
BLIBLI.COM MELALUI TWITTER**

SKRIPSI



disusun oleh

Shodiq Mustofa

16.11.0066

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

**SENTIMEN ANALISIS MENGGUNAKAN SUPPORT VECTOR
MACHINE (SVM) UNTUK PENGUKURAN PELAYANAN
BLIBLI.COM MELALUI TWITTER**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Sistem Informasi



disusun oleh

Shodiq Mustofa

16.11.0066

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

2021

PERSETUJUAN

SKRIPSI

**SENTIMEN ANALISIS MENGGUNAKAN SUPPORT VECTOR
MACHINE (SVM) UNTUK PENGUKURAN PELAYANAN
BLIBLI.COM MELALUI TWITTER**

yang dipersiapkan dan disusun oleh

Shodiq Mustofa

16.11.0066

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 8 Maret 2021

Dosen Pembimbing,

Andi Sunyoto, M.Kom., Dr..

NIK. 190302052

PENGESAHAN**SKRIPSI****SENTIMEN ANALISIS MENGGUNAKAN SUPPORT VECTOR
MACHINE (SVM) UNTUK PENGUKURAN PELAYANAN
BLIBLI.COM MELALUI TWITTER**

yang dipersiapkan dan disusun oleh

Shodiq Mustofa

16.11.0066

telah dipertahankan di depan Dewan Penguji
pada tanggal 13 Maret 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Andi Sunyoto, M.Kom., Dr.
NIK. 190302052

