

**ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN APLIKASI INDRIVE
MENGUNAKAN BIDIRECTIONAL ENCODER REPRESENTATIONS FROM
TRANSFORMERS (BERT)**

SKRIPSI

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



Disusun Oleh:

ASRI KINANTI FEBRIANY

NPM: 19082010052

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

LEMBAR PENGESAHAN

ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN APLIKASI INDRIVE
MENGUNAKAN BIDIRECTIONAL ENCODER REPRESENTATIONS FROM
TRANSFORMERS (BERT)

Disusun Oleh:

Asri Kinanti Febriny

19082010052

Telah disetujui mengikuti Ujian Negara Lisan Gelombang November

Periode 2023 pada Tanggal 27 November 2023

Menyetujui,

Dosen Pembimbing 1

Dosen Pembimbing 2



Eka Dvar Wahyuni, S.Kom, M.Kom

NIP. 19841201 2021212 005



Reisa Permatasari, S.Kom, M.Kom

NIP. 19920514 2022032 007

Mengetahui,

Ketua Program Studi Sistem Informasi

Fakultas Ilmu Komputer

Universitas Pembangunan Nasional "Veteran" Jawa Timur



Agung Brasama Putra, S.Kom, M.Kom

NIP. 19851124 2021211 003

LEMBAR PENGESAHAN

**ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN APLIKASI INDRIVE
MENGUNAKAN BIDIRECTIONAL ENCODER REPRESENTATIONS FROM
TRANSFORMERS (BERT)**

Disusun Oleh:

Asri Kinanti Febriany

19082010052

Telah disetujui mengikuti Ujian Negara Lisan Gelombang November

Periode 2023 pada Tanggal 27 November 2023

Menyetujui,

Dosen Pembimbing 1

Dosen Pembimbing 2



Eka Dyar Wahvuni, S.Kom, M.Kom

NIP. 19841201 2021212 005



Reisa Permatasari, S.Kom, M.Kom

NIP. 19920514 2022032 007

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 : Asri Kinanti Febriany

NPM : 19082010052

Program Studi : Sistem Informasi

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

**ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN APLIKASI INDRIVE MENGGUNAKAN
BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS (BERT)**

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

Reisa Permatasari, S.Kom, M.Kom

NIP. 19920514 2022032 007

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR
FAKULTAS ILMU KOMPUTER
KETERANGAN REVISI

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Asri Kinanti Febriany

NPM : 19082010052

Program Studi : Sistem Informasi

Menyatakan bahwa Judul Skripsi / Tugas Akhir sebagai berikut:

**ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN APLIKASI INDRIVE
MENGUNAKAN BIDIRECTIONAL ENCODER REPRESENTATIONS FROM
TRANSFORMERS (BERT)**

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,



Asri Kinanti Febriany

NPM. 19082010052

Judul : **ANALISIS SENTIMEN BERBASIS ASPEK PADA
ULASAN APLIKASI INDRIVE MENGGUNAKAN
BIDIRECTIONAL ENCODER
REPRESENTATIONS FROM TRANSFORMERS
(BERT)**

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

Pembimbing 2 : **Reisa Permatasari, S.Kom, M.Kom.**

ABSTRAK

Ekonomi digital di Indonesia turut disumbangkan oleh bisnis transportasi online. Salah satu pendatang baru pada bisnis transportasi online ini adalah inDrive. Rating ulasan dari google play store tidak cukup untuk menggambarkan kualitas sebuah aplikasi, oleh karena itu diperlukannya Analisis Sentimen Berbasis Aspek (ABSA) yang pada dasarnya dapat mengidentifikasi aspek dan mengidentifikasi sentimen dari aspek tersebut. Dengan pendekatan LDA (Latent Dirichlet Allocation) didapatkan 3 aspek utama yang ada pada dataaset periode Januari hingga Juli yaitu fitur lelang, sistem aplikasi, tarif & layanan transportasi. Kelebihan fitur lelang mencakup kemudahan pengguna dalam beraktivitas, namun beberapa driver mengeluhkan penggunaan autobid melalui bot atau APK mod. Sistem aplikasi memudahkan pemesanan, meskipun ada keluhan terkait potongan saldo dan kesulitan mendapatkan pesanan akibat penggunaan autobid. Pelanggan menyukai tarif & layanan transportasi, sementara para driver mengeluhkan ambang batas tarif yang rendah. Analisis trend menunjukkan ulasan fitur lelang seimbang, dengan peningkatan keluhan driver pada bulan Juli. Sistem aplikasi mendominasi ulasan negatif, tetapi mengalami penurunan setelah perbaikan pada bulan Maret. Sentimen positif mendominasi pada tarif & layanan transportasi, mencapai puncak pada bulan Juni meskipun turun pada bulan Juli. Implementasi model ABSA menggunakan IndoBERT, dan hasil uji validasi dengan 100 data ulasan baru menunjukkan keberhasilan dengan akurasi 92% untuk pengujian aspek dan 89% untuk pengujian sentimen.

