

BAB III

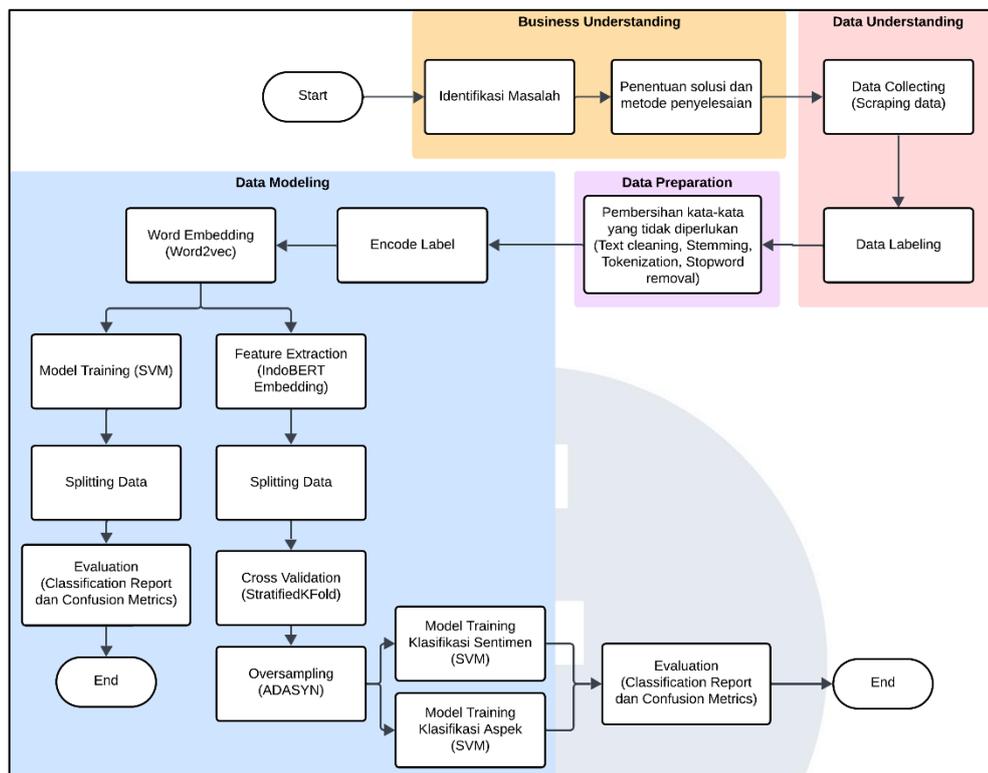
METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Objek penelitian pada penelitian ini berfokus untuk memberikan pemahaman lebih dalam tentang opini pengguna dengan analisis sentimen berbasis aspek yang dilakukan pada ulasan para pengguna di aplikasi Duolingo. Penelitian ini memanfaatkan library pada Python, yaitu *google_play_scraper* untuk melakukan pengumpulan data Google Play Store yang digunakan sebagai media pengambilan data ulasan pengguna aplikasi Duolingo. Kolom ulasan, *rating*, dan tanggal ulasan ditulis yang diberikan oleh pengguna pada aplikasi Duolingo di Google Play Store digunakan untuk mengetahui opini para pengguna serta aspek yang menjadi alasan pengguna membuat ulasan tersebut. Data ulasan yang diambil memiliki rentang waktu dari 1 Januari 2020 hingga 1 Januari 2024. Data ulasan tersebut akan diproses dengan membagi ulasan pengguna menjadi 2 kategori sentimen, yaitu positif dan negatif serta melakukan identifikasi terhadap aspek pada sentimen pengguna. Selain itu, ulasan akan dipisah berdasarkan ulasan tahun sebelum terdapat fitur premium yaitu tahun 2020-2021, tahun 2022 atau saat fitur premium diterapkan, dan tahun setelah dirilisnya fitur premium pada Duolingo yaitu pada tahun 2023-2024. Proses analisis sentimen berbasis aspek dilakukan dengan menggunakan algoritma klasifikasi *Support Vector Machine* (SVM) dan model *Bidirectional Encoder Representations from Transformers* (BERT) yaitu IndoBERT.

