

**ANALISIS SENTIMEN ULASAN APLIKASI ONLINE
TRAVEL AGENT (OTA) DI GOOGLE PLAY STORE
MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOUR**

SKRIPSI

Diajukan untuk memenuhi persyaratan
dalam memperoleh gelar Sarjana Komputer
Program Studi Sistem Informasi



Disusun Oleh:

HASTRI CANTYA DANAHISWARI

19082010056

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
SURABAYA
2023**

SKRIPSI

ANALISIS SENTIMEN ULASAN APLIKASI ONLINE TRAVEL AGENT (OTA) DI GOOGLE
PLAY STORE MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOUR

Disusun Oleh :

HASTRI CANTYA DANAHISWARI

NPM.19082010056

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Skripsi

Program Studi Sistem Informasi Fakultas Ilmu Komputer

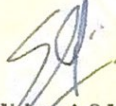
Universitas Pembangunan Nasional "Veteran" Jawa Timur

Pada Tanggal 27 November 2023

Menyetujui,

Dosen Pembimbing:

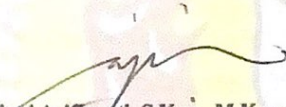
1.



Eka Dyar Wahyuni, S.Kom, M.Kom

NIP.19841201 202121 2 005

2.

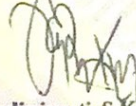


Amalia Anjani Arifiyanti, S.Kom, M.Kom

NIP.19841201 202121 2 005

Dosen Penguji:

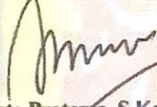
1.



Rizka Hadiwivanti, S.Kom, M.Kom, MBA

NIP 19860727 2018032 001

2.



Arista Pratama, S.Kom, M.Kom

NPT. 171199 10 320052

3.



Anindo Saka Fiti, S.Kom, M.Kom

NPT. 212199 30 325268

Mengetahui,

Dekan Fakultas Ilmu Komputer

Universitas Pembangunan "Veteran" Jawa Timur



Prof. Dr. Ir. Novirina Hendrasarie, MT.

NIP.19681126 199403 2 001

LEMBAR PENGESAHAN

**ANALISIS SENTIMEN ULASAN APLIKASI ONLINE TRAVEL AGENT (OTA) DI
GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA K-NEAREST
NEIGHBOUR**

Disusun Oleh:

Hastri Cantya Danahiswari

19082010056

Telah disetujui mengikuti Ujian Negara Lisan Gelombang November

Periode 2023 pada Tanggal 27 November 2023

Menyetujui,

Dosen Pembimbing 1

Dosen Pembimbing 2

Eka Dyar Wahyuni, S.Kom. M.Kom

NIP. 19841201 2021212 005

Amalia Anjani Arifiyanti, S.Kom, M.Kom

NIP. 19920812 2018032 001

Mengetahui,

Ketua Program Studi Sistem Informasi

Fakultas Ilmu Komputer

Universitas Pembangunan Nasional "Veteran" Jawa Timur

Agung Brastama Putra, S.Kom, M.Kom

NIP. 19851124 2021211 003

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR

FAKULTAS ILMU KOMPUTER

KETERANGAN REVISI

Kami yang bertanda tangan di bawah ini menyatakan bahwa mahasiswa berikut:

Nama : Hastri Cantya Danahiswari

NPM : 19082010056

Program Studi : Sistem Informasi

Telah mengerjakan revisi Ujian Negara Lisan Skripsi pada tanggal 27 November 2023 dengan judul:

**ANALISIS SENTIMEN ULASAN APLIKASI ONLINE TRAVEL AGENT (OTA) DI
GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA K-NEAREST
NEIGHBOUR**

Oleh karenanya mahasiswa tersebut diatas dinyatakan bebas revisi Ujian Negara Lisan Skripsi dan diijinkan untuk membukukan laporan SKRIPSI dengan judul tersebut.

