

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

IMPLEMENTASI BERT UNTUK ANALISIS SENTIMEN

PENGGUNA GAMBAR LINE E-FOOTBALL PADA

ULASAN DAN STORE BERBASIS

MACHINE LEARNING



SKRIPSI

Disusun Oleh :

MUHAMMAD ARYARAKA PRADANA

NIM : 2102020119

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU TEKNIK
UNIVERSITAS BINA INSAN
2025**

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

HALAMAN PENGESAHAN SKRIPSI



**IMPLEMENTASI BERT UNTUK ANALISIS SENTIMEN
PENGGUNA GAME *ONLINE E-FOOTBALL* PADA
ULASAN DI *PLAY STORE* BERBASIS
*MACHINE LEARNING***

Oleh :

Muhammad Aryaraka Pradana

NIM : 2102020119

Pembimbing 1,

**Lubuklinggau, Januari 2025
Pembimbing II,**

Armanto, S.Kom., M.Kom

Fido Rizki, S.Kom., M.Kom

**Menyetujui,
Dekan Fakultas Ilmu Teknik
Universitas Bina Insan**

Dr. Rudi Kurniawan, St., M.Kom

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

HALAMAN PERSETUJUAN TIM PENGUJI



Pada hari.....Tanggal.....Bulan.....Tahun Dua Ribu Dua Puluh Lima telah dilaksanakan sidang skripsi oleh program studi Informatika Universitas Bina Insan Lubuklinggau.

Nama : Muhammad Aryaraka Pradana
NIM : 2102020119
Judul skripsi : Implementasi BERT untuk Analisis Sentimen Pengguna Game *Online E-Football* pada Ulasan di *Play Store* Berbasis *Machine Learning*

Komisi Penguji

1. Ketua : Armanto, S.Kom., M.Kom (.....)
2. Sekretaris : Fido Rizki, S.Kom.,M.Kom (.....)
3. Anggota : Cindi Wulandari S.Kom., M.Kom (.....)

Mengetahui
Ketua Program Studi Informatika
Universitas Bina Insan

Budi Santoso, S.Kom., M.Kom

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

HALAMAN MOTTO DAN PERSEMBAHAN



MOTTO:


- ✧ *Barang siapa keluar untuk mencari sebuah ilmu, maka ia akan berada di jalan Allah hingga ia kembali. – HR Tirmidzi*
- ✧ *Pengetahuan yang baik adalah yang memberikan manfaat, bukan hanya diingat. – Imam Syafi'i.*
- ✧ *There Is No Good And Evil. There Is Only Power And Those Too Weak To Seek It. –Voldemort*

Persembahan kepada :

- ❖ *Ayah dan ibunda tercinta, yang senantiasa memberikan kasih sayang, dukungan, dan doa tiada henti demi keberhasilanku. Segala pencapaian ini tidak akan mungkin tercapai tanpa kehadiran dan perjuangan kalian di setiap langkahku.*
- ❖ *Adik-adikku tersayang, yang menjadi penyemangat sekaligus tempat berbagi kebahagiaan dan harapan.*
- ❖ *Teman-teman Afourmatics seperjuanganku, yang telah berbagi cerita, semangat, dan pengalaman luar biasa selama perjalanan ini. Bersama kalian, setiap tantangan terasa lebih ringan.*

Protected by PDF Anti-Copy Free
(Upgrade to Pro Version to Remove the Watermark)
HALAMAN PERNYATAAN



Saya yang bertanda tang  ah ini :

Nama Mahasiswa : Muhammad Aryaraka Pradana

NIM : 2102020119

Program Studi : Informatika

Menyatakan dengan sesungguhnya bahwa penelitian yang saya susun untuk syarat mendapat gelar sarjana Strata Satu (S-1) pada Universitas Bina Insan Lubuklinggau, merupakan hasil kerja saya sendiri dan tidak menyuruh orang lain yang mengerjakannya. Ada bagian tertentu dalam penulisan penelitian ini yang saya kutip dari hasil karya orang lain dan telah saya tuliskan sumbernya secara jelas sesuai dengan norma, kaidah dan etika penulisan ilmiah. Jika dikemudian hari ternyata terbukti bahwa penelitian dan tugas akhir ini bukan hasil kerja saya sendiri, atau plagiat dalam bagian-bagian tertentu, maka saya bersedia dikenakan sanksi sesuai dengan peraturan perundangan yang berlaku.

Lubuklinggau, Januari 2025

Penulis,

Muhammad Aryaraka Pradana

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

ABSTRACT



Sentiment analysis of user reviews can provide valuable insights into users' perceptions and satisfaction with an application, including the E-Football online game. This study aims to implement the BERT model for sentiment analysis of user reviews for the E-Football game on Google Play Store using a machine learning-based BERT model. The dataset consists of 25,000 reviews categorized into three sentiment classes: Positive, Neutral, and Negative. The reviews underwent preprocessing, including cleaning, tokenization, case folding, stopwords removal, normalization, and stemming. The research methodology involves training the BERT model using the processed data and evaluating it with accuracy, precision, recall, and f1-score metrics. The results show that the BERT model achieves an accuracy of 90%. For the negative class, precision reaches 95% with a recall of 97%. The positive class shows a precision of 87% and recall of 85%. However, the neutral class performs lower, with a precision of 58% and recall of 53%. In conclusion, the BERT model is effective for sentiment analysis in the Indonesian language, particularly for reviews of applications on the Play Store. This study contributes significantly to the development of machine learning-based sentiment analysis methods for the Indonesian language. The implications of this research include opportunities to enhance sentiment analysis performance for the neutral class and apply the model to other domains with different datasets.

Keywords: *Sentiment Analysis, BERT, Google Play Store, E-Football*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

ABSTRAK



Analisis sentimen pada pengguna dapat memberikan wawasan penting terkait persepsi dan kegunaan terhadap suatu aplikasi, termasuk *game online E-Football*. Penelitian ini bertujuan untuk mengimplementasikan model BERT dalam analisis sentimen terhadap ulasan pengguna *game online E-Football* di *Google Play Store* menggunakan model BERT berbasis *machine learning*. Dataset yang digunakan terdiri dari 25.000 ulasan dengan tiga klasifikasi sentimen: Positif, Netral, dan Negatif. Ulasan-ulasan ini diproses melalui proses *Cleaning, Tokenization, Case Folding, Stopwords, Normalization, dan Stemming*. Metode penelitian meliputi pelatihan model BERT menggunakan data yang telah diproses dan dievaluasi dengan metrik akurasi, presisi, *recall*, dan *f1-score*. Hasil penelitian menunjukkan bahwa model BERT menghasilkan akurasi sebesar 90%. Untuk kelas negatif, presisi mencapai 95% dengan *recall* 97%. Kelas positif memiliki presisi 87% dan *recall* 85%. Namun, performa untuk kelas netral lebih rendah, dengan presisi sebesar 58% dan *recall* 53%. Kesimpulan dari penelitian ini adalah model BERT efektif untuk analisis sentimen dalam bahasa Indonesia, khususnya pada ulasan aplikasi di Play Store. Hasil penelitian ini memberikan kontribusi signifikan terhadap pengembangan metode analisis sentimen berbasis *machine learning* untuk bahasa Indonesia. Implikasi dari penelitian ini adalah peluang untuk meningkatkan performa analisis sentimen pada kelas netral dan mengaplikasikan model ini dalam domain lain dengan dataset yang berbeda.

Kata kunci : Analisis Sentimen, BERT, *Google Play Store, E-Football*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

KATA PENGANTAR

Alhamdulillah puji dan syukur penulis ucapkan kepada Allah SWT atas segala rahmad dan karunia-Nya yang telah memberikan kekuatan dan kesempatan, sehingga penulis dapat menyelesaikan skripsi ini dengan maksimal. Untuk diajukan sebagai syarat menyelesaikan pendidikan strata satu (S1) pada Program Studi Informatika Universitas Bina Insan Lubuklinggau.

Kemudian sholawat beserta salam semoga tetap tercurahkan kepada baginda Nabi Muhammad SAW, keluarga, sahabat, serta umatnya hingga akhir zaman. Dalam penulisan skripsi ini penulis telah berusaha sebaik mungkin untuk menyajikan skripsi ini, baik dari segi isi maupun dari segi desain. Penulis menyadari dalam penulisan skripsi ini tentunya masih jauh dari sempurna. Hal ini dikarenakan keterbatasan pengetahuan yang dimiliki. Oleh karena itu dalam rangka melengkapi kesempurnaan dari penulisan skripsi ini diharapkan adanya saran dan kritik yang diberikan bersifat membangun.

Untuk selanjutnya penulis mengucapkan banyak terimakasih kepada pihak-pihak yang telah membantu dalam menyelesaikan skripsi ini, yaitu:

1. Ayah dan Bundaku tercinta yang selalu memberikan doa, dukungan, dan semangat yang tiada henti kepada penulis dalam menyelesaikan skripsi ini.
2. Bapak Dr. H. Sardiyo., MM selaku Rektor Universitas Bina Insan
3. Bapak Dr. Muhamad Akbar, S.T., M.IT selaku Wakil Rektor I Universitas Bina Insan.
4. Bapak Wakhid Nur Mukhlis, M.Pd., M.M selaku Wakil Rektor II Universitas Bina Insan
5. Bapak Dr. Rudi Kurniawan, St., M.Kom Selaku Dekan Fakultas Teknik Universitas Bina Insan Lubuklinggau.
6. Bapak Budi Santoso, M.Kom selaku Ketua Program Studi Informatika Insan Lubuklinggau.
7. Bapak Armanto, M.Kom Selaku Pembimbing I yang telah banyak memberikan bimbingan dan arah dalam penulisan skripsi ini.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

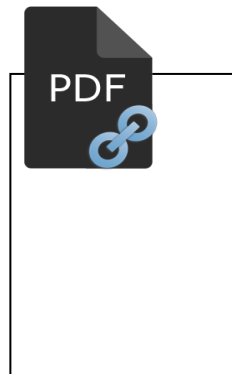
8. Bapak Fido Rizki, M.Kom selaku Pembimbing II yang juga telah banyak memberikan bimbingan dan arahan dalam penulisan skripsi ini.
9. Ibu Cindi Wulandari, M.Pd selaku Pembimbing I yang telah memberikan saran, masukan dan koreksi yang sangat berarti dalam perbaikan ini.
10. Seluruh Staf Dosen dan Karyawan Universitas Bina Insan Lubuklinggau
11. Adik-adikku tersayang yang selalu memberikan keceriaan, dukungan, dan motivasi dalam setiap langkah penulis.
12. Teman-teman *Afourmatics* yang telah memberikan dukungan, semangat, dan kebersamaan selama masa studi dan penulisan skripsi ini.
13. Semua pihak yang telah membimbing, membantu dan mendorong penyelesaian skripsi ini.

Mudah-mudahan Allah SWT membalas semua budi baik yang telah diberikan. semoga penelitian ini dapat berguna bagi pembaca. Akhir kata semoga penelitian ini dapat bermanfaat bagi untuk penelitian selanjutnya.

Lubuklinggau, Januari 2025

Muhammad Aryaraka Pradana

Protected by PDF Anti-Copy Free
(Upgrade to Pro Version to Remove the Watermark)
DAFTAR RIWAYAT HIDUP



Biodata

Nama : Muhammad Aryaraka Pradana
Tempat Lahir : Lubuklinggau
Tanggal Lahir : 31 Juli 2003
Jenis Kelamin : Laki-laki
Agama : Islam
Alamat : Jl. Patimura No.05 RT.01 Kel.Sukajadi
Kec.Lubuklinggau Barat I Kota Lubuklinggau

Pendidikan

- SD : MI Darussalam
- SMP/MTS Sederajat : SMP Negeri 7 Kota Lubuklinggau
- SMA/MAN/SMK Sederajat : SMK Yadika Kota Lubuklinggau

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

DAFTAR ISI



IMPLEMENTASI	i
HALAMAN PENGESAHAN	ii
IMPLEMENTASI	ii
HALAMAN PERSETUJUAN TIM PENGUJI	iii
HALAMAN MOTTO DAN PERSEMBAHAN	iv
HALAMAN PERNYATAAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR RIWAYAT HIDUP	x
DAFTAR ISI	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	4
1.3 Rumusan Masalah	5
1.4 Batasan Masalah	5
1.5 Tujuan dan Manfaat Penelitian	6
1.6 Sistematika Penulisan	8
BAB II TINJAUAN PUSTAKA	10
2.1 Literatur	10
2.2 Penelitian Terdahulu yang Relevan	19
2.3 Kerangka Berpikir	23
BAB III METODOLOGI PENELITIAN	26
3.1 Metode Penelitian	26
3.2 Metode Pengumpulan Data	26
3.3 Metode Analisa	28
3.4 Tempat dan Waktu Penelitian	28
3.5 Alat dan Bahan	29

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

3.6	Metode Pengujian dan Pengolahan Data.....	30
BAB IV	PENELITIAN DAN PEMBAHASAN.....	42
4.1	Gambaran Umum Penelitian.....	42
4.2	Hasil Penelitian.....	43
4.3	Pembahasan.....	63
BAB VI	KESIMPULAN DAN SARAN	65
5.1	Kesimpulan	65
5.2	Saran.....	66
DAFTAR PUSTAKA		
LAMPIRAN		

Protected by PDF Anti-Copy Free
(Upgrade to Pro Version to Remove the Watermark)
DAFTAR TABEL



	Halaman
Table 2.1 Penelitian Terdahulu	19
Table 3.1 Waktu Kegiatan.....	29
Table 3.2 Representasi Confusion Matrix.....	31
Table 3.3 Hasil <i>Cleaning</i>	34
Table 3.4 Hasil <i>Tokenizing</i>	35
Table 3.5 Hasil <i>Case Folding</i>	36
Table 3.6 Hasil <i>Stopwords Removal</i>	37
Table 3.7 Hasil <i>Normalization</i>	38
Table 3.8 Hasil <i>Stemming</i>	39
Table 3.9. <i>Labeling Data</i>	40

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

DAFTAR GAMBAR



Halaman

Gambar 2.1 Tampilan E-Football	14
Gambar 2.2 Tahapan <i>Preprocessing</i>	15
Gambar 2.3 Kerangka Berpikir	23
Gambar 3.1 <i>Scrapping Data</i>	27
Gambar 3.2 Hasil <i>Scrapping Data</i>	28
Gambar 4.1 <i>Source code</i> pengambilan data	43
Gambar 4.2 Hasil Pengambilan Data	44
Gambar 4.3 Hasil Proses <i>Cleaning</i>	46
Gambar 4.4 Hasil proses <i>Normalization</i>	47
Gambar 4.5 Hasil Proses <i>Case Folding</i>	48
Gambar 4.6 Hasil Proses <i>Stopword</i>	50
Gambar 4.7 Hasil proses <i>Tokenizing</i>	51
Gambar 4.8 Hasil proses <i>Stemming</i>	52
Gambar 4.9 Hasil Proses <i>Labeling</i>	54
Gambar 4.10 Visualisasi Grafik <i>Labeling</i>	54
Gambar 4.11 Hasil Proses <i>Encoding</i>	55
Gambar 4.12 Hasil proses <i>Splitting</i>	56
Gambar 4.13 <i>Fine Tuning</i>	59
Gambar 4.14 Evaluasi Model.....	60
Gambar 4.15 <i>Confusion Matrix</i>	61
Gambar 4.16 <i>Classification Report</i>	62
Gambar 4.17 <i>Word Cloud</i>	63

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

DAFTAR LAMPIRAN



1. Lembar Pengajuan Judul Pen
2. Lembar Bimbingan Skripsi Pembimbing I
3. Lembar Bimbingan Skripsi Pembimbing II
4. Lembar Perbaikan Skripsi
5. Lampiran *Scraping Data*



1.1 Latar Belakang

Industri *game online* terus mengalami perkembangan pesat dalam beberapa dekade terakhir seiring dengan kemajuan teknologi dan meningkatnya akses internet. Salah satu genre game yang paling diminati adalah game olahraga, khususnya sepak bola, yang memiliki basis penggemar luas di seluruh dunia. *E-Football*, game ini menawarkan fitur kompetisi *real time*, pembaruan tim berdasarkan musim sepak bola dunia, serta mode game yang mencakup *Single Match*, *Local Match*, dan *Online Match*. Dalam mode ini, pemain dapat berkompetisi melawan pemain lain baik secara lokal maupun *online* [1].

Game yang dirilis oleh KONAMI sejak tahun 1995 dan untuk versi *mobile* nya pertama kali dirilis pada tahun 2017 yang dimana dulu bernama *Pro Evolution Soccer (PES)* dan sekarang berganti nama menjadi *e-Football* ini termasuk dalam salah satu game *online eSports (Electronic Sport)* yang diturnamenkan di berbagai tempat hingga seluruh dunia dengan total hadiah yang terbilang fantastis. *Game e-Football* memperoleh popularitas yang luar biasa dalam kurun waktu dari awal rilisnya hingga sekarang dikarenakan grafis dan *gameplay* mekaniknya yang meningkat setiap tahunnya [2].

