

**PENERAPAN METODE SUPPORT VECTOR MACHINE  
UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER**

**SKRIPSI**



disusun oleh

**Zidna Alhaq**

**18.21.1185**

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

**PENERAPAN METODE SUPPORT VECTOR MACHINE  
UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER**

**SKRIPSI**

untuk memenuhi sebagian persyaratan  
mencapai gelar Sarjana  
pada Program Studi Informatika



disusun oleh

**Zidna Alhaq**

**18.21.1185**

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

# **PERSETUJUAN**

## **SKRIPSI**

### **PENERAPAN METODE SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER**

yang dipersiapkan dan disusun oleh

**Zidna Alhaq**

**18.21.1185**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 07 November 2019

**Dosen Pembimbing,**



**Ali Mustopa, M.Kom**

**NIK. 190302192**

**PENGESAHAN**

**SKRIPSI**

**PENERAPAN METODE SUPPORT VECTOR MACHINE**

**UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER**

yang dipersiapkan dan disusun oleh

**Zidna Alhaq**

**18.21.1185**

telah dipertahankan di depan Dewan Penguji  
pada tanggal 20 November 2019

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Mulia Sulistiyono, M.Kom**  
**NIK. 190302248**

**Arifiyanto Hadinegoro, S.Kom, MT**  
**NIK. 190302289**

**Ali Mustopa, M.Kom**  
**NIK. 190302192**



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
pada tanggal 20 November 2019



**DEWAN PENGUJIAN 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, 25 November 2019



Zidna Alhaq

NIM. 18.21.1185

## MOTTO

*“ Ilmu itu lebih baik daripada harta. Ilmu menjaga engkau dan engkau menjaga harta. Ilmu itu penghukum (hakim) dan harta terhukum. Harta itu kurang apabila dibelanjakan tapi ilmu bertambah bila dibelanjakan. “*

*(Ali bin Abi Thalib)*

*“ Akan kuberikan ilmu yang kumiliki kepada siapapun, asal mereka mau memanfaatkan ilmu yang telah kuberikan itu. “*

*(Imam Syafi'i)*

*“ life just once don't forget to enjoy this life ”*

*(Zidna A)*

## PERSEMBAHAN

Puji Syukur saya panjatkan kepada Allah SWT karena telah berhasil menyelesaikan Skripsi ini. Saya persembahkan Skripsi ini untuk :

- Bapak dan Ibu Tercinta yang tak henti-hentinya mendoakan, merawat, memberi semangat, serta membesarkan saya dengan penuh kasih sayang.
- Untuk adikku terimakasih atas dukungannya selama ini.
- Dosen Pembimbing Bapak Ali Mustopa, M.Kom yang telah membimbing dan mengarahkan dalam pembuatan tugas akhir ini.
- Teman – Teman kos rajawali yang selalu memberi dukungan.
- Teman-teman 18 IFT 01 terimakasih atas doa dan dukungannya selama ini.
- Dosen Universitas Amikom Yogyakarta yang telah memberikan ilmu selama menempuh kuliah.
- Zidna Alhaq atau diri saya sendiri. Saya mengapresiasi diri saya sendiri atas terlaksananya penelitian ini.

Terima Kasih

## KATA PENGANTAR

Puji syukur penulis panjatkan atas kehadiran ALLAH SWT yang telah memberikan rahmat-Nya, sehingga penulis dapat menyelesaikan skripsi ini dengan judul **“Penerapan Metode *Support Vector Machine* Untuk Analisis Sentimen Pengguna Twitter”**

Penulisan Tugas Akhir ini dimaksudkan untuk memenuhi syarat kelulusan program Sarjana Sistem Informasi di Universitas Amikom Yogyakarta.