Surabaya, 5 Desember 2023

Dosen penguji yang memeriksa revisi:

1. Rizka Hadiwiyanti, S.Kom, M.Kom, MBA

NIP 19860727 2018032 001

2. Arista Pratama, S.Kom, M.Kom

NPT 171199 10 320052

3. Anindo Saka Fitri, S.Kom, M.Kom

NPT 212199 30 325268

Mengetahui,

Dosen Pembimbing 1

Dosen Pembimbing 2

Eka Dyar Wahyuni, S.Kom, M.Kom

NIP. 19841201 2021212 005

Amalia Anjani Arifiyanti, S.Kom, M.Kom

NIP. 19920812 2018032 001

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAWA TIMUR
FAKULTAS ILMU KOMPUTER
KETERANGAN REVISI**

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Hastri Cantya Danahiswari

NPM : 19082010056

Program Studi : Sistem Informasi

Menyatakan bahwa Judul Skripsi / Tugas Akhir sebagai berikut:

**ANALISIS SENTIMEN ULASAN APLIKASI ONLINE TRAVEL AGENT
(OTA) DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA K-NEAREST
NEIGHBOUR**

Bukan merupakan plagiat dari Skripsi / Tugas Akhir / Penelitian orang lain dan juga bukan merupakan Produk / Hasil Karya yang Saya beli dari orang lain.

Saya juga menyatakan bahwa Skripsi / Tugas Akhir ini adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional “Veteran” Jawa Timur maupun di Institusi Pendidikan lain. Jika ternyata dikemudian hari pernyataan terbukti benar, maka Saya bertanggung jawab penuh dan siap menerima segala konsekuensi, termasuk pembatalan ijazah dikemudian hari.

Surabaya, 12 Desember 2023

Hormat Saya,



Hastri Cantya Danahiswari

NPM. 19082010056

Judul : ANALISIS SENTIMEN ULASAN APLIKASI ONLINE TRAVEL
AGENT (OTA) DI GOOGLE PLAY STORE MENGGUNAKAN
ALGORITMA K-NEAREST NEIGHBOUR

Pembimbing 1 : Eka Dyar Wahyuni, S.Kom. M.Kom

Pembimbing 2 : Amalia Anjani Arifiyanti, S.Kom, M.Kom

ABSTRAK

Di era digital saat ini, berbagai macam aplikasi Online Travel Agent (OTA) di Google Play Store digunakan oleh Masyarakat Indonesia untuk berbagai macam keperluan seperti berlibur atau travelling. Hingga akhir tahun 2022 aplikasi OTA di Google Play Store seperti aplikasi Traveloka dan Tiket.com telah diunduh oleh 50 juta dan 10 juta pengguna. Penelitian ini bertujuan untuk mempermudah, membantu serta memberikan informasi tentang emosi yang terkandung dalam ulasan tentang aplikasi OTA yang ada pada situs Google Play Store dengan melakukan analisis sentimen dengan algoritma K-Nearest Neighbour sebagai algoritma yang optimal dan melakukan visualisasi berbasis web. Hasil akurasi menunjukkan bahwa nilai $k=6$ mendapat akurasi yang optimal dimana data tidak overfit maupun underfit. Dari 50 data dilakukan pengujian pada sistem visualisasi terdapat 4 data sentiment yang salah dan 46 data sentiment yang benar sehingga memiliki akurasi yang sama, yaitu 92% dan dilengkapi dengan tampilan wordcloud untuk mengetahui kata yang sering muncul dan grafik batang untuk mengetahui distribusi 50 data sentiment yang diuji.

Kata kunci:

Analisis Sentimen, *Online Travel Agent*, *K-Nearest Neighbour*

KATA PENGANTAR

Segala puji syukur penulis panjatkan atas kehadiran Allah SWT karena berkat limpahan rahmat dan hidayah-Nya penulis dapat menyelesaikan skripsi yang berjudul “ANALISIS SENTIMEN ULASAN APLIKASI ONLINE TRAVEL AGENT (OTA) DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOUR” sebagai sasyarat untuk menyelesaikan program studi Sistem Informasi Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Skripsi ini bisa terselesaikan juga berkat dorongan dan bantuan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada pihak yang telah turut serta membantu selama proses pengerjaan skripsi ini. Untuk itu penulis menyampaikan terima kasih kepada:

1. Bapak, Ibu, dan adik yang mendukung dan memberikan doa agar dimudahkan dalam menimba ilmu terutama saat pembuatan skripsi ini.
2. Ibu Eka Dyar Wahyuni, S.Kom. M.Kom selaku pembimbing I dan Ibu Amalia Anjani Arifiyanti, S.Kom, M.Kom selaku pembimbing II yang telah memberikan bimbingan dan arahan terkait penelitian ini.
3. Bapak Mohammad Irwan Afandi, S.T, M. Sc selaku dosen wali yang selalu membantu saya selama kuliah dalam berbagai hal.
4. Seluruh dosen Sistem Informasi yang telah banyak memberikan ilmu yang sangat bermanfaat selama kuliah dan untuk bekal di masa depan.
5. Eric, Aldy, Feby, Ale, dan rekan – rekan kantor PT. Pupuk Indonesia (Persero) yang selalu menyemangati, mendukung, dan membantu dalam proses pengerjaan skripsi.

