

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP LAYANAN  
INTERNET BIZNET MENGGUNAKAN ALGORITMA SUPPORT  
VECTOR MACHINE (SVM)**

**SKRIPSI**



disusun oleh

**Fadhilah Dwi Ananda**

**17.11.1497**

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

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP LAYANAN  
INTERNET BIZNET MENGGUNAKAN ALGORITMA SUPPORT  
VECTOR MACHINE (SVM)**

**SKRIPSI**

untuk memenuhi sebagian persyaratan  
mencapai gelar Sarjana  
pada Program Studi Informatika



disusun oleh

**Fadhilah Dwi Ananda**

**17.11.1497**

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

**PERSETUJUAN**

**SKRIPSI**

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP LAYANAN  
INTERNET BIZNET MENGGUNAKAN ALGORITMA SUPPORT  
VECTOR MACHINE (SVM)**

yang dipersiapkan dan disusun oleh

**Fadhilah Dwi Ananda**

17.11.1497

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 03 November 2020

**Dosen Pembimbing,**

**Yoga Pristyanto, S.Kom, M.Eng**  
**NIK. 190302412**

**PENGESAHAN**

**SKRIPSI**

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP LAYANAN  
INTERNET BIZNET MENGGUNAKAN ALGORITMA SUPPORT  
VECTOR MACHINE (SVM)**

yang dipersiapkan dan isusun oleh

**Fadhilah Dwi Ananda**

**17.11.1497**

telah dipertahankan di depan Dewan Penguji  
pada tanggal 19 Januari 2021

Susunan Dewan Penguji

**Nama Penguji**

**Tanda Tangan**

**Yoga Pristyanto, S.Kom, M.Eng**

**NIK. 190302412**

**Anna Balta, M.Kom**

**NIK. 190302290**

**Windha Mega Pradnya D, M.Kom**

**NIK. 190302185**

Skrripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 13 Februari 2021

**DEKAN FAKULTAS ILMU KOMPUTER UNIVERSITAS AMIKOM  
YOGYAKARTA**

**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, 08 Februari 2021



Fadhilah Dwi Ananda

NIM. 17.11.1497

## **MOTTO**

**“ Jangan Pergi Mengikuti Kemana Jalan akan Berujung. Buat Jalanmu Sendiri dan Tinggalkanlah Jejak ”**

**“ Jangan Pernah Berhenti Bermimpi atau Berharap, Karena Harapanmu akan Mengentarkan Sebuah Keajaiban ”**



## PERSEMBAHAN

Skripsi ini saya persembahkan sangat spesial untuk kedua orangtua saya Bapak dan Ibu yang telah mendidik saya, tiada henti-hentinya selalu memberikan saya dorongan semangat, dukungan, motivasi, kasih sayangnya dan mendo'akan saya hingga mampu menyelesaikan segala kewajiban saya sebagai mahasiswa. Terima kasih untuk selalu mendukung dan memberikan semangat kepada saya dalam mewujudkan cita-cita saya, serta mengajarkan saya untuk selalu hidup dengan sabar dan jujur. Skripsi ini sebagai tanda bahwa perjuangan orangtuaku tiada sia-sia. Semoga Allah selalu memberikan kesehatan dan diberi umur yang panjang.





## KATA PENGANTAR

Segala puji bagi Allah SWT karena atas nikmat dan rahmat-Mya sehingga penulis dapat menyelesaikan penyusunan skripsi ini. Penulisan skripsi ini dilakukan dalam rangka memnuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta. Dalam proses penyelesaian skripsi ini banyak pihak yang telah memberikan bantuan, dukungan, saran dan kritik yang telah penulis terima, oleh karena itu dalam kesempatan ini penulis ingin mengucapkan terima kasih kepada :