Kata kunci : **Analisis Sentimen Berbasis Aspek, IndoBERT, inDrive**

KATA PENGANTAR

Segala puji syukur penulis panjatkan atas kehadiran Allah SWT karena atas berkat dan hidayahnya, penulis mampu menyelesaikan skripsi yang berjudul “Analisis Sentimen Berbasis Aspek Terhadap Ulasan Aplikasi inDrive Menggunakan Bidirectional Encoder Representations from Transformers” sebagai syarat untuk menyelesaikan program studi Sistem Informasi di Universitas Pembangunan Nasional Veteran Jawa Timur.

Dalam upaya menyelesaikan skripsi ini, tidak dapat dilepaskan dari dukungan yang diterima dari berbagai pihak, dan dalam kesempatan ini penulis ingin menyampaikan terima kasih kepada:

1. Kedua orang tua yang senantiasa doa dan dukungannya tidak pernah terputus, sehingga penulis sangat termotivasi untuk segera menyelesaikan skripsi ini.
2. Kakak dan adik saya yang senantiasa memberi dukungan dan pengertian selama proses pengerjaan skripsi ini.
3. Ibu Eka Dyar Wahyuni, S.Kom, M.Kom, dan Ibu Reisa Permatasari, S.Kom, M.Kom, selaku dosen pembimbing yang telah membimbing dengan sabar, memberikan ilmu yang bermanfaat serta memberi dukungan sehingga penulis dapat menyelesaikan skripsi ini.
4. Bapak Nur Cahyo Wibowo, S.Kom, M.Kom, selaku dosen wali yang selama ini memberikan motivasi dan arahan selama menjalani pendidikan di Sistem Informasi UPN Veteran Jatim.
5. Terima kasih kepada Nadia robbi, boling, ibna nirmala, ara, sekar, dariya emira dan teman-teman al-izzah saya yang lain yang senantiasa

mendengarkan keluh kesah dalam proses pengerjaan skripsi serta memberi dukungan dan kerap menjadi kawan seperjuangan skripsi saya.

6. Terima kasih kepada Catur, Hastri, Tiara, Fanni, Harun dan solasifo lainnya yang turut menemani dan memotivasi dalam suka dan duka pengerjaan serta menjadi rekan seperjuangan dalam menyelesaikan skripsi ini.
7. Terima kasih kepada Kak Ingkafi yang senantiasa membantu dan memberi pencerahan dikala saya sedang buntu dan tidak bersemangat dalam mengerjakan skripsi ini sehingga membuat saya termotivasi kembali dan akhirnya dapat menyelesaikan skripsi ini.

Semoga Allah membalas kebaikan pihak-pihak di atas, Dalam skripsi ini penulis menyadari masih banyak sekali kekurangan. Semoga skripsi ini bermanfaat bagi semua pihak.

Surabaya, Desember 2023

Penulis

DAFTAR ISI

ABSTRAK.....	ii
KATA PENGANTAR	iii
DAFTAR ISI.....	v
DAFTAR GAMBAR	ix
DAFTAR TABEL.....	xiii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	5
1.3 Batasan Masalah.....	6
1.4 Tujuan Penelitian	6
1.5 Manfaat Penelitian.....	6
1.6 Sistematika Penulisan	7
BAB II TINJAUAN PUSTAKA.....	9
2.1 Dasar Teori.....	9
2.1.1 InDrive	9
2.1.2 Text Mining.....	10
2.1.3 Natural Language Processing (NLP)	10
2.1.4 Data Preprocessing.....	11

2.1.5 Aspect-Based Sentiment Analysis (Analisis Sentimen Berbasis Aspek)	12
2.1.6 Vader	12
2.1.7 Latent Dirichlet Allocation	13
2.1.8 <i>Random Under Sampling</i>	14
2.1.9 BERT	14
2.1.10 Confusion Matrix	17
2.1.11 Streamlit	18
2.2 <i>Penelitian Terdahulu</i>	19
 BAB III METODOLOGI PENELITIAN	25
3.1 <i>Studi Literatur</i>	26
3.2 <i>Analisis Kebutuhan</i>	26
3.2.1 Kebutuhan Data.....	26
3.2.2 Kebutuhan Software dan Hardware	26
3.3 <i>Data Collecting</i>	27
3.4 <i>Data Preprocessing</i>	27
3.5 <i>Topic Modelling</i>	28
3.6 <i>Labeling</i>	29
3.6.1 Labeling Data Aspek.....	29
3.6.2 Labeling Data Sentimen.....	29
3.7 <i>Klasifikasi</i>	29
3.7.1 Pembagian Data	29

