

**IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE
UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER
TERHADAP PELAYANAN TELKOM DAN BIZNET**

SKRIPSI



disusun oleh

Fadholi Fat Haranto

15.11.8710

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



**IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE
UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER
TERHADAP PELAYANAN TELKOM DAN BIZNET**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai derajat Sarjana S1
pada Program Studi Informatika



disusun oleh

Fadholi Fat Haranto

15.11.8710

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



PERSETUJUAN

SKRIPSI

**IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE
UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER
TERHADAP PELAYANAN TELKOM DAN BIZNET**

yang dipersiapkan dan disusun oleh

Fadholi Fat Haranto

15.11.8710

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 November 2018

Dosen Pembimbing



Bety Wulan Sari, M.Kom.

NIK. 190302254

PENGESAHAN

SKRIPSI

IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PELAYANAN TELKOM DAN BIZNET

yang dipersiapkan dan disusun oleh

Fadholi Fat Haranto

15.11.8710

telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Juli 2019

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Mardhiya Hayaty, S.T., M.Kom

NIK. 190302108

Anggit Dwi Hartanto, M.Kom

NIK. 190302163

Bety Wulan Sari, M.Kom

NIK. 190302254



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 29 Juli 2019

DEKAN FAKULTAS ILMU KOMPUTER



Krisnawati, S.Si., M.T.

NIK. 190302038

PERNYATAAN

Saya yang bertanda tangan 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, 29 Juli 2019



Fadholi Fat Haranto

NIM. 15.11.8710



MOTTO

“Kesempatan bukanlah hal yang kebetulan. Kau harus menciptakannya”

- Chris Grosser

“Jangan menunggu. Takkan pernah ada waktu yang tepat”

- Napoleon Hill

PERSEMBAHAN

Alhamdulillahirrabil'alamin,

Terimakasih ya Allah untuk segala rahmat, nikmat, dan karunia-Mu yang telah mengizinkanku untuk menyelesaikan kuliah di kampus Universitas AMIKOM Yogyakarta.

Satu tanggung jawab telah terpenuhi, namun perjalanan panjang ini masih belum selesai. Titik ini merupakan awal dari perjuanganku yang lain, perjuangan untuk melakukan hal yang lebih besar. Sebagai ucapan terima kasihku atas capaian ini, kupersembahkan karya kecil ini kepada orang-orang yang spesial.

1. Terima kasih untuk Bapak Sutriyono dan Ibu Maryam yang telah memberikan doa, semangat, serta nasehat yang teramat sangat saya butuhkan.
2. Ibu Bety Wulan Sari, M.Kom., selaku dosen pembimbing yang selalu memberikan arahan, waktu, bimbingan serta masukan-masukan yang sangat bermanfaat dan membantu dalam menyelesaikan penyusunan skripsi ini.
3. Terimakasih untuk Hafid Alpin Al Gazni, Khakim Assidiqi Nur Hudaya, Khulafa Pinta Winastya, Rachel Gita Saraswati, dan Triyani Kunderini yang membantu saya berproses dalam hidup.
4. Terimakasih untuk teman-teman Universitas AMIKOM Yogyakarta, terutama Yoga Aditama Ika Nanda yang memberi pengetahuan tentang *Machine Learning*, Serta teman-teman kelas S1 Informatika angkatan 2015 khususnya kelas 04, yang telah menjadi sahabat seperjuangan selama menempuh perkuliahan.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah subhanahu wa ta'ala yang telah melimpahkan nikmat dan rahmat-Nya, sehingga penulis dapat menyelesaikan Skripsi yang berjudul "*Implementasi Algoritma Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom Dan Biznet*" dengan tepat waktu.

Skripsi ini berisi tentang penerapan teknik *data mining* menggunakan algoritma *Support Vector Machine*, dari mulai tahap pengumpulan data, hingga tahap implementasi ke dalam program.

