

**ANALISIS SENTIMEN PADA APLIKASI BIBIT DENGAN
ALGORITMA SUPPORT VECTOR MACHINE
MENGUNAKAN TEKNIK BOOTSTRAPPING**



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

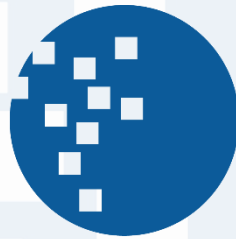
Wilson Manuel

0000041925

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2023

**ANALISIS SENTIMEN PADA APLIKASI BIBIT DENGAN
ALGORITMA SUPPORT VECTOR MACHINE
MENGUNAKAN TEKNIK BOOTSTRAPPING**



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Komputer (S.Kom)

Wilson Manuel

0000041925

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNIK DAN INFORMATIKA

UNIVERSITAS MULTIMEDIA NUSANTARA

TANGERANG

2023

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Wilson Manuel
Nomor Induk Mahasiswa : 00000041925
Program studi : Sistem Informasi

Skripsi dengan judul:

Analisis Sentimen pada Aplikasi Bibit dengan Algoritma Support Vector Machine Menggunakan Teknik Bootstrapping

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas Akhir yang telah saya tempuh.

Tangerang, 16 Juni 2023



Wilson Manuel

UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PERSETUJUAN

Skripsi dengan judul

*Analisis Sentimen pada Aplikasi Bibit dengan Algoritma Support Vector Machine
Menggunakan Teknik Bootstrapping*

Oleh

Nama : Wilson Manuel
NIM : 00000041925
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

Telah disetujui untuk diajukan pada
Sidang Ujian Skripsi Universitas Multimedia Nusantara

Tangerang, 16 Juni 2023

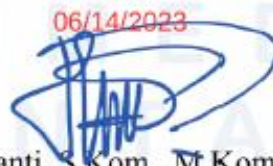
Pembimbing



Dr. Irmawati, S.Kom., MMSI.
0805097703/ 081431

Ketua Program Studi Sistem Informasi

06/14/2023



Ririn Ikana Desanti, S.Kom., M.Kom.

HALAMAN PENGESAHAN

Skripsi dengan judul

*Analisis Sentimen pada Aplikasi Bibit dengan Algoritma Support Vector Machine
Menggunakan Teknik Bootstrapping*

Oleh

Nama : Wilson Manuel
NIM : 00000041925
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

Telah diujikan pada hari Rabu, 21 Juni 2023
Pukul 10.00 s.d 12.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut.

Ketua Sidang



27/6/2023

Ir. Raymond Sunardi Octama, MCIS.
0328046803/023897

Penguji



26/06 2023

Jansen Wiratama, S.Kom., M.Kom.
0409019301/074887

Pembimbing



27/6 '23

Dr. Irmawati, S.Kom., MMSI.
0805097703/ 081431

Ketua Program Studi Sistem Informasi



27/6 '23

Ririn Ikana Desanti, S.Kom., M.Kom.
0313058001/051314

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas academica Universitas Multimedia Nusantara, saya yang bertanda tangan di bawah ini:

Nama : Wilson Manuel
NIM : 00000041925
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika
Jenis Karya : Tesis/Skripsi/Tugas Akhir

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Multimedia Nusantara Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul.

Analisis Sentimen pada Aplikasi Bibit dengan Algoritma *Support Vector Machine* Menggunakan Teknik *Bootstrapping*

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini, Universitas Multimedia Nusantara berhak menyimpan, mengalihmediakan/mengalihformatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 16 Juni 2023

Yang menyatakan,



Wilson Manuel

KATA PENGANTAR

Puji Syukur kepada Tuhan Yang Maha Esa atas kasih dan penyertaannya yang memungkinkan skripsi dengan judul “Analisis Sentimen pada Aplikasi Bibit dengan Algoritma *Support Vector Machine* Menggunakan Teknik Bootstrapping” terselesaikan tepat waktu. Skripsi ini dibuat untuk memenuhi salah satu syarat dalam mencapai gelar Sarjana (S1) Jurusan Sistem Informasi Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Penulis menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan tugas akhir ini, sangatlah sulit bagi penulis untuk menyelesaikan tugas akhir ini. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Dr. Ninok Leksono, M.A., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Ibu Dr. Irmawati, S.Kom., MMSI., sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan baik secara material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
6. Teman-teman seperjuangan yang telah memberikan motivasi dan semangat selama penulisan skripsi ini.