Namun, popularitas sebuah game tidak hanya ditentukan oleh kualitas *gameplay*, tetapi juga oleh kepuasan pengguna yang diekspresikan

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

melalui ulasan dan *rating* di platform distribusi seperti *Google Play Store*.

Google Play Store merupakan platform digital yang berfungsi sebagai toko resmi untuk *Android*, memungkinkan pengguna menelusuri, mengunduh aplikasi, dan memberikan ulasan, serta peringkat untuk konten digital sebagai opini publik [3]. Ulasan ini sering kali mencerminkan persepsi pemain terhadap aspek teknis, *gameplay*, dan kualitas layanan yang diberikan oleh pengembang, serta pengalaman bermain secara keseluruhan. Masalah seperti *bug*, *lag*, dan stabilitas server kerap menjadi keluhan yang mempengaruhi *rating* dan kepuasan pemain, sehingga dapat memengaruhi reputasi dan keberhasilan jangka panjang game tersebut.

Salah satu fitur penting di *Google Play Store* adalah ulasan atau *review*, dimana pengguna dapat memberikan opini dan *rating*. Fitur ini memungkinkan pengguna menilai dan memberikan teks ulasan terhadap aplikasi yang digunakan [4]. Melihat hal tersebut, penelitian ini bertujuan untuk menganalisis ulasan pengguna game *E-Football* di *Play Store* guna memahami sentimen pemain terhadap game ini. Analisis ini penting untuk mengidentifikasi elemen-elemen yang berkontribusi pada kepuasan maupun ketidakpuasan pengguna, sehingga dapat memberikan wawasan berharga bagi pengembang dalam meningkatkan kualitas game, dengan memahami faktor-faktor yang mempengaruhi pengalaman pengguna, pengembang dapat mengambil langkah-langkah strategis untuk meningkatkan performa game dan mempertahankan loyalitas pemain.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Maka, dibutuhkan metode analisis yang efektif untuk memahami sentimen pengguna secara mendalam, khususnya terkait kepuasan dan persepsi mereka terhadap berbagai faktor yang mempengaruhi pengalaman bermain *game online E-Football*. BERT memiliki kemampuan memahami konteks kata dalam ulasan pengguna secara mendalam karena pembacaannya yang bersifat dua arah (*bidirectional*), sehingga dapat memberikan analisis yang lebih presisi terkait perasaan dan opini pemain terhadap game ini.

Dengan menerapkan model BERT pada ulasan di *Google Play Store*, pengembang dapat memperoleh wawasan mendalam tentang berbagai faktor yang mempengaruhi pengalaman pengguna. Sentimen positif dan negatif dalam ulasan dapat diidentifikasi dan dikategorikan, sehingga masalah seperti *bug*, *lag*, atau ketidakstabilan server yang sering dikeluhkan dapat diprioritaskan dalam perbaikan. Di sisi lain, pengembang juga dapat memahami elemen-elemen yang dihargai pengguna, seperti peningkatan fitur atau grafis, yang dapat terus dikembangkan untuk memperkuat kepuasan dan loyalitas pemain. Penelitian ini diharapkan tidak hanya meningkatkan akurasi analisis sentimen ulasan, tetapi juga berkontribusi pada pengembangan metodologi dalam kajian ilmiah terkait analisis sentimen untuk *game online*, khususnya melalui model BERT.

Berdasarkan masalah di atas, penulis melakukan penelitian tugas akhir dengan judul “Implementasi BERT Untuk Analisis Sentimen Pengguna *Game Online E-Football* Pada Ulasan Di *Play Store* Berbasis *Machine*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Learning”, dengan menggunakan *Bidirectional Encoder Representations from Transformers* untuk menganalisis ulasan pengguna pada *game online e-football*. Penelitian ini bertujuan untuk meningkatkan akurasi penelitian sebelumnya dan melakukan beberapa eksperimen untuk menemukan model BERT yang paling akurat dan menentukan tingkat akurasi.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah dijelaskan, terdapat beberapa masalah yang dapat diidentifikasi dalam penelitian ini, yaitu:

1. Kepuasan pengguna yang bervariasi, banyaknya ulasan pengguna di *Google Play Store* menunjukkan adanya perbedaan signifikan dalam tingkat kepuasan pemain terhadap *E-Football*. Sebagian pengguna memberikan ulasan positif terkait *gameplay* dan fitur, sementara sebagian lainnya mengeluhkan masalah teknis seperti *bug*, *lag*, dan ketidakstabilan server. Masalah ini perlu diidentifikasi secara rinci untuk memahami faktor-faktor utama yang memengaruhi kepuasan pemain.
2. Persepsi pengguna terhadap pembaruan game, setiap kali pengembang melakukan pembaruan, umpan balik dari pengguna cenderung bervariasi. Beberapa pengguna menyambut baik pembaruan yang membawa peningkatan *gameplay*, sementara yang lain merasa bahwa pembaruan tersebut justru memperburuk pengalaman bermain. Hal ini mengindikasikan adanya perbedaan persepsi yang perlu dianalisis lebih lanjut.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

3. Kurangnya kajian ilmiah tentang penggunaan analisis sentimen pada *game online*, meskipun terdapat ulasan pengguna yang tersedia, penelitian ilmiah yang menggunakan teknik analisis sentimen pada ulasan pengguna *game online* seperti *E-Football* masih terbatas. Ini menunjukkan adanya kebutuhan untuk melakukan penelitian lebih lanjut dalam bidang ini.

1.3 Rumusan Masalah

Berdasarkan latar belakang yang telah dijabarkan di atas, rumusan masalah yang diangkat dalam penelitian ini adalah bagaimana mengimplementasikan algoritma BERT untuk analisis sentimen ulasan pengguna *game online E-Football*.

1.4 Batasan Masalah

Dalam penelitian ini, penulis memberikan batasan masalah agar lebih terarah. Batasan masalah tersebut sebagai berikut:

1. Data diperoleh dari proses *web scraping* ulasan pengguna *game online E-Football* menggunakan *google play scraper* yang tersedia dalam *Python* dan menggunakan *Google Colab* yang didalamnya tersedia *Jupyter Notebook* sebagai *notebook* dalam projek ini.
2. *Dataset* yang diproses sebanyak 25.000 ulasan pengguna.
3. Data ulasan *game online E-Football* dibagi menjadi tiga jenis klasifikasi sentimen, yaitu : Positif, Netral, dan Negatif.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

4. Model BERT yang digunakan pada penelitian ini adalah model IndoBERT.



1.5 Tujuan dan Manfaat Penelitian

1.5.1 Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut :

1. Tujuan Umum

Mengimplementasikan model BERT untuk menganalisis sentimen dari ulasan pengguna *game E-Football* di *Play Store* guna memahami pola sentimen dan mendapatkan wawasan untuk pengembangan game yang lebih baik.

2. Tujuan Khusus

- a. Mengetahui sentimen masyarakat sebagai pengguna, baik positif maupun negatif, terhadap *game online E-Football*.
- b. Mengetahui performa dan akurasi BERT terhadap sentimen pengguna mengenai kinerja *game online e-Football*.
- c. Untuk mengidentifikasi dan evaluasi efektivitas Algoritma BERT dalam menganalisis aspek-aspek game yang perlu ditingkatkan atau dipertahankan.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

1.5.2 Manfaat Penelitian

Adapun manfaat penelitian ini adalah sebagai berikut :

1. Manfaat Bagi Peneliti

- a. Meningkatkan pemahaman terhadap sentimen pengguna terkait kinerja *game online E-Football* di *Google Play Store*.
- b. Sebagai sarana untuk mengimplementasikan hasil teori yang didapat di perkuliahan dengan penerapan dalam sistem nyata.

2. Manfaat Bagi Pengembang Game

Penelitian ini memberikan wawasan yang mendalam mengenai sentimen pengguna terhadap *E-Football* di *Google Play Store*, sehingga dapat membantu pengembang dalam memahami opini dan keluhan pengguna. Dengan demikian, pengembang dapat memperbaiki kualitas game, meningkatkan kepuasan pemain, dan menyusun strategi pembaruan game yang lebih efektif.

3. Manfaat Bagi Pengguna

Pengguna dapat merasakan peningkatan pengalaman bermain, karena pengembang dapat melakukan perbaikan yang lebih tepat sasaran berdasarkan hasil analisis sentimen dari ulasan mereka.

4. Manfaat Bagi Universitas Bina Insan

Dapat dijadikan sebagai bahan referensi untuk penelitian bagi mahasiswa lain yang berhubungan dengan proses implementasi analisis sentimen.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

1.6 Sistematika Penulisan

Untuk memberikan gambaran secara garis besar dari laporan skripsi, berikut akan diuraikan secara sistematis sistematika penulisannya yang terdiri 5 bab dan masing-masing sub bab dengan relevan terhadap permasalahan yang dibahas secara sistematis, isi dari masing-masing bab tersebut adalah sebagai berikut:

BAB I : PENDAHULUAN

Dalam bab ini berisi latar belakang masalah, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, metode penelitian, dan sistematika penelitan.

BAB II : LANDASAN TEORI

Dalam bab ini berisikan tentang landasan teori yang dipakai oleh penulis yang berhubungan dengan tema penelitian.

BAB III : ANALISA DAN PERANCANGAN

Dalam bab ini berisikan tentang analisa sistem, analisis alur kerja perangkat lunak, metode pengumpulan data, metode pengembangan sistem dan perancangan perangkat lunak untuk analisis sentimen ulasan pengguna game *online E-Football* di *Google Play Store*.



BAB IV : HASIL DAN PEMBAHASAN

Bab ini berisi tentang hasil dan pembahasan.

Protected by PDF Anti-Copy Free

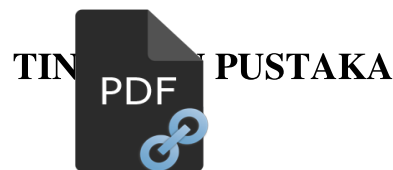
(Upgrade to Pro Version to Remove the Watermark)

BAB V : KESIMPULAN DAN SARAN

Bab ini berisi  dari seluruh penelitian dan saran saran masukan masuk  berguna di masa yang akan datang.

DAFTAR PUSTAKA

LAMPIRAN



2.1 Literatur

2.1.1 *Machine Learning*

Machine learning adalah cabang dari kecerdasan buatan (AI) yang memungkinkan sistem komputer belajar dari data tanpa perlu diprogram secara eksplisit. Menurut [5], *Machine Learning* (ML) atau pembelajaran mesin merupakan pendekatan dalam AI yang banyak digunakan untuk menggantikan atau menirukan perilaku manusia untuk menyelesaikan masalah atau melakukan otomatisasi. Dengan menggunakan algoritma khusus, *machine learning* mengidentifikasi pola atau informasi dari data dan kemudian membuat prediksi atau keputusan berdasarkan pola tersebut.

Teknik yang digunakan dalam *Machine Learning* juga dapat penanganan data besar (*big data*) dengan cerdas untuk memberikan hasil yang tepat [6]. *Machine learning* terbagi menjadi tiga kategori [7], yaitu *Supervised Learning*, *Unsupervised Learning*, *Reinforcement Learning*. Kesimpulannya adalah *Machine learning* memungkinkan komputer belajar dari data untuk mengenali pola dan membuat keputusan secara otomatis.

2.1.2 *Text Mining*

Text mining adalah proses yang bertujuan untuk menemukan informasi atau tren terbaru yang belum dipublikasikan dengan memproses dan menganalisis data dalam jumlah besar. *Text mining* berusaha

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

mengekstrak informasi yang bermanfaat dari sumber data melalui identifikasi dan eksplorasi yang menarik [8]. *Text mining* umumnya mencakup kategorisasi informasi atau teks, mengelompokkan teks, ekstraksi entitas atau konsep, pengembangan dan perumusan taksonomi umum. *Text mining* berkenaan dengan informasi terstruktur atau tekstual ekstraksi informasi yang bermakna dan pengetahuan dari jumlah besar teks [9].

Text mining merupakan proses untuk menggali informasi agar pengguna dapat berinteraksi dengan koleksi dokumen dari waktu ke waktu menggunakan alat analisis. Proses utama dalam *text mining* adalah untuk mengidentifikasi kata-kata yang dapat mewakili isi dokumen dan kemudian dilakukan pengkajian untuk memahami hubungan antar dokumen [10]. Kesimpulannya adalah *text mining* adalah proses mengekstrak informasi dan pola dari data teks besar untuk mendapatkan pengetahuan yang bermakna, dengan menganalisis isi dan hubungan antar dokumen.

2.1.3 Analisis Sentimen

Analisis sentimen adalah bagian dari ilmu data mining yang memiliki tujuan untuk menganalisis dan mengekstrak data tekstual yang berupa pendapat, evaluasi, sikap, emosi, penilaian, dan sentimen seseorang terhadap suatu barang, orang, organisasi, dan masalah [11]. Analisis sentimen ini dapat mengelompokkan polaritas dari teks dalam kalimat atau dokumen untuk mengetahui apakah opini pada kalimat atau dokumen tersebut apakah termasuk positif atau negatif.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Tujuan analisis sentimen adalah untuk mengetahui perilaku opini penulis terhadap suatu topik atau isu. Tingkah laku atau dapat menunjukkan alasan, pendapat atau perasaan penulis terhadap kondisi, kecenderungan, bagaimana penulis ingin mempengaruhi pembaca. Analisis sentimen juga dapat mengungkapkan kesedihan emosional, kegembiraan atau kemarahan [12]. Kecenderungan teks tertentu dinilai berdasarkan konteks dan polaritas yang diperoleh dan dapat menjadi argumen, opini, atau keadaan sentimental teks yang potensial. Apabila semakin banyak data tekstual yang dikumpulkan dalam penelitian analisis sentimen, maka semakin mudah juga menemukan korelasi yang signifikan antara teks dan jenis sentimen [13].

Kesimpulannya adalah analisis sentimen merupakan teknik dalam data mining yang digunakan untuk memahami opini, emosi, dan sikap seseorang terhadap suatu topik melalui pengelompokan polaritas teks menjadi positif atau negatif. Dengan demikian, teknik ini tidak hanya membantu mengidentifikasi pandangan penulis tetapi juga menggali kecenderungan emosional yang terkandung dalam teks. Selain itu, analisis sentimen dapat menemukan korelasi signifikan antara teks dan jenis sentimen yang dihasilkan, terutama ketika jumlah data yang dianalisis semakin besar.

2.1.4 *Game Online*

Game online adalah permainan video yang dimainkan selama beberapa bentuk jaringan komputer, menggunakan komputer pribadi atau

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

konsol video game [14]. *Game online* merupakan permainan yang hanya mampu dioperasikan dan menggunakan jaringan internet dengan beberapa fitur dan tampilan yang berbeda-beda sesuai dengan jenisnya [15]. Pada mulanya *game online* dikenal dengan istilah “Game Jaringan”, yang berupa kumpulan beberapa PC (*personal computer*) yang terhubung satu sama lain sehingga antar user satu dengan yang lain bisa terhubung [16]. Dalam penelitian ini game yang di fokuskan adalah *game e-football*.

2.1.5 E-Football

Salah satu game *mobile* yang populer di kalangan masyarakat Indonesia adalah *e-Football*. *E-Football* adalah game sepak bola *online* yang dikembangkan oleh Konami dan sebelumnya dikenal sebagai *Pro Evolution Soccer* (PES). Terdapat dua raksasa dalam industri game sepak bola, yaitu *e-Football* dan *FC Mobile* [17]. *E-Football* adalah *game online* yang termasuk dalam kategori *sports and casual*, *game online* ini banyak dimainkan oleh orang-orang yang memiliki ketertarikan pada bidang olahraga sepak bola. Hal itu bisa terjadi karena *e-Football* menyediakan berbagai macam pemain dan tim sepak bola dari berbagai negara yang ada di dunia [15]. Game ini beralih ke model *free-to-play* pada tahun 2021, memungkinkan pemain bertanding secara *real-time* melawan pengguna lain di seluruh dunia. *E-Football* juga merupakan salah satu game bergenre olahraga yang dipertandingkan dalam jenis olahraga elektronik atau *e-sport* [18]. *E-Football* tersedia di berbagai platform, termasuk *PlayStation*, *Xbox*,

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

PC, dan perangkat *mobile*. Dengan grafik realistis dan fitur kompetisi *online*, game ini menawarkan pengalaman bermain sepak bola digital yang mendalam dan kompetitif



Gambar 2.1 Tampilan E-Football

2.1.6 Google Play Store

Bagi para pengguna *android*, salah satu tempat untuk mengunduh ratusan ribu aplikasi android adalah *Google Play Store* [19]. Sistem operasi *Android* menyediakan *platform* bernama *Play Store* yang berisi berbagai aplikasi yang dikembangkan oleh para pengembang dan dapat digunakan oleh pengguna sebagaimana mestinya [20]. *Play Store* adalah layanan distribusi digital yang dioperasikan dan dikembangkan oleh *Google*. *Google Play* diluncurkan pada tahun 2012 pada tanggal 6 Maret. *Google Play* adalah kombinasi dari tiga produk *Google* berbeda yang awalnya dibuat

Protected by PDF Anti-Copy Free

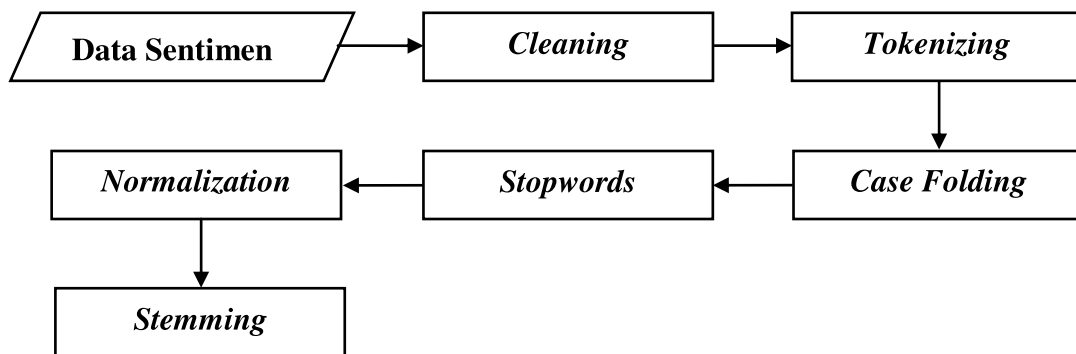
(Upgrade to Pro Version to Remove the Watermark)

secara terpisah yaitu *Android Market*, *Google Music*, dan *Google eBookstore* [21].



