

**AKURASI ANALISIS SENTIMEN PADA ALGORITMA SUPPORT
VECTOR MACHINE DAN RANDOM FOREST**

(Study Kasus : Akun pln_123)

SKRIPSI



disusun oleh

Ibed David Unedo Sihaloho

15.11.9293

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2020**

**AKURASI ANALISIS SENTIMEN PADA ALGORITMA SUPPORT
VECTOR MACHINE DAN RANDOM FOREST**

(Study Kasus : Akun pln_123)

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Progam Studi Informatika



disusun oleh

Ibed David Unedo Sihaloho

15.11.9293

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2020**

PERSETUJUAN

SKRIPSI

**AKURASI ANALISIS SENTIMEN PADA ALGORITMA SUPPORT
VECTOR MACHINE DAN RANDOM FOREST**

(Study Kasus : Akun pln_123)

yang dipersiapkan dan disusun oleh

Ibed David Unedo Sihaloho

15.11.9293

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 2 Maret 2020

Dosen Pembimbing,

Mardhiya Hayaty, S.T., M.Kom.

NIK. 190302105

PENGESAHAN

SKRIPSI

**AKURASI ANALISIS SENTIMEN PADA ALGORITMA SUPPORT
VECTOR MACHINE DAN RANDOM FOREST**

(Study Kasus : Akun pln_123)

yang dipersiapkan dan disusun oleh

Ibed David Unedo Sihaloho

15.11.9293

telah dipertahankan di depan Dewan Penguji

pada tanggal 21 Februari 2020

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Anggit Dwi Hartanto, M.Kom.
NIK. 190302163

Sumarni Adi, S.Kom, M.Cs.
NIK. 190302256

Mardhiya Hayaty, S.T., M.Kom.
NIK. 190302108

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 2 Maret 2020

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,

Meterai
Rp. 6.000

Ibed David Unedo Sihaloho

NIM. 15.11.9293

PERSEMBAHAN

Segala puji syukur atas kehadiran Tuhan YME 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 orang tua saya yang selalu berdoa dan mendukung saya setiap saat tanpa kenal lelah.

Untuk Ibu Mardhiya, yang sudah dengan sabar membimbing saya. Saya ucapkan terimakasih sedalam-dalamnya.

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

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

MOTTO

” The mind shall vanquish the sword ”

- Sima Yi from Dynasty Warriors

” Life is unpredictable, not everything is under control. But as long as we're with the right people we can handle anything ”

- Amy Santiago from Brooklyn Nine-Nine

” You gotta keep on keeping on even with the feeling that you're gonna keep losing. You gotta come back strong ”

- Travi McCoy ft. Bendon Urie on Keep On Keeping On

” No one should be in pain, everyone should love themselves ”

- Gerard Way

KATA PENGANTAR

Puji syukur penulis haturkan atas kehadiran Tuhan Yang Maha Esa yang telah memberikan berkat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “ Akurasi Analisis Sentimen Pada Algoritma Support Vector Machine Dan Random Forest (Study Kasus : Akun Pln_123)” dengan sebaik-baiknya.

Dengan selesainya skripsi ini, penulis mengucapkan banyak terima kasih 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 Mardhiya Hayaty, S.T., 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 saudara penulis yang telah memberikan dukungan terbaiknya selama berkuliah.
6. Makhluk-makhluk Nongki yang sudah menemani dan bersenang-senang bersama penulis selama berkuliah.
7. Sahabat-sahabat penulis yang selalu memberikan dukungan terbaik selama selama berkuliah.
8. Para pembakar MMR yang sudah menemani dan bersenang-senang bersama penulis selama berkuliah.