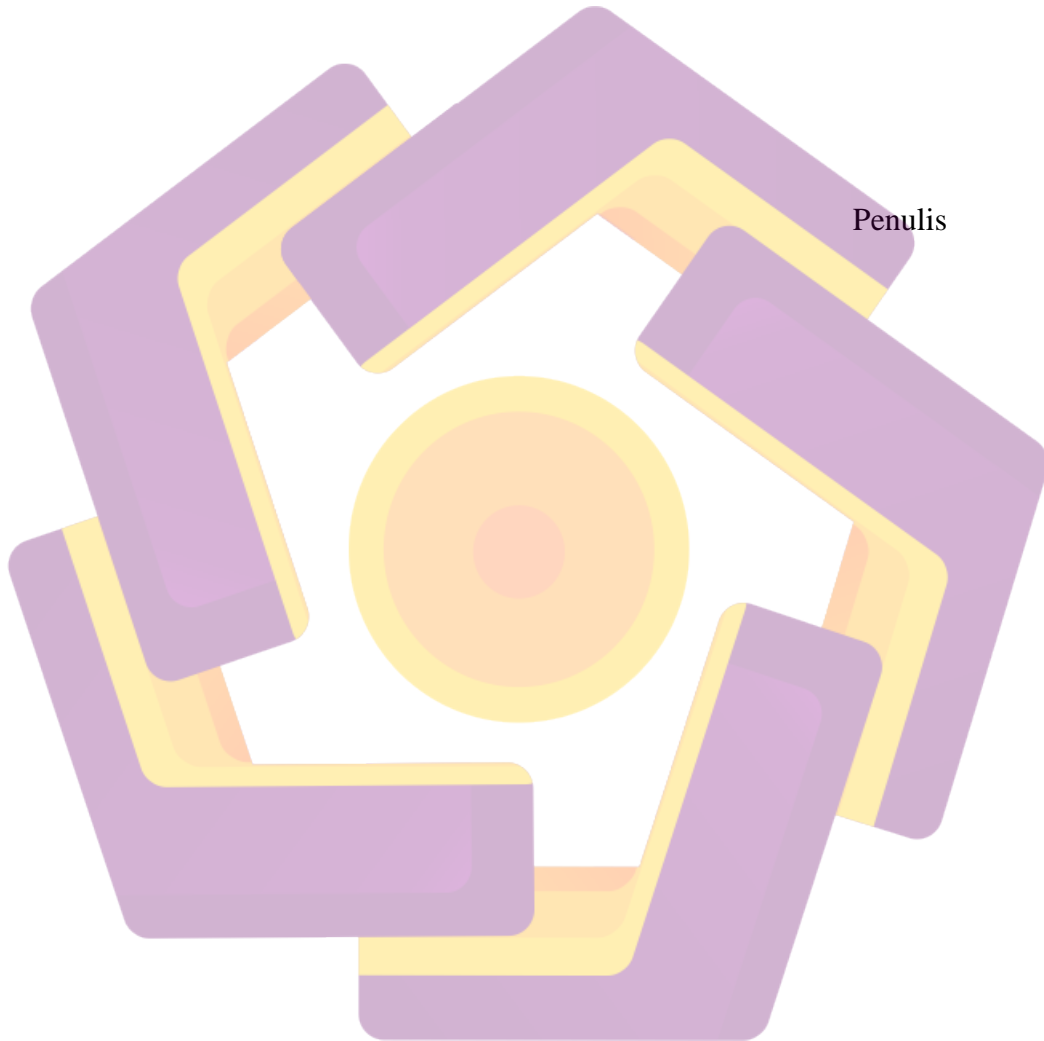
Selesaiannya skripsi ini tidak terlepas dari dukungan berbagai pihak yang telah memberikan dorongan moril maupun spiritual dan juga bimbingan ilmu pengetahuan. Oleh karena itu penulis mengucapkan terimakasih kepada :

1. Allah SWT atas segala berkah, hidayah bimbingan dan keridhoan ilmu-Nya sehingga penulis dapat menyelesaikan tugas akhir ini dengan baik.
2. Nabi Agung Muhammad SAW yang selalu menjadi panutan dan suri tauladan.
3. Kedua orang tua dan adik penulis sebagai sosok luar biasa yang memberikan segala dukungan dan semangat baik moril maupun material.
4. Bapak Prof. Dr. M. Suyanto, MM selaku rektor Universitas Amikom Yogyakarta.
5. Bapak Ali Mustopa, M.Kom selaku dosen pembimbing yang telah membimbing dan mengarahkan dalam pembuatan skripsi ini.
6. Seluruh dosen, staff maupun karyawan Universitas Amikom Yogyakarta.
7. Teman – teman di Universitas Amikom Yogyakarta yang telah memberikan dukungan.