3.2 Metode Penelitian

3.2.1 Alur Penelitian



Gambar 3. 1 Diagram Alur Penelitian

Gambar 3.1 menunjukkan diagram alur penelitian atau *flowchart* yang menjelaskan tentang proses-proses yang dilakukan pada penelitian analisis sentimen berbasis aspek. Langkah pertama adalah mengidentifikasi masalah serta solusi dan metode yang akan dilakukan untuk menyelesaikan permasalahan tersebut. Selanjutnya, pengumpulan data ulasan dan *rating* aplikasi Duolingo di Google Play Store dilakukan. Setelah itu, penyortiran data dilakukan dengan rentang waktu 4 tahun. Data ulasan yang telah disortir akan diberikan *label* untuk sentimen positif dan negatif berdasarkan *rating*, serta penentuan aspek berdasarkan standar *core app quality* Android. Setelah semua ulasan telah diberikan label, hal yang dapat dilakukan adalah melakukan persiapan data, meliputi pembersihan data. Kemudian, data yang telah dilakukan *preparation* akan dilakukan pembangunan model data menggunakan algoritma SVM dan model IndoBERT. Pembangunan model akan dibagi menjadi 2 tahap, yaitu pemodelan tanpa menerapkan IndoBERT untuk *feature extraction* dan pemodelan yang menerapkan IndoBERT. Pemodelan pada data akan dibagi menjadi klasifikasi sentimen dan klasifikasi aspek. Setelah

pembangunan model, tahap yang dilakukan berikutnya adalah evaluasi model menggunakan *confusion matrix* dengan mempertimbangkan *accuracy*, *f1-score*, *recall*, dan *precision*. Hasil dari pengujian akurasi tersebut akan dijadikan hasil akhir terhadap penelitian terkait opini pengguna dan aspek-aspek yang telah diketahui.

3.2.2 Metode Data Mining

Terdapat tiga *framework* utama pada *data mining*. Berikut tabel perbandingan dari *framework* CRISP-DM, KDD, dan SEMMA:

Tabel 3.1 Perbandingan *Framework Data Mining* [59]

Indikator	CRISP-DM	KDD	SEMMA
Proses	<i>Business Understanding</i>	<i>Selection</i>	<i>Sample</i>
	<i>Data Understanding</i>	<i>Preprocessing</i>	<i>Explore</i>
	<i>Data Preparation</i>	<i>Transformation</i>	<i>Modify</i>
	<i>Data Modeling</i>	<i>Data Mining</i>	<i>Model</i>
	<i>Evaluation</i>	<i>Interpretation/Evaluation</i>	<i>Assess</i>
	<i>Deployment</i>	<i>Knowledge</i>	...
Keunggulan	Berfokus pada pemahaman bisnis untuk membantu mengidentifikasi tujuan dari analisis data	Berfokus pada pengetahuan terhadap data	Berfokus pada proses pemodelan
	Dapat digunakan di berbagai jenis permasalahan <i>data mining</i>	Memiliki proses yang lebih singkat	Memiliki proses yang sederhana dan mudah dipelajari
	Memiliki proses yang terstruktur dan komprehensif	Lebih mudah dalam memahami data	Dapat digunakan untuk penerapan model <i>data mining</i>
Kekurangan	Kurang cocok untuk proyek yang terlalu sederhana atau kecil	Membutuhkan mesin pembelajaran yang mendalam dan keahlian statistik	Struktur penelitian kurang jelas untuk pengelolaan proyek
	Kurang fleksibel untuk melakukan proyek	Sensitif dan kurang efektif terhadap data yang tidak lengkap	Tidak memiliki tahapan evaluasi
Aksesibilitas	Memerlukan pemahaman yang cukup terkait data	Memerlukan pemahaman terhadap komputasi tingkat tinggi dan statistik	Mudah diakses oleh peneliti

Indikator	CRISP-DM	KDD	SEMMA
Kompleksitas	Kompleksitas tinggi karena memiliki tahapan yang banyak	Kompleksitas tergantung berdasarkan teknik data mining yang digunakan	Lebih sederhana karena tidak memiliki tahapan yang terlalu banyak.

Berdasarkan tabel perbandingan *framework* tersebut, penelitian ini menggunakan *framework* CRISP-DM (*Cross-Industry Standard Process for Data Mining*) dalam analisis sentimen berbasis aspek, karena berfokus terhadap pemahaman bisnis pada awal proses, sehingga tujuan dari analisis data dapat diidentifikasi, berbagai permasalahan data mining dapat dipecahkan, sehingga memungkinkan untuk mendapatkan hasil yang optimal menggunakan *framework* ini. Selanjutnya, *framework* CRISP-DM memiliki tahapan yang jelas dari awal hingga akhir dalam proses data mining sehingga aspek-aspek yang penting pada proses analisis sentimen tidak terabaikan.

