

**ANALISIS SENTIMEN PADA ULASAN APLIKASI DIGITAL  
KORLANTAS POLRI MENGGUNAKAN MODEL  
*DISTILBERT***

**SKRIPSI**

**Diajukan untuk memenuhi salah satu syarat kelulusan  
di Program Studi Sains Data**



**Disusun Oleh:**

**NABILA RIZKY AMALIA PUTRI**

**20083010018**

**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR  
FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI SAINS DATA  
SURABAYA**

**2024**

**LEMBAR PENGESAHAN**

**ANALISIS SENTIMEN PADA ULASAN DIGITAL KORLANTAS POLRI  
MENGUNAKAN MODEL *DISTILBERT***

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Sains Data  
pada : **Senin, 15 Juli 2024**

**Program Studi S-1 Sains Data**

**Fakultas Ilmu Komputer**

**Universitas Pembangunan Nasional Veteran Jawa Timur  
Surabaya**

Oleh :

**Nabila Rizky Amalia Putri**

**NPM. 20083010018**

Disetujui oleh Tim Penguji Skripsi :

Penguji 1

**Amri Muhaimin, S.Stat., M.Stat., M.S.**  
NIP. 21119950723270

Penguji 2

**Tresna Maulana Fahrudin, S.ST., M.T.**  
NIP. 1993050120220301007

Pembimbing 1

**Trimono, S.Si., M.Si**  
NIP. 199509082022031003

Pembimbing 2

**Aviolla Terza Damaliana, S.Si., M.Stat**  
NIP. 199408022022032015

Fakultas Ilmu Komputer  
Dekan,

**Prof. Dr. Ir. Novirina Hendrasarie, MT**  
NIP. 196811261994032001

Mengetahui,

Program Studi Sains Data  
Fakultas Ilmu Komputer  
Koordinator,

**Dr. Eng. Ir. Dwi Arman Prasetya, ST, MT, IPU.**  
NIP. 198012052005011002

**Surabaya, Juli 2024**

## SURAT PERNYATAAN

Saya yang *BERT*anda tangan di bawah ini:

Nama : Nabila Rizky Amalia Putri  
NPM : 20083010018  
Program Studi : Sains Data

Menyatakan bahwa judul Skripsi / Tugas Akhir sebagai berikut:

### **ANALISIS SENTIMEN PADA ULASAN APLIKASI DIGITAL KORLANTAS POLRI MENGGUNAKAN MODEL *DISTILBERT***

Bukan merupakan plagiat dari Skripsi/ Tugas Akhir/ Penelitian orang lain dan juga bukan merupakan produk/ *software*/ 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 ini terbukti tidak benar, maka Saya bertanggung jawab penuh dan siap menerima segala konsekuensi, termasuk pembatalan ijazah dikemudian hari

Surabaya, 5 Juli 2024

Hormat Saya



Nabila Rizky Amalia Putri

NPM. 20083010018

## ABSTRAK

### ANALISIS SENTIMEN PADA ULASAN APLIKASI DIGITAL KORLANTAS POLRI MENGGUNAKAN MODEL *DISTILBERT*

**Nama Mahasiswa / NPM** : Nabila Rizky Amalia Putri / 20083010018  
**Program Studi** : Sains Data, FASILKOM, UPN Veteran Jatim  
**Dosen Pembimbing 1** : Trimono, S.Si., M.Si  
**Dosen Pembimbing 2** : Aviolla Terza Damaliana, S.Si., M.Stat