Akhir kata penulis berharap semoga laporan tugas akhir ini dapat memberikan manfaat bagi pembaca. Penulis menyadari pembuatan tugas akhir ini kurang dari sempurna. Kritik dan saran yang membangun sangat diharapkan demi kesempurnaan tugas akhir ini.

Yogyakarta, 25 November 2019



## DAFTAR ISI

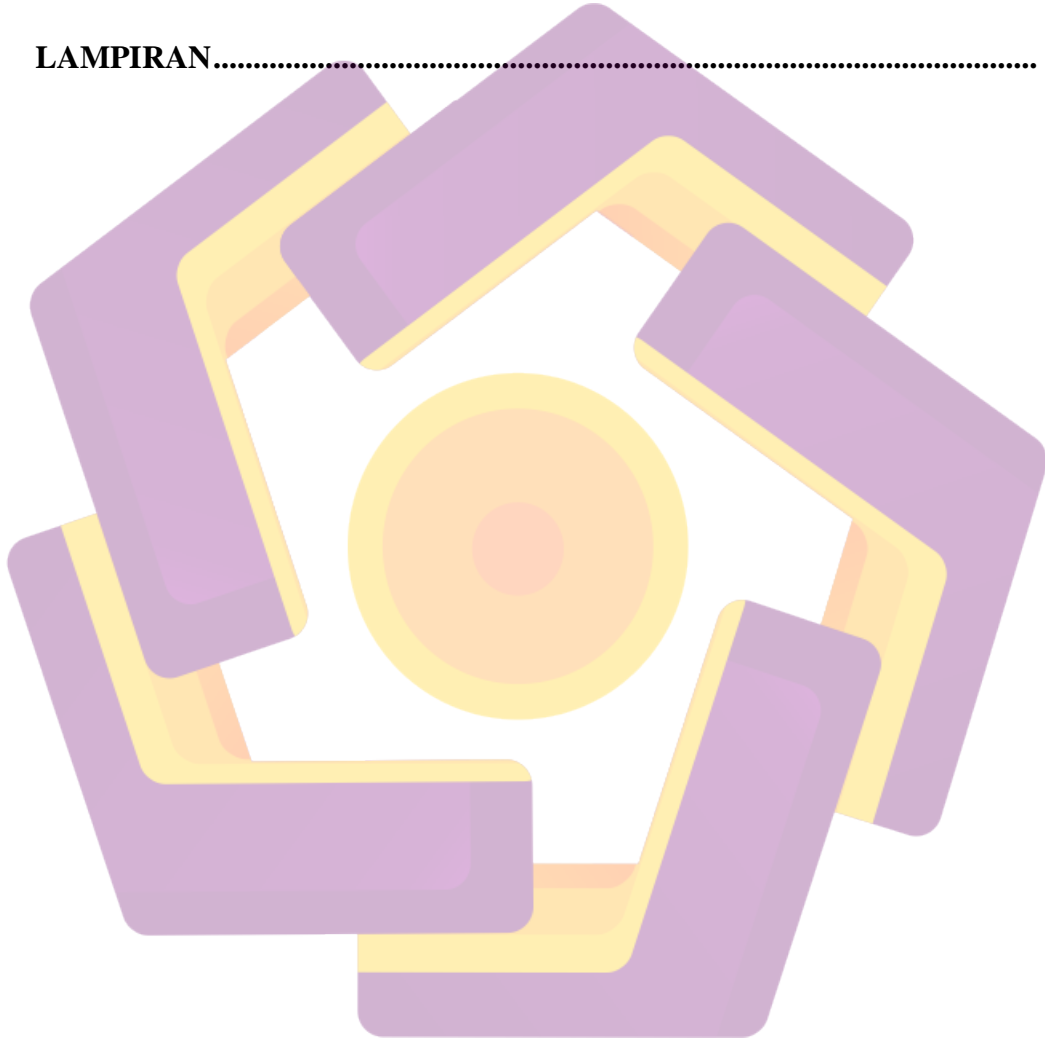
<b>JUDUL</b> .....	<b>i</b>
<b>PERSETUJUAN</b> .....	<b>ii</b>
<b>PENGESAHAN</b> .....	<b>iii</b>
<b>PERNYATAAN</b> .....	<b>iii</b>
<b>MOTTO</b> .....	<b>v</b>
<b>PERSEMBAHAN</b> .....	<b>vi</b>
<b>KATA PENGANTAR</b> .....	<b>vii</b>
<b>DAFTAR ISI</b> .....	<b>ix</b>
<b>DAFTAR GAMBAR</b> .....	<b>xiv</b>
<b>DAFTAR TABEL</b> .....	<b>xvi</b>
<b>INTISARI</b> .....	<b>xviii</b>
<b>ABSTRACT</b> .....	<b>xix</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah.....	4
1.4 Maksud dan Tujuan Penelitian .....	4
1.5 Manfaat Penelitian.....	5
1.6 Metode Penelitian.....	6
1.6.1 Metode Studi Literatur .....	6
1.6.2 Metode Analisis .....	6
1.6.3 Perancangan .....	6
1.6.4 Implementasi.....	7