2.1.7 Text Preprocessing

Text Preprocessing adalah langkah awal dalam membersihkan dan memperbaiki data. Saat mengumpulkan data, seringkali data yang didapatkan tidak terstruktur dan mengandung banyak karakter yang tidak relevan. Tujuan dari *preprocessing* adalah untuk menghilangkan gangguan (*noise*) tersebut agar data menjadi lebih bersih dan siap untuk dianalisis [22]. Tahapan *preprocessing* ini diantaranya:



Gambar 2.2 Tahapan *Preprocessing*

1. Data Sentimen

Data sentimen adalah evaluasi mengenai emosi atau pendapat yang diungkapkan oleh pengguna melalui media sosial. Data sentimen dapat beragam, bisa berupa positif, negatif, atau netral [23].

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

2. *Cleaning*

Merupakan tahapan pembersihan data yang bertujuan untuk menghilangkan tanda baca, simbol, link URL, dan *username* di dalam teks [24].



3. *Tokenizing*

Merupakan tahapan pemotongan string input berdasarkan tiap kata yang menyusunnya. Secara garis besar memecah sekumpulan karakter dalam suatu teks ke dalam satuan kata [25].

4. *Case Folding*

Merupakan tahap untuk mengkonverensi teks dalam dokumen data menjadi *lower case* [26].

5. *Stopwords*

Merupakan penghapusan kata-kata yang tidak relevan seperti kata "kok", "lah", "yang", dan lain-lain. Proses tahapan ini dimana kata-kata kunci diambil dari hasil *tokenizing* [27].

6. *Normalization*

Merupakan perubahan kata ke dalam bentuk awal seperti penggantian kata non baku ke dalam kata baku [28]. Misalnya, mengubah bentuk slang atau singkatan menjadi kata yang lebih formal.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

7. *Stemming*

Merupakan proses mereduksi kata menjadi kata dasarnya dengan menghilangkan imbuhan pada kata dalam dokumen atau mengubah kata kerja menjadi kata benda [29].



2.1.8 Klasifikasi

Klasifikasi merupakan suatu fungsi dari *data mining* untuk memprediksi kelas terhadap objek didalam *database*. Klasifikasi adalah proses dari pembangunan terhadap suatu model yang mengklasifikan suatu objek sesuai dengan atributnya [30].

Klasifikasi adalah salah satu pembelajaran yang paling umum di *data mining*. Klasifikasi didefinisikan sebagai bentuk analisis data untuk mengekstrak model yang akan digunakan untuk memprediksi label kelas. Kelas dalam klasifikasi merupakan atribut dalam satu set data yang paling unik yang merupakan variabel bebas dalam statistik [31]. Teknik klasifikasi berguna untuk memprediksi suatu nilai dari target variabel kategori [32]. Kesimpulannya adalah klasifikasi adalah teknik *data mining* untuk memprediksi kelas objek berdasarkan atributnya dengan membangun model analisis data.

2.1.9 *Bidirectional Encoder Representations from Transformers (BERT)*

BERT (*Bidirectional Encoder Representations from Transformers*) adalah model *machine learning* berbasis *transformer* yang dikembangkan oleh *Google* untuk memahami konteks dari kedua arah dalam sebuah teks.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Tidak seperti model yang hanya membaca dari kiri ke kanan atau sebaliknya, BERT dapat memahami konteks dan ambiguitas dalam teks dengan mengekstrak fitur semantikal dari kata-kata dalam konteks yang berbeda [33]. Pemilihan Algoritma BERT dilakukan karena model ini telah menjadi salah satu model bahasa terkemuka dalam pemrosesan bahasa alami. BERT telah menghasilkan kinerja yang sangat baik dalam tugas-tugas serupa. BERT adalah model representasi yang digunakan secara luas dan efisien [34].

Model ini mencapai kinerja terbaik dalam tugas-tugas pemrosesan bahasa alami baik pada tingkat kalimat maupun token, mengungguli banyak arsitektur yang dirancang khusus untuk tugas-tugas tersebut [35]. BERT membaca keseluruhan kata dalam kalimat secara *bidirectional*, sehingga dapat menangkap makna yang lebih akurat dari sebuah kata berdasarkan konteks kalimatnya. BERT mengintegrasikan konteks secara menyeluruh dengan menganalisis pola yang muncul sebelum dan sesudah kata tersebut [36].

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

2.2 Penelitian Terdahulu yang Relevan

Berikut adalah tabel penelitian terdahulu yang dilakukan peneliti sebelumnya:



Table 2.1 Penelitian Terdahulu

<i>Autor's</i>	Judul	Tahun	Metode	Hasil
Cindy Alifia Putri, Adiwijaya, Said Al Faraby [37]	Analisis Sentimen <i>Review Film</i> Berbahasa Inggris dengan Pendekatan <i>Bidirectional Encoder Representations from Transformers</i>	2020	BERT	Penelitian ini menggunakan 2000 data. 1000 data berlabel positif dan 1000 berlabel negatif. Pada penelitian peneliti pertama menggunakan algoritma <i>naive bayes</i> dan mendapatkan akurasi 48%, lalu menggunakan metode BERT dan mendapatkan akurasi 78%
Manish Munikar, dkk [38]	Ulasan Film Dari <i>Rotten Tomatoes</i> Pada SST <i>Dataset</i>	2019	BERT	Klasifikasi sentimen <i>finegrained</i> dengan akurasi rata-rata 90%

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Jahfal Uno	Analisis Sentimen	2023	BERT	Komputasi
Surya	Ulasan Penggu			menggunakan
Lazuardi,	Aplikasi Joox			10.000 data, hasil
Asep Juarna	Android			analisis sentimen
[39]	Menggunakan			adalah 41,92%
	Metode			<i>true</i> (sentimen)
	<i>Bidirectional</i>			positif, 1,01% <i>true</i>
	<i>Encoder</i>			netral, dan 35,95%
	<i>Representation</i>			<i>true</i> negatif,
	<i>From Transformer</i>			semuanya dari 990
	(BERT)			data uji, dengan
				akurasi F1-score
				berturut-turut 86%,
				51%, dan 76%
				sementara akurasi
				<i>baseline</i> adalah
				83%, 79%, dan
				75%, yang berarti
				ada peningkatan
				akurasi <i>true</i> positif
				sebesar 3,6% dan
				<i>true</i> negatif
				sebesar 1,3%.
Raden Mas	Analisis Sentimen	2021	BERT	Dari 5437 <i>testing</i>
Rizqi Wahyu	<i>Customer Review</i>			data, 5254
Panca	Aplikasi Ruang			komentar yang
Kusuma	Guru dengan			dinyatakan positif,
Atmaja, Wiyli	Metode BERT			16 yang
Yustanti [40]	<i>(Bidirectional</i>			dinyatakan netral

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Encoder

Representation

from Transform



sentimen, dan 167

komentar

dinyatakan negatif
sentimen.

Berdasarkan dari
nilai *F1 Score*
yang didapat dari
presisi dan *recall*
yakni 98.9% dan
nilai akurasi
bernilai 99%.

Zilvi Azus	Implementasi	2024	BERT	Hasil percobaan,
Sriyanti,	Model BERT Pada			skenario 1 dengan
Dhian Satria	Analisis Sentimen			<i>learning rate</i> 2e-5,
Yudha	Pengguna Twitter			<i>batch size</i> 16,
Kartika,	Terhadap Aksi			<i>dropout rate</i> 0.1
Abdul Rezha	Boikot Produk			memiliki akurasi
Efrat Najaf	Israel			yang paling tinggi
[41]				yaitu 85%. Nilai
				presisi 85%, nilai
				<i>recall</i> 85% , dan
				nilai <i>F1-score</i>
				85%. Hasil
				validasi sistem
				mendapatkan
				akurasi 78%.

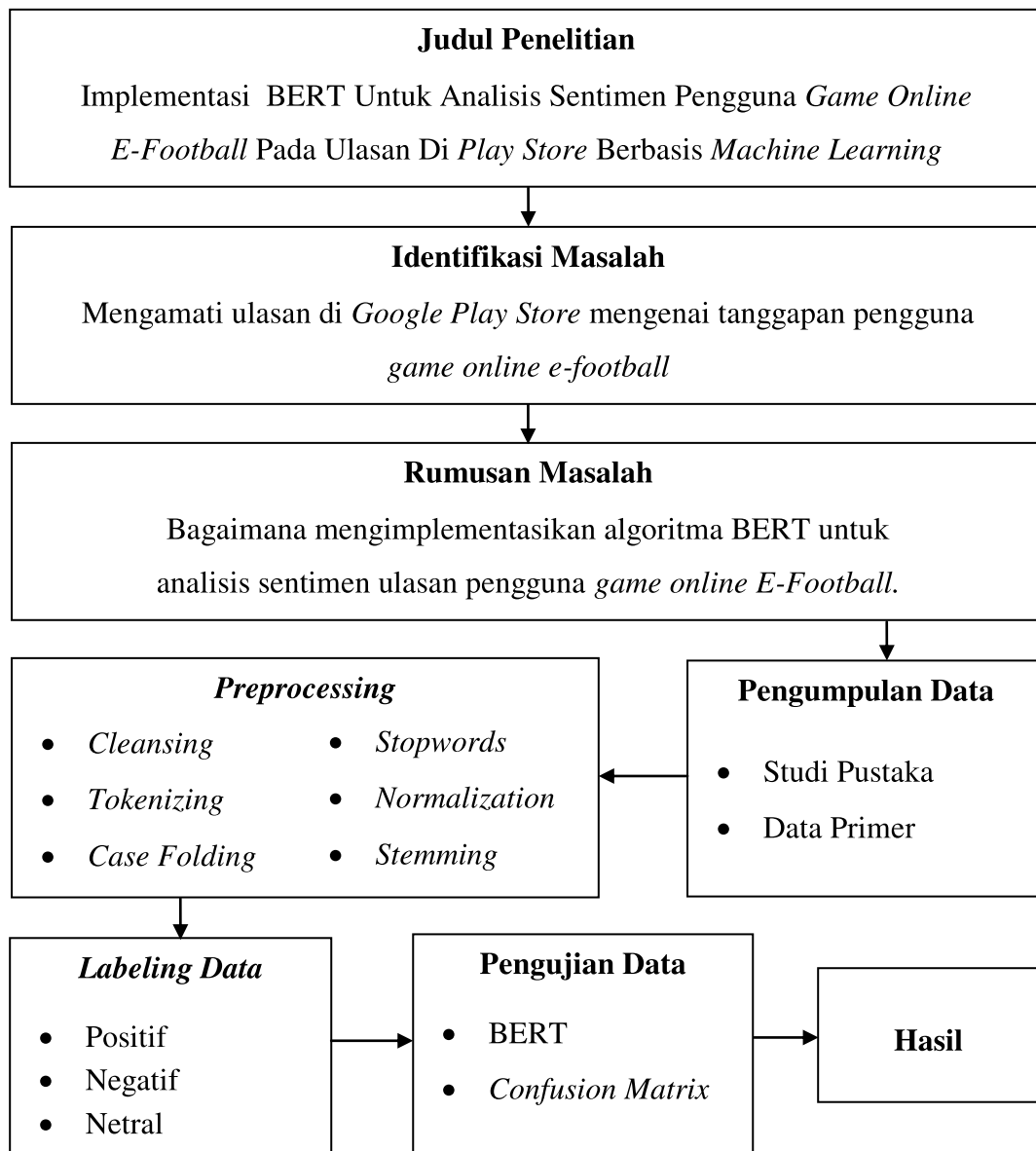
Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Riska Kurnia	Eksplorasi Sentimen	2024	BERT	Dengan analisis
Septiani,	Pengguna Media			sentimen yang
Windu Gata	Sosial Terhadap			ditulis oleh
[42]	Layanan <i>Seab</i>			penulis, diketahui
	Pendekatan Dengan			bahwa dari 500
	Algoritma BERT			data di Twitter (X)
				dan YouTube, 306
				data menunjukkan
				sentimen positif
				dan 144
				menunjukkan
				sentimen negatif.
				Nilai akurasi yang
				dihasilkan deep
				learning BERT
				yaitu 86% dengan
				<i>epoch</i> 3, dan
				memperoleh
				akurasi tertinggi
				dari akurasi
				algoritma lainnya.

2.3 Kerangka Berpikir

Kerangka berpikir adalah urutan pemikiran atau konsep yang disusun secara sistematis untuk menunjukkan hubungan antar variabel atau faktor yang menjadi fokus penelitian. Adapun kerangka berpikir dapat dilihat pada gambar berikut:




Gambar 2.3 Kerangka Berpikir

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Kerangka berpikir pada gambar 3 diatas dapat dijelaskan sebagai berikut :

1. Judul Penelitian

Judul penelitian adalah  singkat yang mencerminkan pokok permasalahan penelitian. Judul ini memberi gambaran umum tentang apa yang akan di teliti.

2. Identifikasi Masalah

Tahap dimana masalah atau isu penelitian diidentifikasi, beserta konteks dan relevansinya.

3. Rumusan Masalah

Merinci masalah penelitian dengan pertanyaan masalah yang konkret. Ini melibatkan merinci variabel- variabel yang akan diamati.

4. Pengumpulan Data

Pengumpulan data dilakukan menggunakan bahasa pemograman *python* yang terdapat pada *google colab*. Data dari *game online e-football* ini diambil berdasarkan ulasan relevan para pengguna yang ada di *Google Play Store*.

5. Preprocessing

Preprocessing bertujuan membersihkan data yang akan digunakan dalam penelitian. Adapun prosedur yang digunakan antaranya adalah *Cleansing, Tokenzing, Case Folding, Stopword, Normalization, Stemming*.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

6. *Labeling Data*

Labeling Data merupakan pemberian label pada data, data ulasan yang sudah dikumpulkan ke dalam kategori label positif, negatif, dan netral.



7. **Pengujian Data dan Evaluasi.**

Tahapan ini merupakan tahapan akhir yang dimana pengujian data dilakukan untuk membuat model menggunakan algoritma klasifikasi berupa *Bidirectional Encoder Representations from Transformers* (BERT) dalam menganalisa *dataset* yang telah melalui tahap *preprocessing*. Sedangkan evaluasi bertujuan untuk mengetahui nilai *accuracy*, *precision*, *F1-Score* dan *recall*.

METODE PENELITIAN



3.1 Metode Penelitian

Penelitian ini menggunakan metode eksperimen dengan fokus pada analisis data yang dikumpulkan dari ulasan pengguna *game online e-football*. Data ini akan diolah melalui klasifikasi menggunakan algoritma BERT (*Bidirectional Encoder Representations from Transformers*).

3.2 Metode Pengumpulan Data

Pengumpulan data dilakukan guna mendapatkan suatu informasi dalam menyelesaikan permasalahan. Berikut merupakan metode pengumpulan data yang digunakan penulis dalam melakukan penelitian:

3.2.1 Studi Pustaka

Studi pustaka merupakan suatu proses pengumpulan data kepustakaan dilakukan dengan cara mengumpulkan materi melalui jurnal, literatur, buku ataupun situs internet sebagai sumber pustaka yang berkaitan dengan penelitian [43], terutama tentang analisis sentimen menggunakan algoritma BERT (*Bidirectional Encoder Representations from Transformers*). Studi pustaka berguna untuk menjadi dasar referensi yang diperlukan untuk menganalisis data ulasan pengguna pada *game online e-football*.