#### Abstrak

Penerapan digitalisasi dalam pelayanan publik oleh Korlantas Polri membantu mempercepat administrasi, memperluas akses, dan meningkatkan kualitas layanan. Diluncurkan pada April 2021, aplikasi Digital Korlantas Polri telah diunduh lebih dari 5 juta kali di Google Play Store dengan rating 3,7 dan sekitar 110 ribu ulasan. Banyak kritik dapat mempengaruhi reputasi aplikasi, sehingga analisis sentimen diperlukan untuk mengelompokkan ulasan pengguna menjadi positif, negatif, atau netral, membantu pengembang mengenali kekurangan aplikasi. Penelitian ini menggunakan *DistilBERT*, model *deep learning* yang distilasi dari *BERT*, untuk mengevaluasi efektivitas analisis sentimen ulasan. Data dikumpulkan dari ulasan pengguna di Google Play Store dalam rentang waktu 1 September 2023 hingga 31 Mei 2024, menghasilkan 8.752 ulasan untuk dataset. Model dievaluasi menjadi tiga rasio data: 60:20:20, 70:15:15, dan 80:10:10. Hasil dari penelitian ini performa terbaik pada rasio 80:10:10, mencapai akurasi 88% dengan menggunakan *hyperparameter* (*batch size* 16, *learning rate*  $2e-5$ , dan *epoch* 10). Peningkatan rasio data latih menunjukkan dampak positif pada kinerja model. Namun, model masih kesulitan mengklasifikasikan sentimen netral, menunjukkan perlunya peningkatan lebih lanjut dalam mendeteksi kelas ini. Hasil penelitian ini diharapkan memberikan wawasan penting bagi pengembangan sistem analisis sentimen yang lebih canggih dan andal di masa depan.

**Kata Kunci:** Digitalisasi, Digital Korlantas Polri, Analisis sentimen, DistilBERT

## ***ABSTRACT***

### **SENTIMENT ANALYSIS ON DIGITAL KORLANTAS POLRI APPLICATION REVIEWS USING THE DISTILBERT MODEL**

**Student Name / NPM** : Nabila Rizky Amalia Putri / 20083010018  
**Study Program** : Sains Data, FASILKOM,UPN Veteran Jatim  
**Advisor 1** : Trimono, S.Si., M.Si  
**Advisor 2** : Aviolla Terza Damaliana, S.Si., M.Stat

#### **Abstract**

The application of digitalisation in public services by Korlantas Polri helps speed up administration, expand access, and improve service quality. Launched in April 2021, the Korlantas Polri Digital app has been downloaded more than 5 million times on the Google Play Store with a rating of 3.7 and around 110 thousand reviews. Many criticisms can affect the app's reputation, so sentiment analysis is needed to classify user reviews into positive, negative, or neutral, helping developers recognise the app's shortcomings. This study uses DistilBERT, a deep learning model distilled from BERT, to evaluate the effectiveness of review sentiment analysis. Data was collected from user reviews on the Google Play Store in the time span of 1 September 2023 to 31 May 2024, resulting in 8,752 reviews for the dataset. The model was evaluated into three data ratios: 60:20:20, 70:15:15, and 80:10:10. The results of this study showed the best performance at 80:10:10 ratio, achieving 88% accuracy using hyperparameters (batch size 16, learning rate  $2e-5$ , and epoch 10). Increasing the ratio of training data shows a positive impact on model performance. However, the model still struggles to classify neutral sentiment, indicating the need for further improvement in detecting this class. The results of this study are expected to provide important insights for the development of more sophisticated and reliable sentiment analysis systems in the future.

**Keywords:** Digitalization, Digital Korlantas Polri, sentiment analysis, DistilBERT.

## KATA PENGANTAR

Puji dan syukur kehadirat ALLAH SWT, atas limpahan Rahmat serta Kasih Sayang-Nya sehingga penulis dapat menyelesaikan Skripsi dengan judul “Analisis Sentimen Pada Ulasan Aplikasi Digital Korlantas Polri Menggunakan Model *Distilbert*” sebagai syarat untuk menyelesaikan Program Studi S1 Sains Data di Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Dalam penyusunan Skripsi ini tidak terlepas dari bantuan berbagai pihak, dan dalam kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Kedua orang tua dan keluarga penulis yang selalu memberikan doa, motivasi, dan dukungan moril dalam menyelesaikan skripsi ini.
2. Bapak Dr. Eng Ir. Dwi Arman Prasetya, S.T., M.T., IPU selaku Koordinator Prodi Sains Data.
3. Bapak Trimono, S.Si., M.Si selaku dosen pembimbing utama Ibu Aviolla Terza Damaliana, S.Si., M.Stat selaku dosen pembimbing kedua yang telah meluangkan waktunya memberikan bimbingan, arahan, dan saran selama penyusunan skripsi.
4. Bapak Amri Muhaimin, S.Stat., M.Stat., M.S dan bapak Tresna Maulana Fahrudin, S.ST., M.T selaku dosen penguji yang telah memberikan saran untuk perbaikan skripsi ini.
5. Teman penulis Hanis, Kiki, Ken, Febrianti, dan Ina yang telah selalu memberikan semangat dan membantu penulis selama mengerjakan skripsi ini.
6. Teman-teman dari Progam Studi Sains Data angkatan 2020 yang telah membantu, mendukung dan memberikan semangat.
7. Semua pihak yang tidak bisa disebutkan, terima kasih telah membantu hingga penulis bisa menyelesaikan skripsi ini.

