



**KLASIFIKASI SENTIMEN ULASAN PENGGUNA MENGGUNAKAN
METODE *SUPPORT VECTOR MACHINE* PADA APLIKASI BIBIT DAN
BAREKSA**

SKRIPSI

DINDA AULIA SETIANINGSIH

2010512035

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA
2024**



**KLASIFIKASI SENTIMEN ULASAN PENGGUNA MENGGUNAKAN
METODE *SUPPORT VECTOR MACHINE* PADA APLIKASI BIBIT DAN
BAREKSA**

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana
Komputer**

DINDA AULIA SETIANINGSIH

2010512035

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA
2024**

PERNYATAAN ORISINALITAS

Skripsi ini merupakan hasil karya sendiri serta semua sumber referensi yang dikutip maupun yang dirujuk telah saya nyatakan benar.

Nama : Dinda Aulia Setianingsih

NIM : 2010512035

Tanggal : 19 April 2024

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan berlaku.

Jakarta, 19 April 2024

Yang menyatakan,



Dinda Aulia Setianingsih

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI
UNTUK KEPENTINGAN AKADEMIS**

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Dinda Aulia Setianingsih

NIM : 2010512035

Fakultas : Ilmu Komputer

Program Studi : S1 Sistem Informasi

Demi pembangunan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Eksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

**KLASIFIKASI SENTIMEN ULASAN PENGGUNA MENGGUNAKAN
METODE *SUPPORT VECTOR MACHINE* PADA APLIKASI BIBIT DAN
BAREKSA**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sesungguhnya.

Dibuat di : Jakarta

Pada Tanggal : 19 April 2024

Yang menyatakan,



Dinda Aulia Setianingsih

LEMBAR PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Dinda Aulia Setianingsih

NIM : 2010512035

Program Studi : SI Sistem Informasi

Judul : **KLASIFIKASI SENTIMEN ULASAN PENGGUNA
MENGUNAKAN METODE *SUPPORT VECTOR MACHINE*
PADA APLIKASI BIBIT DAN BAREKSA**

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



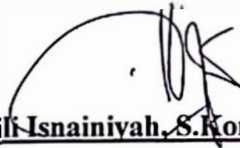
Nur Hafifah Matondang, S.Kom., MM., M.T.I.

Penguji I



Sarika, M.Kom.

Penguji II



Ika Nurlaili Isnainiyah, S.Kom., M. Sc.

Pembimbing I



Nindy Irzavika, S.SI., M.T.

Pembimbing II



Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM.

Dekan



Anita Muliawati, S.Kom., M.TI.

Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 5 Juni 2024

KLASIFIKASI SENTIMEN ULASAN PENGGUNA MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE* PADA APLIKASI BIBIT DAN BAREKSA

Dinda Aulia Setianingsih

ABSTRAK

Pesatnya perkembangan teknologi memberikan dampak pada industri, salah satunya industri keuangan. Pembelian reksa dana, saham, dan aset investasi lainnya yang dulu hanya bisa dilakukan secara konvensional kini dapat dilakukan dengan mudah melalui aplikasi dalam telepon pintar. Aplikasi bibit dan bareksa merupakan aplikasi yang menyediakan layanan pembelian reksa dana dan produk investasi lainnya. Kemudahan yang diberikan oleh aplikasi investasi ini menarik minat banyak kalangan, salah satunya anak muda. Aplikasi bibit dan bareksa banyak menerima kritik dan masukan dari para pengguna. Penelitian ini dilakukan untuk mengetahui lebih dalam terkait ulasan yang diberikan oleh pengguna terhadap aplikasi melalui analisis sentimen juga menghasilkan *website* sederhana dengan *framework flask* yang dapat digunakan untuk mengklasifikasikan ulasan dari input pengguna. Algoritma *support vector machine* digunakan untuk melakukan klasifikasi sentimen positif dan negatif, proses klasifikasi dilakukan dengan dan tanpa seleksi fitur *chi-square* untuk mendapatkan nilai akurasi terbaik, dengan menggunakan kernel linear. Pada aplikasi bibit menggunakan *chi-square* didapatkan nilai akurasi 92%, presisi 94%, *recall* 89%, dan *f1-score* 91%. Sedangkan tanpa *chi-square* didapatkan nilai akurasi sebesar 91%, presisi 94%, *recall* 88%, dan *f1-score* 91%. Pada aplikasi bareksa menggunakan *chi-square* didapatkan akurasi sebesar 86%, presisi 85%, *recall* 87%, dan *f1-score* 86%. Sedangkan tanpa *chi-square* didapatkan akurasi 84%, presisi 83%, *recall* 85%, dan *f1-score* 84%. Hasil klasifikasi sentimen tersebut didapatkan bahwa penggunaan seleksi fitur *chi-square* dalam penelitian ini menghasilkan akurasi yang lebih baik dengan kenaikan 1% sampai 2%.