Pada penelitian ini, 5 tahapan pada CRISP-DM (*Cross-Industry Standard Process for Data Mining*) akan diterapkan sebagai berikut:

1. *Business Understanding*

Tahap *business understanding* bertujuan untuk mengetahui tujuan dilakukannya penelitian dengan manfaat yang akan didapatkan setelah menyelesaikan permasalahan tersebut. Pada tahap ini, penelitian dilakukan untuk mengetahui opini pengguna terhadap aplikasi Duolingo sebagai aplikasi belajar bahasa asing dari tahun 2020 hingga 2024 serta aspek berdasarkan *core app quality* milik *Android*. Selain itu, hasil dari analisis sentimen berbasis aspek dapat dijadikan sebagai pandangan bagi calon pengguna aplikasi Duolingo untuk menjadikan aplikasi Duolingo sebagai media belajar bahasa asing.

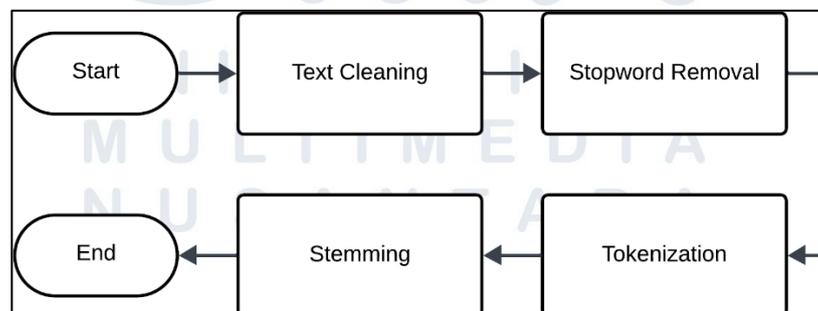
2. *Data Understanding*



Gambar 3. 2 Flowchart Tahap Data Understanding

Gambar 3.2 merupakan tahap *data understanding* yang berisi proses pemahaman data yang akan digunakan dalam penelitian. Langkah yang dilakukan pertama kali dalam tahap ini adalah melakukan pengumpulan data menggunakan *library google-play-scraper* dalam Google Colab. Data diambil secara berkala yaitu 500 per request hingga terambil 30.000 data ulasan dan disortir berdasarkan ulasan dengan relevansi tertinggi menggunakan parameter `Sort.MOST_RELEVANT`, kemudian dilakukan eksplorasi dengan menghapus kolom yang tidak digunakan, seperti nama pengguna, versi aplikasi yang dimiliki pengguna, hingga balasan pengguna lain terhadap ulasan tersebut, serta mengubah nama kolom dan memeriksa kualitas dari data ulasan yang diambil. Kolom yang digunakan adalah kolom berisi tanggal ulasan diunggah, ulasan para pengguna, serta *rating* yang diberikan terhadap aplikasi Duolingo. Kolom yang telah dipilih sebagai data yang akan dilakukan penelitian, selanjutnya dilakukan sortir data ulasan periode 1 Januari 2020 hingga 1 Januari 2024. Selain itu, pada tahap *data understanding* dilakukan pelabelan sentimen serta aspek terhadap data yang melibatkan kolom *rating* dan ulasan. Penentuan aspek terhadap ulasan diambil dari *core app quality Android*, yaitu fungsionalitas, kinerja & stabilitas, kegunaan & aksesibilitas, desain visual & interaksi pengguna, serta keamanan & privasi. Setelah kolom aspek dan sentimen ditambahkan, data tersebut kemudian diekspor ke dalam format CSV berdasarkan kategori tahun.