1.6.5	Pengujian.....	7
1.7	Sistematika Penulisan.....	8
<b>BAB II LANDASAN TEORI.....</b>		<b>9</b>
2.1	Tinjauan Pustaka .....	9
2.2	Analisis Sentimen.....	11
2.2.1	Kelas Sentimen.....	12
2.3	<i>Text Mining</i> .....	12
2.4	<i>Preprocessing</i> .....	13
2.5	<i>Part Of Speech Tagging (POS Tagging)</i> .....	16
2.6	<i>Term Frequency-Invers Document Frequency (TF-IDF)</i> .....	18
2.7	<i>Cosine Similarity</i> .....	19
2.8	<i>Support Vector Machine (SVM)</i> .....	20
2.8.1	<i>Kernel Trick dan non Linear SVM</i> .....	23
2.8.2	<i>Multi-Class Support Vector Machine</i> .....	25
2.9	Validasi dan Evaluasi .....	27
2.8	<i>Flowchart</i> .....	29
<b>BAB III ANALISIS DAN PERANCANGAN.....</b>		<b>33</b>
3.1	Analisis Kebutuhan .....	33
3.1.1	Analisis Kebutuhan Fungsional.....	33
3.1.2	Analisis Kebutuhan Non-Fungsional .....	34
3.2	Gambaran Umum Sistem .....	35
3.3	Analisis Sentimen dengan Algoritma <i>Support Vector Machine</i> .....	37
3.3.1	Pengumpulan Data .....	37

3.3.2	Pelabelan <i>Tweet</i> .....	39
3.3.3	<i>Preprocessing</i> .....	40
3.3.3.1	<i>Tweet Cleaning</i> .....	41
3.3.3.2	<i>Tokenization</i> .....	43
3.3.3.3	Normalisasi Kata.....	46
3.3.3.4	<i>Stopword Remove</i> .....	49
3.3.3.5	<i>POS Tagging</i> .....	52
3.3.3.6	<i>POS Filtering</i> .....	55
3.3.3.7	<i>Stemming</i> .....	58
3.3.3.8	<i>Feature List</i> .....	59
3.3.3.8	<i>Vektor Feature</i> .....	60
3.3.4	<i>Training</i> .....	61
3.3.4.1	Pembobotan TF-IDF.....	61
3.3.4.2	<i>Cosine Similarity</i> .....	68
3.3.4.3	Perhitungan <i>Hyperplane</i> .....	69
3.3.4.4	Evaluasi Model.....	71
3.4	Analisis Kebutuhan Data.....	73
3.4.1	Perancangan Data Twitter.....	73
3.4.2	Perancangan Dataset.....	74
3.5	Perancangan Sistem.....	75
3.5.1	<i>Flowchart</i> .....	75
3.6	Perancangan Antarmuka Pengguna.....	75
3.6.1	Perancangan Antarmuka Halaman Utama.....	75
3.6.2	Perancangan Antarmuka Halaman Prediksi.....	76
3.6.3	Perancangan Antarmuka Data <i>Tweet</i> .....	77

