

**TUGAS AKHIR**

**PENERAPAN *LEVENSHTEIN DISTANCE* TERHADAP *SUPPORT VECTOR MACHINE (SVM)* DALAM ANALISIS SENTIMEN TEKS  
(STUDI KASUS : ULASAN APLIKASI DANA)**



Oleh

**ALVIAN ANANTA BAGASKARA**

**19 1065 1157**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS MUHAMMADIYAH JEMBER**

**2023**

## **TUGAS AKHIR**

### **PENERAPAN *LEVENSHTEIN DISTANCE* TERHADAP *SUPPORT VECTOR MACHINE (SVM)* DALAM ANALISIS SENTIMEN TEKS (STUDI KASUS : ULASAN APLIKASI DANA)**

Diajukan Untuk Melengkapi Tugas Akhir Dan Memenuhi  
Syarat Kelulusan Program Sastra 1 Jurusan Teknik Informatika  
Fakultas Teknik Universitas Muhammadiyah Jember



Oleh

**ALVIAN ANANTA BAGASKARA**

**19 1065 1157**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MUHAMMADIYAH JEMBER**

**2023**

**HALAMAN PERSETUJUAN SIDANG TUGAS AKHIR**

**PENERAPAN *LEVENSHTTEIN DISTANCE* TERHADAP *SUPPORT VECTOR MACHINE (SVM)* DALAM ANALISIS SENTIMEN TEKS  
(STUDI KASUS : ULASAN APLIKASI DANA)**

Oleh:

**Alvian Ananta Bagaskara**

**1910651157**

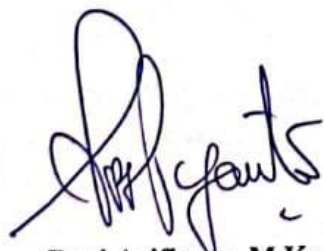
Telah mempertanggung jawabkan Laporan Tugas Akhirnya pada sidang Tugas Akhir tanggal 12 Agustus 2023 sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana Komputer (S.Kom)

di

Universitas Muhammadiyah Jember

Disetujui oleh,

**Pembimbing I**



**Deni Arifianto, M.Kom.**

**NIDN. 0718068103**

**Pembimbing II**



**Rosita Yanuarti S.kom., M. Cs**

**NIDN. 0629018601**

**LEMBAR PENGESAHAN**

**PENERAPAN *LEVENSHTEIN DISTANCE* TERHADAP *SUPPORT VECTOR MACHINE (SVM)* DALAM ANALISIS SENTIMEN TEKS  
(STUDI KASUS : ULASAN APLIKASI DANA)**

Oleh:

**Alvian Ananta Bagaskara**

**1910651157**

Telah mempertanggung jawabkan Laporan Tugas Akhirnya pada sidang Tugas Akhir tanggal 12 Agustus 2023 sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana Komputer (S.Kom)

di

Universitas Muhammadiyah Jember

Disetujui oleh,

Penguji I

Taufiq Timur W. S.Kom., M.Kom

NIDN. 0705078006

Penguji II

Habibatul Azizah Al Faruq M.Pd

NIDN. 0718128901

Mengesahkan,

Dekan Fakultas Teknik

Dr. Ir. Nanang Saiful Rizal, S.T., M.T., IPM

NIDN. 0014027501

Pembimbing I

Deni Arifianto M.Kom

NIDN. 0718068103

Pembimbing II

Rosita Yanuarti S.Kom., M. Cs

NIDN. 0629018601

Mengetahui,

Ketua Program Studi Teknik Informatika

Arifko Wardoyo, S.T., M.Kom

NIDN. 0014027501

**PENERAPAN *LEVENSHTTEIN DISTANCE* TERHADAP *SUPPORT VECTOR MACHINE (SVM)* DALAM ANALISIS SENTIMEN TEKS  
(STUDI KASUS : ULASAN APLIKASI DANA)**

**LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR**

Yang bertanda tangan di bawah in:

Nama : Alvian Ananta Bagaskara  
NIM : 1910651157  
Program Studi : Teknik Informatika  
Perguruan Tinggi : Universitas Muhammadiyah Jember

Menyatakan bahwa Tugas Akhir yang berjudul “Penerapan *Levenshtein Distance* Terhadap *Support Vector Machine (SVM)* Dalam Analisis Sentimen Teks (Studi Kasus : Ulasan Aplikasi Dana)” adalah benar dan belum pernah dibuat orang lain, kecuali yang diacu dalam Daftar Pustaka pada Tugas Akhir ini.

