

**SENTIMEN ANALISIS PADA ULASAN APPS
GOOGLE PLAY BERBASIS BERT**

SKRIPSI



disusun oleh

Kukuh Setyo Budi

17.11.1652

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

**SENTIMEN ANALISIS PADA ULASAN APPS
GOOGLE PLAY BERBASIS BERT**

SKRIPSI

Untuk memenuhi sebagian persyaratan
Mencapai gelar Sarjana pada
Program Studi Informatika



disusun oleh

Kukuh Setyo Budi

17.11.1652

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

PERSETUJUAN

SKRIPSI

SENTIMEN ANALISIS PADA ULASAN APPS GOOGLE PLAY BERBASIS BERT

yang dipersiapkan dan disusun oleh

Kukuh Setyo Budi

17.11.1652

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 07 Juni 2021

Dosen Pembimbing,

Arif Akbarul Huda, S.SI, M.Eng

NIK. 190302287

PENGESAHAN

SKRIPSI

SENTIMEN ANALISIS PADA ULASAN APPS GOOGLE PLAY BERBASIS BERT

yang dipersiapkan dan disusun oleh

Kukuh Setyo Budi

17.11.1652

telah dipertahankan di depan Dewan Penguji
pada tanggal 17 Juni 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Lukman, M.Kom

NIK. 190302151

Ria Andriani, M.Kom

NIK. 190302458

Arif Akbarul Huda, S.Si, M.Eng

NIK. 190302287

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer

Tanggal 17 Juni 2021

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, S.Kom, M.Kom.

NIK. 190302096

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Sragen, 03 Juli 2021




Kukuh Setyo Budi

NIM. 17.11.1652

MOTO

“Weakness of attitude becomes weakness of character.”

- Albert Einstein.

“Quality is much better than quantity. One home run is much better than two doubles. “

- Steve Jobs

“I think if you do something and it turns out pretty good, then you should go do something else wonderful, not dwell on it for too long. Just figure out what’s next.”

- Steve Jobs

“Sometimes when you innovate, you make mistakes. It is best to admit them quickly, and get on with improving your other innovations.”

- Steve Jobs



PERSEMBAHAN

Alhamdulillah dengan kerja keras serta doa, skripsi ini dapat diselesaikan dengan baik. Segala puji dan syukur bagi Allah SWT yang tiada henti memberikan keberkahan. Dengan ini saya mempersembahkan skripsi ini kepada semua pihak yang terlibat secara langsung atau tidak langsung, yaitu untuk :

1. Kedua orang tua dan Adik-adik saya yang selalu mendoakan, selalu menyemangati dan memberikan motivasi tiada henti kepada saya.
2. Dosen pembimbing saya Arif Akbarul Huda S.Si, M.eng. yang telah membimbing saya dari awal sampai akhir pembuatan skripsi.
3. Dosen-dosen Universitas AMIKOM Yogyakarta yang telah memberikan banyak ilmu selama masa kuliah.
4. Teman-teman kelas 17-IF-11 yang telah menemani dan selalu memberikan semangat untuk menyelesaikan skripsi ini.

KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Allah SWT yang telah melimpahkan rahmat serta hidayah-Nya dan shawalat serta salam juga tidak lupa penulis panjatkan kepada junjungan kita Nabi Muhammad SAW yang telah memberikan teladan mulia dalam menuntun umatnya sehingga penulis dapat menyelesaikan skripsi ini.

Skripsi yang berjudul **“SENTIMEN ANALISIS PADA ULASAN APPS GOOGLE PLAY BERBASIS BERT”** ini disusun sebagai salah satu syarat utama untuk menyelesaikan program sarjana pada Universitas AMIKOM Yogyakarta. Penyelesaian skripsi ini juga tidak lepas dari bantuan berbagai pihak, karena itu pada kesempatan ini penulis ingin menyampaikan rasa hormat dan terima kasih kepada :

1. Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom, M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta
3. Ibu Windha Mega P.D M.kom selaku ketua Program Studi Informatika Universitas AMIKOM Yogyakarta.
4. Bapak Arif Akbarul Huda S.Si, M.eng. selaku dosen pembimbing yang selalu memberikan bimbingan, nasehat serta waktunya selama penulisan skripsi ini.