<b>BAB IV IMPLEMENTASI DAN PEMBAHASAN .....</b>	<b>78</b>
4.1 Deskripsi Implementasi .....	78
4.2 Implementasi Pengumpulan Data Tweet.....	78
4.3 Implementasi Pelabelan Data .....	79
4.4 Implementasi <i>Preprocessing</i> .....	79
4.4.1 <i>Tweet Cleaning</i> .....	79
4.4.2 <i>Tokenization</i> .....	80
4.4.3 Normalisasi Kata.....	80
4.4.4 <i>Stopword Remove</i> .....	81
4.4.5 <i>POS Tagging</i> .....	82
4.4.6 <i>POS Filtering</i> .....	82
4.4.7 <i>Stemming</i> .....	82
4.4.8 <i>Feature List</i> .....	83
4.5 Vektor <i>Tweet</i> .....	84
4.5.1 Implementasi <i>TF-IDF Cosine Similarity</i> .....	84
4.5.2 Implementasi Pelatihan <i>Support Vector Machine</i> .....	85
4.6 Implementasi Antarmuka .....	87
4.6.1 Antarmuka Halaman Utama.....	87
4.6.2 Antarmuka Halaman Prediksi .....	87
4.6.1 Antarmuka Halaman Data <i>Tweet</i> .....	88
4.7 Evaluasi Model.....	89
4.7.1 Evaluasi Model Terhadap Hasil Pelatihan .....	90
4.6.2 Evaluasi Model Terhadap Prediksi .....	92