1. Allah SWT, yang telah memberikan pertolongan dan kekuatan dalam proses pembuatan skripsi ini.
2. Kedua orangtua penulis, Bapak Suyana dan Ibu Hesti Sulistiowati yang selalu mendoakan dan selalu memberikan dukungan kepada penulis dalam menyelesaikan skripsi ini.
3. Bapak Yoga Pristyanto, S.Kom, M.Eng selaku dosen pembimbing skripsi, atas waktu yang telah diberikan untuk membimbing, mengarahkan, memberi dukungan, motivasi serta memberikan masukan kepada penulis dalam pengerjaan skripsi ini hingga akhir.
4. Seluruh dosen Fakultas Ilmu Komputer Universitas Amikom Yogyakarta atas segala ilmu pengetahuan yang diberikan kepada penulis. Beserta seluruh staff Administrasi Fakultas Ilmu Komputer, terima kasih atas segala ilmu dan bimbingannya.
5. Sege nap dosen Jurusan Informatika yang telah memberikan bimbingan dan ilmu kepada penulis selama masa studi.



6. Seluruh rekan-rekan studi yang tidak dapat disebutkan satu persatu, terima kasih atas segala kebaikan yang diberikan kepada penulis.
7. Kepada teman-teman satu kosan yang sudah membantu dan memberikan semangat.

Berbagai kekurangan dan kesalahan mungkin pembaca temukan dalam penulisan skripsi ini, untuk itu kritik dan saran sangat diharapkan untuk perbaikan yang akan datang. Semoga apa yang menjadi kekurangan bisa disempurnakan dan semoga skripsi ini dapat bermanfaat dan menambah wawasan serta pengetahuan bagi para pembaca.