Penulis menyadari skripsi ini masih ada kekurangan. Maka, penulis menerima kritik dan saran yang membangun serta teguran dari semua pihak. Penulis menerima dengan lapang dada untuk kesempurnaan karya selanjutnya. Semoga skripsi yang sederhana ini bisa bermanfaat, khususnya bagi penulis dan pembaca yang budiman pada umumnya. Apabila terdapat kesalahan semoga Allah SWT melimpahkan magfirah-Nya. Aamiin.

Sragen, 03 Juli 2021



Kukul Setyo Budi

DAFTAR ISI

SENTIMEN ANALISIS PADA ULASAN APPS GOOGLE PLAY BERBASIS BERT	ii
PERSETUJUAN	iii
PENGESAHAN	iv
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR PERSAMAAN	xiv
INTISARI	xv
ABSTRACT	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Maksud dan Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Metodologi Penelitian	4
1.6.1 Studi Literatur	5
1.6.2 Analisis Permasalahan	5
1.6.3 Pengumpulan Data	5
1.6.4 Perancangan Sistem	5
1.6.5 Implementasi	5
1.6.6 Pengujian	5
1.6.6 Penyusunan Laporan	6
1.7 Sistematika Penulisan	6
BAB II LANDASAN TEORI	7
2.1 Kajian Pustaka	7

2.2 Dasar Teori.....	8
2.2.1 Sentimen Analisis	8
2.2.2 <i>Deep Learning</i>	12
2.2.3 BERT (<i>Bidirectional Encoder Representations from Transformers</i>).....	14
2.2.4 <i>Fine-tuning</i> BERT	16
2.2.5 PyTorch.....	17
2.2.6 IndoNLU <i>Benchmark</i>	18
2.2.7 <i>Confusion Matrix</i>	19
BAB III METODE PENELITIAN	21
3.1 Alat dan Bahan penelitian	21
3.2 Hipotesis	21
3.3 Alur Penelitian	21
3.3.1 <i>Scraping Data</i>	22
3.3.2 <i>Data Preprocessing</i>	22
3.3.3 <i>Labelling</i>	23
3.3.4 Pembagian Data	23
3.3.5 <i>Word Embedding</i> BERT	23
3.3.6 Analisis Hasil.....	23
BAB IV HASIL DAN PEMBAHASAN	25
4.1 Perangkat Pendukung.....	25
4.2 Pengumpulan Data.....	25
4.3 Dataset.....	27
4.4 Persiapan	28
Cloning repository IndoNLU	29
Import Library atau package	29
4.5 <i>Preprocessing</i>	29
4.5.1 <i>Data Cleaning</i>	30

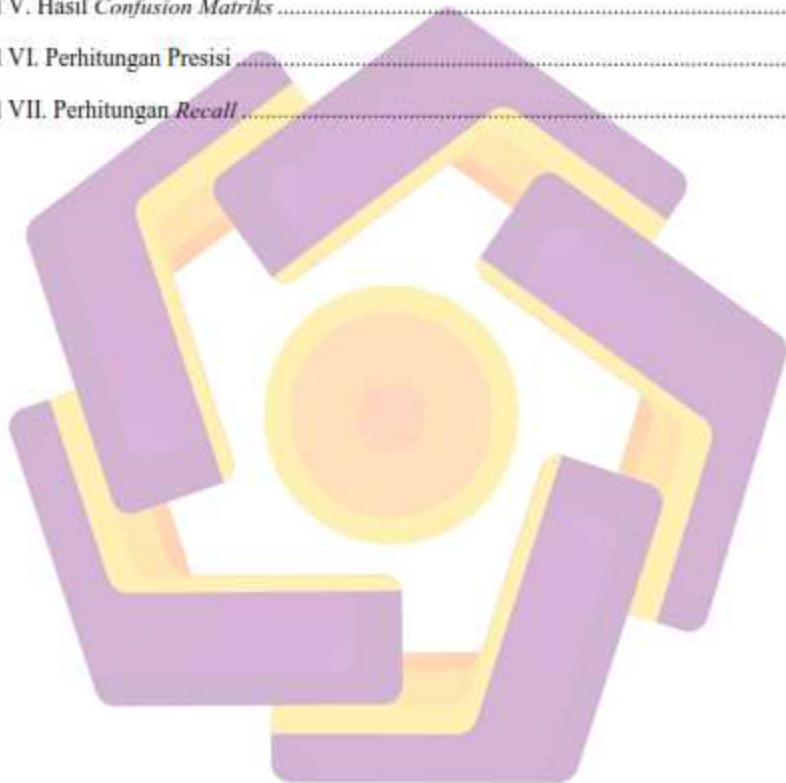
Data <i>Cleaning</i> , Penghapusan tanda baca, dan <i>Whitespace</i>	30
4.5.2 <i>Case Folding</i>	30
<i>Case Folding</i>	31
4.6 <i>Labelling</i>	31
4.6.1 Naïve Bayes Classifier	31
Analisis Sentiment Polarity menggunakan TextBlob	32
<i>Labelling</i> menggunakan Naïve Bayes Classifier	33
4.6.2 <i>Tokenizing</i>	36
Pemeriksaan kinerja <i>Tokenizing</i>	36
Deklarasi <i>Dataset</i>	38
Tahap <i>Tokenizing</i> pada <i>Dataset</i>	38
Memberi Konfigurasi tambahan pada <i>Dataset</i>	38
4.7 <i>Pemodelan</i>	39
<i>Optimizer Adam</i>	39
Menjalankan <i>Training</i>	39
Evaluasi data <i>valid</i>	40
4.8 <i>Testing model</i>	45
Evaluasi <i>Testing</i>	45
Fungsi <i>forward sequence classification</i>	46
4.9 Pengukuran Akurasi <i>Testing</i>	49
Menyalin label <i>pred.txt</i> ke dalam array	50
Perhitungan akurasi, <i>confusion matrix</i> dan <i>classification report</i>	50
Menyajikan hasil prediksi BERT	54
BAB V PENUTUP	57
5.1 Kesimpulan	57
5.2 Saran	57
DAFTAR PUSTAKA	58

