

**ANALISIS SENTIMEN SEPAKBOLA INDONESIA MENGGUNAKAN  
SUPPORT VECTOR MACHINE**

**SKRIPSI**



disusun oleh

**Anang Satria Wicaksono**

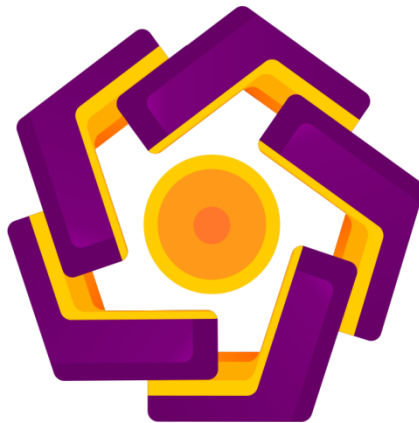
**15.11.9312**

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

**ANALISIS SENTIMEN SEPAKBOLA INDONESIA MENGGUNAKAN  
SUPPORT VECTOR MACHINE**

**SKRIPSI**

untuk memenuhi sebagian persyaratan  
mencapai gelar Sarjana  
pada Program Studi Informatika



disusun oleh

**Anang Satria Wicaksono**

**15.11.9312**

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

**2019**

## **PERSETUJUAN**

### **SKRIPSI**

#### **ANALISIS SENTIMEN SEPAKBOLA INDONESIA MENGGUNAKAN SUPPORT VECTOR MACHINE**

yang dipersiapkan dan disusun oleh

**Anang Satria Wicaksono**

**15.11.9312**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 23 Juli 2019

**Dosen Pembimbing,**



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

**NIK. 190302185**

## PENGESAHAN

### SKRIPSI

#### ANALISIS SENTIMEN SEPAKBOLA INDONESIA MENGGUNAKAN SUPPORT VECTOR MACHINE

yang dipersiapkan dan disusun oleh

**Anang Satria Wicaksono**

**15.11.9312**

telah dipertahankan di depan Dewan Penguji  
pada tanggal 27 Agustus 2019

#### Susunan Dewan Penguji

**Nama Penguji**

**Ike Verawati, M.Kom**  
**NIK. 190302237**

**Erni Seniwati, M.Cs**  
**NIK. 190302231**

**Windha Mega Pradnya D, M.Kom.**  
**NIK. 190302185**

**Tanda Tangan**



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 17 September 2019

**DEKAN 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, 14 September 2019



Anang Satria Wicaksono  
NIM. 15.11.9312

## MOTTO

” فَإِنَّ مَعَ الْعُسْرِ يُسْرًا إِنَّ مَعَ الْعُسْرِ يُسْرًا ”

- Al-Qur'an Surah Al-Insyiraah (94: 5-6)

” *Jika aku mati setelah berjuang sebisaku, aku tidak akan menyesali*

*kematianku* ”

- Monkey D. Luffy

” *The Way Get Started Is To Quit Talking And Begin Doing.* ”

- Walt Disney

” *If you are working on something that you really care about, you don't  
have to be pushed. The vision pulls you.* ”

- Steve Jobs

” *Kegagalan hanya terjadi bila kita menyerah.* ”

- B.J. Habibie

## PERSEMBAHAN

Alhamdulillah segala puji syukur kehadiran Allah SWT yang telah memberikan limpahan rahmat dan karunia-Nya sehingga penelitian ini dapat dilakukan dan diselesaikan dengan sebaik-baiknya. Saya juga ucapkan terimakasih untuk dukungan dan bantuan semua pihak yang membantu selesainya penelitian ini.

Skripsi ini saya persembahkan untuk kampus saya, Universitas AMIKOM Yogyakarta.

Untuk kedua orang tua saya yang telah susah payah mendukung saya tanpa kenal lelah.

Untuk ibu Windha Mega Pradnya, yang sudah dengan sabar membimbing saya. Saya ucapkan terimakasih.