Kata Kunci: Analisis Sentimen, *Support vector machine*, *Chi-square*, Bibit, Bareksa

SENTIMENT CLASSIFICATION OF USER REVIEWS USING *SUPPORT VECTOR MACHINE* METHOD ON BIBIT AND BAREKSA APPLICATIONS

Dinda Aulia Setianingsih

ABSTRACT

The rapid advancement of technology has impacted various industries, including the financial sector. Purchasing mutual funds, stocks, and other investment assets, which were previously done conventionally, can now be easily accomplished through applications on smartphones. Applications such as Bibit and Bareksa provide services for purchasing mutual funds and other investment products. The convenience offered by these investment applications has attracted the interest of various demographics, including the younger generation. Bibit and Bareksa applications have received considerable criticism and feedback from users. This research aims to delve deeper into user reviews of these applications through sentiment analysis and to develop a simple website using the Flask framework that can classify user input reviews. The Support vector machine algorithm with a linear kernel is employed to classify positive and negative sentiments, with and without chi-square feature selection, to achieve the best accuracy. In the Bibit application, using chi-square yielded an accuracy of 92%, precision of 94%, recall of 89%, and f1-score of 91%. Without chi-square, the accuracy was 91%, precision was 94%, recall was 88%, and f1-score was 91%. For the Bareksa application, using chi-square resulted in an accuracy of 86%, precision of 85%, recall of 87%, and f1-score of 86%. Without chi-square, the accuracy was 84%, precision was 83%, recall was 85%, and f1-score was 84%. The classification results indicate that the utilization of chi-square feature selection in this study leads to improved accuracy, with an increase of 1% to 2%.

Keywords: *Sentiment Analysis, Support vector machine, Chi-square, Bibit, Bareksa*

KATA PENGANTAR

Puji syukur dipanjatkan kehadirat Allah SWT karena atas berkat dan rahmat-Nya penulis mampu menyelesaikan skripsi yang berjudul “Klasifikasi Sentimen Ulasan Pengguna Menggunakan Metode *Support vector machine* Pada Aplikasi Bibit dan Bareksa”. Penyelesaian skripsi ini juga tidak lepas dan bimbingan dan dorongan berbagai pihak. Untuk itu, penulis ingin mengucapkan banyak terima kasih kepada:

1. Prof. Dr. Ir. Supriyanto, Selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.
2. Ibu Ika Nurlaili, S.Kom, M.Sc. selaku dosen pembimbing I dan Ibu Nindy Irzavika, S.SI., M.T. selaku dosen pembimbing II atas bimbingannya selama proses pembuatan skripsi.
3. Ayah dan Mama atas doa dan dukungan yang diberikan kepada penulis.
4. Teman – teman satu perjuangan, terkhusus Adzra Sajida, Tsaabitah Anggraini, dan Berliana Septyani Suganda yang sudah memberikan dorongan dan berbagai keluh kesah selama proses penyusunan penelitian.
5. Rekan penulis, Wydia Ningrum, yang telah berkenan meluangkan waktu untuk membantu dalam proses anotasi data di tengah kesibukannya.
6. Serta seluruh pihak yang tidak bisa penulis sebutkan satu persatu tanpa mengurangi rasa hormat.