Yogyakarta, 08 Februari 2021



Fadhilah Dwi Ananda

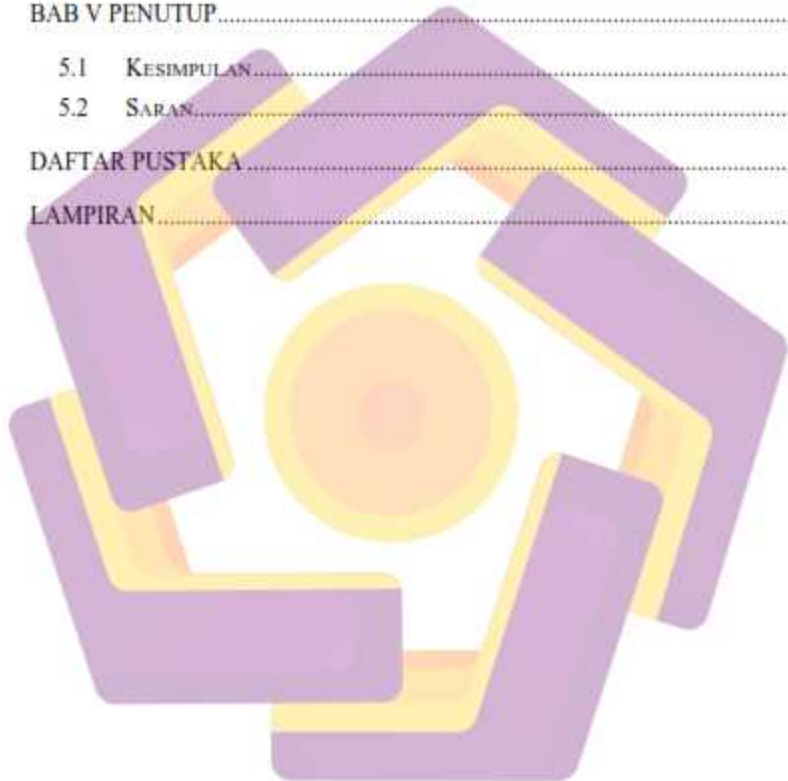
## DAFTAR ISI

ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP LAYANAN INTERNET BIZNET MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM).....	I
PERSETUJUAN .....	III
PENGESAHAN .....	IV
PERNYATAAN .....	V
MOTTO .....	VI
PERSEMBAHAN .....	VII
KATA PENGANTAR .....	VIII
DAFTAR ISI .....	X
DAFTAR TABEL .....	XIV
DAFTAR GAMBAR .....	XV
DAFTAR SOURCE CODE .....	XVI
DAFTAR ISTILAH .....	XVII
INTISARI .....	XVIII
ABSTRACT .....	XIX
BAB I PENDAHULUAN .....	1
1.1 LATAR BELAKANG .....	1
1.2 RUMUSAN MASALAH .....	4
1.3 BATASAN MASALAH .....	4
1.4 TUJUAN PENELITIAN .....	4
1.5 MANFAAT PENELITIAN .....	5
1.6 LUARAN .....	5
1.7 SISTEMATIKA PENULISAN .....	6
BAB II LANDASAN TEORI .....	8

2.1	ANALISIS SENTIMEN .....	8
2.2	TINJAUAN PUSTAKA.....	9
2.2.1	<i>Penelitian Terdahulu</i> .....	9
2.3	LANDASAN TEORI.....	13
2.3.1	<i>Twitter API</i> .....	13
2.3.2	<i>Text Mining</i> .....	14
2.3.3	<i>Preprocessing</i> .....	15
2.3.4	<i>Cleansing</i> .....	15
2.3.5	<i>Case Folding</i> .....	16
2.3.6	<i>Tokenizing</i> .....	16
2.3.7	<i>Stopword Removal</i> .....	17
2.3.8	<i>Stemming</i> .....	17
2.3.9	<i>Pembobotan Kata TF-IDF</i> .....	18
2.3.10	<i>Klasifikasi</i> .....	19
2.3.11	<i>Support Vector Machine</i> .....	19
2.3.12	<i>Linear SVM</i> .....	20
2.3.13	<i>Kernel SVM</i> .....	22
2.3.14	<i>Evaluasi</i> .....	24
<b>BAB III METODE PENELITIAN.....</b>		<b>27</b>
3.1	DESKRIPSI UMUM.....	27
3.2	KERBUHUAN SISTEM.....	28
3.2.1	<i>Hardware</i> .....	28
3.2.2	<i>Software</i> .....	28
3.3	PERANCANGAN DESAIN SISTEM.....	28
3.4	IMPLEMENTASI.....	29
3.5	UJI COBA SISTEM.....	29
3.6	ANALISIS HASIL UJI COBA SISTEM .....	29
3.7	METODE PENGUMPULAN DATA .....	30
3.7.1	<i>Studi Pustaka</i> .....	30
3.7.2	<i>Search Data Twitter</i> .....	30

3.8	ANALISIS DATA.....	32
3.9	ANALISIS PERMASALAHAN .....	32
3.10	PREPROCESSING.....	33
3.10.1	Tokenizing .....	34
3.10.2	Stopword Removal .....	34
3.10.3	Stemming.....	35
3.10.4	Pembobotan Kata TF-IDF.....	35
3.11	KLASIFIKASI.....	38
3.12	CONFUSION MATRIX.....	44
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>		<b>46</b>
4.1	IMPLEMENTASI RESAMPLING TERHADAP DATASET .....	46
4.1.1	Pengumpulan Dataset .....	46
4.1.2	Data Training dan Data Testing.....	48
4.2	PREPROCESSING DATA .....	49
4.2.1	Cleansing .....	50
4.2.2	Case Folding.....	52
4.2.3	Tokenizing.....	53
4.2.4	Stopword Removal .....	55
4.2.5	Stemming.....	56
4.3	PEMBOBOTAN TF-IDF .....	58
4.4	UJI COBA SISTEM.....	58
4.4.1	Klasifikasi Menggunakan Support Vector Machine .....	59
4.4.2	Evaluasi Model Menggunakan Confusion Matrix .....	60
4.4.3	Hasil Accuracy, Precision dan Recall Pengujian 1 .....	60
4.4.4	Hasil Accuracy, Precision dan Recall Pengujian 2 .....	61
4.4.5	Hasil Accuracy, Precision dan Recall Pengujian 3 .....	62
4.4.6	Grafik Hasil Pengujian .....	64
4.4.7	Learning Curve .....	66
4.5	DESAIN DAN IMPLEMENTASI GUI .....	67
4.5.1	Halaman Utama.....	67

