

Lab SCR IT - Analisis Data

***SENTIMENT ANALYSIS* ULASAN PENGGUNA APLIKASI  
*ONLINE DATING* BUMBLE DENGAN ALGORITMA  
K-NEAREST NEIGHBOR(KNN) DAN METODE WORD2VEC**

**SKRIPSI**



Oleh :

**LAVENIA AGLIS ARDANA**

**19081010076**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA  
TIMUR**

**2023**

**LEMBAR PENGESAHAN**

**Judul** : SENTIMENT ANALYSIS ULASAN PENGGUNA APLIKASI  
ONLINE DATING BUMBLE DENGAN ALGORITMA K-  
NEAREST NEIGHBOR(KNN) DAN METODE WORD2VEC

**Oleh** : LAVENIA AGLIS ARDANA

**NPM** : 19081010076

**Telah Diseminarkan Dalam Ujian Skripsi Pada**

**Hari Jum'at, 10 November 2023**

**Mengetahui**

**Dosen Pembimbing**

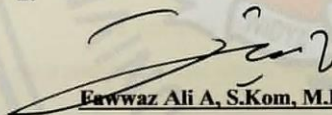
1.



Hendra Maulana, S.Kom, M.Kom

NPT. 201198 31 223248

2.

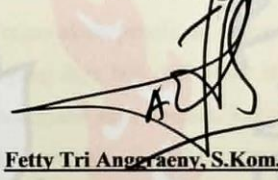


Ewwaz Ali A, S.Kom, M.Kom

NIP. 19920317 2018031 002

**Dosen Penguji**

1.



Fetty Tri Anggraeny, S.Kom, M.Kom

NIP. 19820211 2021212 005

2.



Afina Lina Nurlaili, S.Kom, M.Kom

NIP. 1993121 3202203 2010

**Menyetujui**

**Dekan**

**Fakultas Ilmu Komputer**

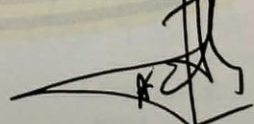


Prof. Dr. Ir. Novirina Hendrasarie, M.T

NIP. 19681126 199403 2 001

**Koordinator Program Studi**

**Informatika**



Fetty Tri Anggraeny, S.Kom, M.Kom

NIP. 19820211 2021212 005

## SURAT PERNYATAAN ANTI PLAGIAT

Saya, mahasiswa Program Studi Informatika UPN "Veteran" Jawa Timur,  
yang bertanda tangan dibawah ini:

Nama : Lavenia Aglis Ardana

NPM : 19081010076

Menyatakan bahwa judul skripsi yang saya ajukan dan kerjakan dengan  
judul:

**"SENTIMENT ANALYSIS ULASAN PENGGUNA APLIKASI ONLINE  
DATING DENGAN ALGORITMA K-NEAREST NEIGHBOR(KNN) DAN  
METODE WORD2VEC"**

Bukan merupakan plagiat dari skripsi/tugas akhir/penelitian orang lain dan  
juga bukan merupakan produk atau *software* yang saya beli dari pihak lain. Saya  
juga menyatakan bahwa skripsi ini adalah pekerjaan saya sendiri, kecuali yang  
dinyatakan dalam daftar pustaka dan tidak pernah diajukan untuk syarat  
memperoleh gelar di UPN "Veteran" Jawa Timur maupun di institusi pendidikan  
lain.

Jika ternyata di kemudian hari pernyataan ini terbukti tidak benar, maka  
saya siap menerima segala konsekuensinya.

Surabaya, 10 November 2023

Hormat saya,

  
  
**Lavenia Aglis Ardana**  
**NPM. 19081010076**

# SENTIMENT ANALYSIS ULASAN PENGGUNA APLIKASI ONLINE DATING BUMBLE DENGAN ALGORITMA K-NEAREST NEIGHBOR(KNN) DAN METODE WORD2VEC