Jakarta, 20 April 2024

Dinda Aulia Setianingsih

DAFTAR ISI

PERNYATAAN ORISINALITAS	iii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS.....	iv
LEMBAR PENGESAHAN.....	v
ABSTRAK	vi
ABSTRACT	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Ruang Lingkup	4
1.5 Luaran Yang Diharapkan	5
1.6 Manfaat Penelitian	5
1.6.1 Bagi Aplikasi.....	5
1.6.2 Bagi Masyarakat.....	6
1.6.3 Bagi Peneliti	6
1.7 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA.....	8
2.1 Analisis Sentimen	8
2.2 Bibit	8
2.3 Bareksa	9
2.4 Reksa dana	10
2.5 Support vector machine	10
2.6 Pembobotan TF-IDF	11
2.7 Chi-square	11
2.8 Data Mining	12
2.9 Pelabelan Data	12

2.10 Preprocessing	13
2.10.1 <i>Case folding</i>	13
2.10.2 <i>Cleansing Data</i>	13
2.10.3 <i>Normalization</i>	13
2.10.4 <i>Stemming</i>	13
2.10.5 <i>Stopword removal</i>	14
2.10.6 <i>Tokenization</i>	14
2.11 Confusion Matrix	14
2.12 Python	16
2.13 Framework Flask	16
2.14 Black Box Testing	17
2.15 Penelitian Terkait	17
BAB III METODOLOGI PENELITIAN	22
3.1 Tahapan Penelitian	22
3.1.1 Identifikasi Masalah	23
3.1.2 Studi Pustaka.....	23
3.1.3 Pengumpulan Data	23
3.1.4 Pelabelan Data.....	24
3.1.5 Text Preprocessing	26
3.1.6 Pembobotan TF-IDF	29
3.1.7 Seleksi Fitur <i>Chi-square</i>	29
3.1.8 Klasifikasi <i>Support vector machine</i>	29
3.1.9 Evaluasi	30
3.1.10 Perancangan Website	30
3.1.11 Implementasi Website	30
3.1.12 Pengujian Black Box.....	30
3.2 Alat Bantu Penelitian	31
3.2.1 Perangkat Keras (Hardware)	31
3.2.2 Perangkat Lunak (Software).....	31
3.3 Jadwal Penelitian	31
BAB IV	34
PEMBAHASAN	34
4.1 Pengumpulan Data	34
4.2 Pelabelan Data	35

4.3	Data Preprocessing	38
4.3.1	<i>Case folding</i>	39
4.3.2	<i>Cleansing</i>	41
4.3.3	<i>Normalization</i>	43
4.3.4	<i>Stemming</i>	46
4.3.5	<i>Stopword removal</i>	48
4.3.6	<i>Tokenization</i>	50
4.4	Pembobotan TF-IDF	52
4.5	Seleksi Fitur <i>Chi-square</i>	57
4.6	Klasifikasi <i>Support vector machine</i>	61
4.6.1	Split Data.....	61
4.6.2	Pemodelan Tanpa Seleksi Fitur.....	62
4.6.3	Pemodelan Dengan Seleksi Fitur <i>Chi-square</i>	62
4.7	Evaluasi	64
4.7.1	Hasil Evaluasi Klasifikasi tanpa Seleksi Fitur <i>Chi-square</i>	64
4.7.2	Hasil Evaluasi dengan Seleksi Fitur <i>Chi-square</i>	66
4.7.3	Analisis Hasil Evaluasi.....	67
4.8	Visualisasi <i>Word cloud</i>	68
4.8.1	<i>Word cloud</i> Ulasan Aplikasi Bibit	68
4.8.2	<i>Word cloud</i> Ulasan Aplikasi Bareksa.....	71
4.8.3	Rekomendasi untuk Pengembang	74
4.9	Perancangan Website	75
4.10	Implementasi Website	76
4.11	Pengujian Black Box	77
BAB V.....		79
KESIMPULAN		79
5.1	Kesimpulan	79
5.2	Saran	80
DAFTAR PUSTAKA		81
DAFTAR RIWAYAT HIDUP.....		84
LAMPIRAN		85