Penulis juga ingin menyampaikan rasa terima kasih yang sedalam-dalamnya kepada semua pihak yang telah membantu dalam pembuatan skripsi ini, khususnya kepada :

1. Ibu Krisnawati, S.Si, MT, selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
2. Bapak Sudarmawan, MT, selaku Ketua Program Studi S1 Informatika.
3. Ibu Windha Mega Pradnya D, M.Kom, selaku Sekretaris Program Studi S1 Informatika.
4. Ibu Bety Wulan Sari, M.Kom, selaku Dosen Pembimbing yang telah membimbing penulis dalam proses pembuatan Skripsi.
5. Seluruh Dosen Universitas AMIKOM Yogyakarta yang telah memberikan ilmu yang bermanfaat.

6. Bapak dan Ibu tersayang yang selalu memberikan doa, dan dukungannya.
7. Teman-teman S1-IF angkatan 2015 kelas 04 yang telah menjadi keluarga selama proses akademik.
8. Semua pihak yang telah banyak membantu dalam penyusunan skripsi ini yang tidak dapat disebutkan satu persatu.

Penulis menyadari sepenuhnya bahwa skripsi ini banyak mengandung kekurangan, untuk itu segala kritik, dan saran yang membangun sangat diharapkan. Akhir kata, penulis berharap semoga makalah ini dapat memberikan manfaat bagi pihak yang memerlukan.