Ika Nur Fajri, M.Kom
NIK. 190302036

Agus Fatkhurohman, M.Kom
NIK. 190302029

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Juni 2021

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, M.Kom
NIK. 190302031

PERNYATAAN

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, 18 juni 2021

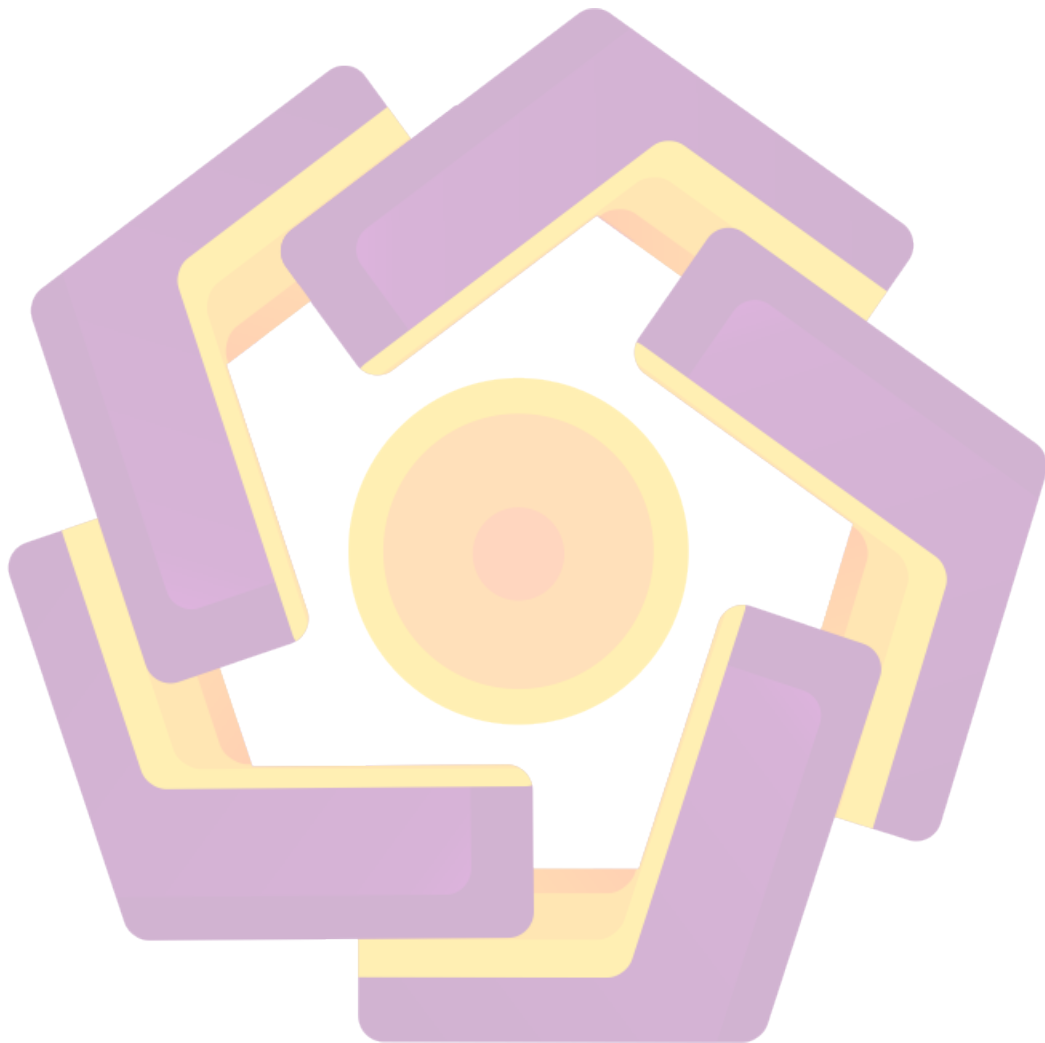


Shodiq Mustofa

NIM. 16.11.0066

MOTTO

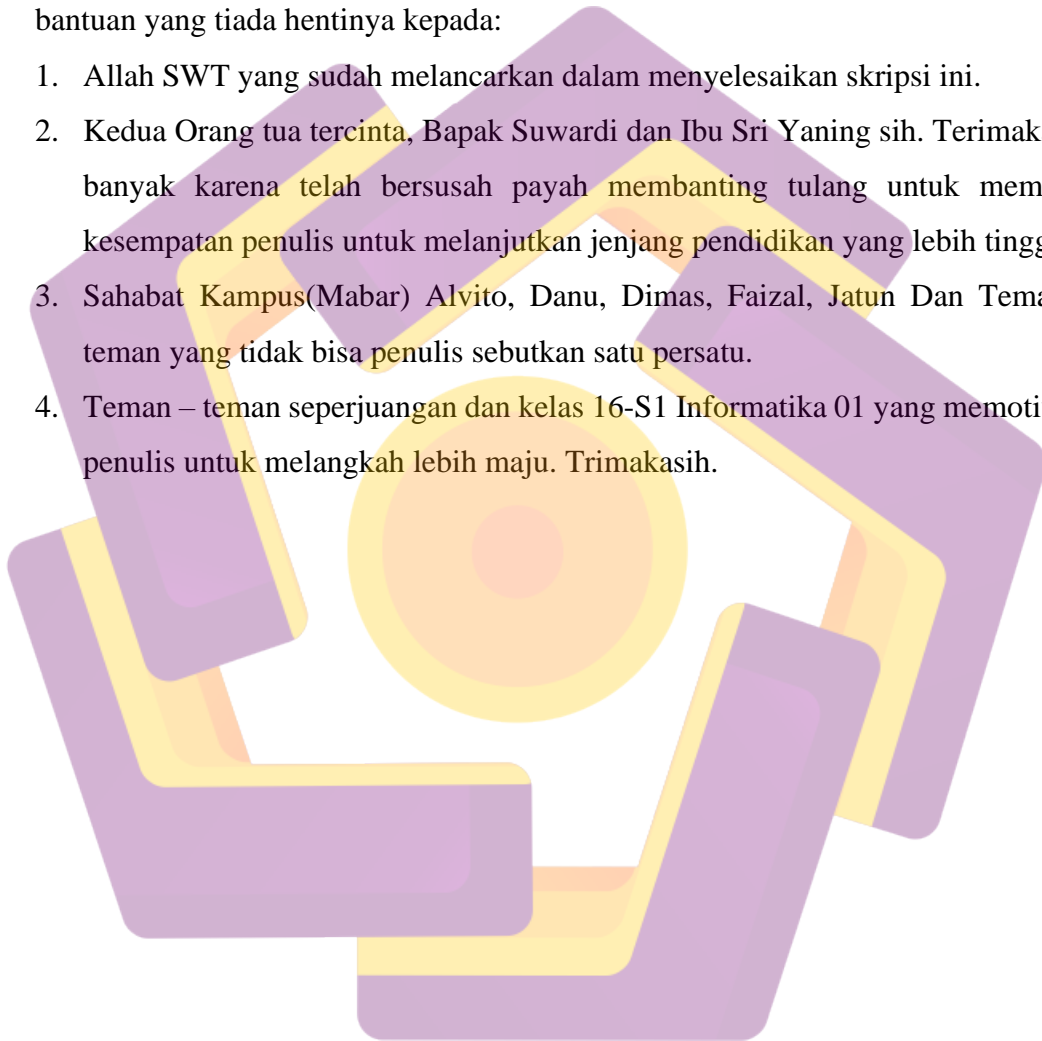
“Keep Dreaming Until It Comes True”