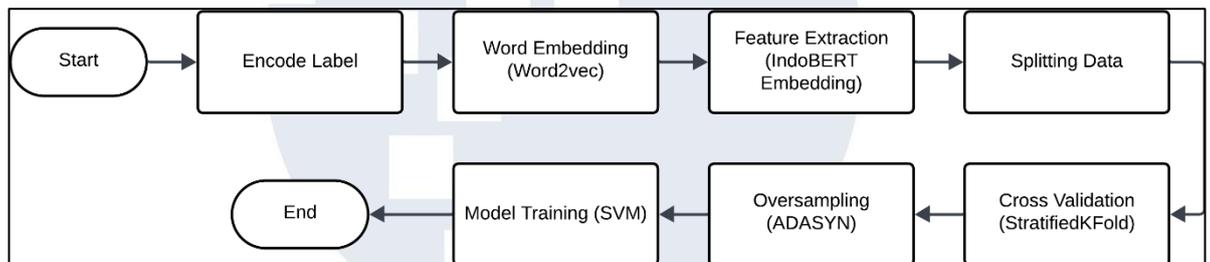
3. *Data Preparation*



Gambar 3. 3 Flowchart Tahap Data Preparation

Gambar 3.3 merupakan tahap *data preparation* yang bertujuan untuk mempersiapkan data sebelum masuk ke tahapan lebih lanjut. Pada penelitian ini, data yang telah dikumpulkan akan dilakukan pembersihan data atau *data cleaning* menggunakan Python pada Google Colab. *Data cleaning* yang dilakukan, meliputi mengubah karakter huruf, menghapus tanda baca dan spasi kosong, pemisahan kata dasar dengan kata awalan dan akhiran, menghilangkan kata-kata yang tidak penting, dan mengubah kata menjadi bentuk dasar (*stemming*). Kemudian, data yang telah selesai dilakukan proses *preparation* dan *pre-processing*, akan dijadikan ke dalam format CSV agar lebih efisien untuk dilakukan proses lanjutan.

4. *Data Modeling*

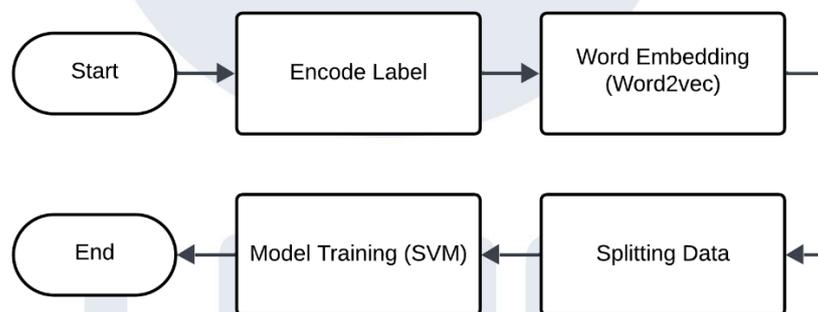


Gambar 3. 4 Flowchart Tahap Data Modeling

Gambar 3.4 merupakan alur dari tahap *data modeling* yang bertujuan untuk membangun model hingga model terbentuk pada *data mining* yang telah ditentukan. Pada tahap ini, hal yang dilakukan pertama kali adalah melakukan *encoding* terhadap label kategori atau mengubah tiap label ke dalam bentuk numerik agar data dapat diproses oleh model. Selanjutnya, dilakukan *word embedding* dengan Word2vec menggunakan *library gensim* serta penggunaan model *Skip-Gram* untuk memaksimalkan probabilitas munculnya kata-kata dengan frekuensi rendah dari kata target yang telah ditokenisasi. Kemudian, dilakukan BERT embeddings yang menggunakan model IndoBERT yang didapat dari *library huggingface* dengan *AutoTokenizer* untuk mengubah teks menjadi token yang dipahami oleh BERT dan *AutoModel* untuk menghasilkan *embeddings*.

Langkah yang dilakukan selanjutnya adalah melakukan pembagian data *train*, *validation*, dan *data test* dengan proporsi 70% untuk data latih, 15%

data validasi, dan 15% data uji. Untuk mengatasi ketidakseimbangan data, *Stratified K-Fold Cross-Validation* digunakan untuk membagi dataset dengan data yang tidak seimbang ke dalam beberapa lipatan yang telah ditetapkan dengan tetap mempertahankan distribusi data asli. Selain menggunakan *Stratified K-Fold Cross-Validation*, teknik *oversampling* ADASYN juga diterapkan pada dataset untuk mengatasi ketidakseimbangan terhadap data dengan mempertimbangkan neighbor terdekat dalam pembuatan sampel baru. Langkah terakhir pada tahap ini adalah melakukan pelatihan data pada model, melakukan pelatihan terhadap data dilakukan menggunakan SVM untuk klasifikasi sentimen dan aspek pada saat *initial training* atau pelatihan awal, *training* dengan *stratified k-fold*, dan *training* akhir setelah penerapan *oversampling* ADASYN pada data. pada tahap ini juga dilakukan perbandingan dengan hanya melakukan training model dengan SVM tanpa menerapkan *feature extraction* menggunakan IndoBERT seperti pada gambar 3.5