**Nama Mahasiswa : Lavenia Aglis Ardana**

**NPM : 19081010076**

**Program Studi : Informatika**

**Dosen Pembimbing : Hendra Maulana, S.Kom., M.Kom**

**Fawwaz Ali Akbar, S.Kom, M.Kom**

---

## ABSTRAK

Komunikasi merupakan salah satu aspek yang penting dalam kehidupan manusia sebagai makhluk sosial. Manusia dapat melakukan komunikasi tanpa perlu bertemu secara langsung dengan memanfaatkan adanya media komunikasi online. Salah satu kegiatan komunikasi yaitu *online dating* dengan aplikasi Bumble. Pengguna aplikasi menyampaikan pengalaman setelah menggunakan aplikasi tersebut dalam bentuk sebuah ulasan. Peneliti melakukan analisis sentimen terhadap ulasan pengguna pada aplikasi Bumble dengan algoritma *K-Nearest Neighbor*(KNN) dan metode ekstraksi fitur yaitu Word2vec. Hasil dari penelitian yang dilakukan yaitu nilai *accuracy*, *precision*, *recall*, dan *f1 score* dari setiap pengujian yang dilakukan pada penelitian ini. Penelitian dilakukan dengan 1200 data ulasan pengguna aplikasi Bumble. Akurasi tertinggi dari percobaan yang dilakukan dengan parameter nilai K yang berbeda-beda pada model algoritma KNN yaitu sebesar 84,75% dengan nilai K yang digunakan yaitu 15. Sedangkan hasil percobaan dengan parameter pada model ekstraksi fitur Word2vec yang berbeda-beda yaitu metode *word vector skip-gram* memiliki nilai rata-rata *accuracy* yang lebih besar yaitu 83,22%. Nilai rata-rata *accuracy* ketika menggunakan metode word vector *Continous Bag of Words*(CBOW) yaitu 80,17%.

**Kata kunci:** *Sentiment analysis, Ulasan, KNN, Word2vec*

## KATA PENGANTAR

Puji syukur kehadiran Allah SWT yang telah memberikan rahmat dan inayah-Nya sehingga penulis dapat menyelesaikan laporan skripsi yang berjudul “*Sentiment Analysis* Ulasan Pengguna Aplikasi *Online Dating* Bumble dengan Algoritma *K-Nearest Neighbor* dan Metode *Word2vec*”.

Terima kasih saya ucapkan kepada seluruh pihak yang telah memberikan banyak dukungan kepada penulis sehingga penulis bisa menyelesaikan laporan skripsi ini dengan baik tepat waktu.

Penulis menyadari, bahwa laporan skripsi ini masih jauh dari kata sempurna baik segi penyusunan, bahasa, maupun penulisannya. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun dari semua pembaca agar penulis bisa menjadi lebih baik lagi di masa mendatang.

Semoga laporan skripsi ini bisa menambah wawasan para pembaca dan bisa bermanfaat untuk perkembangan dan peningkatan ilmu pengetahuan.

Surabaya, 10 November 2023

Penulis

## UCAPAN TERIMAKASIH

Penulis panjatkan puji syukur kehadirat Allah SWT yang telah memberikan keberkahan kepada penulis sehingga dapat menyelesaikan kegiatan penelitian skripsi ini dengan baik. terselesaikannya skripsi ini tidak lepas dari bantuan dan dukungan dari berbagai macam pihak yang telah berkontribusi dalam menyelesaikan penelitian skripsi ini. Secara khusus penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu.

Pada kesempatan kali ini penulis ingin menyampaikan ucapan terima kasih yang sebesar – besarnya kepada seluruh civitas akademika UPN “Veteran” Jawa Timur yang terdiri atas :

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT. selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Prof. Novirina Hendrasarie, S.T, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Fetty Tri Anggraeny, S.Kom, M.Kom. selaku Koordinator Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur serta selaku dosen penguji I yang telah memberikan banyak masukan serta membantu dengan sabar dalam membimbing penulis untuk pengerjaan skripsi.
4. Ibu Afina Lina Nurlaili, S.Kom, M.Kom. selaku dosen penguji II yang juga telah memberikan banyak masukan serta membantu dengan sabar dalam membimbing penulis untuk pengerjaan skripsi.
5. Bapak Dr. Basuki Rahmat, S.Si, M.T. selaku dosen wali penulis yang telah memberikan arahan mulai awal perkuliahan hingga akhir penyelesaian skripsi.
6. Bapak Hendra Maulana S.Kom, M.Kom, selaku dosen pembimbing I yang telah menerima penulis untuk menjadi mahasiswa bimbingan dari sebelum dimulainya pra-skripsi, membimbing serta memberikan arahan

pada penulis dengan sabar dalam proses pengerjaan skripsi selama menjadi dosen pembimbing.