Penulis menyadari bahwa penelitian dan penulisan Skripsi ini masih jauh dari kata kesempurnaan, namun penulis berharap semoga Skripsi ini dapat memberikan kontribusi terhadap perkembangan ilmu pengetahuan, khususnya dalam bidang ilmu Sains Data.

Surabaya, 5 Juli 2024

Penulis

## DAFTAR ISI

LEMBAR PENGESAHAN .....	ii
SURAT PERNYATAAN .....	iii
ABSTRAK.....	iv
<i>ABSTRACT</i> .....	v
KATA PENGANTAR .....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR .....	x
DAFTAR TABEL.....	xii
DAFTAR LAMPIRAN.....	xiv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	5
1.3. Batasan Masalah .....	5
1.4. Tujuan Penelitian .....	5
1.5. Manfaat Penelitian .....	6
BAB II TINJAUAN PUSTAKA .....	7
2.1. Dasar Teori.....	7
2.1.1. Aplikasi <i>Digital Korlantas Polri</i> .....	7
2.1.2. Analisis Sentimen .....	8
2.1.3. <i>Natural Language Processing (NLP)</i> .....	8
2.1.4. <i>Text Preprocessing</i> .....	9
2.1.5. <i>Lexicon Based Labeling</i> .....	10
2.1.6. <i>Transformers</i> .....	10
2.1.7. <i>BERT</i> .....	12
2.1.8. <i>DistilBERT</i> .....	15
2.1.9. <i>Hyperparameter</i> .....	16
2.1.10. <i>Confusion Matrix</i> .....	17
2.2. Penelitian Terdahulu .....	19
BAB III METODOLOGI PENELITIAN .....	24
3.1. Variabel Penelitian dan Sumber Data .....	24
3.2. Langkah Analisis .....	25

3.2.1	Pengumpulan data .....	25
3.2.2	<i>Preprocessing Data</i> .....	26
3.2.3	<i>Data Labelling</i> .....	29
3.2.4	<i>Split Dataset</i> .....	29
3.2.5	Implementasi DistilBERT .....	30
3.2.6	Evaluasi Model .....	30
3.2.7	Analisis Hasil .....	30
3.3.	Diagram Alir Penelitian .....	31
3.4.	Jadwal Penelitian .....	31
BAB IV HASIL DAN PEMBAHASAN .....		33
4.1.	Pengumpulan Data .....	33
4.2.	<i>Preprocessing Data</i> .....	35
4.2.1	<i>Case Folding</i> .....	36
4.2.2	<i>Data Cleaning</i> .....	36
4.2.3	<i>Tokenizing</i> .....	37
4.2.4	<i>Normalization</i> .....	38
4.2.5	<i>Stopword Removal</i> .....	39
4.2.6	<i>Stemming</i> .....	40
4.3.	<i>Labelling</i> .....	41
4.4.	<i>Split Dataset</i> .....	42
4.5.	Implementasi <i>DistilBERT</i> .....	42
4.5.1	Skenario 1 (Rasio Pembagian 60:20:20).....	46
4.5.2	Skenario 2 (Rasio Pembagian 70:15:15).....	48
4.5.3	Skenario 3 (Rasio Pembagian 80:10:10).....	49
4.6.	Evaluasi model.....	51
4.6.1	Skenario 1 (Rasio Pembagian 60:20:20).....	51
4.6.2	Skenario 2 (Rasio Pembagian 70:15:15).....	54
4.6.3	Skenario 3 (Rasio Pembagian 80:30).....	57
BAB V PENUTUP .....		63
5.1.	Kesimpulan .....	63
5.2.	Saran .....	63
DAFTAR PUSTAKA .....		65