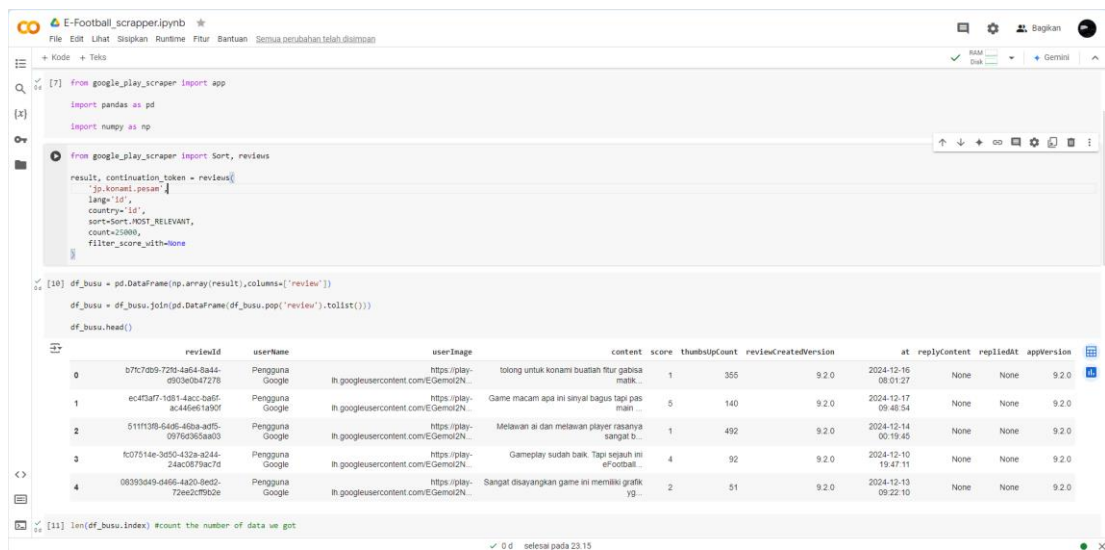
3.2.2 Data Primer

Data primer merupakan data yang diperoleh secara langsung dari objek penelitian. Data dalam penelitian ini diperoleh dari proses *web*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

scraping ulasan pengguna pada game online e-football berbahasa Indonesia sebanyak 25.000 data. Pada pengambilan data dengan jumlah besar karena pada tahap persiapan data banyak data yang akan dihapus karena dianggap terlalu banyak mengandung karakter yang tidak bisa dimasukkan ke dalam model, tidak dapat dilakukan pembersihan (*cleaning*) [40]. Proses pengambilan data dapat dilihat pada gambar 3.1 berikut.



```
from google_play_scraper import app
import pandas as pd
import numpy as np

from google_play_scraper import Sort, reviews
result, continuation_token = reviews(
    "jo.konami.pesman",
    "android",
    "id",
    sort=Sort.MOST_RELEVANT,
    count=25000,
    filter_score_with=None
)

df_busu = pd.DataFrame(np.array(result), columns=['review'])
df_busu = df_busu.join(pd.DataFrame(df_busu.pop('review').tolist()))
df_busu.head()
```

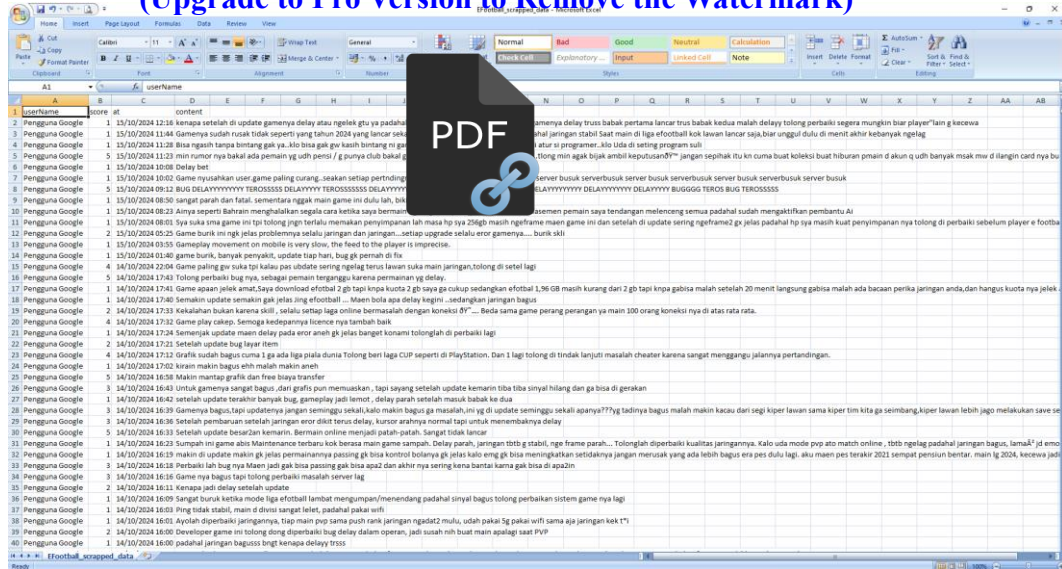
	reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt	appVersion
0	b7c7db0-7251-4a54-8a44-d503a0b47278	Pengguna Google	https://play-ih.googleusercontent.com/EGemol2N...	lokong untuk konami buastah flur gabisa matk...	1	355	9.2.0	2024-12-16 08:01:27	None	None	9.2.0
1	ec4f3a7-1081-4acc-ba6f-ac446e61a90f	Pengguna Google	https://play-ih.googleusercontent.com/EGemol2N...	Game macam apa ini sinyal bagus tapi pas man...	5	140	9.2.0	2024-12-17 09:48:54	None	None	9.2.0
2	511f198-6495-465a-a0f5-09760368a003	Pengguna Google	https://play-ih.googleusercontent.com/EGemol2N...	Melawan ai dan melawan player rasanya sangat b...	1	492	9.2.0	2024-12-14 00:19:45	None	None	9.2.0
3	fc0751e-3050-4329-a244-24a0879ac7fd	Pengguna Google	https://play-ih.googleusercontent.com/EGemol2N...	Gameplay sudah baik Tapi sejauh ini efootball...	4	92	9.2.0	2024-12-10 19:47:11	None	None	9.2.0
4	0839349-0466-4a20-8ed2-72e2c7f562e	Pengguna Google	https://play-ih.googleusercontent.com/EGemol2N...	Sangat disayangkan game ini memiliki grafik yg...	2	51	9.2.0	2024-12-13 09:22:10	None	None	9.2.0

```
len(df_busu.index) #count the number of data we got
```

Gambar 3. 1 Scrapping Data

Setelah proses pengambilan data selesai, hasil data ulasan tersebut disimpan menjadi file CSV. Berikut hasil data yang telah tersimpan dapat dilihat pada gambar 3.2 dibawah ini.

Protected by PDF Anti-Copy Free (Upgrade to Pro Version to Remove the Watermark)



Gambar 3. 2 Hasil Scraping Data

3.3 Metode Analisa

Tahap analisa dan hasil merupakan tahapan untuk melakukan pengolahan data menggunakan Algoritma BERT, tahap analisis dan hasil terdiri dari beberapa tahapan:

1. Klasifikasi BERT, tahap klasifikasi yang dilakukan dengan menghitung akurasi Algoritma BERT.
2. Pengelompokan opini dan proses pengolahan data menggunakan algoritma BERT menghasilkan nilai akurasi.

3.4 Tempat dan Waktu Penelitian

3.4.1 Tempat Penelitian

Penelitian dilakukan melalui proses *web scraping* dari ulasan pengguna *game online E-Football* menggunakan *google-play scraper Package* yang tersedia dalam *Python*.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

3.4.2 Waktu Penelitian

Jadwal penelitian akan mulai bulan September 2024 sampai dengan bulan Februari 2025. Mulai pengajuan judul, penyusunan skripsi hingga ujian skripsi dapat dijelaskan pada tabel 3.1 Berikut ini:

Table 3.1 Waktu Kegiatan

No	Jenis Kegiatan	Waktu Kegiatan (Bulan)					
		Sep 2024	Okt 2024	Nov 2024	Des 2024	Jan 2025	Feb 2025
1	Pengajuan Judul	█					
2	Pengumpulan Data	█	█				
3	Laporan Proposal		█				
4	Bimbingan Proposal			█			
5	Seminar Proposal				█		
6	Revisi Proposal				█		
7	Pembuatan Sistem					█	
8	Pengajuan Sistem					█	
9	Bimbingan Skripsi						█
10	Ujian Skripsi						█

3.5 Alat dan Bahan

3.5.1 Alat

Dalam melakukan penelitian tentunya diperlukan alat untuk mendukung proses penelitian. Adapun alat tersebut sebagai berikut:

1. Laptop
2. Printer
3. *Microsoft Office*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

4. *Web Browser Google Chrome*

5. *Google Colab*

6. Mendeley



3.5.2 Bahan

Adapun bahan yang diperlukan dalam proses penelitian sebagai berikut:

1. Jurnal
2. *E-book*
3. Kertas A4 70 gram
4. Tinta Printer

3.6 Metode Pengujian dan Pengolahan Data

Pada tahap pengolahan data menggunakan Algoritma BERT, langkah-langkah berikut dilakukan:

3.6.1 Metode Pengujian

Dalam penelitian ini metode pengujian dilakukan menggunakan library yang terdapat pada metode BERT yaitu *Confussion Matrix*, *confusion matrix* sangat penting untuk mengevaluasi kinerja model dengan membandingkan hasil prediksi dengan klasifikasi sebenarnya [44]. Dengan menggunakan *confusion matrix*, kita dapat melihat secara detail kinerja suatu sistem klasifikasi dan mengidentifikasi di mana terjadi kesalahan klasifikasi [45]. Pada dasarnya pengujian menggunakan *confussion matrix*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

terdapat 4 (empat) istilah mempresentasikan hasil dari proses klasifikasi.

Seperti yang terlihat pada gambar 2 berikut ini:

Table 3.2 Representasi Confusion Matrix

<i>Class</i>	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
<i>Negative</i>	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Keterangan :

- 1) *True Positive (TP)* Berarti seberapa banyak data yang aktual kelasnya positif, dan model juga memprediksi positif.
- 2) *False Positive (FN)* Berarti seberapa banyak data yang aktual kelasnya negatif, namun model memprediksi positif.
- 3) *True Negative (TN)* Berarti seberapa banyak data yang aktual kelasnya negatif, dan model memprediksi negatif.
- 4) *False Negative (FN)* Berarti seberapa banyak data yang aktual kelasnya positif, namun model memprediksi negatif. [46]

Accuracy, *precision*, dan *recall* dapat dihitung menggunakan informasi yang diberikan dalam *confusion matrix*.

1. *Accuracy*

Accuracy adalah ukuran seberapa tepat model dalam mengklasifikasikan data dengan benar, dan nilai *accuracy* dapat dihitung dengan membagi jumlah prediksi yang benar dengan

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Jumlah total sampel, *Accuracy* bertujuan untuk mengetahui tingkat

kesesuaian antara *Accuracy* sebenarnya dan nilai prediksi . Berikut

adalah rumus untuk mencari nilai *accuracy* :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. *Precision*

Precision adalah rasio identifikasi benar positif terhadap semua hasil identifikasi positif [47], menggambarkan *accuracy* antara data yang diminta dengan hasil prediksi yang diberikan oleh model, nilai *precision* yang besar menunjukkan rendahnya nilai *False Positive*. Berikut adalah rumus untuk mencari nilai *precision* :

$$Precision = \frac{TP}{TP + FP}$$

3. *Recall*

Recall adalah ukuran yang menunjukkan rasio pengamatan positif yang diprediksi dengan benar dibandingkan dengan semua pengamatan yang sebenarnya dalam kelas tersebut. Nilai *recall* yang tinggi menunjukkan rendahnya *False Negative*, yaitu kemungkinan prediksi negatif yang salah oleh model. Berikut adalah rumus untuk mencari nilai *recall* :

$$Recall = \frac{TP}{TP + FN}$$

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

4. *F1-Score*

F1-Score merupakan proses penggabungan *precision* dan *recall* pada matriks evaluasi. Berikut adalah rumus untuk mencari nilai

F1-Score :

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

3.6.2 Pengolahan Data Awal

Pada penelitian ini data yang berhasil didapatkan dari *Google Play Store* pada *game online e-football* akan dilakukan *Preprocessing* dan *Labeling*. Adapun tahapan dalam *Preprocessing* dan *Labeling* adalah sebagai berikut :

3.6.2.1 *Text Preprocessing*

Pada penelitian ini data yang berhasil didapatkan dari ulasan *game online e-football* akan dilakukan *Preprocessing*. Tahap *Preprocessing* dilakukan menggunakan *google collab*. Dengan dilakukannya *text preprocessing* ini akan terbentuk dataset yang bersih, dataset yang terbentuk dari proses ini akan memudahkan dalam pemrosesan sistem [48]. Adapun tahapan-tahapan dalam *Preprocessing* dapat dilihat sebagai berikut :

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

1. *Cleaning*

Proses ini bertujuan meminimalisir *noise* atau data yang tidak mendukung [49]. Ini adalah proses *cleansing* yang digunakan pada beberapa operator untuk penghapusan simbol, karakter, angka.

Table 3.3 Hasil *Cleaning*

Teks	Hasil <i>Cleaning</i>
Sumpah ini game abis	Sumpah ini game abis
Maintenance terbaru kok berasa	Maintenance terbaru kok berasa
main game sampah. Delay parah,	main game sampah Delay parah
jaringan tbtb g stabil, nge frame	jaringan tbtb g stabil nge frame
parah... Tolonglah diperbaiki	parah Tolonglah diperbaiki
kualitas jaringannya. Kalo uda	kualitas jaringannya Kalo uda
mode pvp ato match online , tbtb	mode pvp ato match online tbtb
ngelag padahal jaringan bagus,	ngelag padahal jaringan bagus
lama jd emosi jenuh mainnya	lama jd emosi jenuh mainnya

2. *Tokenizing*

Bertujuan untuk memecah teks menjadi kata-kata yang berurutan [50], tahap ini melakukan proses membagi teks menjadi unit-unit yang lebih kecil seperti kata-kata, frasa atau kalimat. Ini membantu dalam pemrosesan teks lebih lanjut karena masing-masing unit ini dapat diproses secara terpisah.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Table 3.4 Hasil *Tokenizing*

Teks	Hasil <i>Tokenizing</i>
Sumpah ini game abis	['Sumpah', 'ini', 'game', 'abis',
Maintenance terbaru kok berasa	'Maintenance', 'terbaru', 'kok',
main game sampah Delay parah	'berasa', 'main', 'game', 'sampah',
jaringan tbtb g stabil nge frame	'Delay', 'parah', 'jaringan', 'tbtb',
parah Tolonglah diperbaiki	'g', 'stabil', 'nge', 'frame', 'parah',
kualitas jaringannya Kalo uda	'Tolonglah', 'diperbaiki',
mode pvp ato match online tbtb	'kualitas', 'jaringannya', 'Kalo',
ngelag padahal jaringan bagus	'mode', 'pvp', 'ato', 'match',
lama jd emosi jenuh mainnya	'online', 'tbtb', 'ngelag', 'padahal',
	'jaringan', 'bagus', 'lama', 'jd',
	'emosi', 'jenuh', 'mainnya']

3. *Case Folding*

Case Folding adalah proses penyeragaman bentuk huruf serta penghilangan tanda baca [51]. Fitur *Case Folding* ini dapat secara otomatis mengubah semua huruf pada teks menjadi huruf kecil semua atau menjadi huruf kapital semua. Pada penelitian ini peneliti mengubahnya menjadi huruf kecil semua.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Table 3.5 Hasil *Case Folding*

Tek	Hasil <i>Case Folding</i>
['Sumpah', 'ini', 'abis', 'Maintenance', 'terbaru', 'kok', 'berasa', 'main', 'game', 'sampah', 'Delay', 'parah', 'jaringan', 'tbtb', 'g', 'stabil', 'nge', 'frame', 'parah', 'tolonglah', 'diperbaiki', 'kualitas', 'jaringannya', 'Kalo', 'uda', 'mode', 'pvp', 'ato', 'match', 'online', 'tbtb', 'ngelag', 'padahal', 'jaringan', 'online', 'tbtb', 'ngelag', 'padahal', 'bagus', 'lama', 'jd', 'emosi', 'jenuh', 'jaringan', 'bagus', 'lama', 'jd', 'mainnya']	['sumpah', 'ini', 'game', 'abis', 'maintenance', 'terbaru', 'kok', 'berasa', 'main', 'game', 'sampah', 'delay', 'parah', 'jaringan', 'tbtb', 'g', 'stabil', 'nge', 'frame', 'parah', 'tolonglah', 'diperbaiki', 'kualitas', 'jaringannya', 'kalo', 'uda', 'mode', 'pvp', 'ato', 'match', 'online', 'tbtb', 'ngelag', 'padahal', 'jaringan', 'online', 'tbtb', 'ngelag', 'padahal', 'bagus', 'lama', 'jd', 'emosi', 'jenuh', 'mainnya']

4. *Stopwords Removal*

Tahapan ini menerapkan model *stoplist* (menghilangkan token yang kurang efisien) atau *word list* (menggambil token yang efisien) [52].