DAFTAR GAMBAR

Gambar 2.1 Sentimen Analisis	8
Gambar 2.2 <i>Feedforward neural network</i>	13
Gambar 2.3 <i>Pre-training</i> dan <i>Fine-Tuning</i>	15
Gambar 2.4 Representasi <i>input</i> BERT. <i>Input embeddings</i> adalah jumlah token <i>embeddings</i> , <i>segmentation embeddings</i> dan <i>position embeddings</i>	16
Gambar 2.5 <i>fine-tuning hyperparameter</i>	17
Gambar 3.1 Flowchart	22
Gambar 3.2 <i>Output Confusion Matrix</i>	24
Gambar 4.1 <i>Scraping</i> data menggunakan <i>Webharvy</i>	26
Gambar 4.2. Ulasan bersifat SPAM	27
Gambar 4.3. Dataset	28
Gambar 4.4 Hasil klasifikasi dengan nilai <i>sentiment polarity</i>	33
Gambar 4.5 Hasil <i>Labelling</i> Naïve Bayes	35
Gambar 4.6 Detail hasil <i>labelling</i> Naïve Bayes Classifier	36
Gambar 4.7 Keluaran <i>Tokenizing</i>	37
Gambar 4.8 <i>Output</i> yang ditampilkan dari hasil <i>training</i>	42
Gambar 4.9 Visualisasi proses <i>training</i> dan evaluasi BERT <i>epoch</i> 2, 4, 6 dan 8.	43
Gambar 4.11 <i>Output Variabel Logits</i>	49
Gambar 4.12 <i>Output torch.topk()</i>	49
Gambar 4.13 <i>Output</i> akurasi, <i>classification report</i> , dan grafik <i>confusion matrix</i> BERT	51
Gambar 4.14 <i>Output</i> akurasi, <i>classification report</i> , dan grafik <i>confusion matrix</i> BERT dengan tambahan fungsi <i>stopword</i> pada tahapan <i>preprocessing</i>	53
Gambar 4.15 Visualisasi prediksi BERT	54