6. Seluruh teman-teman Sistem Informasi angkatan 2019 yang telah memberi semangat untuk menyelesaikan skripsi ini, terutama untuk kelas A.
7. Seluruh pihak yang tidak bisa saya sebutkan satu persatu karena telah memberikan bantuan sekecil apapun.

Semoga Allah SWT membalas kebaikan pihak-pihak diatas. Dalam pengerjaan skripsi ini penulis menyadari masih banyak sekali kekurangan. Semoga skripsi ini bermanfaat bagi semua pihak.

DAFTAR ISI

ABSTRAK	vi
KATA PENGANTAR.....	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL.....	xiv
BAB I PENDAHULUAN	2
1.1 Latar Belakang	2
1.2 Rumusan Masalah	6
1.3 Batasan Masalah	7
1.4 Tujuan Penelitian	7
1.5 Manfaat Penelitian.....	8
1.6 Sistematika Penulisan.....	8
BAB II TINJAUAN PUSTAKA	10
2.1 Dasar Teori	10
2.1.1 Text Mining	10
2.1.2 Preprocessing Text	11
2.1.3 Analisis Sentimen.....	11
2.1.4 CRISP-DM.....	13
2.1.5 Vader	14
2.1.6 Roberta.....	14
2.1.7 Tiket.com.....	15
2.1.8 Traveloka	15
2.1.9 Google Play Store.....	16

	2.1.10 K-Nearest Neighbour.....	16
	2.1.11 Confusion Matrix	17
	2.2 Penelitian Terdahulu.....	20
BAB III	METODOLOGI PENELITIAN.....	24
	3.1 Business Understanding	25
	3.2 Data Understanding.....	25
	3.3 Data Preparation.....	26
	3.4 Pemodelan (Modelling).....	29
	3.5 Evaluasi (Evaluation)	29
	3.6 Deployment (Visualisasi Web)	30
BAB IV	KESIMPULAN DAN SARAN	33
	4.1 Business Understanding	33
	4.1.1 Kebutuhan Hardware dan Software.....	33
	4.2 Data Understanding.....	34
	4.2.1 Pengumpulan Data (Web Scrapping)	34
	4.3 Data Preparation.....	39
	4.3.1 Merge CSV Ulasan Aplikasi Traveloka dan Tiket.com	39
	4.3.2 Filtering Kata Traveloka dan Tiket.com.....	40
	4.3.3 Cleansing	41
	4.3.4 Case Folding	43
	4.3.5 Tokenization	44
	4.3.6 Stopword.....	45
	4.3.7 Stemming.....	46

4.3.8	Labelling Menggunakan Library Roberta.....	47
4.3.9	Labelling Menggunakan Library Vader.....	48
4.3.10	Labelling Manual	51
4.3.11	Perbandingan Labelling	54
4.3.12	EDA (Explotory Data Analysis).....	55
4.4	Pemodelan (Modelling).....	60
4.4.1	Hold Out	60
4.4.2	Pemodelan Dengan Algoritma K-Nearest Neighbour	60
4.5	Evaluasi	62
4.5.1	Confusion Matrix	62
4.5.2	Grafik Evaluasi Model KNN	63
4.6	Deployment (Web).....	65
4.6.1	Import Model Ke Dalam Bentuk .pkl (berkas).....	66
4.6.2	Pembuatan Page HTML Menggunakan VsCode	66
4.6.3	Import dan build menggunakan flask	70
4.6.4	Deployment dan Visualisasi.....	71
BAB V	PENUTUP	74
5.1	Kesimpulan.....	74
5.2	Saran	75
LAMPIRAN	77
DAFTAR PUSTAKA	82