Untuk para pembaca, saya ucapkan terima kasih sebesar-besarnya. Semoga menjadi amal jariyah atas ilmu yang saya bagikan.

Dan untuk semua pihak yang telah banyak membantu dalam penyusunan skripsi ini yang tidak dapat saya sebutkan satu persatu.

## KATA PENGANTAR

Puji syukur penulis haturkan atas kehadiran Allah Subhanahu wa Ta'ala yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “Analisis Sentimen Sepakbola Indonesia Menggunakan Support Vector Machine” dengan sebaik-baiknya. Tidak lupa shalawat serta salam penulis haturkan kepada junjungan Nabi Muhammad Shallallahu ‘alaihi wa Sallam beserta keluarga dan para sahabatnya.

Dengan selesainya skripsi ini, penulis mengucapkan banyak terimakasih kepada:

1. Bapak Prof. Dr. M. Suyanto selaku Rektor Universitas AMIKOM Yogyakarta
2. Bapak Sudarmawan, S.T, M.T. selaku Dekan Fakultas Sains dan Teknologi, dan Ketua Program Studi S1 Informatika serta Wali Dosen penulis.
3. Ibu Windha Mega Pradnya D, M.Kom. selaku pembimbing yang telah dengan sabar membimbing dan banyak memberikan kontribusi bagi penulis dalam pembuatan skripsi ini.
4. Para Dosen dan Staff Universitas AMIKOM Yogyakarta yang telah banyak memberi bantuan informasi, pengalaman dan kontribusi lainnya selama penulis berkuliah di kampus ini hingga terselesaikannya skripsi.
5. Orang tua dan keluarga penulis yang telah memberikan dukungan terbaiknya selama berkuliah.



6. Ayu Indra, penyemangat saat proses penulisan skripsi dari awal hingga selesai.
7. Sahabat Rodi, yang telah berjuang bersama dalam proses penulisan skripsi dari awal sampai selesai.
8. Ibu Lia Ayu Ivanjelita M.kom. selaku dosen yang membuat saya semangat saat kuliah.

Penulis menyadari dengan betul bahwa dalam pembuatan skripsi ini masih banyak kekurangan dan kelemahan. Oleh karena itu penulis berharap untuk semua pihak yang telah membaca dan memahami penelitian ini untuk dapat menyampaikan kritik dan saran yang membangun sehingga dapat menambah kesempurnaan skripsi ini.

Akhir kata, semoga skripsi ini dapat bermanfaat bagi pihak terkait dan pembaca pada umumnya.