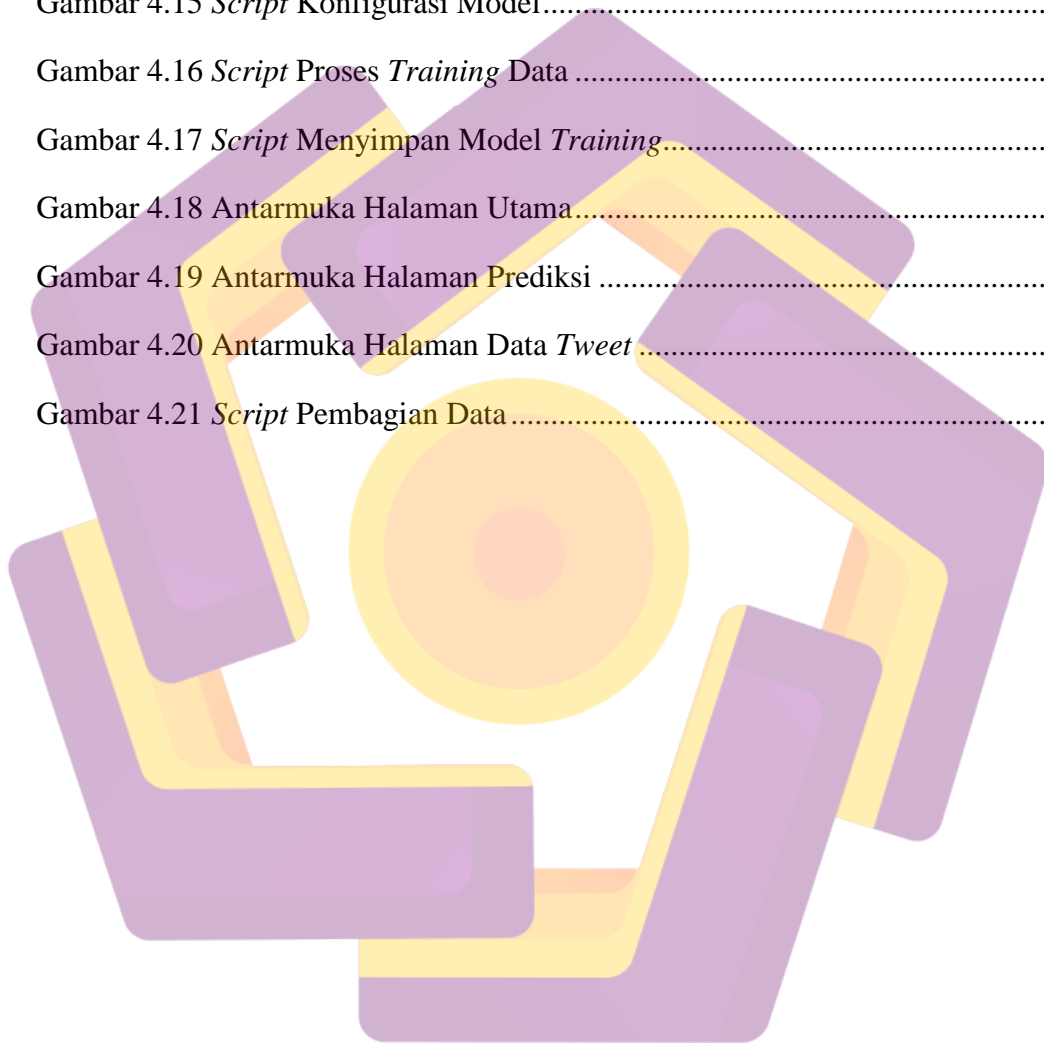
<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>96</b>
5.1 Kesimpulan.....	96
5.2 Saran.....	96
<b>DAFTAR PUSTAKA .....</b>	<b>98</b>
<b>LAMPIRAN.....</b>	<b>100</b>



## DAFTAR GAMBAR

Gambar 2.1 (a) Pencarian <i>Hyperplane</i> (b) <i>Hyperplane</i> terbaik .....	21
Gambar 2.2 Ilustrasi <i>One Against One</i> .....	26
Gambar 2.3 Ilustrasi <i>One Against All</i> .....	27
Gambar 2.4 Ilustrasi <i>K-fold Cross Validation</i> .....	28
Gambar 2.5 Contoh <i>Flowchart</i> .....	33
Gambar 3.1 Gambaran Umum Sistem .....	35
Gambar 3.2 <i>Preprocessing</i> .....	41
Gambar 3.3 Perancangan File <i>Training</i> .....	74
Gambar 3.4 Perancangan File <i>Prediksi</i> .....	74
Gambar 3.5 Gambaran Umum Sistem .....	75
Gambar 3.6 Rancangan Antarmuka Halaman Utama .....	76
Gambar 3.7 Rancangan Antarmuka Hasil <i>Prediksi</i> .....	76
Gambar 3.8 Rancangan Antarmuka Data <i>Tweet</i> .....	77
Gambar 4.1 <i>Script</i> Pengumpulan Data <i>Tweet</i> .....	78
Gambar 4.2 Contoh File Data <i>Training</i> .....	79
Gambar 4.3 <i>Script Tweet Cleaning</i> .....	80
Gambar 4.4 <i>Script Tokenization</i> .....	80
Gambar 4.5 Kamus Kata Baku .....	81
Gambar 4.6 <i>Script</i> Normalisasi Kata .....	81
Gambar 4.7 <i>Script Stopword Remove</i> .....	81
Gambar 4.8 <i>Script POS Tagging</i> .....	82
Gambar 4.9 <i>Script POS Filtering</i> .....	82
Gambar 4.10 <i>Script Stemming</i> .....	82