DAFTAR GAMBAR

Gambar 3.1 Tahapan penelitian	22
Gambar 4.1 Syntax <i>scraping</i> bibit.....	34
Gambar 4.2 Data hasil <i>scraping</i> bibit.....	35
Gambar 4.3 Data hasil <i>scraping</i> bareksa.....	35
Gambar 4.4 Pie chart kiri persebaran sentimen bibit dan pie chart kanan persebaran sentimen bareksa.....	38
Gambar 4.5 <i>Word cloud</i> sentimen positif dataset bibit	69
Gambar 4.6 Bar chart frekuensi sentimen positif dataset bibit	69
Gambar 4.7 <i>Word cloud</i> sentimen negatif dataset bibit	70
Gambar 4.8 Bar chart frekuensi sentimen negatif dataset bibit	71
Gambar 4.9 <i>Word cloud</i> sentimen positif dataset bareksa	72
Gambar 4.10 Bar chart sentimen positif dataset bareksa	72
Gambar 4.11 <i>Word cloud</i> sentimen negatif dataset bareksa	73
Gambar 4.12 Bar chart frekuensi sentimen negatif dataset bareksa	73
Gambar 4.13 <i>Data flow diagram sentiment analyzer</i>	76
Gambar 4.14 Tampilan antarmuka website <i>sentiment analyzer</i>	76
Gambar 4.15 Tampilan <i>database</i> website <i>sentiment analyzer</i>	77

DAFTAR TABEL

Tabel 1.1. Perbandingan aplikasi bibit, ajaib, dan bareksa	2
Tabel 2.1 Confusion matrix.....	14
Tabel 2.2 Penelitian terdahulu.....	17
Tabel 3.1 Ulasan pada google play store.....	24
Tabel 3.2 Hasil Analisis Anotator	25
Tabel 3.3 Sebelum dan Sesudah <i>Case folding</i>	26
Tabel 3.4 Sebelum dan Sesudah <i>Cleansing</i>	27
Tabel 3.5 Sebelum dan Sesudah <i>Normalization</i>	27
Tabel 3.6 Sebelum dan Sesudah <i>Stemming</i>	28
Tabel 3.7 Sebelum dan Sesudah <i>Stopword removal</i>	28
Tabel 3.8 Sebelum dan Sesudah <i>Tokenization</i>	29
Tabel 3.9 Jadwal Penelitian.....	32
Tabel 4.1 Hasil anotasi aplikasi bibit	35
Tabel 4.2 Hasil anotasi aplikasi bareksa	37
Tabel 4.3 Implementasi <i>case folding</i> dataset bibit	39
Tabel 4.4 Implementasi <i>case folding</i> dataset bareksa	40
Tabel 4.5 Implementasi <i>cleansing</i> dataset bibit	41
Tabel 4.6 Implementasi <i>cleansing</i> dataset bareksa	42
Tabel 4.7 Implementasi normalisasi dataset bibit	44
Tabel 4.8 Implementasi normalisasi dataset bareksa	45
Tabel 4.9 Implementasi <i>stemming</i> dataset bibit	46
Tabel 4.10 Implementasi <i>stemming</i> dataset bareksa	47
Tabel 4.11 Implementasi <i>stopword</i> dataset bibit.....	48
Tabel 4.12 Implementasi <i>stopword</i> dataset bareksa.....	49
Tabel 4.13 Implementasi <i>tokenization</i> dataset bibit.....	50
Tabel 4.14 Implementasi <i>tokenization</i> dataset bareksa	51
Tabel 4.15 Contoh dokumen pembobotan <i>tfidf</i>	52
Tabel 4.16 Contoh perhitungan <i>tfidf</i>	53
Tabel 4.17 <i>Tfidf</i> setelah normalisasi	56
Tabel 4.18 Hasil penghitungan nilai <i>chi-square</i>	58
Tabel 4.19 Total feature bibit	60

Tabel 4.20 Total feature bareksa	61
Tabel 4.21 Hasil split data.....	61
Tabel 4.22 Paramater terbaik GridSearchCV.....	62
Tabel 4.23 Pemodelan dengan seleksi fitur <i>chi-square</i> dataset bibit	63
Tabel 4.24 Pemodelan dengan <i>chi-square</i> pada dataset bareksa	63
Tabel 4.25 Hasil confusion matrix bibit tanpa chi sqaure.....	64
Tabel 4.26 Hasil confusion matrix bareksa tanpa <i>chi-square</i>	65
Tabel 4.27 Hasil confusion matrix bibit dengan chi sqaure.....	66
Tabel 4.28 Hasil confusion matrix bareksa dengan <i>chi-square</i>	67
Tabel 4.29 Perbandingan hasil evaluasi aplikasi bibit	68
Tabel 4.30 Perbandingan hasil evaluasi aplikasi bareksa	68
Tabel 4.31 Test case black box testing.....	78