Penulis menyadari bahwa masih terdapat banyak kekurangan pada skripsi ini sehingga kritik dan saran dari para pembaca sangat diharapkan. Kritik dan saran tersebut akan dijadikan sebagai sarana penulis untuk mengintrospeksi diri agar penulis dapat memperbaikinya pada penulisan-penulisan karya ilmiah selanjutnya. Akhir kata, semoga skripsi ini dapat menjadi sumber informasi baru bagi para

pembaca serta menjadi acuan bagi penelitian-penelitian berikutnya terkait topik analisis sentimen.

Tangerang, 16 Juni 2023



Wilson Manuel



ANALISIS SENTIMEN PADA APLIKASI BIBIT DENGAN ALGORITMA SUPPORT VECTOR MACHINE MENGUNAKAN TEKNIK BOOTSTRAPPING

Wilson Manuel

ABSTRAK

Peristiwa *COVID-19* menyebabkan roda ekonomi dunia lumpuh, tidak terkecuali Indonesia. Masyarakat yang memiliki pekerjaan harus menghadapi Pemutusan Hubungan Kerja (PHK) dari perusahaan. Sejak saat itu, investasi menjadi sebuah aktivitas yang semakin populer dikalangan masyarakat Indonesia untuk mendapatkan penghasilan tambahan. Saat ini terdapat banyak penipuan investasi yang dapat merugikan para investor. Oleh karena itu, penelitian ini dilakukan untuk mengetahui mayoritas sentimen dalam bentuk ulasan pengguna pada salah satu aplikasi investasi yang terkenal di Indonesia yaitu Bibit. Pada penelitian ini, algoritma yang digunakan adalah *Naïve Bayes*, *Decision Tree*, *Support Vector Machine*, dan *Random Forest*. Berdasarkan keempat algoritma tersebut, model dengan performa terbaik akan dioptimalkan lebih lanjut dengan menggunakan teknik *bootstrapping* dan *GridSearchCV*. Hasil penelitian menyatakan bahwa sentimen mayoritas ulasan masyarakat terhadap aplikasi Bibit adalah positif. Model dengan performa terbaik adalah *Support Vector Machine* dengan akurasi pada data *testing* sebesar 85%. Setelah dilakukan teknik *bootstrapping*, akurasi model meningkat menjadi 89,57%. Berdasarkan hasil tersebut, performa terbaik diperoleh algoritma *Support Vector Machine* dengan teknik *bootstrapping*.

Kata kunci: *Bootstrapping*, Sentimen, *Support Vector Machine*, Ulasan

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

SENTIMENT ANALYSIS OF BIBIT APP WITH SUPPORT VECTOR MACHINE USING BOOTSTRAPPING TECHNIQUES

Wilson Manuel

ABSTRACT (English)

COVID-19 event has paralyzed the world economy, including Indonesia. People who usually have jobs now have to face layoffs from the company. Since then, investing has become an increasingly popular activity among Indonesians to earn additional income. However, in fact there are currently many investment fraud actions that can harm investors. Therefore, this research was conducted to find out the majority of sentiments in the form of user reviews on one of the well-known investment applications in Indonesia, namely Bibit. 4 algorithms that used in this research are Naïve Bayes, Decision Tree, Support Vector Machine, and Random Forest. Based on these four algorithms, the model with the best performance will be further optimized with bootstrapping and GridSearchCV. The results of the study stated that the sentiment of the majority of public reviews of Bibit app was positive. The model with the best performance is the Support Vector Machine with an accuracy of 85% on testing data. After the bootstrapping technique was carried out, the model accuracy increased to 89.57%. Based on these results, the best performance is obtained by the Support Vector Machine algorithm with the bootstrapping technique.

Kata kunci: *Bootstrapping, Reviews, Sentiment, Support Vector Machine*



DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	v
KATA PENGANTAR	vi
ABSTRAK	viii
<i>ABSTRACT (English)</i>	ix
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR RUMUS	xvii
DAFTAR LAMPIRAN	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Batasan Masalah	6
1.4 Tujuan dan Manfaat Penelitian	6
1.4.1 Tujuan Penelitian	6
1.4.2 Manfaat Penelitian	7
1.5 Sistematika Penulisan	7
BAB II LANDASAN TEORI	9
2.1 Tinjauan Teori	9
2.1.1 Analisis Sentimen	9
2.1.2 Text Mining	10
2.2 Framework, Algoritma, dan Teknik dalam Penelitian	10
2.2.1 Decision Tree	10
2.2.2 Naïve Bayes	11
2.2.3 Support Vector Machine	12
2.2.4 Random Forest	13