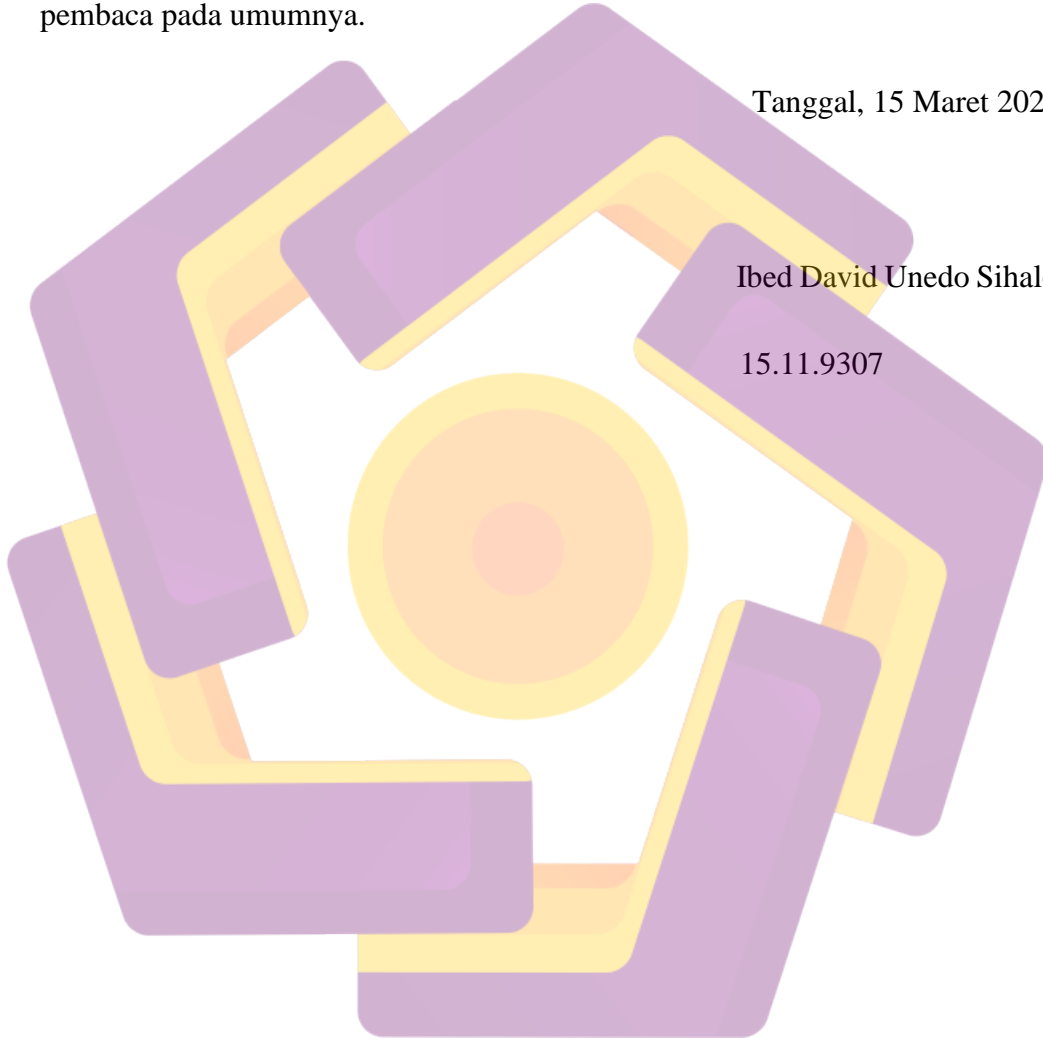
Penulis menyadari dengan betul bahawa dalam pembuatan skripsi ini masih banyak sekali kekurangan dan kelemahan. Oleh karena itu penulis berharap untuk semua pihak yang telah membaca dan memahami penelitian ini untuk dapat menyampaikan kritik dan saran yang membangun sehingga dapat menambah kesempurnaan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