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Table 3.6 Hasil *Stopwords Removal*

Tek	Hasil <i>Stopwords Removal</i>
['sumpah', 'ini', 'abis', 'maintenance', 'terbaru', 'kok', 'berasa', 'main', 'game', 'sampah', 'delay', 'parah', 'jaringan', 'tbtb', 'g', 'stabil', 'nge', 'frame', 'parah', 'tolonglah', 'diperbaiki', 'kualitas', 'jaringannya', 'kalo', 'uda', 'mode', 'pvp', 'ato', 'match', 'online', 'tbtb', 'ngelag', 'padahal', 'jaringan', 'bagus', 'lama', 'jd', 'emosi', 'jenuh', 'mainnya']	['sumpah', 'game', 'abis', 'maintenance', 'terbaru', 'berasa', 'main', 'game', 'sampah', 'delay', 'parah', 'jaringan', 'tbtb', 'stabil', 'nge', 'frame', 'parah', 'tolonglah', 'diperbaiki', 'kualitas', 'jaringan', 'mode', 'pvp', 'match', 'online', 'tbtb', 'ngelag', 'jaringan', 'bagus', 'emosi', 'jenuh', 'mainnya']


5. *Normalization*

Normalization tahapan untuk memperbaiki ejaan yang salah pada setiap token atau mengubah menjadi kata baku [53]. Berikut adalah hasil dari proses *Normalization* :

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Table 3.7 Hasil Normalization

Teks		Hasil Normalization
['sumpah', 'game', 'abis', 'maintenance', 'terbaru', 'berasa', 'main', 'game', 'sampah', 'delay', 'parah', 'jaringan', 'tbtb', 'stabil', 'nge', 'frame', 'parah', 'tolonglah', 'diperbaiki', 'kualitas', 'jaringan', 'mode', 'pvp', 'match', 'online', 'tbtb', 'ngelag', 'jaringan', 'bagus', 'emosi', 'jenuh', 'mainnya']		['sumpah', 'game', 'habis', 'maintenance', 'terbaru', 'terasa', 'main', 'game', 'sampah', 'delay', 'parah', 'jaringan', 'tiba-tiba', 'stabil', 'nge', 'frame', 'parah', 'tolong', 'perbaiki', 'kualitas', 'jaringan', 'mode', 'pvp', 'match', 'online', 'tiba-tiba', 'ngelag', 'jaringan', 'bagus', 'emosi', 'jenuh', 'main']

6. Stemming

Merupakan proses mengubah kata dalam dokumen menjadi kata dasarnya. *Stemming* dilakukan untuk menyeragamkan bentuk kata [54]. Contohnya, kata “terasa”, “berasa”, dan “perasa” akan disistem menjadi kata dasar “rasa”. Berikut hasil dari proses *stemming* :

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Table 3.8 Hasil *Stemming*

Tekst	Hasil <i>Stemming</i>
['sumpah', 'game', 'habis', 'maintenance', 'terbaru', 'terasa', 'main', 'game', 'sampah', 'delay', 'parah', 'jaringan', 'tiba-tiba', 'stabil', 'nge', 'frame', 'parah', 'tolong', 'perbaiki', 'kualitas', 'jaringan', 'mode', 'pvp', 'match', 'online', 'tiba-tiba', 'ngelag', 'jaringan', 'bagus', 'emosi', 'jenuh', 'main']	['sumpah', 'game', 'habis', 'maintenance', 'baru', 'rasa', 'main', 'game', 'sampah', 'delay', 'parah', 'jaringan', 'tiba', 'stabil', 'nge', 'frame', 'parah', 'tolong', 'perbaiki', 'kualitas', 'jaringan', 'mode', 'pvp', 'match', 'online', 'tiba', 'lag', 'jaringan', 'bagus', 'emosi', 'jenuh', 'main']

3.6.2.2 Labeling

Pada tahap *labeling* data dilakukan dengan cara melihat komentar di google play store pada *game online e-football*, untuk melihat hasil dari komentar dan ulasan tersebut apakah positif, negatif atau netral dapat dilihat pada tabel 10 berikut :

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Table 3.9. Labeling Data

Sentimen	Hasil <i>Preprocessing</i>	Label
"Gamenya sangat grafisnya juga memuaskan"	['bagus', 'grafis', 'puas']	Positif
"Performanya standar, tidak buruk tapi juga tidak istimewa."	['performa', 'standar', 'tidak', 'buruk', 'tidak', 'istimewa']	Netral
"Delay parah, susah gerakin pemainnya."	['delay', 'parah', 'susah', 'pemain']	Negatif
"Gamenya ngelag banget bikin emosi mainnya"	['lag', 'parah', 'emosi']	Negatif

3.6.3 Pengolahan Data Menggunakan BERT

1. *Train dan Test Data Splitting*

Pada tahap ini data dibagi menjadi dua, yaitu 80% data latih dan 20% data uji untuk mengukur kinerja model. Proses ini melibatkan pembagian *dataset* menjadi dua bagian utama, satu untuk melatih model (*train data*) dan satu untuk uji data (*test data*).

2. *Input Formating*

Pada Algoritma BERT mengacu pada cara teks disusun dan disiapkan sebelum diberikan ke model untuk pelatihan atau inferensi.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

3. *Fine Tuning*

Fine tuning adalah proses melatih ulang Algoritma BERT pada tugas tertentu dengan menggunakan data tugas yang lebih kecil. Tujuan dari *fine tuning* adalah untuk mengadaptasi representasi BERT yang sudah diperoleh pada data umum besar untuk tugas atau domain tertentu.

4. Evaluasi

Pada tahap evaluasi, dilakukan analisis terhadap performa model IndoBERT menggunakan metrik seperti *confusion matrix*.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

BAB IV

PENELITIAN DAN PEMBAHASAN



4.1 Gambaran Umum Penelitian

Pada bab ini, analisis sentimen terhadap game *online E-Football* di *Play Store* bertujuan untuk menggambarkan bagaimana pengguna merespons permainan ini melalui ulasan yang mereka tulis. Ulasan pengguna mencerminkan pengalaman dan persepsi mereka terhadap kualitas *gameplay*, grafis, fitur, dan performa dari game tersebut. Analisis sentimen bertujuan untuk mengidentifikasi apakah ulasan-ulasan tersebut bersifat positif, netral, atau negatif, memberikan wawasan yang lebih dalam tentang persepsi publik terhadap game ini. Melalui pendekatan *machine learning* dengan model BERT, analisis ini dapat memproses data ulasan dalam jumlah besar secara efektif, mengelompokkan sentimen, dan memberikan kesimpulan yang akurat mengenai kepuasan pengguna terhadap *E-Football*.

Sentimen yang terkandung dalam ulasan *Play Store* memainkan peran penting dalam pengembangan game, karena dapat menjadi indikator bagaimana pengembang harus menyesuaikan atau memperbarui fitur-fitur dalam game untuk memenuhi harapan pemain. Penggunaan BERT (*Bidirectional Encoder Representations from Transformers*) dalam analisis sentimen memungkinkan pemahaman yang lebih baik terhadap konteks dalam setiap ulasan, yang penting untuk menilai sentimen secara lebih akurat.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Dengan model ini, ulasan yang ditulis dalam bahasa Indonesia dapat dianalisis dengan baik, hitungkan aspek-aspek seperti konteks kalimat dan penggunaan kata yang bisa mengubah arti dari ulasan itu sendiri. Analisis semacam ini sangat berguna bagi pengembang game untuk mendapatkan umpan balik yang dapat dijadikan acuan dalam meningkatkan kualitas game.

4.2 Hasil Penelitian

4.2.1 Pengumpulan Data

Pengumpulan data dilakukan menggunakan *library snsrape* yang ada pada bahasa pemrograman *python*. Langkah pertama dalam pengumpulan data adalah menginstal *library snsrape* di *google colab*, dengan menggunakan *source code* berikut ini "`!pip install-q snsrape`". Kemudian dapat dilakukan pengambilan data sesuai jumlah yang penulis inginkan. Berikut adalah gambar *source code* yang penulis gunakan untuk melakukan pengambilan data atau *Scraping*.

```
#Scrape desired number of reviews
#Run kode ini jika ingin scrape data dengan jumlah tertentu. Ganti (misal, ingin scrape sejumlah 1000, maka ganti kode , count = 1000 )

from google_play_scraper import Sort, reviews

result, continuation_token = reviews(
    'jp.konami.pesam&hl=id',
    lang='id', # defaults to 'en'
    country='id', # defaults to 'us'
    sort=Sort.MOST_RELEVANT, # defaults to Sort.MOST_RELEVANT you can use Sort.NEWEST to get newst reviews
    count=25000, # defaults to 100
    filter_score_with=None # defaults to None(means all score) Use 1 or 2 or 3 or 4 or 5 to select certain score
)
```

Gambar 4.1 *Source code* pengambilan data

Pada penelitian ini, data yang diperoleh melalui *source code* tersebut berjumlah 25.000 data. Data yang dikumpulkan terdapat 4 kolom yaitu

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

username, *score*, *at* dan *content*. Kemudian disimpan dengan format csv.

Hasil pengumpulan data csv yang ada pada gambar berikut ini.



	userName	score	at	content
0	Muhammad Saputra	5	2024-12-07 04:21:04	Sejujurnya kalo untuk grafik bagus kok, tapi k...
1	Tiqáh (Naa)	4	2024-11-19 10:41:26	game comfort aku sih. grafik keren, tapi agak ...
2	Zahra Ain	5	2024-12-09 13:27:00	saya sangat suka dengan game ini, grafik nya b...
3	Hyoorin	5	2024-12-01 23:28:05	Semuanya sudah bagus dari cerita, grafik, batt...
4	Jeanne Christabel Krisnugroho	4	2024-12-06 04:32:31	bagus banget buat gamenya, dari segi lore, cha...

Gambar 4.2 Hasil Pengambilan Data

4.2.2 Preprocessing


Tahap *preprocessing* adalah dikumpulkan menjadi format yang diperlukan. Tahap *preprocessing* mengubah data yang berhasil dilakukan dengan bahasa pemrograman *python* yang dijalankan di *google colab*. Adapun langkah-langkah tahap *preprocessing* yang dilakukan dalam penelitian ini adalah sebagai berikut:

- 1) *Cleaning*
- 2) *Case Folding*
- 3) *Tokenizing*
- 4) *Normalisasi*
- 5) *Stopword*
- 6) *Stemming*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

1. *Cleaning*

Cleaning yaitu  pembersihan data ulasan dari kata yang tidak dibutuhkan seperti *HTML* *name* (*@username*), *emoticon*, *hashtag*(#), angka, url, dan lain-lain. Adapun *source code* bahasa pemrograman *python* yang digunakan untuk proses *cleaning* adalah sebagai berikut :

```
import re

# Fungsi cleaning (membersihkan teks dalam kolom content)
def clean_text_and_filter(df, text_field, new_text_field_name):
    # Mengubah teks menjadi huruf kecil (Case Folding)
    df[new_text_field_name] = df[text_field].str.lower()

    # Membersihkan teks tanpa menghapus angka
    df[new_text_field_name] = df[new_text_field_name].apply(
        lambda elem: re.sub(r"(@[A-Za-z0-9]+)|([^\0-9A-Za-z \t])|http.+?|&quot;;", "", elem)
    )
    # Hapus emoji dan karakter HTML
    df[new_text_field_name] = df[new_text_field_name].apply(lambda elem: re.sub(r'\s+',
    ', elem).strip()) # Menghapus spasi ganda
    df[new_text_field_name] = df[new_text_field_name].apply(lambda elem:
    re.sub(r'\b(\w+)(\1)+\b', r'\1', elem)) # Menghapus pengulangan kata
    df[new_text_field_name] = df[new_text_field_name].apply(
        lambda elem: re.sub(r'[^\x00-\x7F]+', "", elem) # Menghapus semua karakter non-
    ASCII (emoji)
    )

    return df

# Proses pembersihan teks dari kolom 'content'
df = clean_text_and_filter(df, text_field='content',
new_text_field_name='cleaned_content')

# Menampilkan hasil pembersihan
print(df[['content', 'cleaned_content']].head())
```

Adapun hasil dari proses *Cleaning* data ulasan yang sudah didapatkan dapat dilihat pada gambar berikut ini :

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

```
content \
0 kenapa setelah di update gamenya delay atau ng...
1 Gamenya sudah rusak tidak seperti yang tahun 2...
2 Bisa ngasih tanpa bintang gak yaklo bisa gak g...
3 min rumor nya bakal ada pemain yg udh pensi / ...
4 delay bet

cleaned_content
0 kenapa setelah di update gamenya delay atau ng...
1 gamenya sudah rusak tidak seperti yang tahun 2...
2 bisa ngasih tanpa bintang gak yaklo bisa gak g...
3 min rumor nya bakal ada pemain yg udh pensi g ...
4 delay bet
```

Gambar 4.3 Hasil Proses *Cleaning*

2. Normalization

Pada tahap ini merupakan tahap untuk mengubah kalimat yang tidak baku atau *slangword* menjadi kalimat baku. Adapun *source code* bahasa pemrograman *python* yang digunakan untuk melakukan proses normalisasi dapat dilihat berikut ini:

```
import pandas as pd

# Membaca file CSV yang berisi kata-kata normalisasi
normalized_word = pd.read_csv("kamus_normalisasi_baru.csv", encoding='latin1',
on_bad_lines='skip')

# Membuat dictionary dari file CSV (kolom pertama sebagai kata dan kolom kedua
sebagai kata normalisasi)
normalized_word_dict = {}
for index, row in normalized_word.iterrows():
    if row[0] not in normalized_word_dict:
        normalized_word_dict[row[0]] = row[1]

# Fungsi untuk melakukan normalisasi
def normalized_term(document):
    # Memastikan document dipecah menjadi daftar kata
    words = document.split()
    # Menormalisasi setiap kata dalam dokumen
    return " ".join([normalized_word_dict[term] if term in normalized_word_dict else term
for term in words])
```

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

```
# Menormalisasi teks dalam kolom 'cleaned_content' dan simpan hasilnya dalam kolom
baru 'normalized_content'
df['normalized_content'] = df['cleaned_content'].apply(normalized_term)

# Menampilkan hasil
print(df[['cleaned_content', 'normalized_content']].head())
```

Adapun hasil dari proses *Normalization data* ulasan yang sudah didapatkan dapat dilihat pada gambar berikut ini :

```
cleaned_content \
0 kenapa setelah di update gamenya delay atau ng...
1 gamenya sudah rusak tidak seperti yang tahun 2...
2 bisa ngasih tanpa bintang gak yaklo bisa gak g...
3 min rumor nya bakal ada pemain yg udh pensi g ...
4 delay bet

normalized_content
0 kenapa setelah di baru gamenya lambat atau nge...
1 gamenya sudah rusak tidak seperti yang tahun 2...
2 bisa memberi tanpa bintang tidak yaklo bisa ti...
3 kak rumor nya bakal ada pemain yang udah pensi...
4 lambat banget
```

Gambar 4.4 Hasil proses *Normalization*

3. Case Folding

Case Folding merupakan proses yang digunakan untuk menyeragamkan huruf yang terdapat pada ulasan, dari huruf kapital menjadi huruf kecil (*Lower Case*). Adapun *source code* yang digunakan untuk melakukan proses *case folding* adalah sebagai berikut:

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

```
# Fungsi untuk case folding
def case_folding(df, text_field, new_text_field_name):
    if text_field not in df.columns:
        raise KeyError(f"Kolom '{text_field}' tidak ditemukan dalam DataFrame.")
    df[new_text_field_name] = df[text_field].astype(str).str.lower() # Konversi ke string
    dan ubah ke huruf kecil
    return df

# Proses case folding pada kolom 'normalized_content'
try:
    df = case_folding(df, text_field='normalized_content',
new_text_field_name='case_folded_content')
    # Menampilkan hasil case folding
    print(df[['normalized_content', 'case_folded_content']])
except KeyError as e:
    print(e)
```

Adapun hasil dari proses *Case Folding* ulasan yang sudah didapatkan dapat dilihat pada gambar berikut ini :

```
                                normalized_content
0      kenapa setelah di baru gamenya lambat atau nge...
1      gamenya sudah rusak tidak seperti yang tahun 2...
2      bisa memberi tanpa bintang tidak yaklo bisa ti...
3      kak rumor nya bakal ada pemain yang udah pensi...
4      lambat banget
...
24995  perhatikan tolong konami jangan pakai energi m...
24996  game ini cukup asyik dan grafisnya juga cukup ...
24997  tolong kalau mau bertanding required energinya...
24998  untuk keseluruhan sudah bagus tapi sekarang me...
24999  gamenya sangat bagus tinggal jursey ajah yang ...

                                case_folded_content
0      kenapa setelah di baru gamenya lambat atau nge...
1      gamenya sudah rusak tidak seperti yang tahun 2...
2      bisa memberi tanpa bintang tidak yaklo bisa ti...
3      kak rumor nya bakal ada pemain yang udah pensi...
4      lambat banget
...
24995  perhatikan tolong konami jangan pakai energi m...
24996  game ini cukup asyik dan grafisnya juga cukup ...
24997  tolong kalau mau bertanding required energinya...
24998  untuk keseluruhan sudah bagus tapi sekarang me...
24999  gamenya sangat bagus tinggal jursey ajah yang ...
```