PERSEMBAHAN

Dengan berucap syukur Alhamdulillah, segala puji bagi Allah SWT yang tiada henti memberikan keberkahan, skripsi ini dapat diselesaikan. Saya persembahkan skripsi ini untuk semua yang telah memberikan doa, dukungan dan bantuan yang tiada hentinya kepada:

1. Allah SWT yang sudah melancarkan dalam menyelesaikan skripsi ini.
2. Kedua Orang tua tercinta, Bapak Suwardi dan Ibu Sri Yaning sih. Terimakasih banyak karena telah bersusah payah membanting tulang untuk memberi kesempatan penulis untuk melanjutkan jenjang pendidikan yang lebih tinggi.
3. Sahabat Kampus(Mabar) Alvito, Danu, Dimas, Faizal, Jatun Dan Teman - teman yang tidak bisa penulis sebutkan satu persatu.
4. Teman – teman seperjuangan dan kelas 16-S1 Informatika 01 yang memotivasi penulis untuk melangkah lebih maju. Trimakasih.



KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Allah SWT yang telah menganugerahkan rahmat serta hidayah-Nya sehingga penulis dapat menyelesaikan penulisan skripsi dengan judul **“Sentimen Analisis Menggunakan Support Vector Machine (Svm) Untuk Pengukuran Pelayanan Blibli.Com Melalui Twitter”**.

Skripsi ini disusun sebagai salah satu syarat utama untuk menyelesaikan program sarjana pada Universitas Amikom Yogyakarta. Selama mengikuti pendidikan program sarjana, penyelesaian skripsi ini juga tidak lepas dari bantuan berbagai pihak, karena itu pada kesempatan ini penulis ingin menyampaikan rasa hormat dan terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto, MM selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Krisnawati, S.Si., M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Bapak Sudarmawan, M.T. selaku Ketua Program Studi Informatika Universitas Amikom Yogyakarta.
4. Bapak Andi Sunyoto, M.Kom., Dr. selaku Dosen Pembimbing yang selalu memberikan saran dan masukan serta support dalam menyelesaikan skripsi ini.
5. Bapak Agus Fatkhurohman, M.Kom dan Bapak Ika Nur Fajri, M.Kom selaku Dosen Penguji atas kritik dan saran yang diberikan selama pengujian dan perbaikan skripsi.
6. Bapak /Ibu Dosen Universitas Amikom Yogyakarta yang telah memberikan penulis ilmu yang bermanfaat.