LAMPIRAN.....	70
BIODATA PENULIS .....	73

## DAFTAR GAMBAR

<b>Gambar 2.1.</b> Arsitektur Transformers .....	11
<b>Gambar 2.2.</b> Arsitektur Encoder dan Decoder pada Transformers .....	11
<b>Gambar 2.3.</b> Arsitektur BERT.....	13
<b>Gambar 2.4.</b> Tahapan pre-training dan fine-tuning .....	14
<b>Gambar 2.5.</b> Arsitektur DistilBERT.....	15
<b>Gambar 3.1.</b> Diagram alur Preprocessing Data .....	26
<b>Gambar 3.2.</b> Alur Metodologi Penelitian .....	31
<b>Gambar 4.1.</b> Hasil Case Folding .....	36
<b>Gambar 4.2.</b> Hasil Cleaning .....	37
<b>Gambar 4.3.</b> Hasil Tokenizing.....	38
<b>Gambar 4.4.</b> hasil Normalization.....	39
<b>Gambar 4.5.</b> Hasil Stopword Removal.....	40
<b>Gambar 4.6.</b> Hasil Stemming .....	41
<b>Gambar 4.7.</b> Hasil Labelling .....	41
<b>Gambar 4.8.</b> Hasil Sentimen.....	41
<b>Gambar 4.9.</b> Grafik Akurasi Training dan Test rasio 60:20:20 dengan <i>epoch 15</i> dan <i>learning rate 1e-5</i> .....	47
<b>Gambar 4.10.</b> Grafik Akurasi Training dan Test rasio 60:20:20 dengan <i>epoch 10</i> dan <i>learning rate 2e-5</i> .....	48
<b>Gambar 4.11.</b> Grafik Akurasi Training dan Test rasio 70:15:15 dengan <i>epoch 15</i> dan <i>learning rate 1e-5</i> .....	48
<b>Gambar 4.12.</b> Grafik Akurasi Training dan Test rasio 70:15:15 dengan <i>epoch 10</i> dan <i>learning rate 2e-5</i> .....	49
<b>Gambar 4.13.</b> Grafik Akurasi Training dan Test rasio 80:10:10 dengan <i>epoch 15</i> dan <i>learning rate 1e-5</i> .....	50
<b>Gambar 4.14.</b> Grafik Akurasi Training dan Test rasio 80:10:10 dengan <i>epoch 10</i> dan <i>learning rate 2e-5</i> .....	50
<b>Gambar 4.15.</b> Hasil Performa model dengan rasio 60:20:20 dengan <i>epoch 15</i> dan <i>learning rate 1e-5</i> .....	52
<b>Gambar 4.16.</b> <i>Confusion Matrix</i> dengan rasio 60:20:20 dengan <i>epoch 15</i> dan <i>learning rate 1e-5</i> .....	52