Gambar 4.5 Hasil Proses *Case Folding*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

4. *Stopword*

Stopword adalah penyaringan dan penghapusan kata yang tidak penting dari analisis. Pada penelitian ini penulis menggunakan *stopword* bahasa Indonesia yang di dapatkan dari library *NLTK* yang terdapat pada bahasa pemrograman *Python* kemudian ditambah dengan *stopword* yang penulis buat sendiri. Adapun *source code* bahasa pemrograman *python* yang digunakan untuk melakukan proses *stopword* adalah sebagai berikut :

```
import nltk
from nltk.corpus import stopwords

# Fungsi untuk menghapus stopwords
def remove_stopwords(df, text_field, new_text_field_name):
    nltk.download('stopwords')
    stop = stopwords.words('indonesian')
    df[new_text_field_name] = df[text_field].apply(
        lambda x: ''.join([word for word in x.split() if word not in stop])
    )
    return df

# Proses menghapus stopwords pada kolom 'case_folded_content'
df = remove_stopwords(df, text_field='case_folded_content',
new_text_field_name='no_stopwords_content')

# Menampilkan hasil penghapusan stopwords
print(df[['case_folded_content', 'no_stopwords_content']].head())
```

Adapun hasil dari proses *Stopword* ulasan yang sudah didapatkan dapat dilihat pada gambar berikut ini :

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

```
case_folded_content
0 kenapa setelah di baru gamenya lambat atau nge...
1 gamenya sudah tidak seperti yang tahun 2...
2 bisa memberi yang tidak yaklo bisa ti...
3 kak rumor nya pemain yang udah pensi...
4 lambat banget

no_stopwords_content
0 gamenya lambat ngelek gtu ya jaringannya lanca...
1 gamenya rusak 2024 lancar kebanyak lambat jari...
2 bintang yaklo kasih bintang ni gamegame permai...
3 kak rumor nya pemain udah pensi g kelompok g d...
4 lambat banget
```

Gambar 4.6 Hasil Proses *Stopword*

5. *Tokenizing*

Tokenizing adalah proses memecahkan atau memotong kata pada dokumen menjadi *term-term* berdasarkan spasi. Adapun *source code* bahasa pemrograman *python* yang digunakan untuk melakukan proses *Tokenizing* dapat dilihat berikut ini :

```
import nltk
from nltk.tokenize import word_tokenize

# Fungsi untuk tokenisasi
def tokenize_text(df, text_field, new_text_field_name):
    nltk.download('punkt')
    df[new_text_field_name] = df[text_field].apply(lambda x: word_tokenize(x))
    return df

# Tokenisasi teks dalam kolom 'normalized_content'
df = tokenize_text(df, text_field='normalized_content',
                  new_text_field_name='tokenized_content')

# Menampilkan hasil tokenisasi
print(df[['normalized_content', 'tokenized_content']].head())
```

Adapun hasil dari proses *Tokenizing* ulasan yang sudah didapatkan dapat dilihat pada gambar berikut ini :

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

```
no_stopwords_content
0 gamenya lambat ngelek gtu ya jaringannya lanca...
1 gamenya rusak lancar sebanyak lambat jari...
2 bintang yaklo tang ni gamegame permai...
3 kak rumor nya ah pensi g kelompok g d...
4 lambat banget

tokenized_content
0 [gamenya, lambat, ngelek, gtu, ya, jaringannya...
1 [gamenya, rusak, 2024, lancar, sebanyak, lamba...
2 [bintang, yaklo, kasih, bintang, ni, gamegame,...
3 [kak, rumor, nya, pemain, udah, pensi, g, kelo...
4 [lambat, banget]
```

Gambar 4.7 Hasil proses *Tokenizing*

6. *Stemming*

Stemming adalah proses untuk menghilangkan awalan atau akhiran kata yang terdapat kata sambung, kata depan, kata ganti, menjadi kata dasar. Proses *Stemming* dilakukan dengan *library sastrawi* yang terdapat pada bahasa pemrograman *python*. Adapun *source code* bahasa pemrograman *python* yang digunakan untuk melakukan proses *stemming* adalah sebagai berikut:

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

# Fungsi untuk stemming
def stem_text(df, text_field, new_text_field_name):
    stemmer = StemmerFactory().create_stemmer() # Membuat objek stemmer
    df[new_text_field_name] = df[text_field].apply(
        lambda tokens: [stemmer.stem(token) for token in tokens] # Melakukan stemming
        pada setiap token
    )
    return df

# Memastikan kolom 'tokenized_content' sudah berisi list token
if df['tokenized_content'].apply(lambda x: isinstance(x, list)).all():
    # Stemming teks dalam kolom 'tokenized_content'
    df = stem_text(df, text_field='tokenized_content',
        new_text_field_name='stemmed_content')
```

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

```
# Menampilkan hasil stemming
print(df[['tokenized_content', 'stemmed_content']].head())
else:
    print("Kolom 'tokenized_content' berisi list. Periksa proses tokenisasi terlebih dahulu.")
```



Adapun hasil dari proses *Stemming* ulasan yang sudah didapatkan dapat dilihat pada gambar berikut ini :

```
                                tokenized_content
0  [gamenya, lambat, ngelek, gtu, ya, jaringannya...
1  [gamenya, rusak, 2024, lancar, kebanyakan, lamba...
2  [bintang, yaklo, kasih, bintang, ni, gamegame,...
3  [kak, rumor, nya, pemain, udah, pensi, g, kelo...
4                                     [lambat, banget]

                                stemmed_content
0  [gamenya, lambat, ngelek, gtu, ya, jaring, lan...
1  [gamenya, rusak, 2024, lancar, banyak, lambat,...
2  [bintang, yaklo, kasih, bintang, ni, gamegame,...
3  [kak, rumor, nya, main, udah, pensi, g, kelomp...
4                                     [lambat, banget]
```

Gambar 4.8 Hasil proses *Stemming*

7. Labeling

Pada bagian ini, proses *labeling* dilakukan untuk menentukan sentimen dari setiap ulasan pengguna game online *E-Football* di *Play Store*. *Labeling* dilakukan dengan memanfaatkan kamus sentimen (*Lexicon*) dalam bahasa Indonesia, yang diperoleh melalui sumber terbuka di *GitHub* [55]. Kamus ini terdiri dari daftar kata-kata dengan asosiasi sentimen positif dan negatif, masing-masing disertai bobot atau nilai tertentu yang mencerminkan seberapa kuat kata tersebut memengaruhi sentimen. Adapun *source code* bahasa pemrograman

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)
python yang digunakan untuk melakukan proses *labeling* adalah sebagai

berikut:



```
# Membaca kamus positif dan negatif ke dalam kolom 'word' dan 'weight'
positive_words_df = pd.read_csv('content/positive_update.tsv')
negative_words_df = pd.read_csv('/content/negative_update.tsv')

# Mengubah kamus menjadi dictionary untuk kemudahan pencarian
positive_words = dict(zip(positive_words_df['word'], positive_words_df['weight']))
negative_words = dict(zip(negative_words_df['word'], negative_words_df['weight']))

# Fungsi untuk memberikan label sentimen berdasarkan kata positif dan negatif dengan bobot
def label_sentiment_with_weight(text):
    total_score = 0 # Total bobot sentimen untuk setiap ulasan

    # Menghitung total bobot kata positif dan negatif dalam teks
    for word in text.split():
        if word in positive_words:
            total_score += positive_words[word] # Menambahkan bobot kata positif
        elif word in negative_words:
            total_score += negative_words[word] # Menambahkan bobot kata negatif

    # Menentukan label berdasarkan total bobot
    if total_score > 0:
        return 'Positif'
    elif total_score < 0:
        return 'Negatif'
    else:
        return 'Netral'

# Pastikan nilai di kolom 'stemmed_content' berupa string
df['stemmed_content'] = df['stemmed_content'].apply(lambda x: ''.join(x) if isinstance(x, list) else x)

# Menerapkan fungsi labeling sentimen
df['sentiment'] = df['stemmed_content'].apply(label_sentiment_with_weight)


# Menampilkan beberapa baris pertama setelah diberi label dengan bobot
print(df[['content', 'stemmed_content', 'sentiment']].head())
```

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

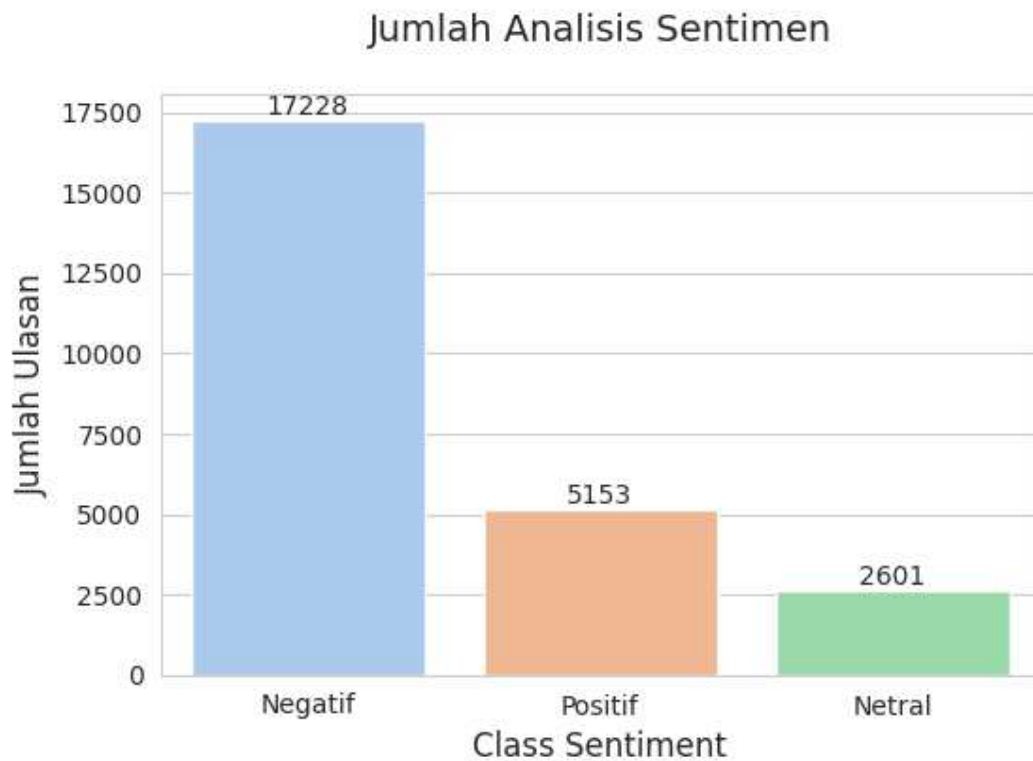
Adapun hasil dan visualisasi dari proses *labeling* ulasan yang sudah

didapatkan dapat dilihat pada gambar berikut ini :



	content \	
0	kenapa setelah di... enya delay atau ng...	
1	Gamenya sudah rusak tidak seperti yang tahun 2...	
2	Bisa ngasih tanpa bintang gak ya..klo bisa gak...	
3	min rumor nya bakal ada pemain yg udh pensi / ...	
4	Delay bet	
	stemmed_content	sentiment
0	kenapa telah di baru gamenya lambat atau ngele...	Positif
1	gamenya sudah rusak tidak seperti yang tahun 2...	Positif
2	bisa beri tanpa bintang tidak yaklo bisa tidak...	Negatif
3	kak rumor nya bakal ada main yang udah pensi g...	Negatif
4	lambat bet	Positif


Gambar 4.9 Hasil Proses *Labeling*



Gambar 4.10 Visualisasi Grafik *Labeling*

Protected by PDF Anti-Copy Free

8. *Label Encoding* (Upgrade to Pro Version to Remove the Watermark)

Setelah proses  menghasilkan tiga kategori sentimen berupa *string* yaitu Positif, Netral, dan Negatif, tahap selanjutnya adalah melakukan *encoding label* menggunakan *Label Encoder* dari *library Scikit-learn*. Proses ini bertujuan untuk mengubah label sentimen berbentuk *string* menjadi nilai numerik, yang lebih mudah diproses oleh algoritma *machine learning*. Adapun *source code* bahasa pemrograman *python* yang digunakan untuk melakukan proses *label Encoding* adalah sebagai berikut:

```
# Import LabelEncoder dari scikit-learn
from sklearn.preprocessing import LabelEncoder
# Inisialisasi LabelEncoder
label_encoder = LabelEncoder()

# Encode label string menjadi label numerik
df['encoded_label'] = label_encoder.fit_transform(df['sentiment'])

# Verifikasi hasil encoding
print(df[['sentiment', 'encoded_label']].head())
```

Adapun hasil dari proses *label encoding* dapat dilihat pada gambar berikut ini :

	sentiment	encoded_label		sentiment	encoded_label
0	Negatif	0	24987	Negatif	0
1	Negatif	0	24988	Negatif	0
2	Positif	2	24989	Positif	2
3	Negatif	0	24990	Positif	2
4	Netral	1	24991	Negatif	0

Gambar 4.11 Hasil Proses *Encoding*


Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

4.2.3 Pengujian Sistem

4.2.3.1 Klasifikasi dan Evaluasi



Setelah dilakukan  data, proses selanjutnya yaitu klasifikasi.

Data yang sudah diberi label kemudian dibagi menjadi dua, yaitu data *training* dan data *testing*. Proses klasifikasi dan evaluasi pada penelitian ini dilakukan menggunakan metode BERT.

1. Pembagian Data *Training* dan Data *Testing*

Langkah awal klasifikasi menggunakan metode BERT adalah membagi data menjadi dua, yaitu data *training* dan data *testing*. Pada penelitian ini penulis menggunakan perbandingan 80% : 20%. Adapun *source code* bahasa pemrograman *python* yang digunakan untuk melakukan proses pembagian data atau *Splitting* adalah sebagai berikut:

```
from sklearn.model_selection import train_test_split

# Split dataset menjadi 80% untuk pelatihan dan 20% untuk pengujian
X = df['stemmed_content'] # Fitur (text)
y = df['sentiment'] # Target (label sentimen)

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Menampilkan ukuran data setelah split
print(f" Jumlah data pelatihan: {len(X_train)}")
print(f" Jumlah data pengujian: {len(X_test)}")
```

Hasil proses *Splitting* dapat dilihat pada gambar dibawah ini :

```
Jumlah data pelatihan: 19984
Jumlah data pengujian: 4998
```

Gambar 4.12 Hasil proses *Splitting*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

2. Klasifikasi dan Evaluasi menggunakan BERT

Metode BERT melakukan pelatihan *Machine* dengan cara mendeteksi ciri-ciri pada tiap kelas yang dibentuk pada data *training*. Hasil pelatihan pada metode BERT kemudian dievaluasi menggunakan data *testing*. Proses evaluasi dilakukan menggunakan *library* yang terdapat pada metode BERT yaitu *confusion matrix* untuk mencari nilai *accuracy*, *precision*, dan *recall*.

3. Tokenisasi dengan IndoBERT

Sebelum masuk ke *modeling*, kita perlu mengubah teks menjadi format yang bisa diterima oleh model IndoBERT. Biasanya ini dilakukan dengan menggunakan *tokenizer* dari *transformers*.