2.2.5	<i>Bootstrapping</i>	14
2.2.6	<i>TF-IDF</i>	14
2.2.7	<i>Confusion Matrix</i>	14
2.2.8	CRISP-DM	16
2.3	Tools yang digunakan	18
2.3.1	Bibit	18
2.3.2	Google Play Store	19
2.3.3	Google Colaboratory	20
2.3.4	Visual Studio Code	20
2.3.5	Excel	21
2.4	Penelitian Terdahulu	22
BAB III METODOLOGI PENELITIAN		28
3.1	Gambaran Umum Objek Penelitian	28
3.2	Metode Penelitian	28
3.2.1	Alur Penelitian	28
3.2.2	Metode Pengembangan Sistem / Metode Data Mining	30
3.3	Teknik Pengumpulan Data	32
3.3.1	Populasi dan Sampel	32
3.3.2	Periode Pengambilan Data	33
3.4	Variabel Penelitian	33
3.5	Teknik Analisis Data	33
BAB IV ANALISIS DAN HASIL PENELITIAN		35
4.1	<i>Business Understanding</i>	35
4.1.1	Proses Bisnis	35
4.1.2	Kebutuhan Sistem	36
4.2	<i>Data Understanding</i>	36
4.3	<i>Data Preparation</i>	38
4.3.1	<i>Labeling</i>	38
4.3.2	<i>Data cleaning</i>	40
4.3.3	<i>Data pre-processing</i>	41
4.4	<i>Modeling</i>	46
4.4.1	Tanpa Optimasi	47

4.4.2	Optimasi	61
4.5	<i>Evaluation</i>	70
4.6	<i>Deployment</i>	73
4.7	Hasil dan Pembahasan	76
BAB V SIMPULAN DAN SARAN		79
5.1	Simpulan	79
5.2	Saran	80
DAFTAR PUSTAKA		81
LAMPIRAN		85



DAFTAR TABEL

Tabel 2.1 Penelitian terdahulu.....	22
Tabel 4.1 Perbandingan performa algoritma pada data training	62
Tabel 4.2 Perbandingan performa algoritma pada data validation	62
Tabel 4.3 Perbandingan performa algoritma pada data testing.....	62
Tabel 4.4 Perbandingan performa ketiga model SVM.....	71
Tabel 4.5 Perbandingan TP, TN, FP, dan FN pada confusion matrix ketiga model SVM	73
Tabel 4.6 Perbandingan performa algoritma SVM pada penelitian-penelitian lain	77
Tabel 4.7 Perbandingan akurasi model SVM pada penelitian sebelumnya.....	77
Tabel 4.8 Perbandingan akurasi model SVM pada penelitian ini dengan training 70% dan testing 30%.....	77



DAFTAR GAMBAR

Gambar 2.1 Kategori analisis sentimen	9
Gambar 2.2 Bentuk Decision Tree (DT).....	11
Gambar 2.3 Cara kerja algoritma Support Vector Machine (SVM)	12
Gambar 2.4 Algoritma Decision Tree (DT).....	13
Gambar 2.5 Tabel confusion matrix.....	15
Gambar 2.6 Alur metode CRISP-DM.....	17
Gambar 2.7 Logo Bibit	18
Gambar 2.8 Logo Google Play.....	19
Gambar 2.9 Logo Google Colaboratory.....	20
Gambar 2.10 Logo Visual Studio Code	20
Gambar 2.11 Logo Microsoft Excel	21
Gambar 3.1 Diagram alur penelitian	28
Gambar 4.1 Proses pengambilan data ulasan Google Play Store.....	36
Gambar 4.2 Penyaringan data dari tahun 2020 sampai 2022.....	37
Gambar 4.3 Fitur head pada Python untuk menampilkan 5 data teratas.....	38
Gambar 4.4 Hasil file CSV data ulasan aplikasi Bibit yang diambil	38
Gambar 4.5 Pemberian label pada data ulasan.....	39
Gambar 4.6 Penghapusan ulasan dengan label netral	39
Gambar 4.7 Grafik perkembangan ulasan aplikasi Bibit dari tahun 2020-2023...	40
Gambar 4.8 Proses data cleaning	40
Gambar 4.9 Proses mengunduh package NLTK.....	41
Gambar 4.10 Tokenizing.....	42
Gambar 4.11 Stop words Bahasa Indonesia dari package NLTK.....	42
Gambar 4.12 Penambahan kata pada stop words.....	42
Gambar 4.13 Pengaplikasian stop words pada dataset.....	42
Gambar 4.14 Stemming pada dataset.....	43
Gambar 4.15 Penghapusan data ulasan dengan list kosong	43
Gambar 4.16 Pengaturan ulang indeks data dengan reset_index	44
Gambar 4.17 Fungsi info() untuk melihat struktur dan jumlah data.....	44
Gambar 4.18 Wordcloud ulasan positif (kiri) dan negatif (kanan)	45
Gambar 4.19 Grafik distribusi sentimen (kiri) dan score (kanan)	46
Gambar 4.20 Modul metrics	46
Gambar 4.21 Proses train_test_split.....	47
Gambar 4.22 Fungsi value_counts untuk melihat jumlah data	48
Gambar 4.23 Distribusi data testing.....	49
Gambar 4.24 Jumlah ulasan yang akan digunakan pada data testing	49
Gambar 4.25 Aplikasi TF-IDF pada data training dan validation	49
Gambar 4.26 Algoritma NB dan akurasinya.....	50
Gambar 4.27 Hasil classification_report algoritma NB.....	51
Gambar 4.28 Confusion matrix algoritma NB pada data validation.....	51