DAFTAR GAMBAR

Gambar 1.1 Online Travel Agencies Terpopuler (Katadata, 2022)	2
Gambar 1.2 Ulasan Tiket.com di Google Play Store	4
Gambar 1.3 Ulasan Traveloka di Google Play Store	5
Gambar 2.1 Tahapan dalam analisis sentimen.....	12
Gambar 3.1 Alur Metodologi Penelitian	24
Gambar 3.1 Metodologi Penelitian CRISP-DM	24
Gambar 3.2 Alur Preprocessing Data.....	26
Gambar 3.3 Ilustrasi Visualisasi Web	32
Gambar 4.1 Source Code Pengumpulan Data Ulasan Tiket.com	34
Gambar 4.2 Source Code Pengumpulan Data Ulasan Traveloka.....	35
Gambar 4.3 Merge CSV Ulasan Aplikasi	39
Gambar 4.4 Filtering Kata Traveloka dan Tiketcom	40
Gambar 4.5 Source Code Cleansing	42
Gambar 4.6 Source Code Case Folding	43
Gambar 4.7 Source Code Tokenization.....	44
Gambar 4.8 Source Code Stopword.....	46
Gambar 4.9 Source Code Stemming.....	46
Gambar 4.10 Source Code Library Roberta	47
Gambar 4.11 Source Code Library Translator.....	49
Gambar 4.12 Source Code Library Vader	50
Gambar 4.13 Perbandingan data label positive dan negative	56
Gambar 4.14 Flow Metode Random Undersampling	57
Gambar 4.15 Flow Metode Random Oversampling	58
Gambar 4.16 Word Cloud Ulasan OTA.....	58
Gambar 4.17 Word Cloud Positive Ulasan OTA.....	59
Gambar 4.18 Word Cloud Negative Ulasan OTA	59
Gambar 4.19 Source Code Word Cloud.....	59
Gambar 4.20 Source Code Hold Out	60
Gambar 4.21 Source Code Confusion Matrix	62
Gambar 4.22 Confusion Matrix	62
Gambar 4.23 Source Code Evaluasi Pengujian Model	63
Gambar 4.24 Grafik Evaluasi Model KNN	64
Gambar 4.26 Source Code Import Model ke dalam bentuk .pkl	66
Gambar 4.27 Source Code Halaman Index	67
Gambar 4.28 Source Code Halaman Upload.....	68
Gambar 4.29 Source Code Halaman Script.....	69

Gambar 4.30 Source Code Halaman Head	69
Gambar 4.31 Source Code Halaman Navbar	69
Gambar 4.32 Source Code app.py	70
Gambar 4.33 Tampilan Halaman Index	71
Gambar 4.34 Tampilan Halaman Upload Sebelum Submit	72
Gambar 4.35 Tampilan Halaman Upload Sesudah Submit	72

DAFTAR TABEL

Tabel 2.1 Confusion Matrix.....	18
Tabel 2.2 Rumus Pengukuran Evaluasi.....	18
Tabel 2.3 Perhitungan Confusion Matrix	19
Tabel 2.4 Tabel Penelitian Terdahulu	20
Tabel 4.1 Tabel Ulasan Aplikasi Tiket.com	37
Tabel 4.2 Tabel Ulasan Aplikasi Tiket.com	39
Tabel 4.3 Proses Cleansing.....	43
Tabel 4.4 Proses Case Folding.....	44
Tabel 4.5 Proses Tokenization.....	45
Tabel 4.6 Proses Stopword	46
Tabel 4.7 Proses Stemming	47
Tabel 4.8 Contoh Hasil Labelling Roberta	48
Tabel 4.9 Proses Translate Sebelum Labelling Library Vader	50
Tabel 4.10 Contoh Hasil Labelling Roberta	51
Tabel 4.11 Contoh Hasil Menggunakan Library Roberta	52
Tabel 4.12 Contoh Hasil Menggunakan Library Vader	53
Tabel 4.13 Perbandingan Hasil Labelling Library Roberta dan Vader	55
Tabel 4.14 Hasil Pengujian dan Perbandingan Skenario Ulasan	61
Tabel 4.15 Hasil Percobaan Nilai K Skenario Oversampling.....	65
Tabel 4.16 Pengujian Pada Sistem Dengan Hasil Salah.....	73
Tabel 4.17 Pengujian Pada Sistem Dengan Hasil Benar	73