Yogyakarta, 19 juni 2021

Shodiq Mustofa
NIM. 16.11.0066

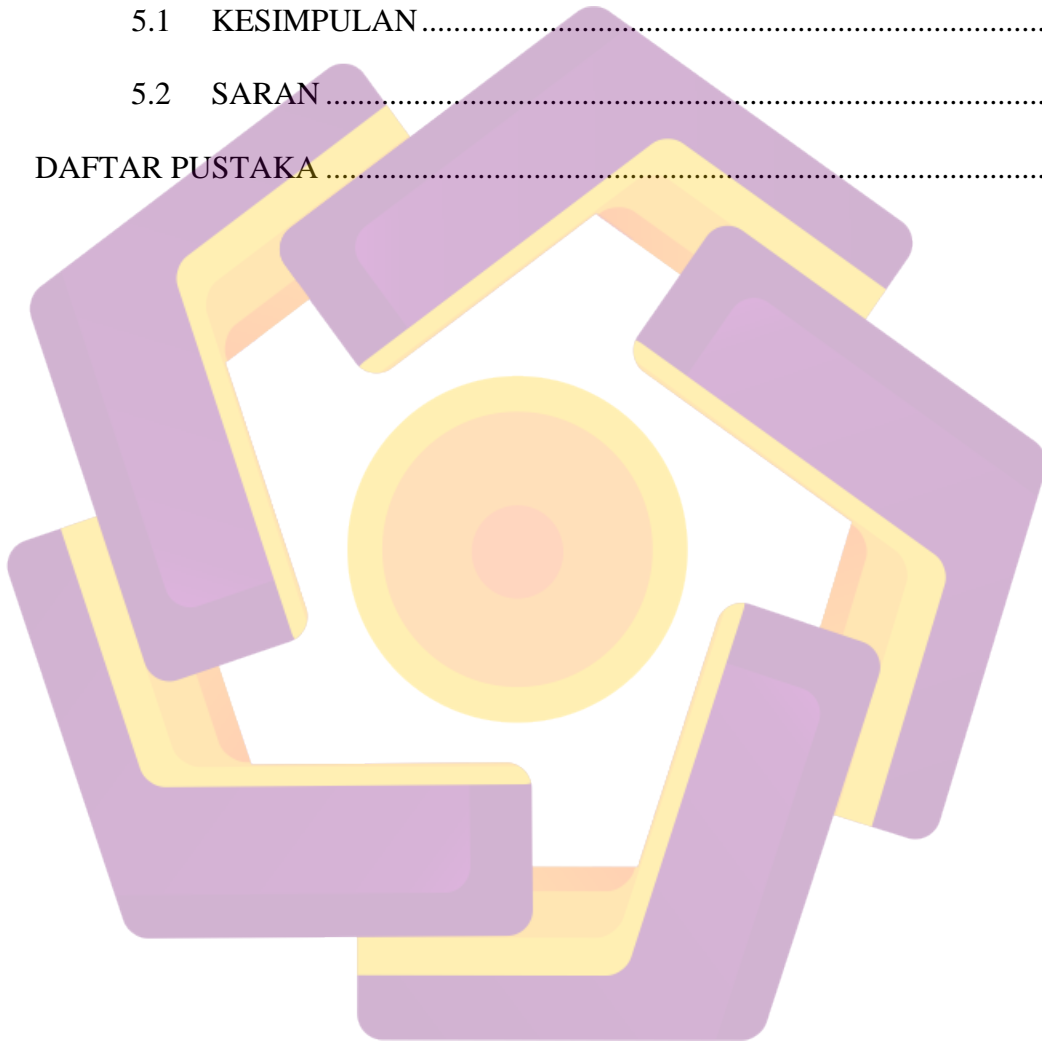
DAFTAR ISI

JUDUL	I
PERSETUJUAN	II
PENGESAHAN	III
PERNYATAAN	IV
MOTTO	V
PERSEMBAHAN	VI
KATA PENGANTAR	VII
DAFTAR ISI	VIII
DAFTAR TABEL	XI
DAFTAR GAMBAR	XII
DAFTAR ISTILAH	XIII
INTISARI	XIV
ABSTRACT	X
BAB 1 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
1.6.1 Metode Pengumpulan Data	4
1.6.1.1 Studi Pustaka	4