```
from transformers import BertTokenizer

# Memuat tokenizer untuk IndoBERT
tokenizer = BertTokenizer.from_pretrained('indolem/indobert-base-uncased')

# Fungsi untuk tokenisasi teks
def tokenize_function(text):
    return tokenizer(text, padding='max_length', truncation=True, max_length=128)

# Tokenisasi data pelatihan dan pengujian
train_encodings = tokenizer(list(X_train), padding=True, truncation=True,
max_length=128)
test_encodings = tokenizer(list(X_test), padding=True, truncation=True,
max_length=128)
```

4. Fine Tuning Model IndoBERT

Pada tahap ini adalah memuat model IndoBERT dan melakukan *fine tuning* menggunakan data pelatihan. Adapun *source*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

code bahasa pemrograman *python* yang digunakan untuk melakukan

proses *fine tuning* sebagai berikut:

```
from transformers import BertTokenizer, BertForSequenceClassification, Trainer,
TrainingArguments
import torch
import pandas as pd
from sklearn.preprocessing import LabelEncoder
import re

# Load tokenizer dan model
tokenizer = BertTokenizer.from_pretrained('indolem/indobert-base-uncased')
model = BertForSequenceClassification.from_pretrained('indolem/indobert-base-
uncased', num_labels=3)

# Dataset untuk BERT
class SentimentDataset(torch.utils.data.Dataset):
    def __init__(self, encodings, labels):
        self.encodings = encodings
        self.labels = labels

    def __getitem__(self, idx):
        item = {key: torch.tensor(val[idx]) for key, val in self.encodings.items()}
        item['labels'] = torch.tensor(self.labels[idx]) # pastikan label dalam bentuk tensor
numerik
        return item
    def __len__(self):
        return len(self.labels)

# Siapkan dataset untuk pelatihan
train_dataset = SentimentDataset(encodings, data['encoded_label'].values)

# Argumen untuk pelatihan
training_args = TrainingArguments(
    output_dir='./results',           # Direktori output
    num_train_epochs=3,              # Jumlah epoch
    per_device_train_batch_size=8,    # Ukuran batch per perangkat
    per_device_eval_batch_size=16,    # Ukuran batch saat evaluasi
    warmup_steps=500,                # Langkah untuk warmup
    weight_decay=0.01,               # Pengurangan bobot
    logging_dir='./logs',            # Direktori log
    logging_steps=10,
)

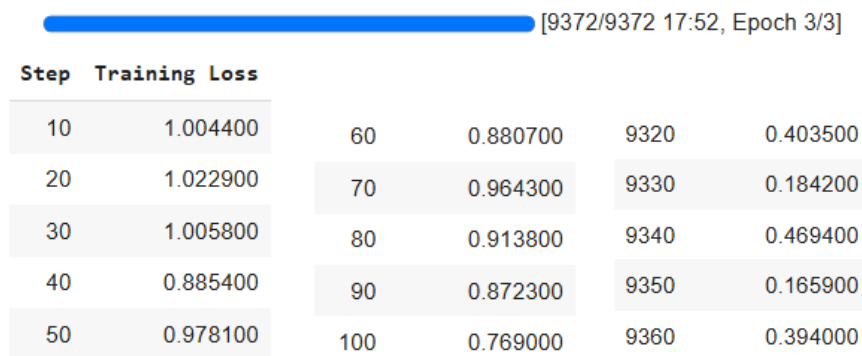
# Inisialisasi Trainer dengan model, argumen, dan dataset
trainer = Trainer(
```

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

```
model=model, # Model BERT
args=training_args, # Argumen pelatihan
train_dataset=train_dataset # Dataset untuk pelatihan
)
# Melakukan pelatihan
trainer.train()
```

Adapun hasil dari proses *Fine Tuning* Model IndoBERT dapat dilihat pada gambar berikut ini:



Gambar 4.13 *Fine Tuning*

5. Evaluasi Model

Setelah pelatihan selesai, kita dapat mengevaluasi model pada data pengujian untuk melihat kinerjanya. Berikut adalah *source code* evaluasi model :

```
eval_results = trainer.evaluate(eval_dataset=test_dataset)
print(eval_results) # Melihat hasil evaluasi (loss, accuracy, dll.)
```

Adapun hasil dari proses evaluasi model dapat dilihat pada gambar berikut ini:

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

[313/313 00:15]

```
{'eval_loss': 1.851107007232056, 'eval_runtime': 15.8335,
```

```
'eval_samples_per_second': 315.0, 'eval_steps_per_second': 19.768, 'epoch': 3.0}
```

Gambar 1.1 Evaluasi Model

6. Confusion Matrix

Confusion matrix adalah Visualisasi yang sangat berguna untuk mengevaluasi performa model klasifikasi. Berikut adalah *source code Confusion matrix*:

```
from sklearn.metrics import confusion_matrix

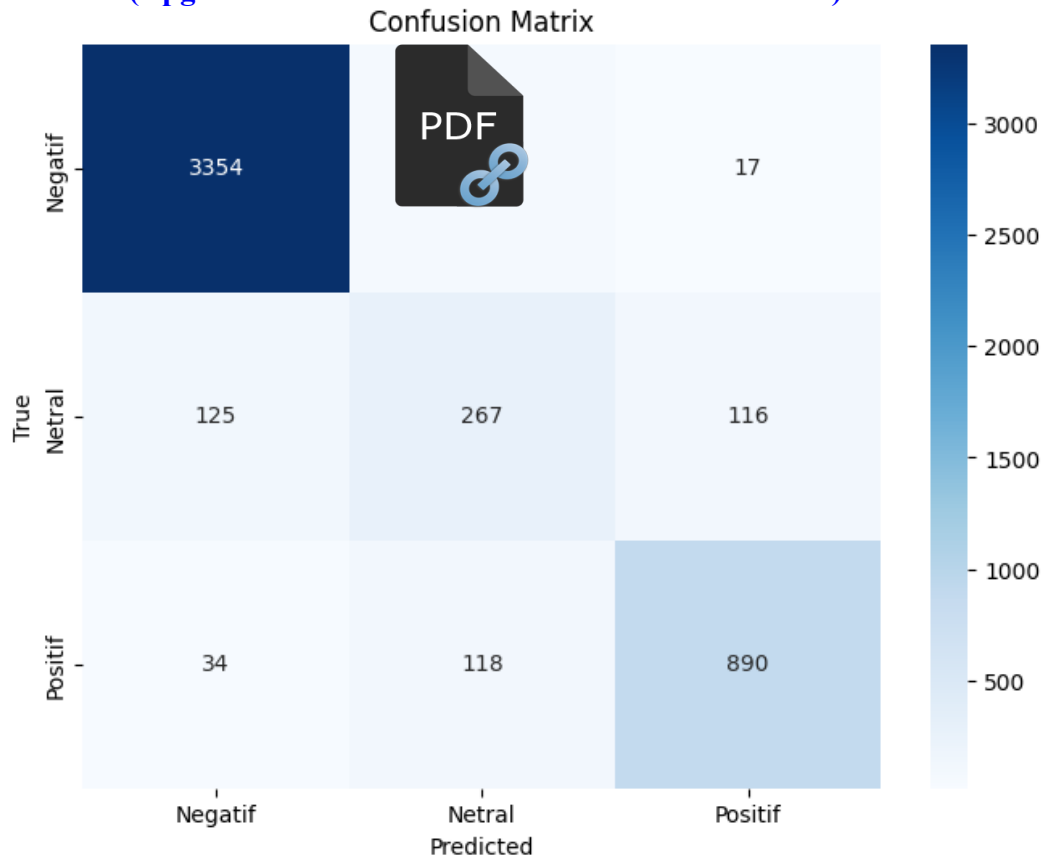
# Hitung confusion matrix
cm = confusion_matrix(test_labels, predictions)
print("Confusion Matrix:")
print(cm)
import matplotlib.pyplot as plt
import seaborn as sns

# Visualisasi dengan Seaborn
plt.figure(figsize=(8, 6))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=label_encoder.classes_,
            yticklabels=label_encoder.classes_)
plt.xlabel('Predicted')
plt.ylabel('True')
plt.title('Confusion Matrix')
plt.show()
```

Adapun hasil dari tampilan gambar dari *Confusion Matrix* dapat dilihat pada gambar berikut ini:

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)



Gambar 4.15 *Confusion Matrix*

Selanjutnya adalah *source code* untuk menampilkan ringkasan kinerja model untuk setiap kelas, seperti *accuracy*, *presisi*, *recall*, dan *F1-score*.

```
from sklearn.metrics import classification_report  
  
# Laporan klasifikasi  
report = classification_report(test_labels, predictions, target_names=label_encoder.classes_)  
print("Classification Report:")  
print(report)
```

Adapun hasil nya dapat dilihat pada gambar berikut ini:

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

Classification Report:

	precision	recall	f1-score	support
Negatif	0.97	0.96	0.96	3447
Netral	0.53	0.55	0.55	508
Positif	0.85	0.86	0.86	1042
accuracy			0.90	4997
macro avg	0.80	0.78	0.79	4997
weighted avg	0.90	0.90	0.90	4997

Gambar 4.16 Classification Report

7. Word Cloud

Word cloud adalah representasi visual dari kata-kata yang sering muncul dalam suatu teks, di mana ukuran setiap kata mencerminkan frekuensi kemunculannya. Semakin besar ukuran kata dalam visual, maka semakin umum kata tersebut dalam dokumen [56]. Berikut adalah *source code Word Cloud*:

```
wordcloud = WordCloud(  
    width=800,  
    height=400,  
    background_color='black',  
    colormap='inferno', # Tema warna modern 'inferno'  
    max_words=100,  
    contour_width=3,  
    contour_color='black'  
)  
.generate(all_text)  
  