Yogyakarta, 29 Juli 2019



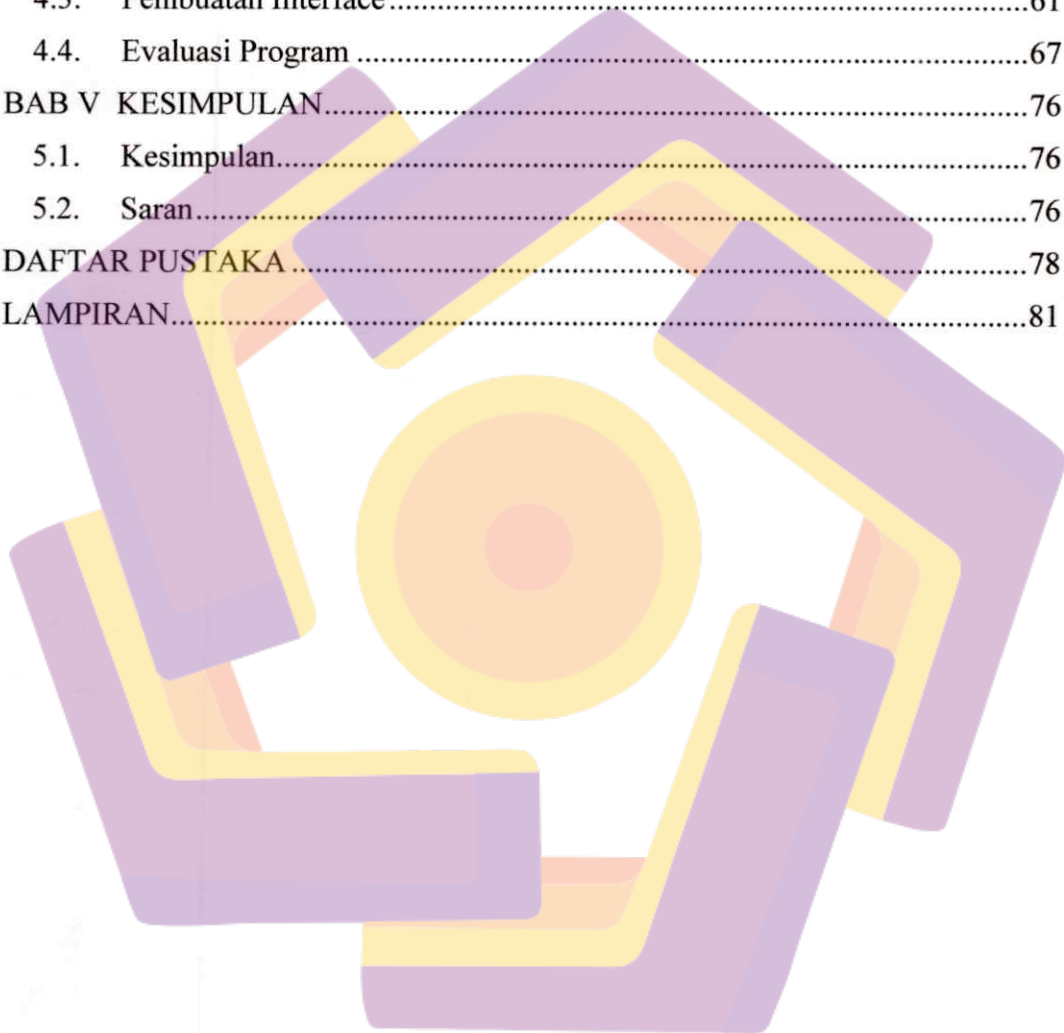
Fadholi Fat Haranto

DAFTAR ISI

PERSETUJUAN.....	ii
PENGESAHAN.....	iii
PERNYATAAN	iv
MOTTO	v
PERSEMBAHAN.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiv
INTISARI	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah.....	2
1.4. Tujuan Penelitian.....	3
1.5. Manfaat Penelitian.....	3
1.6. Metode Penelitian.....	4
1.6.1. Metode Pengumpulan Data.....	4
1.6.2. Metode Analisis	5
1.6.3. Metode Perancangan.....	6
1.6.4. Metode Implementasi.....	7
1.6.5. Metode Pengujian	7
1.7. Sitematika Penulisan	7
BAB II LANDASAN TEORI.....	9
2.1. Tinjauan Pustaka	9
2.2. Dasar Teori.....	11
2.2.1. Pengertian Data Mining	11
2.2.2. Pengertian Analisis Sentimen	12
2.2.3. Pre-processing.....	13
2.2.4. Term Frequency-Inverse Document Frequency(TF-IDF)	13

2.2.5.	Cosine Similarity	14
2.2.6.	Support Vector Machine	14
2.2.7.	K-fold Cross Validation	17
2.2.8.	Confusion matrix	18
BAB III ANALISIS DAN PERANCANGAN		20
3.1.	Analisis Masalah	20
3.2.	Solusi Yang Dipilih	21
3.3.	Analisis Kebutuhan	22
3.3.1.	Analisis Kebutuhan Fungsional	22
3.3.2.	Analisis Kebutuhan Non-fungsional	22
3.4.	Analisa Sentimen dengan Klasifikasi Algoritma SVM	23
3.4.1.	Data Cleaning	23
3.4.2.	Case Folding	25
3.4.3.	Tokenization	27
3.4.4.	Stopword	28
3.4.5.	TF-IDF	30
3.4.6.	Cosine Similarity	36
3.4.7.	Support Vector Machine	43
3.4.8.	K-fold Cross Validation	46
3.4.9.	Confusion Matrix	46
3.5.	Perancangan Aplikasi	47
3.5.1.	Diagram Alir Program	47
3.5.2.	Proses Klasifikasi Support Vector Machine	48
3.6.	Perancangan Crawling Data Twitter	49
3.7.	Perancangan File <i>Training</i>	50
3.8.	Langkah-Langkah Pengembangan Aplikasi	50
BAB IV IMPLEMENTASI DAN PEMBAHASAN		55
4.1.	Deskripsi Implementasi	55
4.2.	Implementasi Perancangan Script dan Sistem	55
4.2.1.	Pengumpulan Data Tweet	55
4.2.2.	Pelabelan Data <i>Training</i>	57
4.2.3.	Pembuatan Script Akses Data Training	58