4.5.2	<i>Halaman Upload Data</i> .....	67
4.5.3	<i>Halaman Preprocessing</i> .....	68
4.5.4	<i>Halaman Pembobotan Kata TF-IDF</i> .....	69
4.5.5	<i>Halaman Klasifikasi SVM</i> .....	70
4.5.6	<i>Halaman Tes Model</i> .....	71
BAB V PENUTUP.....		72
5.1	KESIMPULAN.....	72
5.2	SARAN.....	72
DAFTAR PUSTAKA.....		74
LAMPIRAN.....		77



## DAFTAR TABEL

Tabel 2. 1 Studi Literatur Sejenis.....	12
Tabel 2. 2 Contoh <i>Case Folding</i> .....	16
Tabel 2. 3 Contoh <i>Tokenizing</i> .....	16
Tabel 2. 4 Cotoh <i>Stopword Removal</i> .....	17
Tabel 2. 5 Contoh <i>Stemming</i> .....	18
Tabel 2. 6 <i>Confusion Matrix</i> .....	25
Tabel 3. 1 <i>Input</i> Dokumen <i>Tweet</i> .....	31
Tabel 3. 2 Proses <i>Tokenizing</i> .....	34
Tabel 3. 3 Proses <i>Stopword Removal</i> .....	35
Tabel 3. 4 Proses <i>Stemming</i> .....	35
Tabel 3. 5 Proses Pembobotan Kata TF-IDF .....	37
Tabel 3. 6 Data <i>input</i> variabel $x_1$ dan $x_2$ .....	40
Tabel 3. 7 Visualisasi data kedalam grafik <i>hyperplane</i> .....	42
Tabel 3. 8 Data <i>testing</i> untuk klasifikasi data ke dalam <i>hyperplane</i> .....	43
Tabel 3. 9 Proses <i>Confusion Matrix</i> .....	44
Tabel 4. 1 Data Hasil <i>Crawling</i> .....	48
Tabel 4. 2 Hasil Proses <i>Cleansing</i> .....	51
Tabel 4. 3 Hasil Proses <i>Case Folding</i> .....	52
Tabel 4. 4 Hasil Proses <i>Tokenizing</i> .....	54
Tabel 4. 5 Hasil Proses <i>Stopword Removal</i> .....	55
Tabel 4. 6 Hasil Proses <i>Stemming</i> .....	57
Tabel 4. 7 Jumlah Data <i>Training</i> dan <i>Testing</i> .....	59
Tabel 4. 8 Hasil Proses <i>Confusion Matrix</i> Pengujian 1 (Kernel Linear) .....	60
Tabel 4. 9 Hasil Proses <i>Confusion Matrix</i> Pengujian 1 (Kernel RBF) .....	61
Tabel 4. 10 Hasil Proses <i>Confusion Matrix</i> Pengujian 2 (Kernel Linear) .....	61
Tabel 4. 11 Hasil Proses <i>Confusion Matrix</i> Pengujian 2 (Kernel RBF) .....	62
Tabel 4. 12 Hasil Proses <i>Confusion Matrix</i> Pengujian 3 (Kernel Linear) .....	62
Tabel 4. 13 Hasil Proses <i>Confusion Matrix</i> Pengujian 3 (Kernel RBF) .....	63