Demikian Surat Pernyataan ini dibuat dengan sebenar-benarnya dan apabila pernyataan ini tidak sesuai, penulis bersedia mendapat sanksi dari akademik.

Jember, 12 Agustus 2023



Alvian Ananta Bagaskara

**PENERAPAN *LEVENSHTEIN DISTANCE* TERHADAP *SUPPORT VECTOR MACHINE (SVM)* DALAM ANALISIS SENTIMEN TEKS  
(STUDI KASUS : ULASAN APLIKASI DANA)**

Alvian Ananta Bagaskara<sup>1</sup>, Deni Arifianto<sup>2</sup>, Rosita Yanuarti<sup>3</sup>

Program Studi Teknik Informatika, Fakultas Teknik

Universitas Muhammadiyah Jember

Email : alvianananta74@gmail.com

**ABSTRAK**

Google *Play store* merupakan sebuah aplikasi toko *online* penyedia aplikasi, buku maupun musik pada *android*. Google *Play store* memberikan kebebasan memberi komentar, masukan maupun kritikan oleh para pengguna kepada pihak *developer* aplikasi yang pengguna konsumsi. Namun ada beberapa ulasan yang tidak sesuai dikarenakan *rating* yang diberikan tidak sesuai dengan ulasan yang tertulis, yang menyebabkan beberapa kebingungan atau keambiguan tentang hal mana yang benar adanya. Maka dari itu dilakukanlah analisis sentimen menggunakan *Support Vector Machine (SVM)*. Namun masalah berikutnya adalah penggunaan bahasa yang bebas tak dapat terelakkan dalam pengulasan sebuah aplikasi. Banyak dari para pengguna yang menyelipkan kata-kata tidak baku, penyingkatan kata, dan kesalahan ejaan atau *typo* yang akhirnya lolos pada tahap *preprocessing text*. Maka dari itu dilakukan normalisasi kata menggunakan *Levenshtein distance*, sehingga dapat diketahui perbedaan hasil akurasi sebelum dan sesudah menerapkan normalisasi kata menggunakan *Levenshtein Distance*. Tahapan klasifikasi meliputi pengumpulan data sampel yang bersumber dari ulasan aplikasi Dana pada penelitian in, pelabelan data, *preprocessing text*, penerapan algoritma *Levenshtein distance*, TF-IDF, serta proses klasifikasi sentimen menggunakan *Support Vector Machine*. Hasil pengujian dengan menerapkan normalisasi kata menggunakan *Levenshtein Distance* menunjukkan peningkatan sebesar 2% yang awalnya memperoleh nilai 83% tanpa menggunakan *Levenshtein Distance* menjadi 85% setelah menerapkan *Levenshtein Distnsnce*.

**Kata Kunci** : Analisis Sentimen, *Support Vector Machine*, *Levenshtein distance*.

**PENERAPAN LEVENSHTTEIN DISTANCE TERHADAP SUPPORT  
VECTOR MACHINE (SVM)DALAM ANALISIS SENTIMEN TEKS  
(STUDI KASUS : ULASAN APLIKASI DANA)**

Alvian Ananta Bagaskara<sup>1</sup>, Deni Arifianto<sup>2</sup>, Rosita Yanuarti<sup>3</sup>

Program Studi Teknik Informatika, Fakultas Teknik

Universitas Muhammadiyah Jember

Email : alvianananta74@gmail.com

**ABSTRACT**

*Google Play store is an online store application that provides applications, books dan music on Danroid. Google Play store provides freedom to provide comments, input dan criticism by users to the application developers that users consume. However, there are some reviews that are not appropriate because the rating given does not match the written review, which causes some confusion or ambiguity about which things are true, therefore sentiment analysis using Support Vector Machine (SVM) is carried out. But the next problem is the inevitable use of free language in reviewing an application. Many of the users slipped in nonstandanard words, word abbreviations, dan spelling mistakes or typos that eventually passed the text preprocessing stage. Therefore, word normalization using Levenshtein Distance is carried out, so that the difference in accuracy results before dan after applying word normalization using Lecenshtein Distance can be known. The classification stages include collecting sample data sourced from Dana application reviews in this research, labeling data, preprocessing text, applying the Levenshtein Distance algorithm, TF-IDF, dan the sentiment classification process using Support Vector Machine. The test results by applying word normalization using Levenshtein Distance showed an increase of 2% which initially obtained a value of 83% without using Levenshtein Distance to 85% after applying Levenshtein Distnsnce.*

**Keywords** : Sentiment Analysis, Support Vector Machine, Levenshtein distance.

## KATA PENGANTAR

Alhamdulillah puji Allah atas segala yang telah dilimpahkan serta karunianya yang sangat luar biasa dalam hidup ini kepada penulis berupa nikmat untuk selalu bersujud dan meminta ampunya kepada-Nya, serta berkat rahmat-Nya penulis diberikan kemudahan dalam menyelesaikan studi di kampus Universitas Muhammadiyah Jember.