Gambar 3. 5 Tahap modeling tanpa IndoBERT

5. Evaluation

Setelah proses pemodelan terhadap data telah selesai dilakukan, tahap selanjutnya adalah melakukan evaluasi pada hasil *training model* dengan melakukan pengujian akurasi. Pada penelitian ini, pengujian akurasi menggunakan *confusion matrix* dengan mempertimbangkan *accuracy*, *f1-score*, *recall*, dan *precision*. Hasil dari akurasi yang telah dilakukan pada pemodelan terhadap data ulasan akan menjadi pemahaman yang lebih mendalam terkait opini serta aspek para pengguna pada periode penggunaan aplikasi tahun 2020 hingga 2024.

3.2.3 Metode Pengolahan Data

Tabel 3.2 Perbandingan Algoritma Pengolahan Data

Indikator	<i>Support Vector Machine (SVM)</i>	<i>Bidirectional Encoder Representations from Transformers (BERT)</i>
Keunggulan	<ol style="list-style-type: none"> 1. Memiliki kinerja yang baik dan menghasilkan akurasi tinggi 2. Dapat mengatasi masalah <i>overfitting</i> 3. Fleksibel terhadap model yang digunakan 	<ol style="list-style-type: none"> 1. Memahami konteks kata secara mendalam 2. Performa yang tinggi 3. Menggunakan representasi kontekstual
Kekurangan	<ol style="list-style-type: none"> 1. <i>Noise</i> dapat mempengaruhi kinerja 2. Memerlukan dataset yang terstruktur 3. Kurang efektif untuk dataset yang tidak seimbang 	<ol style="list-style-type: none"> 1. Memerlukan dataset yang besar dan komputasi tinggi 2. Sangat kompleks 3. Membutuhkan GPU untuk inferensi cepat
Aspek	<i>Kernel, Hyperparameter, Decision Boundary</i>	<i>Transformer-based deep learning</i>
Fleksibilitas	Fleksibel dalam pemilihan kernel yang sesuai	Sangat fleksibel dalam NLP
Kompleksitas	Kompleksitas tingkat tinggi	Kompleksitas tingkat tinggi

Pada penelitian ini, algoritma *Support Vector Machine (SVM)* dan model *Bidirectional Encoder Representations from Transformers (BERT)* akan dikombinasikan dalam analisis sentimen terhadap ulasan aplikasi Duolingo di Google Play Store. Berdasarkan tabel 3.2 yang menjelaskan perbandingan kinerja algoritma dan model, algoritma *Support Vector Machine (SVM)* memiliki keunggulan, yaitu dapat menghasilkan akurasi yang tinggi dalam mengklasifikasikan kelas yang berbeda dengan kinerja yang baik terhadap penelitian. Selain itu, algoritma SVM dapat mengatasi permasalahan *overfitting* yang biasa terjadi pada penelitian dengan memaksimalkan margin antar kelas, dan dalam melakukan pemodelan antara fitur dan label, algoritma SVM cenderung sangat fleksibel sehingga memungkinkan SVM untuk menangani korelasi yang kompleks. Sementara itu, keunggulan yang dimiliki oleh model *Bidirectional Encoder Representations from Transformers (BERT)* adalah dapat memahami konteks secara mendalam karena melakukan analisis teks dalam dua arah. Kemudian, model BERT dapat memahami makna kata yang berbeda dalam konteks yang berbeda, sehingga pemrosesan teks pada analisis sentiment menjadi lebih akurat.