Yogyakarta, 14 September 2019

Anang Satria Wicaksono

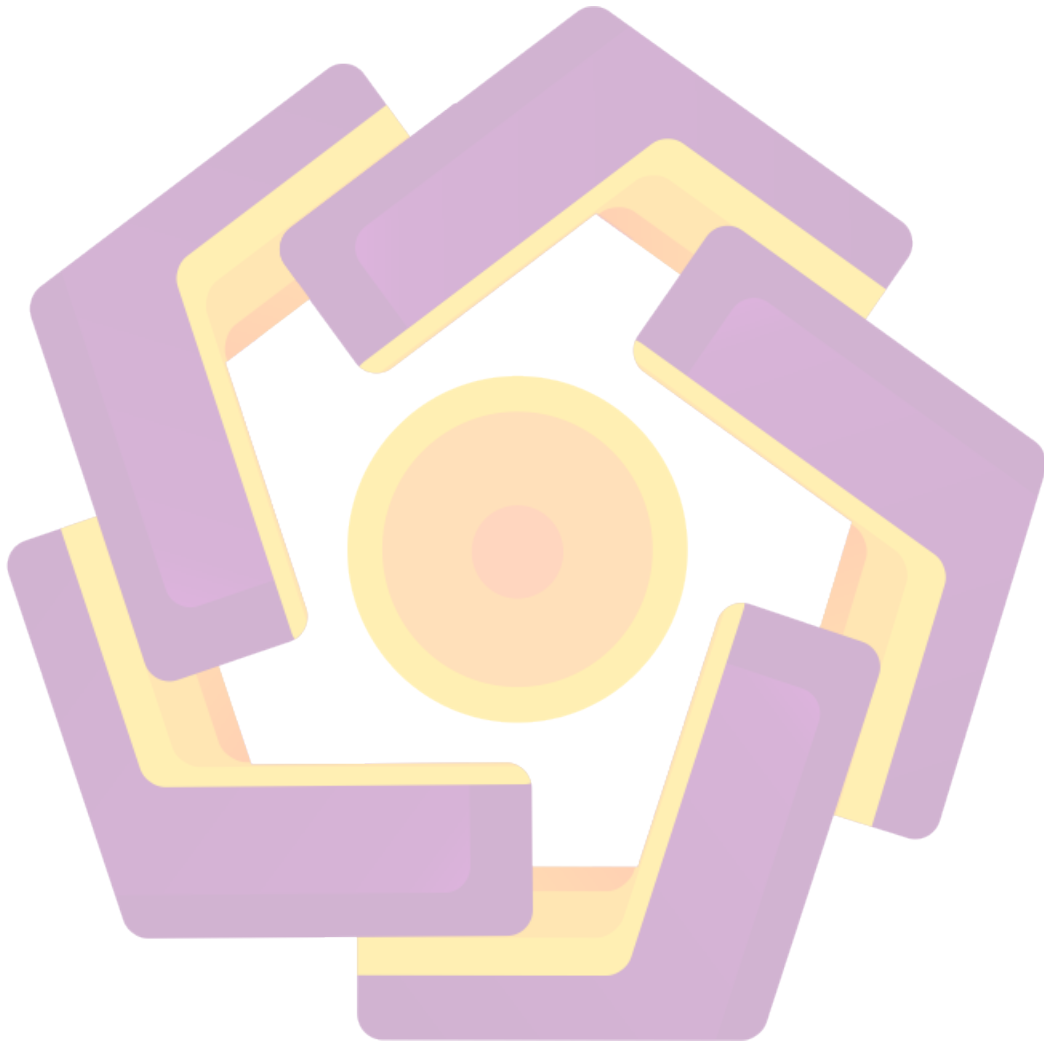
15.11.9312

## DAFTAR ISI

JUDUL .....	ii
PERSETUJUAN .....	iii
PENGESAHAN .....	iv
PERNYATAAN.....	v
MOTTO .....	vi
PERSEMBAHAN.....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR .....	xiv
INTISARI.....	xvi
ABSTRACT.....	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian .....	4
1.5 Manfaat Penelitian .....	4
1.6 Metode Penelitian.....	5
1.7 Sistematika Penulisan .....	6
BAB II LANDASAN TEORI.....	8
2.1 Tjauan Pustaka .....	8
2.2 Landasan Teori.....	9
2.2.1 Text Mining.....	9

2.2.2	Analisis Sentimen .....	10
2.2.3	Pre-Processing.....	10
2.2.4	Term Frequency – Inverse Document Frequency .....	12
2.2.5	Klasifikasi .....	14
2.2.6	Support Vector Machine .....	15
2.2.7	Evaluasi.....	20
2.2.8	Validasi .....	22
<b>BAB III METODE PENELITIAN.....</b>		<b>23</b>
3.1	Alat dan Bahan.....	23
3.1.1	Alat Penelitian.....	23
3.1.2	Bahan Penelitian.....	24
3.2	Alur Penelitian .....	24
3.2.1	Pengumpulan Data .....	27
3.2.2	Pre-Processing Data .....	27
3.2.3	TF-IDF .....	36
3.2.4	Klasifikasi .....	42
3.2.5	Evaluasi.....	53
3.2.6	validasi .....	53
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>		<b>54</b>
4.1	Eksperimen.....	54
4.1.1	Pengumpulan Data .....	54
4.1.2	Pre-Processing Data .....	55
4.1.3	Pembagian Data .....	67
4.1.4	Pembobotan.....	68
4.1.5	Implementasi Support Vector Machine .....	70
4.1.6	Evaluasi .....	71
4.1.7	Validasi .....	72
4.2	Hasil dan Pembahasan.....	73
4.2.1	Hasil Evaluasi.....	74
4.2.2	Hasil Validasi .....	77