3.7.2 Word Embedding	29
3.7.3 Klasifikasi Sentimen Menggunakan BERT	30
3.7.4 Klasifikasi Aspek Menggunakan BERT	30
3.8 <i>Evaluasi</i>	30
3.9 <i>Exploratory Data Analysis dan Analisis Trend</i>	31
3.10 <i>Visualisasi dan Validasi</i>	31
BAB IV HASIL DAN PEMBAHASAN	32
4.1 <i>Analisis Kebutuhan</i>	32
4.1.1 <i>Kebutuhan Data</i>	32
4.1.2 <i>Kebutuhan Software dan Hardware</i>	32
4.2 <i>Data Collecting</i>	33
4.3 <i>Data Preprocessing</i>	36
4.4 <i>Topic Modeling</i>	45
4.4.1 <i>Penentuan Jumlah Topik</i>	46
4.4.2 <i>Penentuan Nama Aspek</i>	48
4.5 <i>Labeling</i>	49
4.5.1 <i>Labeling Data Aspek</i>	49
4.5.2 <i>Labeling Data Sentimen</i>	51
4.6 <i>Klasifikasi</i>	56
4.6.1 <i>Pembagian Data</i>	57
4.6.2 <i>Word Embedding</i>	57
4.6.3 <i>Klasifikasi Sentimen Menggunakan BERT</i>	62

4.6.4 Klasifikasi Aspek Menggunakan BERT	63
4.7 <i>Evaluasi</i>	65
4.7.1 Klasifikasi Sentimen Menggunakan BERT	65
4.7.2 Klasifikasi Aspek Menggunakan BERT	68
4.8 <i>Exploratory Data Analysis dan Analisis Trend</i>	71
4.8.1 Exploratory Data Analysis	71
4.8.2 Analisis Trend	72
4.9 <i>Visualisasi dan Validasi</i>	87
4.9.1 Visualisasi	87
4.9.2 Validasi	88
 BAB V PENUTUP.....	 91
5.1 <i>Kesimpulan</i>	91
5.2 <i>Saran</i>	92
 DAFTAR PUSTAKA	 94
 LAMPIRAN.....	 98

DAFTAR GAMBAR

Gambar 1.1 Tingkat Penggunaan Transportasi Online di Indonesia	2
Gambar 2.1 Ilustrasi 12 Layer Dalam Model BERT	15
Gambar 3.1 Alur Metodologi Penelitian	25
Gambar 3.2 Alur Data Preprocessing	27
Gambar 4.1 <i>Source Code</i> Pengambilan Data	33
Gambar 4.2 Wordcloud Sebelum Tahap Data Preprocessing	35
Gambar 4.4 <i>Source Code</i> Data Cleaning	38
Gambar 4.5 <i>Source Code</i> Normalisasi	40
Gambar 4.6 <i>Source Code</i> Stopwords Removal	42
Gambar 4.7 <i>Source Code</i> Stemming	44
Gambar 4.8 <i>Source Code Coherence Score</i>	47
Gambar 4.9 Grafik Hasil Nilai Koherensi.....	47
Gambar 4.10 <i>Source Code</i> Mengekstrak Kata Kunci dari Topik Utama Dataset.	48
Gambar 4.11 <i>Source Code</i> Penentuan <i>Dominant Topic</i> pada Tiap baris di Dataset	50
Gambar 4.12 <i>Source Code</i> Menerjemahkan Data.....	51
Gambar 4.13 <i>Source Code</i> Pelabelan Vader	52
Gambar 4.14 <i>Source Code</i> Pelabelan TextBlob.....	52
Gambar 4.14 Grafik Persebaran Jumlah Sentimen pada Tiap Aspek	56
Gambar 4.15 Ilustrasi Tokenisasi.....	57
Gambar 4.16 <i>Source Code</i> Tokenisasi	58
Gambar 4.17 Ilustrasi Menambahkan [CLS] dan [SEP].....	58
Gambar 4.18 <i>Source Code</i> Menambahkan [CLS] dan [SEP] Tokens	58