1.6.1.2 Survei	4
1.6.1.3 Eksperimen.....	5
1.6.2 Metode Analisis	5
1.6.3 Metode Perancangan	7
1.6.4 Metode Implementasi.....	7
1.6.5 Metode Testing.....	7
1.7 SISTEMATIKA PENULISAN	7
BAB II LANDASAN TEORI	9
2.1 TINJAUAN PUSTAKA	9
2.2 PENGERTIAN DATA MINING.....	10
2.3 PENGERTIAN SENTIMENT ANALYSIS	11
2.4 PENGERTIAN TEXT MINING	12
2.5 PENGERTIAN PRE-PROCESSING	13
2.6 TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF).....	14
2.7 PENGERTIAN SUPPORT VECTOR MACHINE	16
2.8 COSINE SIMILARITY	19
2.9 K-FOLD CROSS VALIDATION	20
2.10 CONFUSION MATRIX	21
2.11 BAHASA PEMROGRAMAN PYTHON	23
2.12 LANGKAH-LANGKAH PENGEMBANGAN APLIKASI.....	23
BAB III METODE PENELITIAN.....	29
3.1 ANALISIS MASALAH.....	29

3.2	SOLUSI YANG DIPILIH.....	30
3.3	ANALISIS KEBUTUHAN.....	30
3.3.1	Analisis Kebutuhan Fungsional	31
3.3.2	Analisis Kebutuhan Non-Fungsional	31
3.4	ANALISA ALGORITMA	32
3.4.1	Cleaning Data.....	32
3.4.2	Case Folding	33
3.4.3	Tokenization.....	36
3.4.4	Stemming	38
3.4.5	TF-IDF	40
3.4.6	Consine Similarity.....	45
3.4.7	Support Vector Machine	49
3.5	PERACANGAN APLIKASI	52
3.5.1	Diagram Alur Preprocessing.....	52
3.5.2	Proses Klasifikasi SVM	53
3.6	PERANCANGAN FILE TRAINING.....	54
BAB IV HASIL DAN PEMBAHASAN		56
4.1	DESKRIPSIS IMPLEMENTASI	56
4.2	IMPLEMENTASI CRAWLING DATA	56
4.3	IMPLEMENTASI PEMBUATAN DATA TRAINING.....	58
4.4	IMPLEMENTASI PROSES TRAINING DATA	59
4.5	IMPLEMENTASI PROSES PREDIKSI DAN KLASIFIKASI	60
4.6	PEMBUATAN INTERFACE.....	62

4.6.1 HALAMAN UTAMA.....	63
4.6.2 HALAMAN RESULT	63
4.7 EVALUASI PROGRAM.....	67
BAB V PENUTUP.....	67
5.1 KESIMPULAN.....	67
5.2 SARAN	67
DAFTAR PUSTAKA	69



DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian Terkait Sistem yang Dibangun.....	10
Table 2. 2 Confusion Matrix	21
Table 3.1 Contoh Data Tweet Sebelum Dilakukan Cleaning Data.....	32
Tabel 3.2 Contoh Data Tweet Setelah Dilakukan Cleaning Data.....	32
Table 3.3 Contoh Data Tweet Setelah Case Folding	35
Table 3.4 Contoh Proses Tokenization Dari Cleaning Data	36
Table 3.5 Contoh Proses Stemming	38
Table 3.6 Feature List	39
Table 3.7 Contoh Proses Term Frequency Positive	41
Table 3.8 Contoh Hasil Perhitungan Tf-Idf Positive	43
Table 3.9 Contoh Hasil Perhitungan Tf-Idf Negative.....	44
Table 3.10 Contoh Proses Perhitungan Consine Similarity Feature Positive	47
Table 3.11 Contoh Proses Perhitungan Consine Similarity Feature Negative.....	48
Table 3.12 Contoh Hasil Proses Consine Similarity	49
Table 3.13 Contoh Hasil Perhitungan Persamaan Positif Terhadap Data Training.....	51
Table 4.1 Table Confusion Matrix	65
Table 4.3 Hasil Evaluasi Cross Validation SVM	66