# Menampilkan WordCloud  
plt.figure(figsize=(10, 5))  
plt.imshow(wordcloud, interpolation='bilinear')  
plt.axis('off')  
plt.title("WordCloud of User Reviews")  
plt.show()
```

Adapun hasil tampilan gambar dari *Word Cloud* dapat dilihat pada gambar berikut ini:

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

menunjukkan bahwa dari total 24.982 data ulasan, terdapat 17.228 ulasan dengan sentimen negatif, 4.998 ulasan dengan sentimen positif, dan 2.601 ulasan dengan sentimen netral. Setelah proses *labeling* selesai, data kemudian dibagi menjadi data *training* dan data *testing* dengan perbandingan 80:20. Dari pembagian ini, diperoleh 19.984 data untuk *training* dan 4998 data untuk *testing*.

Metode BERT kemudian diterapkan pada data tersebut untuk melakukan klasifikasi sentimen. Model yang dihasilkan memiliki akurasi sebesar 90%, dengan nilai presisi untuk kelas negatif mencapai 95% dan *recall* kelas negatif mencapai 97%. Untuk kelas positif, nilai presisi mencapai 87% dan *recall* 85%. Kelas netral menunjukkan performa yang sedikit lebih rendah, dengan presisi 58% dan *recall* 53%. Meskipun demikian, secara keseluruhan model memberikan hasil yang sangat baik, dengan *macro average F1-score* yang mencerminkan keseimbangan yang cukup baik dalam menangani seluruh kelas sentimen, dan *weighted average F1-score* yang menunjukkan model memberikan kinerja yang lebih baik pada kelas-kelas yang memiliki jumlah lebih besar, seperti negatif.

KESIMPULAN DAN SARAN



5.1 Kesimpulan

Kesimpulan dari penelitian yang telah dilakukan mengenai analisis sentimen pengguna game *online E-Football* di *Google Play Store* menggunakan *machine learning* berbasis BERT adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini diperoleh dengan memanfaatkan *Google Play Scraper* sebanyak 25.000 data ulasan. Setelah dilakukan *preprocessing*, data yang diambil merupakan ulasan dari *Google Play Store* selama periode tertentu.
2. Hasil *labeling* data menghasilkan tiga kelas sentimen, yaitu kelas positif, negatif, dan netral.
3. Proses klasifikasi dan penerapan algoritma BERT dilakukan dengan membagi data menjadi data *training* sebanyak 80% dan data *testing* sebanyak 20%. Hasil pengujian model menunjukkan akurasi sebesar 90%, yang mengindikasikan bahwa model berhasil mengklasifikasikan 90% dari total data dengan benar. Selain itu, model juga menunjukkan performa yang sangat baik pada kelas Negatif, dengan nilai presisi mencapai 95%, yang berarti 95% dari semua prediksi kelas Negatif benar-benar sesuai dengan kelas tersebut. *Recall* untuk kelas Negatif mencapai 97%, yang menunjukkan bahwa model berhasil menangkap 97% dari seluruh contoh yang sebenarnya merupakan kelas Negatif.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

5.2 Saran

Untuk meningkatkan dan mengembang pada penelitian selanjutnya, ada beberapa saran yang mungkin dilakukan pada peneliti selanjutnya agar mendapatkan hasil yang lebih maksimal yaitu

1. Penelitian selanjutnya disarankan untuk menggunakan *Google Colab Pro* dan koneksi internet yang memadai untuk melakukan penelitian ini. Dikarenakan jika hanya menggunakan *Google Colab Free* pada beberapa proses di bagian *preprocessing* dan *Training* data akan memerlukan waktu yang cukup lama untuk *loading data* pada beberapa proses tersebut.
2. Peneliti selanjutnya disarankan memasukan dan menambahkan metode klasifikasi yang lain agar dapat membandingkan hasil akurasi.
3. Penelitian selanjutnya disarankan untuk mempelajari lebih dalam lagi mengenai judul penelitian yang berhubungan ataupun mirip dengan judul penelitian ini dikarenakan pada beberapa literatur atau pokok pembahasan yang ingin didapatkan itu cukup sulit untuk dicari dan dipelajari.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

DAFTAR PUSTAKA

- [1] R. Adolph, “No Title,” pp. 1–23, 2016.
- [2] S. Pokhrel, “No Title,” *EAI*, vol. 15, no. 1, pp. 37–48, 2024.
- [3] R. D. Wahyuni and A. ..., “Penggunaan Metode Lexicon Untuk Analisis Sentimen pada Ulasan Aplikasi KAI Access di Google Play Store,” *J. Rekayasa Inf.*, vol. 11, no. 2, pp. 134–145, 2022.
- [4] E. Daryfayi, P. Daulay, and I. Asror, “Sentimen Analisis pada Ulasan Google Play Store Menggunakan Metode Naïve Bayes,” *e-Proceeding Eng.*, vol. 7, no. 2, pp. 8400–8410, 2020.
- [5] A. Nata and S. Suparmadi, “Analisis Sistem Pendukung Keputusan Dengan Model Klasifikasi Berbasis Machine Learning Dalam Penentuan Penerima Program Indonesia Pintar,” *J. Sci. Soc. Res.*, vol. 5, no. 3, p. 697, 2022, doi: 10.54314/jssr.v5i3.1041.
- [6] E. Retnoningsih and R. Pramudita, “Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python,” *Bina Insa. Ict J.*, vol. 7, no. 2, p. 156, 2020, doi: 10.51211/biict.v7i2.1422.
- [7] I. M. Faiza, G. Gunawan, and W. Andriani, “Tinjauan Pustaka Sistematis: Penerapan Metode Machine Learning untuk Deteksi Bencana Banjir,” *J. Minfo Polgan*, vol. 11, no. 2, pp. 59–63, 2022, doi: 10.33395/jmp.v11i2.11657.
- [8] S. Widaningsih, U. Suryakencana, A. Suheri, and U. Suryakencana, “Klasifikasi Jurnal Ilmu Komputer Berdasarkan Pembagian Web of,” vol. 2018, no. March, pp. 23–24, 2018.
- [9] H. Hashimi, A. Hafez, and H. Mathkour, “Selection criteria for text mining approaches,” *Comput. Human Behav.*, vol. 51, pp. 729–733, 2015, doi: 10.1016/j.chb.2014.10.062.
- [10] M. Afdal and L. R. Elita, “Penerapan Text Mining Pada Aplikasi Tokopedia Menggunakan Algoritma K-Nearest Neighbor,” *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 8, no. 1, p. 78, 2022, doi: 10.24014/rmsi.v8i1.16595.
- [11] F. S. Pamungkas and I. Kharisudin, “Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter,” *Pros. Semin. Nas. Mat.*, vol. 4, pp. 1–7, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/prisma/article/view/45038>
- [12] Y. Tajul Arifin AMIK BSI Bogor Jl Merdeka No and B. Jawa Barat, “Komparasi Fitur Seleksi Pada Algoritma Support Vector Machine Untuk Analisis Sentimen Review,” *Informatika*, vol. 3, no. September, pp. 191–199, 2016, [Online]. Available: www.tripadvisor.co.id
- [13] F. Septianingrum, J. H. Jaman, and U. Enri, “Analisis Sentimen Pada Isu Vaksin Covid-19 di Indonesia dengan Metode Naive Bayes Classifier,” *J. Media Inform. Budidarma*, vol. 5, no. 4, p. 1431, 2021, doi: 10.30865/mib.v5i4.3260.

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

- [14] K. Surbakti, "Pengaruh Game Online Terhadap Remaja," *J. Curere*, vol. 1, no. 1, p. 20, 2017, [Online]. Available: <http://www.portaluniversitasidonegara.ac.id:5388/ojsystem/index.php/CURE/RE/article/view/20/22>
- [15] A. A. SHELEMO, *No Title*, vol. 13, no. 1. 2023.
- [16] khabibur rohman, "Majalah Jurnal Perempuan dan Anak Pusat Studi Gender dan Anak (PSGA) IAIN Tulungagung," *Martabat J. Peremp. dan Anak*, vol. 2, pp. 1–19, 2018.
- [17] M. Haris, A. Suharso, E. H. Nurkifli, P. S. Informatika, U. S. Karawang, and T. Timur, "ANALISIS SENTIMEN PADA GAME EFOOTBALL DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA INDOBERT," vol. 8, no. 6, pp. 12108–12121, 2024.
- [18] U. Nini, S. Berlin, L. A. R. Laliyo, M. R. Katili, and M. K. Umar, "Game E-Football Sebagai Media Audiovisual Terhadap E-Football Game As Audiovisual Media on Learning Outcomes of," vol. 5, no. 2, pp. 162–174, 2023.
- [19] S. Kurniawan, "Analysis of Google Play Store Datasets Using the Exploratory Data Analysis (EDA) Method," no. December, 2022, doi: 10.13140/RG.2.2.14192.12808.
- [20] A. J. N. Kisma, C. R. A. Widiawati, and S. Suliswaningsih, "Analysis of applications in Playstore based on Rating and Type using Naive Bayes and Logistic Regression," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 10, no. 2, pp. 174–184, 2023, [Online]. Available: <http://jurnal.mdp.ac.id>
- [21] Indra, "Mari Kita Bahas Aplikasi Sejuta Umat! Google Play Store," *Idmetafora.Com*, p. 1, 2022, [Online]. Available: <https://idmetafora.com/news/read/898/Mari-Kita-Bahas-Aplikasi-Sejuta-Umat-Google-Play-Store.html>
- [22] M. K. Insan, U. Hayati, and O. Nurdiawan, "Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di," *J. Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 478–483, 2023.
- [23] "teknokrat.ac.id," p. 141.
- [24] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.
- [25] K. Anwar, "Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone Menggunakan Naive Bayes," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 4, pp. 148–155, 2022, doi: 10.30865/klik.v2i4.315.
- [26] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 9, pp. 4305–4313, 2022.
- [27] A. Gaizka, A. R. Dzikrillah, and E. Sinduningrum, "Analisis Sentimen Masyarakat Sebelum Dan Sesudah Terpilihnya Gibran Sebagai Cawapres Prabowo Menggunakan Naive Bayes," *KLIK Kaji. Ilm. Inform. dan*

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

- Komput.*, vol. 4, no. 6, pp. 2830–2841, 2024, doi: 10.30865/klik.v4i6.1876.
- [28] D. A. Alzahra, U. Enri, and Y. U. Maidah, “Analisis Sentimen Ulasan Pengguna Klik Indomare di Google Play Menggunakan Support Vector Machine,” *Innov. J. Soc. Sci. Technol.*, vol. 3, no. 4, pp. 2173–2185, 2023, [Online]. Available: <https://j-innovative.org/index.php/ijst/article/view/3715>
- [29] S. Wardani, “Analisis Sentimen Data Presiden Jokowi Dengan Preprocessing Normalisasi Dan Stemming Menggunakan Metode Naive Bayes Dan Svm,” *J. Din. Inform.*, vol. 5, no. November, pp. 1–13, 2015.
- [30] E. Susilowati, M. K. Sabariah, and A. A. Gozali, “Implementasi Metode Support Vector Machine untuk Melakukan Klasifikasi Kemacetan Lalu Lintas Pada Twitter,” *E-Proceeding Eng.*, vol. 2, no. 1, pp. 1478–1484, 2015.
- [31] D. Sartika and D. I. Sensuse, “Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian,” *Jatisi*, vol. 1, no. 2, pp. 151–161, 2017, [Online]. Available: <https://jurnal.mdp.ac.id/index.php/jatisi/article/view/78>
- [32] A. S. Fitriani, “Penerapan Data Mining Menggunakan Metode Klasifikasi Naïve Bayes untuk Memprediksi Partisipasi Pemilihan Gubernur,” *JTAM (Jurnal Teor. dan Apl. Mat.)*, vol. 3, no. 2, pp. 98–104, 2019.
- [33] A. Pardamean and H. F. Pardede, “Tuned bidirectional encoder representations from transformers for fake news detection,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 22, no. 3, pp. 1667–1671, 2021, doi: 10.11591/ijeecs.v22.i3.pp1667-1671.
- [34] rahayu deny danar dan alvi furwanti Alwie, A. B. Prasetyo, R. Andespa, P. N. Lhokseumawe, and K. Pengantar, *Tugas Akhir Tugas Akhir Bandeng*, vol. 2, no. 1. 2020.
- [35] N. Putu, V. D. Saraswati, N. Yudistira, and P. P. Adikara, “Analisis Sentimen terhadap Perundangan Siber pada Twitter menggunakan Algoritma Bidirectional Encoder Representations from Transformer (BERT),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 2, pp. 909–916, 2023, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12345>
- [36] A. Nayla, C. Setianingsih, and B. Dirgantoro, “Deteksi Hate Speech Pada Twitter,” *eProceeding Eng.*, vol. 10, no. 1, p. 256, 2023.
- [37] C. A. Putri, “Analisis Sentimen Review Film Berbahasa Inggris Dengan Pendekatan Bidirectional Encoder Representations from Transformers,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 6, no. 2, pp. 181–193, 2020, doi: 10.35957/jatisi.v6i2.206.
- [38] M. Munikar, S. Shakya, and A. Shrestha, “Fine-grained Sentiment Classification using BERT,” *Int. Conf. Artif. Intell. Transform. Bus. Soc. AITB 2019*, vol. 1, pp. 1–5, 2019, doi: 10.1109/AITB48515.2019.8947435.
- [39] J. U. S. Lazuardi and A. Juarna, “Analisis Sentimen Ulasan Pengguna

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

- Aplikasi Joox Pada Android Menggunakan Metode Bidirectional Encoder Representation From Transformer (Bert),” *J. Ilm. Inform. Komput.*, vol. 28, no. 3, pp. 251–260, 2023, doi: 10.23960/ik.2023.v28i3.10090.
- [40] R. Mas, R. W. Panca, K. ... and W. Yustanti², “Analisis Sentimen Customer Review Aplikasi Ruang Guru dengan Metode BERT (Bidirectional Encoder Representations from Transformers),” *Jeisbi*, vol. 02, no. 3, pp. 55–62, 2021.
- [41] Z. A. Sriyanti, D. S. Y. Kartika, and A. R. E. Najaf, “Implementasi Model Bert Pada Analisis Sentimen Pengguna Twitter Terhadap Aksi Boikot Produk Israel,” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, pp. 2335–2342, 2024, doi: 10.23960/jitet.v12i3.4743.
- [42] R. K. Septiani and W. Gata, “Eksplorasi Sentimen Pengguna Media Sosial Terhadap Layanan Seabank : Pendekatan Dengan Algoritma Bert,” pp. 886–897.
- [43] M. A. Muzani, M. I. A. Sukri, S. N. Fauziah, A. Fatkhurohman, and D. Ariatmanto, “Data Mining Untuk Klasifikasi Produk Menggunakan Algoritma K-Nearest Neighbor Pada Toko Online,” *Pros. SISFOTEK ISSN 2597-3584*, vol. 5, no. 1, pp. 141–145, 2021, [Online]. Available: <http://seminar.iaii.or.id/index.php/SISFOTEK/article/view/273>
- [44] H. C. Husada and A. S. Paramita, “Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM),” *Teknika*, vol. 10, no. 1, pp. 18–26, 2021, doi: 10.34148/teknika.v10i1.311.
- [45] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021.
- [46] S. Kumar, “Understanding Confusion Matrix,” *Towar. Data Sci.*, 2018, [Online]. Available: <https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>
- [47] G. Rininda, I. Hartami Santi, and S. Kirom, “Penerapan Svm Dalam Analisis Sentimen Pada Edlink Menggunakan Pengujian Confusion Matrix,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 5, pp. 3335–3342, 2024, doi: 10.36040/jati.v7i5.7420.
- [48] L. Ardiani, H. Sujaini, and T. Tursina, “Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak,” *J. Sist. dan Teknol. Inf.*, vol. 8, no. 2, p. 183, 2020, doi: 10.26418/justin.v8i2.36776.
- [49] H. S. Ginting, K. M. Lhaksmana, and D. T. Murdiansyah, “Klasifikasi Sentimen Terhadap Bakal Calon Gubernur Jawa Barat 2018 di Twitter,” *e-Proceeding Eng.*, vol. 5, no. 1, pp. 1793–1802, 2018.
- [50] F. Imanuel Tanesab, I. Sembiring, H. Dwi Purnomo, F. Teknologi Informasi, U. Kristen Satya Wacana, and S. -Jawa Tengah, “Implementasi Support Vector Machine (Svm) Untuk Sentimen Analisis Terhadap Kinerja

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)

LAMPIRAN

UNIVERSITAS BINA INSAN
Jalan Jendral
Kec. Lubuk Kumpang Kec. Lubuklinggau Selatan 1 Kota Lubuklinggau Provinsi Sumatera Selatan

ENDIDIKAN DWI TUNGGAL PALEMBANG
UNIVERSITAS BINA INSAN

PDF

**Formulir Pengajuan Judul Skripsi
Program Studi Informatika**


Nama : Muhammad Aryaraka Pradana
NIM : 2102020119
Alamat : Jl.Patimura No.05 RT.01 Kec.Lubuklinggau Barat 1 Kel.Sukajadi
No.Hp : 0895-6210-18738



Rumusan Masalah 1 : Bagaimana cara mengimplementasikan model BERT untuk menganalisis sentimen dari ulasan pengguna game online E-Football di Play Store?
Judul 1 : Implementasi BERT Untuk Analisis Sentimen Pengguna Game Online E-Football Pada Ulasan di Play Store berbasis Machine Learning.

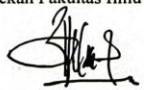
Rumusan Masalah 2 : Bagaimana performa algoritma BERT dalam mengklasifikasikan persepsi pengguna aplikasi Duolingo belajar bahasa menjadi kelas positif dan negatif?
Judul 2 : Implementasi BERT Untuk Analisis Sentimen Pengguna Aplikasi Duolingo Pada Ulasan di Play Store berbasis Machine Learning.

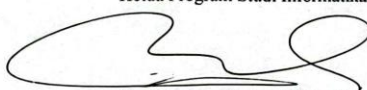
Rumusan Masalah 3 : Bagaimana menerapkan metode Algoritma BERT dalam sentiment analisis hasil web scrapping data komentar dari Google Play Store?
Judul 3 : Analisis Sentimen Menggunakan Algoritma BERT Terhadap Komentar Aplikasi Discord Di Play Store.

Diusulkan Judul Nomor : 1(satu) / 2(Dua) / 3(Tiga)*

Lubuklinggau, Oktober 2024
Mahasiswa yang mengusulkan,

(Muhammad Aryaraka Pradana)

Menyetujui Dosen Pembimbing,
Pembimbing 1 (Armanto, S.Kom., M.Kom) 
Pembimbing 2 (Fido Risky, S.Kom., M. Kom) 

Mengesahkan
Dekan Fakultas Ilmu Teknik 
(Dr.Rudi Kuriawan, S.T., M.Kom)

Mengetahui
Ketua Program Studi Informatika 
(Budi Santoso, S.Kom., M.Kom)

0733-4553932 (Rektorat Universitas Bina Insan)
0733-3280300 (Pascasarjana)
0812-1826-6228 (Marketing UNIVBI)
0852-3151-5800 (Admin UNIVBI)
Admin@univbinaisan.ac.id
univbinaisan.ac.id - pasca.univbinaisan.ac.id

Lampiran 1. Pengajuan Judul Skripsi

Protected by PDF Anti-Copy Free
(Upgrade to Pro Version to Remove the Watermark)



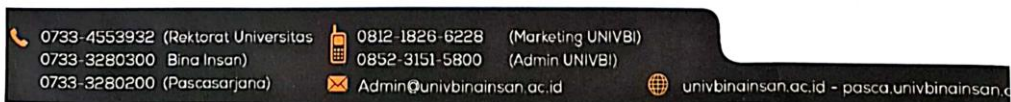
LEMBAR BIMBINGAN SKRIPSI

Nama : MUHAMMAD ARYARAKA PRADANA
 Nim : 2102020119
 Program Studi : Informatika
 Pembimbing 1 : Armanto, M.Kom
 Pembimbing 2 : Fido Rizki, S.Kom., M.Kom
 Judul : Implementasi BERT Untuk Analisis Sentimen Pengguna Game Online E-Football Pada Ulasan Di Play Store Berbasis Machine Learning

NO	TANGGAL	TOPIK	KOMENTAR PEMBIMBING	TANDA TANGAN PEMBIMBING	
				1	2
1	6/1/25	Penulisan	- Sesuaikan dengan format pada buku panduan		<i>Pr</i>
2	8/1/25	Halaman	- Sesuaikan penomoran - Sesuaikan nomor tabel dan gambar		<i>Pr</i>
3	10/1/25	ISI	- Lengkapi semua hasil dan pembahasannya		<i>Pr</i>
4	16/1/25	Acc	Lengkap		<i>Pr</i>

Lubuklinggau, Januari 2025
Ketua Program Studi Informatika

(Budi Susanto, M.Kom)



Lampiran 2. Lembar Bimbingan Pembimbing II

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)



LEMBAR Bimbingan SKRIPSI

Nama : MUHAMMAD ... PRADANA
Nim : 2102020119
Program Studi : Informatika
Pembimbing 1 : Armanto, M.Kom
Pembimbing 2 : Fido Rizki, S.Kom.,M.Kom
Judul : Implementasi BERT Untuk Analisis Sentimen Pengguna Game Online E-Football Pada Ulasan Di Play Store Berbasis Machine Learning

NO	TANGGAL	TOPIK	KOMENTAR PEMBIMBING	TANDA TANGAN PEMBIMBING	
					2
1	17/01/2025		no skripsi sudah selesai ACC Silahkan ulas kompre-		

Lubuklinggau, Januari 2025
Ketua Program Studi Informatika

(Budi Susanto, M.Kom)

0733-4553932 (Rektorat Universitas) 0812-1826-6228 (Marketing UNIVBI)
0733-3280300 Bina Insan 0852-3151-5800 (Admin UNIVBI)
0733-3280200 (Pascasarjana) Admin@univbinainsan.ac.id univbinainsan.ac.id - pasca.univbinainsan

Lampiran 3. Lembar Bimbingan Pembimbing I

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)



LEMBAR PERBAIKAN SEMINAR PROPOSAL SKRIPSI

Nama Mahasiswa : Muhammad Aryaraka Pradana
NIM : 2102020119
Jenjang Pendidikan : Strata I (S1)
Fakultas : Teknik
Program Studi : Informatika
Konsentrasi :
Judul : IMPLEMENTASI BERT UNTUK ANALISIS SENTIMEN
PENGGUNA GAME ONLINE E-FOOTBALL PADA ULASAN DI
PLAY STORE BERBASIS MACHINE LEARNING

No	Dosen Penguji	Komentar Perbaikan	Tanda Tangan Ujian	Tanda Tangan Revisi
1	Armando, M.Kom			
2	Fido Rizki, M.Kom			
3	Cinda Wulandari, M.Kom	Sesuaikan dengan catatan yg ada di proposal pada saat ujian berlangsung		 3/1/2023

Lubuklinggau,2024
Ketua Program Studi Informatika

(Budi Santoso, M.Kom)

0733-4553932 (Rektorat Universitas) 0812-1826-6228 (Marketing UNIVBI)
0733-3280300 (Bina Insan) 0852-9151-5800 (Admin UNIVBI)
0733-3280200 (Pascasarjana) Admin@univbinaisan.ac.id univbinaisan.ac.id - pasca.univbinaisan.ac.id

Dipindai dengan
CamScanner

Lampiran 4. Lembar Perbaikan Seminar Proposal Skripsi

Protected by PDF Anti-Copy Free

(Upgrade to Pro Version to Remove the Watermark)



The screenshot shows a Jupyter Notebook titled "E-Football_scrapper.ipynb". The code in the notebook is as follows:

```
[7]: from google_play_scraper import app
import pandas as pd
import numpy as np

from google_play_scraper import Sort, reviews

result, continuation_token = reviews(
    'jp.konami.pesem',
    lang='id',
    country='id',
    sort=Sort.MOST_RELEVANT,
    count=20000,
    #filter_score_with=None
)

[10]: df_busu = pd.DataFrame(np.array(result), columns=['review'])
df_busu = df_busu.join(pd.DataFrame(df_busu.pop('review').tolist()))
df_busu.head()
```

The output of the code is a DataFrame with the following columns: reviewId, userName, userImage, content, score, thumbUpCount, reviewCreatedVersion, at, replyContent, repliedAt, appVersion. The data is as follows:

	reviewId	userName	userImage	content	score	thumbUpCount	reviewCreatedVersion	at	replyContent	repliedAt	appVersion
0	b7c70d9-7294-a6f4-8a44-690360b47276	Pengguna Google	https://play.googleusercontent.com/EGemolZ...	tolong untuk konami buatlah fitur gabisa matik...	1	355	9.2.0	2024-12-16 08:01:27	None	None	9.2.0
1	ec4f3d7-10814acc-8a0f-8d466d1a90f	Pengguna Google	https://play.googleusercontent.com/EGemolZ...	Game macam apa ini sinyal bagus tapi pas main ...	5	140	9.2.0	2024-12-17 09:48:54	None	None	9.2.0
2	51113b-649c-46ba-adf5-0976d365aa03	Pengguna Google	https://play.googleusercontent.com/EGemolZ...	Melawan ai dan melawan player rasanya sangat b...	1	492	9.2.0	2024-12-14 00:19:45	None	None	9.2.0
3	fc0751ae-3d50-432a-a244-2aac0879ac7d	Pengguna Google	https://play.googleusercontent.com/EGemolZ...	Gameplay sudah baik. Tapi sejauh ini eFootball ...	4	92	9.2.0	2024-12-10 19:47:11	None	None	9.2.0
4	0835049-0466-4a20-8ed2-72e6c0f90ca	Pengguna Google	https://play.googleusercontent.com/EGemolZ...	Sangat disayangkan game ini memiliki grafik yg	2	51	9.2.0	2024-12-13 09:22:10	None	None	9.2.0

The notebook also shows a cell with the command `len(df_busu.index)` and the output `count the number of data we got`, indicating that the data has been successfully loaded.

Lampiran 5. Scraping Data