Gambar 4.29 Performa algoritma NB pada data testing	52
Gambar 4.30 Confusion matrix algoritma NB pada data testing	52
Gambar 4.31 Algoritma DT dan akurasinya	53
Gambar 4.32 Classification_report algoritma DT.....	54
Gambar 4.33 Confusion matrix algoritma DT pada data validation	54
Gambar 4.34 Performa algoritma DT pada data testing	55
Gambar 4.35 Confusion matrix algoritma DT pada data testing	55
Gambar 4.36 Algoritma SVM dan akurasinya.....	56
Gambar 4.37 Classification_report algoritma SVM	56
Gambar 4.38 Confusion matrix algoritma SVM pada data validation.....	57
Gambar 4.39 Performa algoritma SVM pada data testing	58
Gambar 4.40 Confusion matrix algoritma SVM pada data testing.....	58
Gambar 4. 41 Algoritma RF dan akurasinya	59
Gambar 4.42 Classification_report algoritma RF	60
Gambar 4.43 Confusion matrix algoritma RF pada data validation	60
Gambar 4.44 Performa algoritma RF pada dataset testing	61
Gambar 4.45 Confusion matrix algoritma RF pada dataset testing	61
Gambar 4.46 Sintaks bootstrapping	64
Gambar 4.47 Jumlah data setelah bootstrapping.....	64
Gambar 4.48 Proses TF-IDF pada data yang akan dioptimalkan.....	65
Gambar 4.49 Proses training dan validation algoritma SVM	65
Gambar 4.50 Performa algoritma SVM dengan bootstrapping pada data validation	66
Gambar 4.51 Confusion matrix algoritma SVM dengan bootstrapping pada data validation.....	66
Gambar 4.52 Hasil performa algoritma SVM dengan teknik bootstrapping pada data testing	66
Gambar 4.53 Confusion matrix algoritma SVM dengan teknik bootstrapping pada data testing	67
Gambar 4.54 Sintaks GridSearchCV pada algoritma Support Vector Machine...	68
Gambar 4.55 Parameter dan akurasi training terbaik pada GridSearchCV	68
Gambar 4.56 Akurasi model SVM dengan optimasi GridSearchCV pada data validation.....	69
Gambar 4.57 Performa model SVM dengan optimasi pada data validation.....	69
Gambar 4.58 Confusion matrix model SVM dengan optimasi pada data validation	69
Gambar 4.59 Performa algoritma SVM dengan optimasi GridSearchCV pada data testing	70
Gambar 4.60 Confusion matrix algoritma SVM dengan optimasi GridSearchCV pada data testing	70
Gambar 4.61 Perbandingan performa model SVM (atas), bootstrap (tengah), dan bootstrap dengan optimasi (bawah)	72

Gambar 4.62 Perbandingan confusion matrix model SVM (kiri), bootstrap (tengah), dan bootstrap dengan optimasi (kanan)	72
Gambar 4.63 Sintaks untuk mengunduh file model menggunakan pickle.....	74
Gambar 4.64 Sintaks untuk mengunduh file hasil proses TF-IDF.....	74
Gambar 4.65 Sintaks untuk menjalankan streamlit di localhost	75
Gambar 4.66 Tampilan website	75
Gambar 4.67 Tampilan website pada sentimen negatif	76
Gambar 4.68 Tampilan website pada sentimen positif	76



DAFTAR RUMUS

Rumus 2.1 Rumus algoritma Naïve Bayes (NB)	12
Rumus 2.2 Precision	15
Rumus 2.3 Recall	16
Rumus 2.4 F1-score	16
Rumus 2.5 Accuracy	16



DAFTAR LAMPIRAN

Lampiran A Hasil Turnitin.....	85
Lampiran B Formulir Konsultasi Skripsi.....	96