Gambar 4.19 Source Code Menambahkan Padding	59
Gambar 4.20 Ilustrasi Konversi Tokens ke ID.....	59
Gambar 4.21 Source Code Mengonversi Tokens ke ID	60
Gambar 4.22 Ilustrasi Sentence Embedding	60
Gambar 4.23 Source Code Menambahkan Sentence Embedding.....	60
Gambar 4.24 Ilustrasi Positional Embedding	61
Gambar 4.25 Source Code Menambahkan Positional Embedding	61
Gambar 4.26 Source Code Fine-tuning Sentimen	62
Gambar 4.27 Source Code Fine-tuning Aspek	64
Gambar 4.28 <i>Confusion Matrix</i> Skenario Terbaik Klasifikasi Sentimen	67
Gambar 4.29 Training Validation Accuracy dan Loss.....	67
Gambar 4.30 <i>Confusion Matrix</i> Skenario Terbaik Klasifikasi Aspek	70
Gambar 4.31 Training Validation Accuracy dan Loss.....	70
Gambar 4.32 Grafik Persebaran Jumlah Sentimen pada Tiap Aspek	71
Gambar 4.33 Wordcloud Data Setelah Melalui Tahapan Preprocessing	72
Gambar 4.34 Grafik Analisis Trend Aspek per Bulan	73
Gambar 4.35 Grafik Analisis Trend Sentimen per Bulan	73
Gambar 4.36 Grafik Analisis Trend Aspek Fitur Lelang per Bulan	74
Gambar 4.37 Wordcloud Sentimen Positif pada Aspek Fitur Lelang.....	75
Gambar 4.38 Grafik 10 Kata yang Paling Sering Muncul dengan Sentimen Positif	76
Gambar 4.37 Wordcloud Sentimen Negatif pada Aspek Fitur Lelang	76
Gambar 4.38 Grafik 10 Kata yang Paling Sering Muncul dengan Sentimen Negatif	77

Gambar 4.39 Wordcloud Sentimen Netral pada Aspek Fitur Lelang	77
Gambar 4.40 Grafik 10 Kata yang Paling Sering Muncul dengan Sentimen Netral	78
Gambar 4.41 Grafik Analisis Trend Aspek Sistem Aplikasi per Bulan	79
Gambar 4.42 Wordcloud Sentimen Positif pada Aspek Sistem Aplikasi	80
Gambar 4.43 Grafik 10 Kata yang Paling Sering Muncul dengan Sentimen Positif	80
Gambar 4.44 Wordcloud Sentimen Negatif pada Aspek Sistem Aplikasi	81
Gambar 4.45 Grafik 10 Kata yang Paling Sering Muncul dengan Sentimen Negatif	81
Gambar 4.46 Wordcloud Sentimen Netral pada Aspek Sistem Aplikasi.....	82
Gambar 4.47 Grafik 10 Kata yang Paling Sering Muncul dengan Sentimen Netral	83
Gambar 4.48 Grafik Analisis Trend Aspek Tarif & Layanan Transportasi per Bulan	83
Gambar 4.49 Wordcloud Sentimen Positif pada Aspek Tarif & Layanan Transportasi.....	84
Gambar 4.50 Grafik 10 Kata yang Paling Sering Muncul dengan Sentimen Positif	85
Gambar 4.51 Wordcloud Sentimen Negatif pada Aspek Tarif & Layanan Transportasi.....	85
Gambar 4.52 Grafik 10 Kata yang Paling Sering Muncul dengan Sentimen Negatif	86

Gambar 4.53 Wordcloud Sentimen Negatif pada Aspek Tarif & Layanan Transportasi.....	87
Gambar 4.54 Grafik 10 Kata yang Paling Sering Muncul dengan Sentimen Netral	87
Gambar 4.55 Source Code Load Model.....	88
Gambar 4.56 Hasil Pengembangan Website	88

DAFTAR TABEL

Tabel 2.1 Confusion Matrix	18
Tabel 4.1 Contoh Hasil Data Collecting	35
Tabel. 4.4 Hasil Data Cleaning	39
Tabel 4.5 Hasil Normalisasi	40
Tabel 4.6 Hasil Stopwords Removal.....	42
Tabel 4.7 Hasil Stemming	44
Tabel 4.8 Contoh Isi Corpus.....	46
Tabel 4.9 Contoh TF-IDF	46
Tabel 4.10 Interpretasi Aspek	48
Tabel 4.11 Contoh Pelabelan Apek Berdasarkan Hasil <i>Dominant Topic</i>	50
Tabel 4.12 Tampilan Data Setelah Dilabeli Vader dan TextBlob.....	53
Tabel 4.13 Data Setelah Dilabeli Vader TextBlob, dan Manual.....	54
Tabel 4.14 Skenario Pengujian Sentimen.....	63
Tabel 4.15 Skenario Pengujian Aspek	64
Tabel 4.16 <i>Classification Report</i> Klasifikasi Sentimen.....	65
Tabel 4.17 <i>Classification Report</i> Klasifikasi Aspek.....	68