Atas segala upaya, bimbingan, dan arahan dari semua pihak, penulis mengucapkan banyak terima kasih. Dalam kesempatan ini penulis mengucapkan terima kasih sebanyak-banyaknya kepada:

1. Allah SWT atas segala rahmat, nikmat, dan hidayah-Nya, terima kasih atas kehendak-Mu telah memberikan gelar Sarjana Komputer.
2. Bapak Dr. Ir. Nanang Saiful Rizal, S.T., M.T., IPM, selaku Dekan Fakultas Teknik Universitas Muhammadiyah Jember.
3. Bapak Ari Eko Wardoyo, M. Kom., selaku Ketua Program Studi Teknik Informatika Universitas Muhammadiyah Jember.
4. Bapak Deni Arifianto M.Kom., selaku dosen Pembimbing I dan Ibu Rosita Yanuarti S.Kom., M. Cs., selaku dosen Pembimbing II yang telah meluangkan waktu untuk memberikan arahan dan membimbing penulis dalam menyelesaikan tugas akhir ini.
5. Bapak Taufiq Timur W. S.Kom., M.Kom., selaku dosen Penguji I dan Ibu Habibatul Azizah Al Faruq M.Pd., selaku Penguji II yang telah memberikan saran dan masukan dalam penyelesaian tugas akhir ini.
6. Bapak dan Ibu dosen Program Teknik Informatika Universitas Muhammadiyah Jember yang telah memberikan ilmunya kepada penulis.
7. Ayah, Mama dan keluarga saya yang selalu mendoakan, memberi tuntunan, dan dukungan serta senantiasa tetap di hati serta tak akan terlupa segala nasihat, ilmu, dan segalanya yang telah diberikan kepada saya akan selalu membekas hingga saat ini dan seterusnya.
8. Anggota *Playground*, yang terdiri dari Fatih Janata, Ferdy Mahendra, Fathur Roji Gunawan, Danialvin, Adit, Hernawan Agung, Rency, Berlian,



Virginia, Fiqi yang senantiasa ada, memberikan semangat, dukungan, serta kebahagiaan yang luar biasa selama menemani penulis.

9. Wdانا Afrilia Ayu sebagai support system yang selalu memberi dukungan, semangat, dan mendengarkan segala keluh kesah penulis.
10. Gatot Sussanto S.kom., Tia Monica Regianti S.Kom., Wina Ayundasari S.Kom., Elvira Nugrah, Ivan Fahriz, Tarisa Suci Amalia, Mikhail Kusuma Raharja Rega Sukmawanti, Dini Izzatus Zahro, Iwang Moeslem, Falah Yafi, Nur Cahyo, Singgih Galih, M. Iskandar rasyid, M. Hadyan Risqi Firmansyah, Yoga Wahyu Kusuma, Muhammad Faaza Billah, Zelda Ababil, Zulkifli Aminullah, Sagtya Rengga Surya Pratama, Varian Cyrus Valde Makalew, Muhammad Asril Azim, M. Fadhil Al Hikam Tirta Bayu, Ginda Ali Ridho Bestariyanto, dan Agung Maulana Pratama yang selalu mendukung dan berjuang bersama untuk mencapai target kuliah.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih banyak kekurangan dan masih jauh dari kata sempurna, untuk itu penulis mengharapkan kritik dan saran yang bersifat membangun demi kesempurnaan skripsi ini. Skripsi ini diajukan dan disusun dalam rangka memenuhi ujian akhir S1 Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember.

Demikian pengantar sederhana dari penulis, semoga berkesan di hati pembaca dan bermanfaat bagi semuanya.