Gambar 4.11 <i>Script Feature List</i> .....	83
Gambar 4.12 <i>Script Menyimpan Feature List</i> .....	84
Gambar 4.13 <i>Script Fungsi TF-IDF Cosine Similarity</i> .....	85
Gambar 4.14 <i>Script Menghitung Bobot Per Tweet</i> .....	85
Gambar 4.15 <i>Script Konfigurasi Model</i> .....	86
Gambar 4.16 <i>Script Proses Training Data</i> .....	86
Gambar 4.17 <i>Script Menyimpan Model Training</i> .....	87
Gambar 4.18 Antarmuka Halaman Utama.....	87
Gambar 4.19 Antarmuka Halaman Prediksi .....	88
Gambar 4.20 Antarmuka Halaman Data <i>Tweet</i> .....	88
Gambar 4.21 <i>Script Pembagian Data</i> .....	89





## DAFTAR TABEL

Tabel 2.1 Tabel Perbedaan antara Tinjauan Pustaka dan Penelitian.....	10
Tabel 2.2 Ilustrasi <i>Tweet Cleaning</i> .....	14
Tabel 2.3 Ilustrasi <i>Tweet Case Folding</i> dari <i>Cleaning</i> .....	14
Tabel 2.4 Ilustrasi <i>Tweet Tokenization</i> dari <i>Case Folding</i> .....	15
Tabel 2.5 Ilustrasi <i>Tweet Stemming</i> dari <i>Tokenization</i> .....	16
Tabel 2.6 Label pada <i>POS Tagging</i> .....	17
Tabel 2.7 <i>Confussion Matrix</i> .....	28
Tabel 2.8 <i>Flowchart</i> .....	30
Tabel 3.1 <i>Tweet</i> yang Telah Diambil .....	37
Tabel 3.2 <i>Tweet</i> Berlabel.....	39
Tabel 3.3 <i>Tweet Cleaning</i> .....	41
Tabel 3.4 <i>Tokenization</i> .....	44
Tabel 3.5 Normalisasi Kata.....	46
Tabel 3.6 <i>Stopword Remove</i> .....	50
Tabel 3.7 <i>POS Tagging</i> .....	53
Tabel 3.8 <i>POS Filtering</i> .....	55
Tabel 3.9 <i>Stemming</i> .....	58
Tabel 3.10 <i>Feature List</i> .....	60
Tabel 3.11 <i>Vektor Feature</i> .....	61
Tabel 3.12 Term Positif.....	62
Tabel 3.13 Term Negatif .....	63
Tabel 3.14 Term Netral .....	65
Tabel 3.15 TF <i>Tweet</i> Ke-9 .....	66

Tabel 3.16 DF <i>Tweet</i> Ke-9.....	66
Tabel 3.17 IDF <i>Tweet</i> Ke-9.....	67
Tabel 3.18 TF-IDF <i>Tweet</i> Ke-9.....	67
Tabel 3.19 TF-IDF <i>Cosine Similarity Tweet</i> Ke-9.....	69
Tabel 3.20 Klasifikasi SVM.....	72
Tabel 3.21 <i>Confusion Matrix</i> .....	72
Tabel 3.22 <i>Accuracy, Precision, dan Recall</i> .....	73
Tabel 4.1 Evaluasi <i>K-Fold Cross Validation</i> .....	90
Tabel 4.2 Evaluasi Prediksi.....	92
Tabel 4.3 <i>Confusion Matrix</i> Prediksi .....	93
Tabel 4.4 Perhitungan Evaluasi SVM.....	93
Tabel 4.5 <i>Confusion Matrix</i> SVM .....	95
Tabel 4.6 <i>Accuracy, Precision, dan Recall</i> .....	95