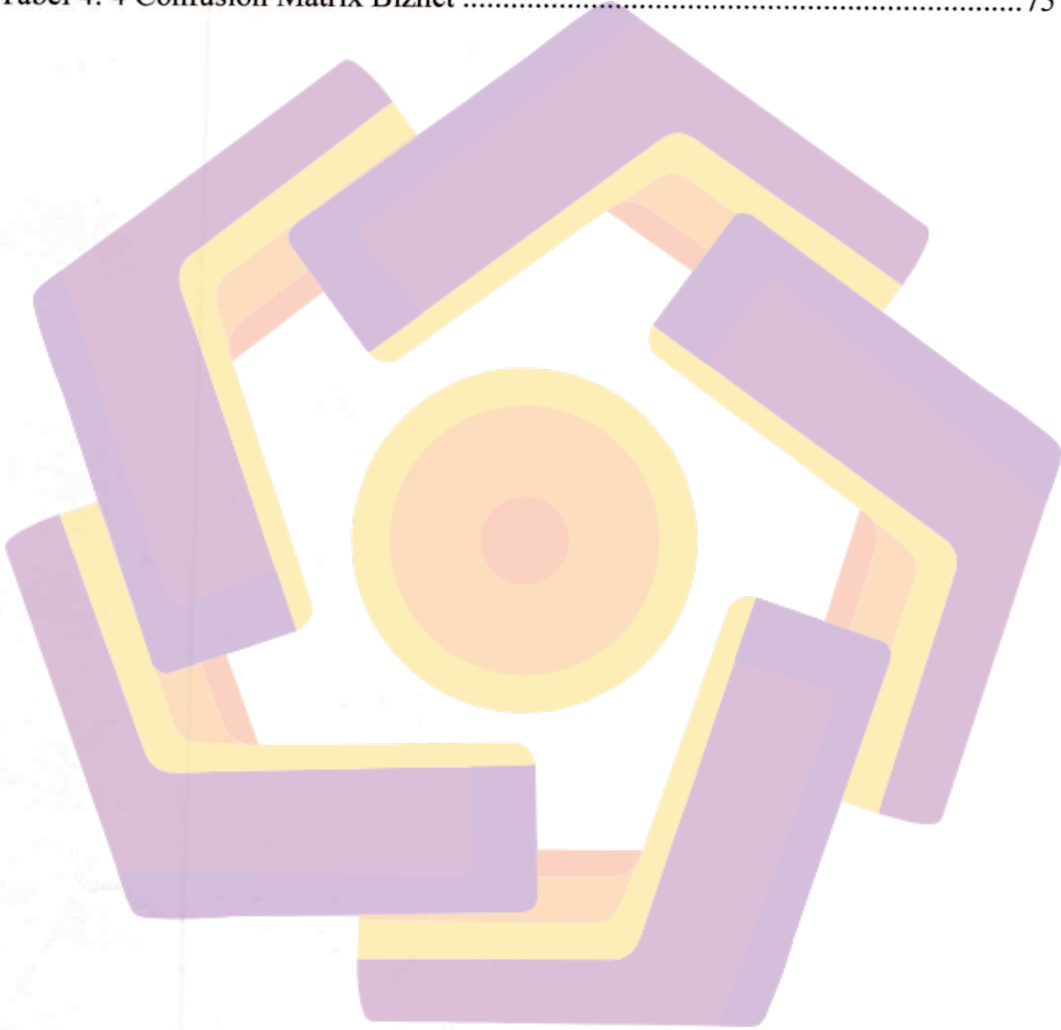
4.2.4.	Pembuatan Script Cleaning dan Case Folding.....	58
4.2.5.	Pembuatan Script Tokenization	59
4.2.6.	Pembuatan Script Stopword.....	60
4.2.7.	Implementasi Pembuatan Fungsi Klasifikasi SVM	60
4.2.8.	Implementasi Pembuatan Prediksi Sentimen.....	60
4.3.	Pembuatan Interface	61
4.4.	Evaluasi Program	67
BAB V KESIMPULAN.....		76
5.1.	Kesimpulan.....	76
5.2.	Saran.....	76
DAFTAR PUSTAKA.....		78
LAMPIRAN.....		81