<b>Gambar 4.17.</b> Hasil Performa model dengan rasio 60:20:20 dengan <i>epoch</i> 10 dan <i>learning rate</i> 2e-5 .....	53
<b>Gambar 4.18.</b> <i>Confusion Matrix</i> dengan rasio 60:20:20 dengan <i>epoch</i> 10 dan <i>learning rate</i> 2e-5 .....	54
<b>Gambar 4.19.</b> Hasil Performa model dengan rasio 70:15:15 dengan <i>epoch</i> 15 dan <i>learning rate</i> 1e-5 .....	55
<b>Gambar 4.20.</b> <i>Confusion Matrix</i> dengan rasio 70:15:15 dengan <i>epoch</i> 15 dan <i>learning rate</i> 1e-5 .....	55
<b>Gambar 4. 21.</b> Hasil Performa model dengan rasio 70:15:15 dengan epoch 10 dan <i>learning rate</i> 2e-5 .....	56
<b>Gambar 4.22.</b> <i>Confusion Matrix</i> dengan rasio 70:15:15 dengan epoch 10 dan <i>learning rate</i> 2e-5 .....	57
<b>Gambar 4.23.</b> Hasil Performa model dengan rasio 80:10:10 dengan <i>epoch</i> 15 dan <i>learning rate</i> 1e-5 .....	58
<b>Gambar 4.24.</b> <i>Confusion Matrix</i> dengan rasio 80:10:10 dengan <i>epoch</i> 15 dan <i>learning rate</i> 1e-5 .....	58
<b>Gambar 4.25.</b> Hasil Performa model dengan rasio 80:10:10 dengan epoch 10 dan <i>learning rate</i> 2e-5 .....	59
<b>Gambar 4.26.</b> <i>Confusion Matrix</i> dengan rasio 80:10:10 dengan epoch 10 dan <i>learning rate</i> 2e-5 .....	60

## DAFTAR TABEL

<b>Tabel 2.1.</b> Confusion Matrix .....	17
<b>Tabel 2.2.</b> Tabel Studi Literatur yang Relevan dalam Penelitian.....	19
<b>Tabel 3.1.</b> Data scrapping Ulasan (1) .....	24
<b>Tabel 3.2.</b> Data scrapping Ulasan (2) .....	24
<b>Tabel 3.3.</b> Contoh Case Folding.....	26
<b>Tabel 3.4.</b> Contoh Cleaning.....	27
<b>Tabel 3.5.</b> Contoh Tokenizing.....	27
<b>Tabel 3.6.</b> Kamus Normalisasi .....	28
<b>Tabel 3.7.</b> Contoh Stopword Removal .....	28
<b>Tabel 3.8.</b> Contoh Stemming.....	29
<b>Tabel 3.9.</b> Jadwal Penelitian.....	32
<b>Tabel 4.1.</b> Kode Scrapping Data .....	33
<b>Tabel 4.2.</b> Data scrapping Ulasan (1) .....	34
<b>Tabel 4.3.</b> Data scrapping Ulasan (2) .....	34
<b>Tabel 4.4.</b> Filter untuk data bulan Setember – Mei .....	34
<b>Tabel 4.5.</b> Hasil Proses Filtering Data.....	35
<b>Tabel 4.6.</b> Kode Case Folding .....	36
<b>Tabel 4.7.</b> kode Cleaning.....	36
<b>Tabel 4.8.</b> Kode Tokenizing .....	37
<b>Tabel 4.9.</b> kode normalization.....	38
<b>Tabel 4.10.</b> kode Stopword Removal .....	39
<b>Tabel 4.11.</b> Kode Stemming.....	40
<b>Tabel 4.12.</b> Rasio pembagian dataset .....	42
<b>Tabel 4.13.</b> Menetapkan Seed acak .....	43
<b>Tabel 4.14.</b> Inisialisasi Tokenizer.....	43
<b>Tabel 4.15.</b> Kode Persiapan data dengan TensorFlow .....	43
<b>Tabel 4.16.</b> Inisialisasi model DistilBERT .....	44
<b>Tabel 4.17.</b> Implementasi DistilBERT .....	44
<b>Tabel 4.18.</b> Training model DistilBERT .....	46
<b>Tabel 4.19.</b> Hyperparameter DistilBERT .....	46

<b>Tabel 4.20.</b> Hasil Evaluasi Model DistilBERT dengan epoch 10 learning rate 1e-5 .....	61
<b>Tabel 4.21.</b> Hasil Evaluasi Model DistilBERT dengan epoch 10 learning rate 2e-5. .....	61

## DAFTAR LAMPIRAN

<b>Lampiran 1.</b> Hasil Uji Plagiasi .....	70
<b>Lampiran 2.</b> Data penelitian .....	71
<b>Lampiran 3.</b> Source Code.....	72