Selain keunggulan, tentunya kedua algoritma tersebut memiliki kelemahannya masing-masing. *Support Vector Machine (SVM)* memiliki sensitifitas terhadap noise yang cukup tinggi karena akan menghasilkan *overfitting* pada *data training* jika terdapat *noise* yang signifikan pada data, hasil yang didapatkan tidak optimal ketika dataset memiliki kompleksitas yang tinggi atau merupakan data yang tidak terstruktur, karena SVM tidak dapat menemukan margin yang memisahkan kelas. Selain itu, jika satu kelas memiliki jumlah sampel yang lebih banyak, maka memungkinkan SVM mengoptimalkan kinerja terhadap kelas dengan jumlah sampel yang lebih banyak. Lain halnya dengan SVM, kelemahan dari *Bidirectional Encoder Representations from Transformers (BERT)* yaitu memerlukan dataset yang besar agar BERT dapat memahami diimplementasikan secara efektif. Selain itu, BERT memiliki banyak lapisan dan parameter yang membuat proses perhitungan dan analisis jauh lebih kompleks dibandingkan dengan SVM, serta membutuhkan GPU agar proses pelatihan bisa menjadi cepat. Oleh karena itu, untuk mengoptimalkan hasil akhir pada penelitian analisis sentimen berbasis aspek terhadap ulasan aplikasi Duolingo di Google Play Store, algoritma SVM dan model BERT akan digunakan untuk memanfaatkan keunggulan yang dimiliki oleh algoritma dan model tersebut. Selain itu, dengan menggunakan algoritma SVM sebagai klasifikasi sentimen dan model BERT untuk representasi teks, risiko terjadinya *overfitting* akan berkurang sehingga mendapatkan model yang lebih stabil. Kemudian, penggunaan algoritma dan model ini akan menghasilkan akurasi yang lebih akurat dibandingkan menggunakan algoritma tunggal. Dengan memanfaatkan keunggulan dan memperhatikan kelemahan dari algoritma dan model tersebut, hasil akhir yang diharapkan adalah dapat mengetahui opini dan aspek terhadap ulasan pengguna di aplikasi Duolingo dengan akurasi yang tinggi.

3.3 Teknik Pengumpulan Data

3.3.1 Populasi dan Sampel

Populasi pada penelitian akan diambil dari Google Play Store pada aplikasi Duolingo. Populasi yang diambil merupakan berbagai ulasan yang

telah ditulis oleh pengguna pada aplikasi Duolingo di Google Play Store. Sampel diambil dengan menggunakan teknik sampling yang *non-probability sampling*, yaitu *purposive sampling*. *Purposive sampling* adalah teknik yang bertujuan untuk melakukan pemilihan sampel secara sengaja berdasarkan tujuan dari penelitian yang dilakukan sehingga akan mendapatkan data yang lebih akurat dan menghasilkan hasil akhir dari penelitian yang lebih optimal [60]. Sampel yang diambil untuk dianalisis pada penelitian, menggunakan *library* yang terdapat pada Python, yaitu *google-play-scraper*. Data sampel diambil dengan jumlah 30,000 ulasan dan menggunakan filter `Sort.Most_Relevant`, sehingga ulasan yang dihasilkan hanya ulasan pengguna yang relevan dan akan kembali disortir berdasarkan kategori tahun yang telah ditentukan, yaitu 2020-2021, 2022, dan 2023-2024.

3.3.2 Periode Pengambilan Data

Pada penelitian ini, data ulasan aplikasi Duolingo di Google Play Store diambil pada periode 1 Januari 2020 hingga 1 Januari 2024. Pemilihan rentang waktu tersebut karena terjadi perubahan layanan aplikasi dari gratis menjadi aplikasi yang berlangganan, sehingga perubahan sentimen pengguna terhadap pembaruan fitur serta layanan aplikasi Duolingo dapat terlihat secara signifikan, yaitu tahun sebelum dirilisnya fitur premium, tahun saat fitur premium mulai diterapkan, dan tahun setelah fitur premium dirilis. Oleh karena itu, data ulasan yang diperoleh adalah selama 4 tahun. Data ulasan tersebut nantinya akan dilakukan perbandingan perkembangan sentimen beserta dengan aspek pengguna aplikasi Duolingo. Total ulasan yang diambil dari Google Play Store adalah 30,000 ulasan yang telah disortir untuk ulasan paling relevan. Setelah itu, 30,000 data ulasan akan di sortir kembali sesuai dengan periode waktu 1 Januari 2020 hingga 1 Januari 2024 yang dibagi ke dalam 3 kategori waktu.

3.4 Teknik Analisis Data

3.4.1 Variabel Penelitian

Variabel terikat atau variabel dependen adalah variabel yang terpengaruh atau bergantung kepada variabel independen pada penelitian [61]. Pada penelitian ini, yang menjadi variabel terikat adalah hasil dari label sentimen positif dan negatif serta aspek pada sentimen pengguna.

Variabel bebas atau independen adalah variabel yang mempengaruhi perubahan variabel terikat [61]. Pada penelitian ini, yang menjadi variabel bebas adalah data ulasan yang ditulis oleh pengguna tentang aplikasi Duolingo pada Google Play Store, yaitu variabel *content* atau *review text*.