## DAFTAR GAMBAR

Gambar 2. 1 <i>Hyperplane</i> terbaik yang memisahkan antara dua kelas positif (+1) dan negatif (-1).....	20
Gambar 2. 2 <i>Kernel SVM</i> untuk memisahkan data secara <i>linier</i> .....	22
Gambar 2. 3 Contoh transformasi untuk data yang tidak dapat dipisahkan secara <i>linier</i> .....	23
Gambar 3. 1 Desain Sistem.....	27
Gambar 3. 2 Proses Pengumpulan Data.....	31
Gambar 3. 3 Pembagian Data.....	32
Gambar 3. 4 Diagram Alir <i>Preprocessing</i> .....	33
Gambar 3. 5 Diagram Alir Pembobotan TF-IDF.....	36
Gambar 3. 6 Diagram alir klasifikasi SVM.....	38
Gambar 3. 7 Plot <i>hyperplane</i> fungsi klasifikasi.....	42
Gambar 3. 8 Visualisasi garis <i>hyperplane</i> data <i>testing</i> .....	43
Gambar 4. 1 Visualisasi Persebaran Data.....	49
Gambar 4. 2 Grafik Hasil Pengujian Akurasi.....	64
Gambar 4. 3 Grafik Hasil <i>Precision</i> .....	64
Gambar 4. 4 Grafik Hasil Recall.....	65
Gambar 4. 5 Grafik hasil dari Learning Curve SVM Kernel Linear.....	66
Gambar 4. 6 Grafik hasil dari Learning Curve SVM Kernel RBF.....	66
Gambar 4. 7 Tampilan Halaman utama.....	67
Gambar 4. 8 Tampilan <i>Upload Data</i> .....	68
Gambar 4. 9 Tampilan Proses <i>Preprocessing</i> .....	69
Gambar 4. 10 Tampilan Proses TF-IDF.....	70
Gambar 4. 11 Tampilan Klasifikasi SVM.....	70
Gambar 4. 12 Tampilan Tes Model.....	71

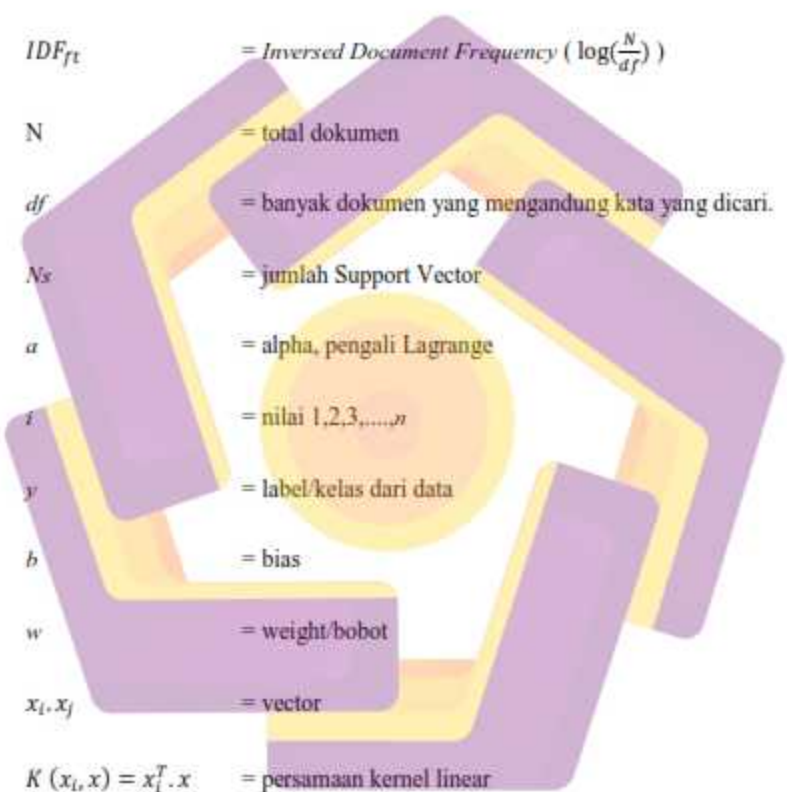


## DAFTAR SOURCE CODE

Source Code 4. 1 Koneksi Token.....	47
Source Code 4. 2 <i>Stream Twitter API</i> .....	47
Source Code 4. 3 <i>Cleansing</i> .....	50
Source Code 4. 4 <i>Case Folding</i> .....	52
Source Code 4. 5 <i>Tokenizing</i> .....	53
Source Code 4. 6 <i>Stopword Removal</i> .....	55
Source Code 4. 7 <i>Stemming</i> .....	56
Source Code 4. 8 Pembobotan TF-IDF.....	58
Source Code 4. 9 Klasifikasi Support Vector Machine.....	59
Source Code 4. 10 <i>Confusion Matrix</i> .....	60