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

Tanggal, 15 Maret 2020

Ibed David Unedo Sihaloho

15.11.9307

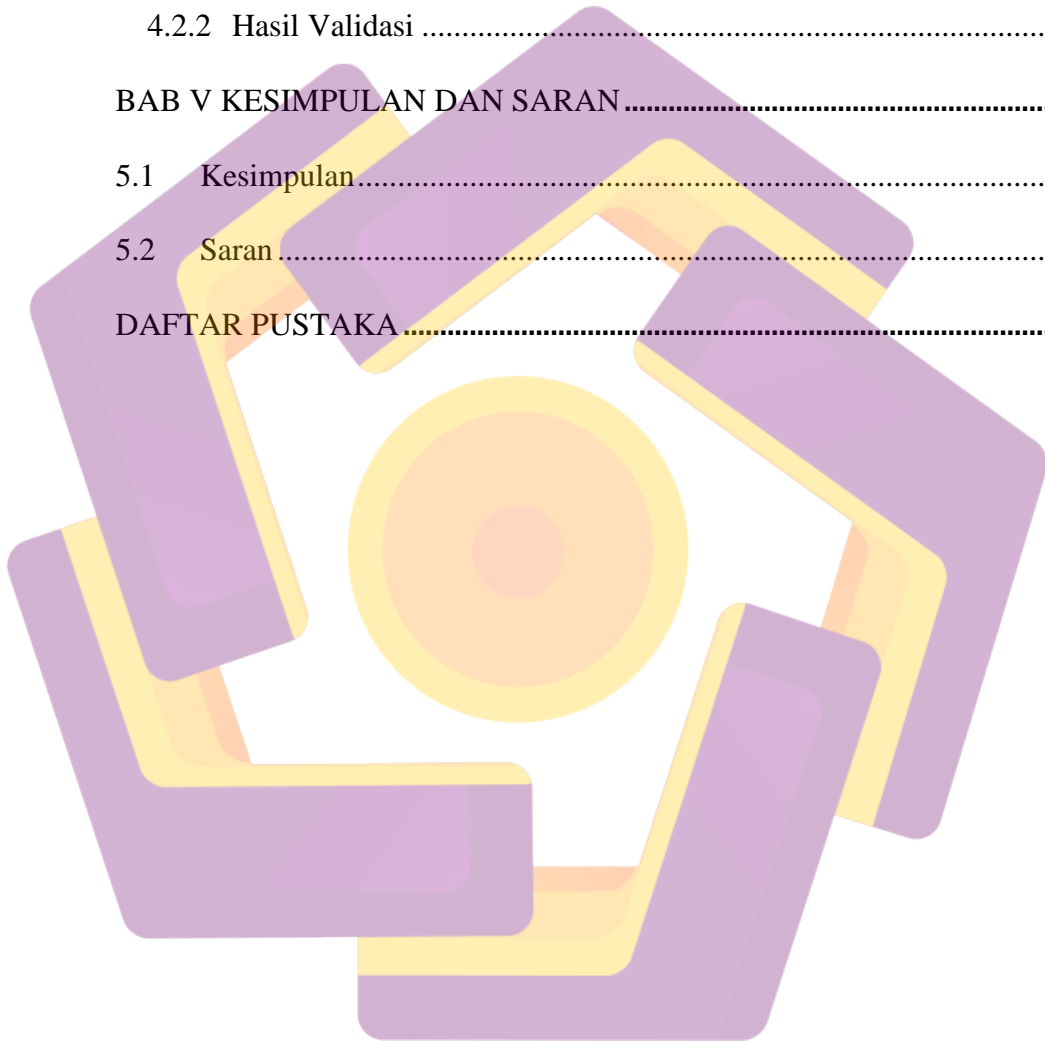


DAFTAR ISI

JUDUL.....	I
PERSETUJUAN.....	III
PENGESAHAN.....	IV
DAFTAR ISI.....	V
DAFTAR GAMBAR.....	XIII
DAFTAR TABEL.....	XIV
INTISARI.....	XV
ABSTRACT.....	XVI
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Metodologi Penelitian.....	4
1.6.1 Metode Pengumpulan Data.....	4
1.6.2 Metode Analisis.....	4
1.6.3 Metode Perancangan.....	4
1.6.4 Metode Implementasi.....	5
1.6.5 Metode Testing.....	5
1.7 Sistematika Penulisan.....	5

BAB II LANDASAN TEORI.....	7
2.1 Tinjauan Pustaka	7
2.2 Dasar Teori	9
2.2.1 Data Mining	9
2.2.2 Supervised Learning	9
2.2.3 Text Mining	10
2.2.4 Sentimen Analisis	11
2.2.5 Pre-Processing	11
2.2.6 Klasifikasi	12
2.2.7 TF-IDF	13
2.2.8 Support Vector Machine	13
2.2.9 Random Forest	19
2.2.10 Validasi dan Evaluasi	21
BAB III METODE PENELITIAN.....	24
3.1 Alat dan Bahan	24
3.1.1 Alat Penelitian	24
3.1.2 Bahan	24
3.2 Alur Penelitian	24
BAB IV HASIL DAN PEMBAHASAN.....	27
4.1 Eksperimen	27
4.1.1 Pengumpulan Data	27
4.1.2 Pre-Processing Data	29
4.1.3 Pembagian Data	37
4.1.4 Pembobotan	38

4.1.5 Pelatihan dan Pengujian	40
4.1.6 Evaluasi	41
4.1.7 Validasi.....	44
4.2 Hasil dan Pembahasan	45
4.2.1 Hasil Evaluasi	45
4.2.2 Hasil Validasi	47
BAB V KESIMPULAN DAN SARAN.....	51
5.1 Kesimpulan.....	51
5.2 Saran.....	51
DAFTAR PUSTAKA.....	53