DAFTAR TABEL

Tabel 2. 1 Perbandingan Penelitian Terkait Dengan Sistem Yang Dibangun	10
Tabel 3. 1 Data Telkom Sebelum Dilakukan Cleaning Data.....	23
Tabel 3. 2 Data Telkom Setelah Dilakukan Cleaning Data	24
Tabel 3. 3 Data Biznet Sebelum Dilakukan Cleaning Data.....	24
Tabel 3. 4 Data Biznet Setelah Dilakukan Cleaning Data	25
Tabel 3. 5 Data Telkom Setelah Dilakukan Case Folding.....	26
Tabel 3. 6 Data Biznet Setelah Dilakukan Case Folding.....	26
Tabel 3. 7 Data Telkom Setelah Dilakukan Tokenization.....	27
Tabel 3. 8 Data Biznet Setelah Dilakukan Tokenization.....	27
Tabel 3. 9. Data Telkom Setelah Dilakukan Stopword	28
Tabel 3. 10 Data Biznet Setelah Dilakukan Stopword	29
Tabel 3. 11 Feature List Telkom.....	30
Tabel 3. 12 Feature List Biznet.....	30
Tabel 3. 13 TF-IDF Telkom Positif.....	32
Tabel 3. 14 TF-IDF Telkom Negatif.....	33
Tabel 3. 15 TF-IDF Biznet Positif.....	34
Tabel 3. 16 TF-IDF Biznet Negatif.....	35
Tabel 3. 17 Cosine Similarity Telkom Positif	38
Tabel 3. 18 Cosine Similarity Telkom Negatif.....	39
Tabel 3. 19 Cosine Similarity Biznet Positif	40
Tabel 3. 20 Cosine Similarity Biznet Negatif.....	41
Tabel 3. 21 Hasil Cosine Similarity Telkom	42
Tabel 3. 22 Hasil Cosine Similarity Biznet	42
Tabel 3. 23 Persamaan w_1 w_2 dan b pada label Telkom	44
Tabel 3. 24 Persamaan w_1 , w_2 , dan b pada label Biznet.....	44
Tabel 3. 25 Persamaan Positif Terhadap Data Training Telkom.....	45
Tabel 3. 26 Persamaan Negatif Terhadap Data Training Telkom	45
Tabel 3. 27 Persamaan Positif Terhadap Data Training Biznet.....	45
Tabel 3. 28 Persamaan Negatif Terhadap Data Training Biznet	45
Tabel 3. 29 Predik Pada Data Training Telkom	46

Tabel 3. 30 Predik Pada Data Training Biznet46
Tabel 3. 31 Confusion Matrix Data Sampel47
Tabel 4. 1 Hasil 10 Fold Cross Validation Telkom67
Tabel 4. 2 Hasil 10 Fold Cross Validation Biznet71
Tabel 4. 3 Confusion Matrix Telkom75
Tabel 4. 4 Confusion Matrix Biznet75



DAFTAR GAMBAR

Gambar 2. 1 Hyperplane pada Support Vector Machine	15
Gambar 2. 2 Pemisahan Dua Kelas Dengan Margin Maksimum	16
Gambar 2. 3 Contoh Confussion Matrix.....	18
Gambar 3. 1 Diagram Alir Program	48
Gambar 3. 2 Proses Klasifikasi Support Vector Machine	49
Gambar 3. 3 Konfigurasi Autentikasi File Crawling	50
Gambar 3. 4 Langkah-Langkah Klasifikasi	51
Gambar 3. 5 Langkah-Langkah Training Data	51
Gambar 3. 6 Langkah-Langkah Testing data.....	54
Gambar 4. 1 Script Crawling Data Twitter	56
Gambar 4. 2 Pelabelan Training	58
Gambar 4. 3 Script Akses File Data Training.....	58
Gambar 4. 4 Script Cleaning dan Case Folding.....	59
Gambar 4. 5 Script Tokenization	59
Gambar 4. 6 Script Stopword.....	60
Gambar 4. 7 Script Pembobotan dan Klasifikasi	60
Gambar 4. 8 Script Prediksi Sentimen	61
Gambar 4. 10 Halaman Model.....	62
Gambar 4. 11 K-10 Fold Cross Validation Telkom.....	62
Gambar 4. 13 K-10 Fold Cross Validation Biznet.....	63
Gambar 4. 15 Halaman Sentimen	64
Gambar 4. 16 Pie Chart Hasil Prediksi Telkom.....	64
Gambar 4. 17 Pie Chart Hasil Prediksi Biznet.....	65
Gambar 4. 18 Crawling Data	66
Gambar 4. 19 Grafik Performa Model Hasil 10 Fold Cross Validation Telkom... 71	
Gambar 4. 20 Grafik Performa Model Hasil 10 Fold Cross Validation Biznet..... 74	