7. Bapak Fawwaz Ali Akbar, S.Kom, M.Kom, selaku dosen pembimbing II yang telah banyak membimbing dan memberikan arahan pada penulis selama pengerjaan skripsi.
8. Seluruh dosen dan staff pada program studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur atas segala ilmu pengetahuan yang diberikan kepada penulis selama masa perkuliahan. Semoga penulis dapat mengamalkan ilmu pengetahuan yang telah diberikan dan akan menjadi bekal yang berguna bagi penulis untuk masa depan.

Lalu penulis juga mengucapkan terima kasih kepada pihak yang telah membantu dalam pengerjaan skripsi ini. Penulis mengucapkan terima kasih sebesar-besarnya kepada :

1. Keluarga penulis, Bapak penulis yaitu Agung Kurniawan, Ibu penulis yaitu Sulistia Ningsih, dan adik penulis yaitu Kirey yang telah memberikan dukungan penuh secara materi dan non-materi sehingga penulis dapat menyelesaikan perkuliahan dari awal hingga terselesaikannya skripsi ini dengan baik.
2. Risda Rosdiana, Annisa Rahmadhani, Alifta Putri, dan Aniisah Eka yang telah saling mengingatkan dan memberikan semangat dalam proses pengerjaan skripsi hingga terselesaikan dengan baik.
3. Seluruh teman-teman angkatan 2019 Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur yang telah berjuang bersama dalam proses perkuliahan dari awal hingga selesai.

## DAFTAR ISI

LEMBAR PENGESAHAN .....	ii
SURAT PERNYATAAN ANTI PLAGIAT .....	iii
ABSTRAK .....	iv
KATA PENGANTAR .....	v
UCAPAN TERIMAKASIH.....	vi
DAFTAR ISI.....	viii
DAFTAR GAMBAR .....	xi
DAFTAR TABEL.....	xiii
DAFTAR KODE PROGRAM.....	xiv
BAB I .....	1
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 Masalah.....	5
BAB II.....	6
TINJAUAN PUSTAKA .....	6
2.1 Penelitian Terdahulu.....	6
2.2 Analisis Sentimen.....	8
2.3 Ulasan Pengguna .....	9
2.4 <i>Online Dating</i> .....	10
2.5 Aplikasi Bumble.....	11
2.6 Algoritma Klasifikasi .....	11
2.7 Algoritma <i>Long Short Term Memory</i> (LSTM).....	12



2.8 Algoritma <i>Naïve Bayes</i> .....	13
2.9 Algoritma C4.5.....	14
2.10 Algoritma <i>K-Nearest Neighbor</i> .....	14
2.11 Metode Ekstraksi Fitur TF-IDF.....	15
2.12 Metode Ekstraksi Fitur <i>Word2vec</i> .....	16
2.13 Bahasa Pemrograman.....	19
2.14 Python.....	20
2.15 <i>Data Preprocessing</i> .....	20
2.16 Evaluasi Performa.....	22
BAB III.....	25
METODOLOGI PENELITIAN.....	25
3.1 Alur Penelitian.....	25
3.2 Studi Literatur.....	26
3.3 Identifikasi Masalah.....	26
3.4 <i>Data Collection</i> .....	26
3.5 <i>Data Labelling</i> .....	27
3.6 <i>Data Preprocessing</i> .....	27
3.7 Ekstraksi Fitur dengan <i>Word2vec</i> .....	37
3.9 Data Splitting.....	39
3.7 Klasifikasi dengan Algoritma <i>K-Nearest Neighbor</i> .....	39
3.8 Skenario Pengujian.....	41
3.9 Evaluasi Performa.....	43
BAB IV.....	44
HASIL DAN PEMBAHASAN.....	44
4.1 <i>Data Collection</i> .....	44
4.2 <i>Data Labelling</i> .....	45