BAB V PENUTUP.....	81
5.1 KESIMPULAN.....	81
5.2 SARAN .....	82
DAFTAR PUSTAKA .....	83



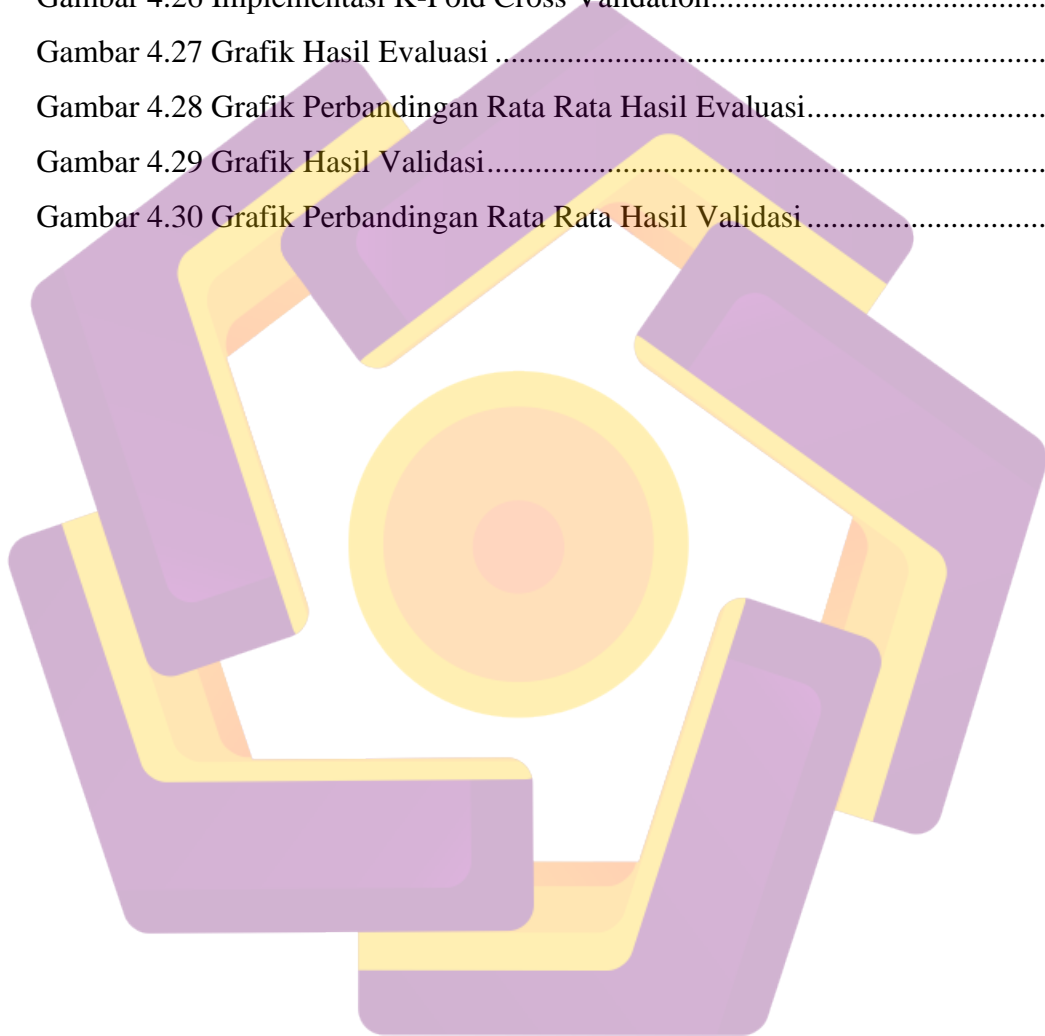
## DAFTAR TABEL

Tabel 2.1 Konversi Emoticon .....	11
Tabel 3.1 Spesifikasi Perangkat keras dan Perangkat Lunak.....	23
Tabel 3.2 Tahapan Model Preprocessing .....	27
Tabel 3.3 Contoh Proses Case Folding .....	29
Tabel 3.4 Contoh Proses Convert Emoticon .....	30
Tabel 3.5 Contoh Proses Cleaning .....	31
Tabel 3.6 Contoh Proses SlangWord .....	32
Tabel 3.7 Contoh Proses Stemming .....	33
Tabel 3.8 Contoh Proses Stopword.....	34
Tabel 3.9 Contoh Proses Filtering.....	35
Tabel 3.10 Contoh Proses Tokenizing .....	35
Tabel 3.11 Contoh proses pembobotan Term Frequency .....	37
Tabel 3.12 Contoh proses pembobotan Inverse Document Frequency.....	39
Tabel 3.13 Contoh hasil pembobotan TF-IDF .....	41
Tabel 3.14 Penentuan data latih dan data uji.....	44
Tabel 3.15 Data latih dalam format Support Vector.....	45
Tabel 3.16 Matriks Kernelisasi $xix$ .....	50
Tabel 4.1 Hasil Sentimen .....	73
Tabel 4.2 Hasil Evaluasi Model 1 .....	74
Tabel 4.3 Hasil Evaluasi Model 2 .....	75
Tabel 4.4 Hasil Validasi Model 1 .....	77
Tabel 4.5 Hasil Validasi Model 2 .....	78

## DAFTAR GAMBAR

Gambar 2.1 (a) Pencarian Hyperplane (b) Hyperplane Terbaik .....	15
Gambar 2.2 Fungsi $\Phi$ memetakan data ke ruang vector yang berdimensi tinggi .	18