INTISARI

Sosial media merupakan suatu media yang dapat digunakan untuk berekspresi oleh penggunanya. Twitter merupakan salah satu sosial media yang banyak digunakan di Indonesia, pengguna twitter dapat berekspresi dan beraspirasi tanpa adanya batasan. *Tweet* yang berupa ekspresi dan aspirasi yang ditulis oleh pengguna twitter dapat digunakan untuk ulasan sebuah produk atau layanan.

Pada skripsi ini, peneliti menggunakan teknik *text mining* menggunakan algoritma *Support Vector Machine* untuk analisis sentimen pengguna twitter terhadap pelayanan Telkom dan Biznet. Data yang digunakan pada analisis sentimen ini adalah data yang berkaitan dengan pelayanan Telkom dan Biznet, jumlah dataset yang digunakan adalah 500 *tweet* yang berasal dari *crawling* data twitter, terdapat 250 *tweet* yang dijadikan dataset pada masing-masing objek, dari 250 *tweet* dataset Telkom terdapat 103 *tweet* positif dan 147 *tweet* negatif, serta 250 *tweet* dataset Biznet terdapat 88 *tweet* positif dan 162 *tweet* negatif. Data tersebut nantinya akan digunakan untuk data *training* dan data *testing* dalam proses pembuatan model menggunakan algoritma *Support Vector Machine*. Metode yang digunakan untuk pengujian model adalah *Confusion Matrix* serta *K-Fold Cross Validation* sebagai metode untuk membagi data *training* dan data *testing* sesuai lipatan yang digunakan.

Hasil pengujian yang diperoleh menggunakan metode *K-Fold Cross Validation* dan *Confusion Matrix* pada model yang dibuat menggunakan algoritma *Support Vector Machine* yang menghasilkan nilai *accuracy*, *precision*, *recall*, dan *F1-score* sebesar 79,6%, 76,5%, 72,8%, dan 74,6% untuk Telkom, serta 83,2%, 78,8%, 71,6%, dan 75% untuk Biznet. Peneliti menyarankan menggunakan atau menambah metode lain untuk mendapatkan kinerja model yang lebih baik.

Kata Kunci : Analisis Sentimen, *Support Vector Machine*, Biznet, Telkom, dan Review Layanan

ABSTRACT

Social media is a media that can be used for expression by its users. Twitter is one of the social media that is widely used in Indonesia, twitter users can express and aspire without restrictions. Tweets in the form of expressions and aspirations written by Twitter users can be used to review a product or service.

In this thesis, researchers use text mining techniques using the Support Vector Machine algorithm for twitter user sentiment analysis of Telkom and Biznet services. The data used in this sentiment analysis are data relating to Telkom and Biznet services, the number of datasets used is 500 tweets derived from twitter crawling data, there are 250 tweets that are used as datasets for each object, of the 250 tweets of the Telkom dataset available 103 Positive tweets and 147 negative tweets, and 250 tweets of Biznet datasets, there were 88 positive tweets and 162 negative tweets. The data will later be used for training data and data testing in the modeling process using the Support Vector Machine algorithm. The method used for testing the model is Confusion Matrix and K-Fold Cross Validation as a method for dividing training data and testing data according to the folds used.

The test results obtained using the K-Fold Cross Validation and Confusion Matrix methods on the model are made using the Support Vector Machine algorithm that produces accuracy, precision, recall, and F1-score values of 79,6%, 76,5%, 72,8%, and 74,6% for Telkom, as well as 83,2%, 78,8%, 71,6%, and 75% for Biznet. Researchers suggest using or adding other methods to get better model performance.

Keywords : *Sentiment Analysis, Support Vector Machine, Biznet, Telkom, and Service Review*