## DAFTAR ISTILAH



$W_{dt}$	= bobot dokumen ke-d terhadap kata ke-t
$TF_{dt}$	= banyaknya kata yang dicari pada sebuah dokumen
$IDF_{ft}$	= <i>Inversed Document Frequency</i> ( $\log(\frac{N}{df})$ )
$N$	= total dokumen
$df$	= banyak dokumen yang mengandung kata yang dicari.
$N_s$	= jumlah Support Vector
$\alpha$	= alpha, pengali Lagrange
$i$	= nilai 1,2,3,...,n
$y$	= label/kelas dari data
$b$	= bias
$w$	= weight/bobot
$x_i, x_j$	= vector
$K(x_i, x) = x_i^T \cdot x$	= persamaan kernel linear
<i>Hyperplane</i>	= garis yang memisahkan dua buah garis pada input space
<i>Margin</i>	= batas
<i>Training &amp; Testing</i>	= pelatihan & pengujian

## INTISARI

Media sosial saat ini merupakan media komunikasi yang sering digunakan oleh kalangan masyarakat Indonesia dalam menyampaikan sebuah opini. Salah satu media yang sering digunakan oleh masyarakat adalah *twitter*. *Twitter* merupakan media sosial yang memberikan banyak informasi yang sangat beragam melalui *tweet*, dari informasi yang ditulis melalui *tweet* tersebut terdapat data yang dapat diolah menjadi analisis sentimen. Penelitian ini bertujuan untuk membuat sistem analisis sentimen masyarakat berupa sentimen positif atau negatif.

Pada penelitian ini, peneliti menggunakan teknik *text mining* dengan menerapkan algoritma *Support Vector Machine* yang dipergunakan untuk analisis sentimen pengguna *twitter* terhadap layanan internet Biznet. *Kernel* yang digunakan dalam penelitian ini adalah *kernel Linear* dan *kernel RBF*. Pengujian dilakukan dengan 3 skenario, pada skenario 1 menggunakan 800 data, skenario 2 menggunakan 900 data dan skenario 3 menggunakan 1000 data, untuk pembagiannya yaitu 90% data *training* dan 10% data *testing* dari masing-masing skenario.

Pembuatan sistem dengan metode *Support Vector Machine* terbukti bahwa algoritma tersebut dapat menganalisis sentimen secara otomatis. Hasil akurasi diperoleh menggunakan metode *Confusion Matrix* pada model yang dibuat menggunakan algoritma *Support Vector Machine* yang memberikan hasil nilai akurasi tertinggi berada pada pengujian 3 untuk masing-masing *kernel*. Pada *kernel linear* tingkat akurasi 90% dan pada *kernel RBF* tingkat akurasi 88%.

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

## ABSTRACT

*Social media is currently a communication medium that is often used by Indonesians to convey an opinion. One of the media that is often used by the community is twitter. Twitter social media which provides a lot of very diverse information through tweets, from information written through these tweets to data that can be processed into sentiment analysis. This study aims to create a system of public sentiment analysis in the form of positive or negative sentiments.*

*In this study, researchers used text mining techniques by implementing the Support Vector Machine algorithm which was used to analyze Twitter user sentiment on Biznet internet services. The kernels used in this research are Linear kernel and RBF kernel. The test is carried out with 3 scenarios, in scenario 1 using 800 data, scenario 2 using 900 data and scenario 3 using 1000 data, for the division that is 90% training data and 10% testing data from each scenario.*

*Systems with the Support Vector Machine method prove that the algorithm can analyze automatically. The results were obtained using the Confusion Matrix method on a model created using the Support Vector Machine algorithm which gave the results of the accuracy value set in test 3 for each kernel. In the linear kernel, the accuracy rate is 90% and the RBF kernel has an accuracy rate of 88%.*

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