DAFTAR GAMBAR

Gambar 2.1 (A) Pencarian Hyperplane, (B) Hyperlane Terbaik	14
Gambar 2.2 Fungsi Φ Memetakan Data Ke Ruang Vector	18
Gambar 2.3 Ilustrasi <i>K-Fold Cross Validation</i>	22
Gambar 3.1 Alur Penelitian	25
Gambar 4.1 Instalasi <i>GetOldtweets3</i>	27
Gambar 4.2 Pengambilan Data	28
Gambar 4.3 <i>Import Package</i> Dan File <i>.Csv</i>	29
Gambar 4.4 Implementasi <i>Casefolding</i>	30
Gambar 4.5 Hasil <i>Casefolding</i>	30
Gambar 4.6 <i>Cleaning Data</i>	31
Gambar 4.7 Hasil <i>Cleaning</i>	32
Gambar 4.8 <i>Slangword Removal</i>	33
Gambar 4.9 Hasil <i>Slangword Removal</i>	34
Gambar 4.10 <i>Stemming</i>	34
Gambar 4.11 Hasil <i>Stemming</i>	35
Gambar 4.12 <i>Stopwords Removal</i>	35
Gambar 4.13 Hasil <i>Stopwords</i>	36
Gambar 4.14 <i>Tokenizing</i>	36
Gambar 4.15 Hasil <i>Tokenizing</i>	37
Gambar 4.16 Pembagian Data	37
Gambar 4.17 <i>TF-IDF</i>	38
Gambar 4.18 Hasil Implementasi <i>TF-IDF</i>	39
Gambar 4.19 Implementasi <i>SVM</i>	40
Gambar 4.20 Implementasi <i>Random Forest</i>	41
Gambar 4.21 <i>Confusion Matrix</i>	42
Gambar 4.22 Hasil <i>Confusion Matrix SVM</i>	43
Gambar 4.23 Validasi Dengan <i>K-Fold</i> , $K=10$	45

DAFTAR TABEL

TABEL 2.1 PERBANDINGAN PENELITIAN.....	8
TABEL 2.2 CONFUSION MATRIX	23
TABEL 4.1 PENGUJIAN AKURASI <i>GINI IMPURITY</i>	43
TABEL 4.2 PERBANDINGAN CONFUSION MATRIX.....	46
TABEL 4.3 HASIL AKURASI SVM DENGAN VALIDASI	47
TABEL 4.4 HASIL AKURASI RANDOM FOREST DENGAN VALIDASI	48



INTISARI

Analisis sentimen adalah salah satu teknik yang paling aktif dan banyak digunakan dalam *Machine Learning*. Bidang ini melakukan studi mengenai opini orang-orang, sentimen, evaluasi, tingkah laku dan emosi terhadap suatu entitas seperti produk, layanan, organisasi, individu, permasalahan, topik, acara dan sebagainya.

Tujuan dasarnya adalah untuk mengklasifikasikan polaritas dokumen, kalimat atau fitur yang diberikan. Penelitian melakukan analisis Sentimen pada dataset akun pln_123 dengan Twitter feed untuk meneliti dua pengklasifikasi: Support Vector Machine (SVM) dan Random Forest, dan menentukan yang paling akurat berdasarkan pada hasil yang diperoleh untuk tweet polaritas positif dan negatif.

Hasil pada perhitungan akurasi menunjukkan bahwa metode Random Forest mendapatkan akurasi yang lebih tinggi dibandingkan metode klasifikasi Support Vector Machine, tanpa validasi K-Fold Cross Validation, Random Forest mendapatkan hasil sebesar akurasi 71.19% , sedangkan menggunakan validasi K-Fold Cross Validation mendapatkan hasil sebesar 71.95%.

Kata Kunci : *machine learning*, sentimen, klasifikasi, support vector machine, random forest

ABSTRACT

Sentiment analysis is one of the most active and widely used techniques in Machine Learning. This field conducts studies on people's opinions, sentiments, evaluations, behavior and emotions towards an entity such as products, services, organizations, individuals, problems, topics, events and others.

The basic purpose is to classify the polarity of a given document, sentence or feature. Researchers conduct sentiment analysis on datasets pln_123 account to examine Twitter feeds two classifiers: Support Vector Machine (SVM) and Random Forest, and hence determine the most accurate based on the results obtained to tweet positive and negative polarity.

The results on the accuracy calculation show that the Random Forest method gets higher accuracy than the Support Vector Machine classification method, without K-Fold Cross Validation, Random Forest gets the results of an accuracy of 71.19%, while using K-Fold Cross Validation gets a result of 71.95 %.

Keywords : machine learning, sentiment, classification, support vector machine, random forest