DAFTAR GAMBAR

Gambar 2.1 Sample Penyebaran Permasalahan Dua Dimensi.....	17
Gambar 2. 2 Hyperplane pada Support Vector Machine	17
Gambar 2. 3 Alur atau Langkah-Langkah Analisa Sentimen	24
Gambar 2. 4 Langkah-Langkah Preprocessing	24
Gambar 2. 5 Langkah – Langkah Training Data.....	26
Gambar 2. 6 Langkah-Langkah Testing Data	27
Gambar 3.1 Diagram Alur Preprocessing	52
Gambar 3.2 Diagram Alur Proses Klasifikasi SVM	53
Gambar 3.3 Data Training.....	55
Gambar 4.1 Perintah Crawling Data	57
Gambar 4.2 Data file JSON Twitter.....	58
Gambar 4.3 Data Training.....	59
Gambar 4.4 Akses Data Training	59
Gambar 4.5 Proses Preprocessing	60
Gambar 4.6 Proses Pembobotan dan Klasifikasi	60
Gambar 4.7 Proses Download Data Twitter	61
Gambar 4.8 Proses Cleaning Data Twitter.....	61
Gambar 4.9 Proses Stemming	62
Gambar 4.10 Proses Klasifikasi	62
Gambar 4.11 Halaman Utama.....	63
Gambar 4.12 Halaman Result	64

INTISARI

Pada era yang sudah memasuki *e-commerce* 4.0, masyarakat sudah bergantung dengan kemudahan untuk bisa melakukan transaksi secara daring. Hal ini ditandai dengan semakin banyaknya jumlah marketplace yang berkembang. Setiap marketplace menggunakan sosial media untuk menjangkau user secara langsung dan juga meningkatkan pelayanan. Salah satu sosial media yang biasa digunakan adalah twitter. Untuk menilai kualitas pelayanan bisa dilakukan dengan menilai komentar yang diberikan oleh pengguna dan diolah menggunakan algoritma *Support Vector Machine*.

Algoritma *Support Vector Machine* merupakan salah satu metode pembelajaran mesin untuk melakukan klasifikasi. Klasifikasi dilakukan untuk mengelompokkan data berdasarkan kelas yang ditentukan yaitu positif dan negatif. Nilai akurasi nantinya akan menjadi tolak ukur untuk mencari model pengujian terbaik.

Pada proses analisis sentimen digunakan TF-IDF untuk melakukan pembobotan vector. vector berisi data twitter yang berisi teks yang akan diolah menggunakan algoritma *Support Vector Machine*. Tahap evaluasi pada penelitian ini menggunakan *K-Fold Cross Validation* dengan jumlah partisi data 10-fold. Hasil evaluasi yang dilakukan menadaptkan nilai akurasi yang cukup baik yaitu sebesar 75.7%.

Kata Kunci: Analisis Sentimen, Klasifikasi, *Support Vector Machine*

ABSTRACT

In the era that has entered e-commerce 4.0, people are already dependent on the ease to be able to make transactions online. This is characterized by a growing number of marketplaces. Each marketplace uses social media to reach users directly and also improve services. One of the commonly used social media is Twitter. To assess the quality of service can be done by assessing comments given by users and processed using the Vector Machine Support algorithm.

Support Vector Machine Algorithm is one of the machine learning methods to do classification. Classification is done to group data based on the specified class of positives and negatives. The accuracy value will then be the benchmark for finding the best test model.

In the sentiment analysis process, TF-IDF is used to perform vector weighting. the vector contains twitter data containing text that will be processed using the Support Vector Machine algorithm. The evaluation phase of this study used K-Fold Cross Validation with a 10-fold number of data partitions. The evaluation result is a fairly good accuracy value of 75.7%.

Keyword: *Sentiment Analysis, Classification, Support Vector Machine*