## MOTTO

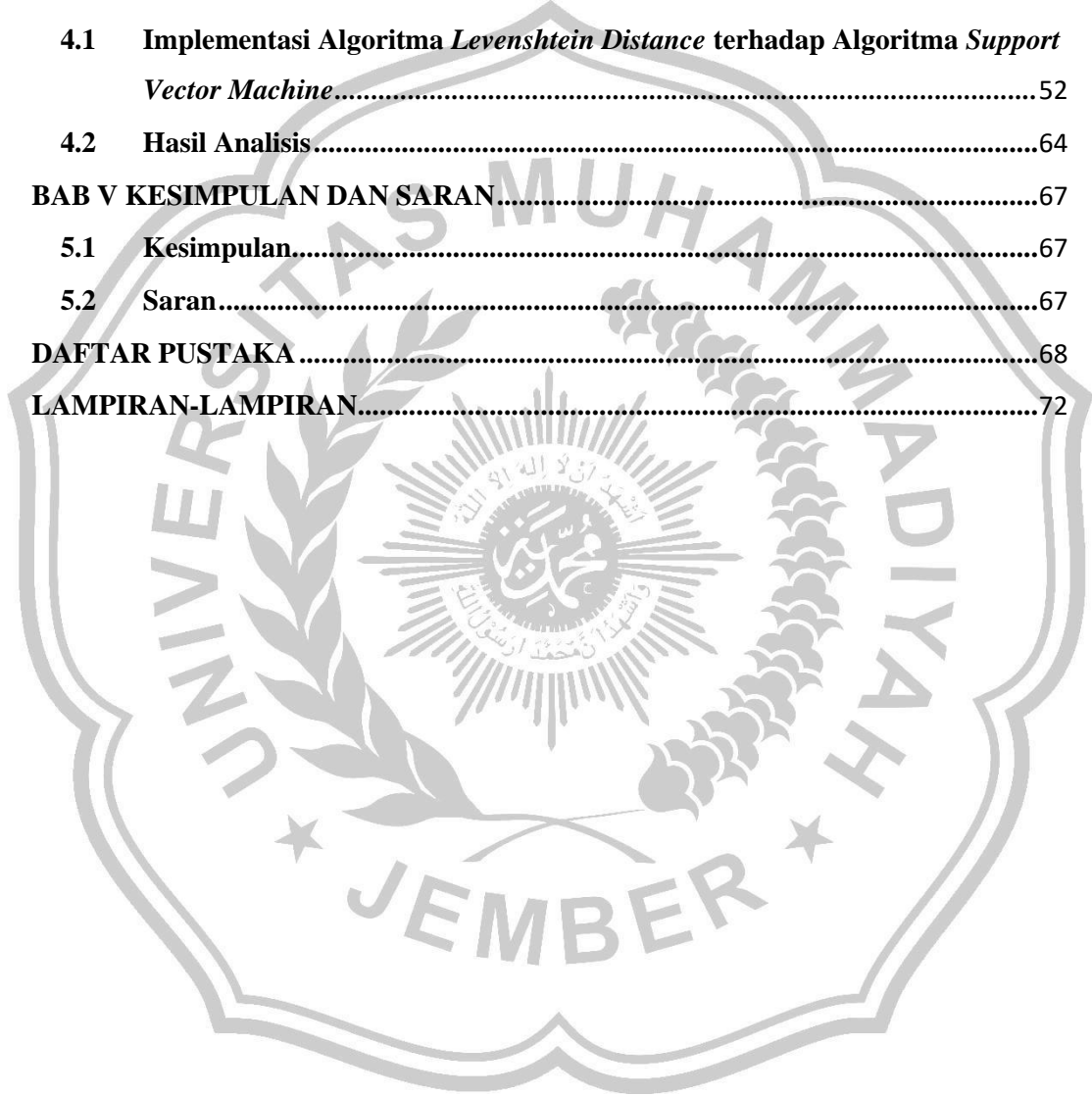
“Semua hal terasa sulit sebelum menjadi mudah.”