4.3 <i>Data Preprocessing</i> .....	47
4.4 Ekstraksi Fitur dengan <i>Word2vec</i> .....	57
4.5 <i>Data Splitting</i> .....	61
4.6 Klasifikasi dengan Algoritma KNN .....	62
4.7 Evaluasi Performa .....	64
4.8 Model Ekstraksi Fitur dengan <i>Word2vec</i> pada Skenario Pengujian 2 .....	66
4.9 Klasifikasi dengan KNN Skenario Pengujian 2 .....	67
4.10 Hasil Pengujian Skenario 1 .....	67
4.11 Hasil Pengujian Skenario 2 .....	75
BAB V .....	92
PENUTUP .....	92
5.1 Kesimpulan .....	92
5.2 Saran .....	94
DAFTAR PUSTAKA .....	95

## DAFTAR GAMBAR

Gambar 2.1 Arsitektur Skip-gram.....	18
Gambar 2.2 Arsitektur CBOW.....	19
Gambar 3.1 Alur Penelitian.....	25
Gambar 3.2 Alur pada Tahap <i>Data Preprocessing</i> .....	28
Gambar 3.3 Alur pada Tahap <i>Case Folding</i> .....	28
Gambar 3.4 Alur pada Tahap <i>Remove Duplicates</i> .....	29
Gambar 3.5 Alur pada Tahap <i>Tokenizing</i> .....	30
Gambar 3.6 Alur pada Tahap <i>Stopwords</i> .....	30
Gambar 3.7 Alur pada Tahap <i>Stemming</i> .....	31
Gambar 3.8 Alur pada Tahap Ekstraksi Fitur dengan <i>Word2vec</i> .....	38
Gambar 3.9 Alur pada Tahap <i>Data Splitting</i> .....	39
Gambar 3.10 Alur pada Fungsi KNN .....	40
Gambar 3.11 Alur pada Tahap Klasifikasi dengan Algoritma KNN.....	41
Gambar 3.12 Alur pada Skenario Pengujian 1 .....	42
Gambar 3.13 Alur pada Skenario Pengujian 2.....	42
Gambar 3.14 Evaluasi Performa .....	43
Gambar 4.1 Tampilan Data pada Tahap <i>Data Collection</i> .....	45
Gambar 4.2 Tampilan Data pada Tahap <i>Data Labelling</i> .....	46
Gambar 4.3 Vektor Kata dengan <i>Word2vec</i> .....	59
Gambar 4.4 Visualisasi Vektor dengan <i>Library UMAP</i> .....	59
Gambar 4.5 Tampilan Data Ulasan Aplikasi Bumble, Sentimen, dan Vektor.....	60
Gambar 4.6 Vektor Setiap Data Ulasan .....	61
Gambar 4.7 <i>Confusion Matrix</i> Nomor Skenario 1 .....	68
Gambar 4.8 <i>Confusion Matrix</i> Nomor Skenario 2 .....	69
Gambar 4.9 <i>Confusion Matrix</i> Nomor Skenario 3 .....	70
Gambar 4.10 <i>Confusion Matrix</i> Nomor Skenario 4 .....	70
Gambar 4.11 <i>Confusion Matrix</i> Nomor Skenario 5 .....	71
Gambar 4.12 <i>Classification Report</i> Nomor Skenario 1 .....	72
Gambar 4.13 <i>Classification Report</i> Nomor Skenario 2 .....	72
Gambar 4.14 <i>Classification Report</i> Nomor Skenario 3 .....	73