Gambar 2.3 <i>Confusion Matrix</i> .....	20
Gambar 2.4 Ilustrasi k-Fold Cross Validation .....	22
Gambar 3.1 Diagram alir tahapan umum eksperimen .....	26
Gambar 3.2 Diagram alir tahapan pre-processing model 1.....	28
Gambar 3.3 Diagram alir tahapan pre-processing model 2.....	29
Gambar 3.4 Diagram alir tahapan SVM .....	43
Gambar 4.1 Instalasi <i>GetOldTweets3</i> .....	54
Gambar 4.2 Pengambilan data .....	55
Gambar 4.3 <i>Import</i> file CSV dengan <i>library</i> Pandas.....	55
Gambar 4.4 Implementasi Case Folding.....	56
Gambar 4.5 Hasil Implementasi Case Folding .....	56
Gambar 4.6 Implementasi Convert Emoticon.....	57
Gambar 4.7 Implementasi Convert Emoticon.....	58
Gambar 4.8 Hasil Implementasi <i>Convert Emoticon</i> .....	58
Gambar 4.9 Implementasi <i>cleaning</i> .....	59
Gambar 4.10 Hasil Implementasi <i>Cleaning</i> .....	60
Gambar 4.11 Implementasi <i>Slangword</i> .....	61
Gambar 4.12 Implementasi <i>SlangWord 2</i> .....	61
Gambar 4.13 Hasil Implementasi <i>SlangWord</i> .....	62
Gambar 4.14 Implementasi <i>Stemming</i> .....	62
Gambar 4.15 Hasil Implementasi <i>Stemming</i> .....	63
Gambar 4.16 Implementasi <i>stopwords removing</i> .....	64
Gambar 4.17 Hasil Implementasi <i>Stopword Removing</i> .....	64
Gambar 4.18 Implementasi <i>filtering</i> .....	65
Gambar 4.19 Hasil implementasi <i>filtering</i> .....	65
Gambar 4.20 Implementasi <i>tokenizing</i> .....	66