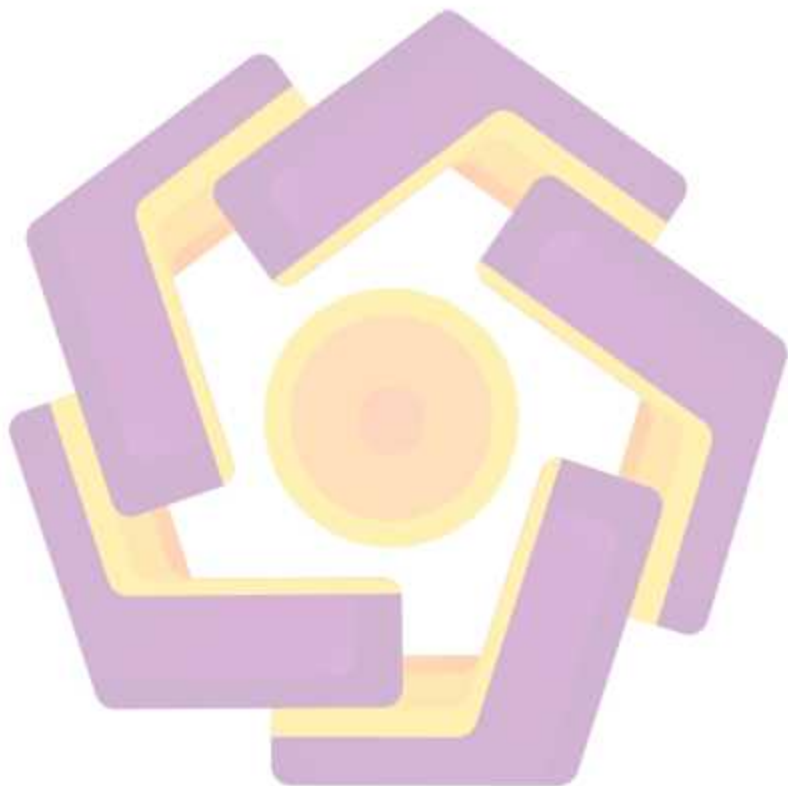
DAFTAR TABEL

Tabel I. Confusion Matrix	19
Tabel II. Perangkat dan <i>tools</i> yang digunakan.....	25
Tabel III. Sebelum dan Sesudah Preprocessing	31
Tabel IV. Detail hasil <i>train</i> setiap <i>epoch</i>	44
Tabel V. Hasil <i>Confusion Matriks</i>	55
Tabel VI. Perhitungan Presisi	55
Tabel VII. Perhitungan <i>Recall</i>	56



DAFTAR PERSAMAAN

Rumus Perhitungan Akurasi	19
Rumus Perhitungan Presisi	20
Rumus Perhitungan Recall.....	20



INTISARI

Google Play merupakan layanan distribusi digital yang berfungsi sebagai toko aplikasi resmi untuk sistem operasi Android. Kita bisa mendapatkan banyak informasi tentang aplikasi yang terpublish pada Google Play melalui ulasan yang diberikan oleh pengguna sebelumnya. Analisis Sentimen merupakan salah satu cabang dari Natural Language Processing (NLP) yang dapat membantu dalam menentukan kualitas suatu aplikasi dari ulasan yang telah diberikan pengguna. Penelitian ini menggunakan dataset hasil *scraping* pada ulasan Google Play berbahasa Indonesia sebagai bahan untuk menguji model yang digunakan dalam menjalankan tugas *analysis sentiment*.

Penelitian ini menggunakan model *Deep Learning* yaitu BERT atau *Bidirectional Encoder Representation from Transformer* sebagai metode *word embedding* untuk mempersentasikan kata menjadi vector. Model BERT akan dilakukan pengujian dalam menganalisis sentiment pada teks. Dataset yang digunakan merupakan hasil *scraping* menggunakan *software* WebHarvy pada ulasan Google Play berbahasa Indonesia. Pada tahapan *labelling* data pada penelitian ini menggunakan metode Naïve Bayes Classifier.

Dari percobaan yang dilakukan dengan menggunakan data sebanyak 3.963 data yang dibagi menjadi 60% data *training*, 20% data *valid*, dan 20% data *test*, model BERT dapat mendapatkan nilai akurasi sebesar 86%, dan mendapatkan akurasi terendah dengan nilai 82%.

Keyword : *Sentimen Analisis, BERT, Deep Learning, NLP*

ABSTRACT

Google Play is a digital distribution service that serves as the official app store for the Android operating system. We can get a lot of information about applications published on Google Play through reviews given by previous users. Sentiment Analysis is a branch of Natural Language Processing (NLP) that can help determine the quality of an application from reviews that have been given by users. This study uses the dataset scraping results in Google Play reviews Indonesian language as the material for the test model used in performing sentiment analysis tasks.

This study uses a Deep Learning model, namely BERT or Bidirectional Encoder Representation from Transformer as a word embedding method to represent words into vectors. The BERT method will be tested in analyzing sentiment in the text. The dataset used is the result of scraping using WebHarvy software on Google Play reviews in Indonesian. At the data labeling stage in this study using the Naïve Bayes Classifier method.

From experiments conducted using 3,963 data, which is divided into 60% training data, 20% valid data, and 20% test data, the BERT model can get an accuracy value of 86%, and get the lowest accuracy with a value of 82%.

Keyword : *Sentiment Analysis, BERT, Deep Learning, NLP*