## DAFTAR ISI

HALAMAN PERSETUJUAN SIDANG TUGAS AKHIR.....	ii
LEMBAR PENGESAHAN.....	iii
LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR.....	iv
ABSTRAK.....	v
ABSTRACT.....	vi
KATA PENGANTAR.....	vii
MOTTO.....	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	4
1.5 Batasan Penelitian.....	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 <i>Text Mining</i> .....	6
2.2 <i>Klasifikasi</i> .....	7
2.3 <i>Analisis Sentimen</i> .....	7
2.4 <i>Support Vector Machine (SVM)</i> .....	8
2.5 <i>Levenshtein distance</i> .....	13
2.6 <i>Term Frequency-Invers Document Frequency</i> .....	14
2.7 <i>K-fold cross validation</i> .....	15
2.8 <i>Confusion matrix</i> .....	16
2.9 <i>Python</i> .....	17
2.10 <i>Jupyter notebook</i> .....	17
2.11 <i>Penelitian Terdahulu</i> .....	17
BAB III METODE PENELITIAN.....	36
3.1 <i>Diagram Alur Penelitian</i> .....	36

3.2	Studi Awal .....	37
3.3	Pengumpulan Data.....	37
3.4	Tahap Analisis.....	38
3.5	Tahap Implementasi .....	45
3.6	Tahap Evaluasi.....	50
<b>BAB IV IMPLEMENTASI DAN HASIL .....</b>		<b>52</b>
4.1	Implementasi Algoritma <i>Levenshtein Distance</i> terhadap Algoritma <i>Support Vector Machine</i> .....	52
4.2	Hasil Analisis.....	64
<b>BAB V KESIMPULAN DAN SARAN.....</b>		<b>67</b>
5.1	Kesimpulan.....	67
5.2	Saran.....	67
<b>DAFTAR PUSTAKA .....</b>		<b>68</b>
<b>LAMPIRAN-LAMPIRAN.....</b>		<b>72</b>



## DAFTAR TABEL

Tabel 3. 1 Contoh <i>Scrapping</i> Data Ulasan Aplikasi DANA.....	37
Tabel 3. 2 Tahapan <i>Cleansing</i> .....	38
Tabel 3. 3 Tahapan <i>Case folding</i> .....	39
Tabel 3. 4 Tahapan <i>Tokenizing</i> .....	40
Tabel 3. 5 Tahapan <i>Filtering</i> .....	40
Tabel 3. 6 Tahapan <i>Stemming</i> .....	41
Tabel 3. 7 Tahapan <i>Levenshtein distance</i> .....	42
Tabel 3. 8 Tahapan <i>Levenshtein distance</i> .....	42
Tabel 3. 9 Tahapan <i>Levenshtein distance</i> .....	43
Tabel 3. 10 Hasil Tahapan <i>Levenshtein distance</i> .....	43
Tabel 3. 11 Pembobotan TF-IDF .....	44
Tabel 3. 12 Pelabelan Data <i>Training</i> .....	45
Tabel 3. 13 Format Data <i>Input</i> .....	46
Tabel 3. 14 Perhitungan untuk mencari panjang vektor.....	47
Tabel 3. 15 Penerapan <i>Confusion matrix</i> pada Contoh Data .....	51
Tabel 4. 1 Hasil <i>Labeling</i> Data .....	53
Tabel 4. 2 Potongan Hasil <i>Cleansing</i> .....	54
Tabel 4. 3 Potongan Hasil <i>Case folding</i> .....	54
Tabel 4. 4 Potongan Hasil <i>Tokenizing</i> .....	55
Tabel 4. 5 Potongan Hasil <i>Filtering</i> .....	56
Tabel 4. 6 Potongan Hasil <i>Stemming</i> .....	57
Tabel 4. 7 Potongan Hasil TF-IDF.....	60
Tabel 4. 8 Rekapitulasi Hasil Akurasi.....	64

## DAFTAR GAMBAR

Gambar 3. 1 Tahapan Penelitian .....	36
Gambar 3. 2 Tahapan Analisis .....	38
Gambar 4. 1 Potongan <i>Source Code Scrapping Data</i> .....	52
Gambar 4. 2 Hasil <i>Scrapping Data</i> .....	52
Gambar 4. 3 Data Label Positif .....	53
Gambar 4. 4 Data Label Negatif .....	53
Gambar 4. 5 Daftar kata yang melalui proses <i>Levenshtein distance</i> .....	59
Gambar 4. 6 Bagan hasil akurasi SVM + LD dan hanya SVM saja pada 2- <i>fold cross validation</i> .....	61
Gambar 4. 7 Bagan hasil akurasi SVM + LD dan hanya SVM saja pada 4- <i>fold cross validation</i> .....	62
Gambar 4. 8 Bagan hasil akurasi SVM + LD dan hanya SVM saja pada 5- <i>fold cross validation</i> .....	63
Gambar 4. 9 Bagan hasil akurasi SVM + LD dan hanya SVM saja pada 10- <i>fold cross validation</i> .....	64