Gambar 4.21 Hasil implementasi tokenizing.....	67
Gambar 4.22 Implementasi Pembagian Data.....	67
Gambar 4.23 Hasil Implentasi TF-IDF .....	69
Gambar 4.24 Implementasi SVM .....	70
Gambar 4.25 Implementasi Evaluasi .....	71
Gambar 4.26 Implementasi K-Fold Cross Validation.....	72
Gambar 4.27 Grafik Hasil Evaluasi .....	76
Gambar 4.28 Grafik Perbandingan Rata Rata Hasil Evaluasi.....	76
Gambar 4.29 Grafik Hasil Validasi.....	79
Gambar 4.30 Grafik Perbandingan Rata Rata Hasil Validasi .....	79



## INTISARI

Di dalam penelitian ini, peneliti melakukan analisis sentimen terhadap PSSI dalam mengelola sepakbola Indonesia. Pada penelitian ini juga akan membandingkan akurasi antara 2 model preprocessing data. pada model 1 menggunakan case folding, cleaning, stemming, stopwords removal, filtering, tokenization. Sedangkan pada model 2 menggunakan case folding, cleaning, stemming, stopwords removal, filtering, tokenization, ditambahkan convert emoticon dan slangword.

Data yang telah melalui tahap preprocessing kemudian akan dilakukan perhitungan bobot TF-IDF, data yg sudah diberi bobot data akan diklasifikasikan menjadi kelas sentimen positif dan negative menggunakan metode Support Vector machine. Kemudian dilakukan evaluasi dan validasi menggunakan Confusion Matriks dan K-fold cross validation.

Hasil dari penelitian yang sudah dilakukan menunjukkan sentimen terhadap PSSI cenderung negatif, dan perhitungan akurasi menunjukkan bahwa model 1 lebih unggul dibanding dengan mendapat tingkat akurasi 88.70% menggunakan perhitungan confusion matrix dan 88.29% menggunakan perhitungan k-fold Cross Validation, sedangkan model 2 mendapat akurasi 86.65% menggunakan perhitungan confusion matriks dan 87.59% menggunakan perhitungan k-fold Cross Validation

**Kata Kunci:** sentimen analisis, preprocessing, klasifikasi, support vector machine, k-fold cross validation Confusion Matriks, K-fold cross validation.



## **ABSTRACT**

*In this study, researchers conducted a sentiment analysis of PSSI in managing Indonesian soccer. This research will also compare the accuracy between 2 data preprocessing models. in model 1 using case folding, cleaning, stemming, stopwords removal, filtering, tokenization. Whereas the model 2 uses case folding, cleaning, stemming, stopwords removal, filtering, tokenization. added convert emoticon and slang words.*

*Data that has been through the preprocessing stage will then be calculated TF-IDF weights, data that have been given data weights will be classified into positive and negative sentiment classes using the Support Vector machine method. Then evaluate and validate using Confusion Matrix and K-fold cross-validation.*

*The results of the research that have been done show that the sentiment towards PSSI tends to be negative, and the accuracy calculation shows that model 1 is superior compared to the accuracy level of 88.70% using the confusion matrix calculation and 88.29% using the k-fold Cross Validation calculation, while model 2 gets an accuracy of 86.65 % uses the confusion matrix calculation and 87.59% uses the k-fold Cross Validation calculation..*

**Keyword:** *Sentiment analysis, preprocessing, classification, support vector machine, k-fold cross validation, Confusion Matrix.*