Gambar 4.15 <i>Classification Report</i> Nomor Skenario 4 .....	74
Gambar 4.16 <i>Classification Report</i> Nomor Skenario 5 .....	74
Gambar 4.17 <i>Confusion Matrix</i> Nomor Skenario 6 .....	77
Gambar 4.18 <i>Confusion Matrix</i> Nomor Skenario 7 .....	77
Gambar 4.19 <i>Confusion Matrix</i> Nomor Skenario 8 .....	78
Gambar 4.20 <i>Confusion Matrix</i> Nomor Skenario 9 .....	79
Gambar 4.21 <i>Confusion Matrix</i> Nomor Skenario 10 .....	79
Gambar 4.22 <i>Confusion Matrix</i> Nomor Skenario 11 .....	80
Gambar 4.23 <i>Confusion Matrix</i> Nomor Skenario 12 .....	81
Gambar 4.24 <i>Confusion Matrix</i> Nomor Skenario 13 .....	81
Gambar 4.25 <i>Confusion Matrix</i> Nomor Skenario 14 .....	82
Gambar 4.26 <i>Confusion Matrix</i> Nomor Skenario 15 .....	83
Gambar 4.27 <i>Classification Report</i> Nomor Skenario 6 .....	84
Gambar 4.28 <i>Classification Report</i> Nomor Skenario 7 .....	84
Gambar 4.29 <i>Classification Report</i> Nomor Skenario 8 .....	85
Gambar 4.30 <i>Classification Report</i> Nomor Skenario 9 .....	85
Gambar 4.31 <i>Classification Report</i> Nomor Skenario 10 .....	86
Gambar 4.32 <i>Classification Report</i> Nomor Skenario 11 .....	86
Gambar 4.33 <i>Classification Report</i> Nomor Skenario 12 .....	87
Gambar 4.34 <i>Classification Report</i> Nomor Skenario 13 .....	88
Gambar 4.35 <i>Classification Report</i> Nomor Skenario 14 .....	88
Gambar 4.36 <i>Classification Report</i> Nomor Skenario 15 .....	89

## DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i> .....	23
Tabel 3.1 Contoh <i>Data Labelling</i> .....	27
Tabel 3.2 Contoh Perbandingan Data pada Tahap <i>Case Folding</i> .....	32
Tabel 3.3 Contoh Perbandingan Data pada Tahap <i>Remove Duplicates</i> .....	32
Tabel 3.4 Contoh Perbandingan Data pada Tahap <i>Tokenizing</i> .....	33
Tabel 3.5 Contoh Perbandingan Data pada Tahap <i>Stopwords</i> .....	35
Tabel 3.6 Contoh Perbandingan Data pada Tahap <i>Stemming</i> .....	37
Tabel 4.1 Perbandingan Data Ulasan Bumble pada Tahap <i>Case Folding</i> .....	48
Tabel 4.2 Perbandingan Data Ulasan Bumble pada Tahap <i>Remove Duplicates</i> ....	49
Tabel 4.3 Perbandingan Data Ulasan Bumble pada Tahap <i>Tokenizing</i> .....	49
Tabel 4.4 Perbandingan Data Ulasan Bumble pada Tahap <i>Stopwords</i> .....	52
Tabel 4.5 Perbandingan Data Ulasan Bumble pada Tahap <i>Stemming</i> .....	54
Tabel 4.6 Perbandingan Data Ulasan Bumble Sebelum dan Sesudah <i>Preprocessing</i> .....	56
Tabel 4.7 Skenario Pengujian 1 .....	68
Tabel 4.8 <i>Accuracy, Precision, Recall, dan F1 Score</i> pada Skenario Pengujian 1.....	75
Tabel 4.9 Skenario Pengujian 2 .....	76
Tabel 4.10 <i>Accuracy, Precision, Recall, dan F1 Score</i> pada Skenario Pengujian 2 .....	90

## DAFTAR KODE PROGRAM

Kode Program 4.1 <i>Data Collection</i> .....	44
Kode Program 4.2 <i>Data Labelling</i> .....	46
Kode Program 4.3 <i>Data Preprocessing</i> .....	47
Kode Program 4.4 Ekstraksi Fitur dengan <i>Word2vec</i> .....	58
Kode Program 4.5 <i>Data Splitting</i> .....	61
Kode Program 4.6 Klasifikasi dengan <i>K-Nearest Neighbor</i> .....	62
Kode Program 4.7 Evaluasi Performa.....	64
Kode Program 4.8 Ekstraksi Fitur dengan <i>Word2vec</i> Skenario Pengujian 2.....	66
Kode Program 4.9 Klasifikasi dengan KNN Skenario Pengujian 2.....	67