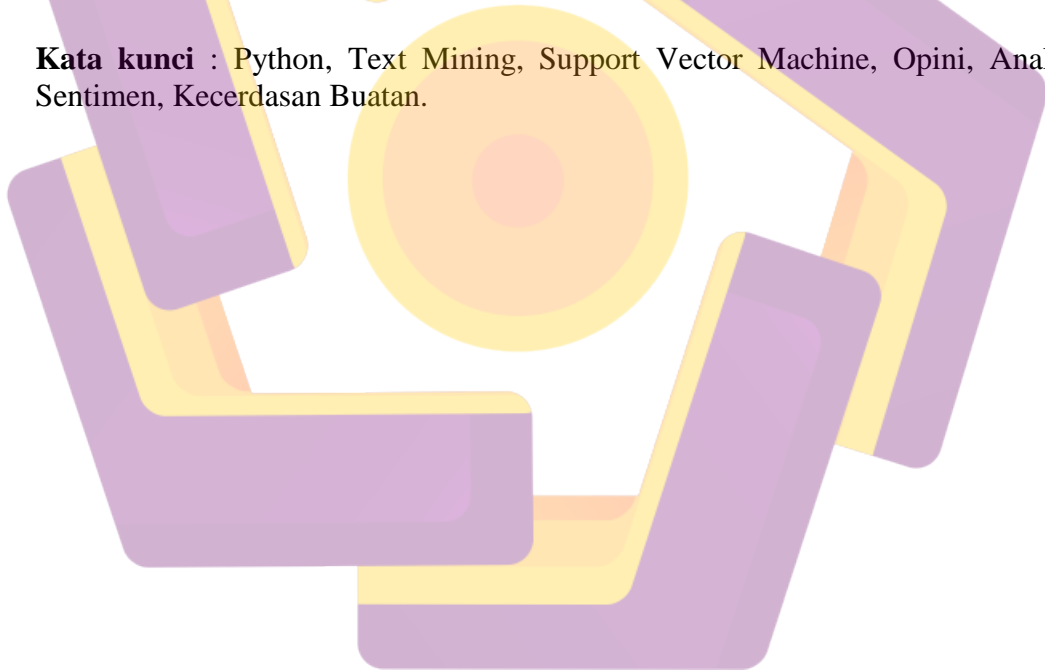
## INTISARI

Twitter merupakan salah satu media sosial yang digunakan untuk menyalurkan opini dan membahas berbagai topik yang ada di sekitar. Salah satu topik yang sering dibahas yaitu marketplace. Bukalapak merupakan salah satu *marketplace* paling populer di Indonesia.

Bukalapak memberikan penggunanya untuk dapat melakukan transaksi secara cepat dan aman. Respon yang diberikan oleh pengguna tersebut dapat berupa respon positif, negatif, dan netral. Oleh karena itu diperlukan suatu metode yang dapat digunakan untuk mengetahui opini pengguna Bukalapak pada sosial media Twitter. Untuk menyelesaikan masalah tersebut maka diperlukan suatu metode yang dapat mengkategorikan opini tersebut.

*Support Vector Machine* merupakan salah satu metode dalam *text mining* yang dapat mengkategorikan opini-opini tersebut. Data yang diperoleh dari Twitter akan dilabeli dan dianalisis menggunakan metode SVM untuk mengklasifikasikan opini tersebut. Hasil dari klasifikasi menggunakan metode SVM diperoleh tingkat akurasi sebesar 93%.

**Kata kunci** : Python, Text Mining, Support Vector Machine, Opini, Analisis Sentimen, Kecerdasan Buatan.



## **ABSTRACT**

*Twitter is one of the social media that is used to express an opinions and discuss various topics around. One topic that is often discussed is marketplace. Bukalapak is one of the most popular marketplace in Indonesia.*

*Bukalapak provides its users with the ability to make transactions quickly and securely. Responses given by these users can be positive, negative and neutral responses. Therefore a method that can be used to find out the opinion of Bukalapak users on Twitter social media is needed. To solve this problem, we need a method that can categorize these opinions.*

*Support Vector Machines is one method of extracting text that can categorize these opinions. Data obtained from Twitter will be labeled and analyzed using the SVM method to classify these opinions. The results of the classification using the SVM method obtained an accuracy rate of 93%.*

**Keywords:** *Python, Text Mining, Support Vector Machine, Opinion, Sentiment Analysis, Artificial Intelligence*