3.5 Teknik Pengujian atau Validasi Sistem

3.5.1 Tools Analisis Data

Tabel 3.3 Perbandingan *Tools* Analisis Data [62][63]

Indikator	Google Colaboratory	Jupyter Notebook	R Studio
Bahasa Pemrograman	Python dan mendukung bahasa lain dengan setup.	Julia, Python, R	R
Keunggulan	<ol style="list-style-type: none"> 1. Gratis dan menyediakan akses ke GPU dan TPU 2. Berbasis <i>cloud</i> 3. Mendukung integrasi dengan Google Drive 4. Dapat disimpan ke GitHub dengan sekali klik. 	<ol style="list-style-type: none"> 1. Dapat digunakan di berbagai bidang dan cukup populer. 2. Cepat dalam hal komputasi performa. 3. Visualisasi yang interaktif 4. Memiliki dokumentasi yang baik untuk teks, kode, dan visualisasi. 	<ol style="list-style-type: none"> 1. Berfokus untuk analisis data. 2. Memiliki sejumlah besar paket dan pustaka untuk analisis statistik. 3. Baik pada debugger untuk bahasa R
Kekurangan	<ol style="list-style-type: none"> 1. Membutuhkan koneksi internet yang baik. 2. Keterbatasan runtime. 	<ol style="list-style-type: none"> 3. Beberapa tugas memerlukan lebih banyak kode. 	<ol style="list-style-type: none"> 1. Kurang umum digunakan pada analisis selain statistik. 2. Kurang fleksibel. 3. Sulit dipelajari oleh pengguna.
Fleksibilitas	Fleksibel dan bisa dikonfigurasi untuk bahasa lain	Fleksibel untuk berbagai bahasa pemrograman.	Hanya mengoptimalkan untuk bahasa R
Kompleksitas	Kompleksitas tingkat rendah.	Kompleksitas tingkat rendah.	Kompleksitas tingkat tinggi.

Berdasarkan tabel perbandingan 3.3, penelitian ini menggunakan *tools* Google Colaboratory dengan bahasa pemrograman yang dipakai adalah Python. Google Colaboratory dipilih sebagai *tools* dibandingkan dengan Jupyter Notebook karena mudah digunakan dan dipelajari untuk pemula, serta menyediakan akses GPU dan TPU secara gratis yang membuat pemrosesan lebih cepat dan efisien dalam menjalankan model, serta berbasis *cloud* dan tidak memerlukan instalasi tambahan. Selain itu, pengguna dapat dengan mudah menggunakan Google Colaboratory karena memiliki kompleksitas yang rendah dibandingkan dengan R Studio.

3.5.2 *Accuracy, Precision, Recall, dan F-1 Score*

Teknik analisis data yang digunakan pada penelitian ini adalah alat pengukuran nilai berdasarkan hasil evaluasi *confusion matrix*, yaitu *Accuracy, Precision, Recall, dan F-1 Score*. Cara kerja *accuracy* adalah dengan menjumlahkan total data dalam dataset, setelahnya akan dibagi dengan jumlah data dengan nilai positif dan diprediksi sebagai positif (TP) serta jumlah data dengan nilai negatif dan diprediksi sebagai negatif (TN) pada data sentimen positif dan negatif. *Precision* dapat dibagi menjadi dua kategori, yaitu *precision positive* dan *precision negatif*. *Precision positive* dihitung dengan membagi jumlah data dengan nilai positif dan diprediksi sebagai positif (TP) dengan jumlah data sentimen yang masuk pada kategori positif (TP+FP). Sementara itu, *precision negative* dihitung dengan membagi jumlah data dengan nilai negatif dan diprediksi sebagai negatif (TN) dengan jumlah data sentimen yang masuk pada kategori negatif (TN+FN). Sama seperti *Precision, Recall* akan dihitung berdasarkan dua kategori sentimen dan akan dibagi sesuai dengan jumlah data yang diprediksi positif (TP+FN), serta jumlah data yang masuk ke dalam kategori negatif (FP+TN). *F-1 Score* dihitung berdasarkan hasil nilai *precision* dan *recall* yang bertujuan untuk mengukur keseimbangan yang dimiliki antara *precision* dan *recall* dan dapat memberikan pengukuran kinerja yang baik pada model secara keseluruhan [41], [43].