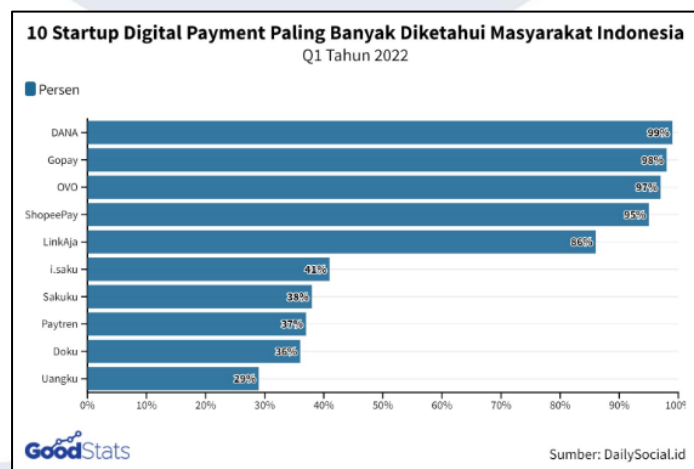


BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Gambaran umum objek penelitian ini berfokus pada tingkat kualitas aplikasi yang diberikan oleh perusahaan aplikasi uang elektronik (*e-wallet*) yaitu OVO, Gopay dan DANA di Indonesia pada tahun 2023 sampai Februari 2024. Data yang digunakan dalam penelitian analisis sentimen ini diambil melalui aplikasi Google Play Store yang terdapat ulasan-ulasan pengguna setiap aplikasi uang elektronik. Ulasan-ulasan yang berisikan pendapat pengguna terhadap ketiga aplikasi tersebut akan dianalisis untuk mengetahui tingkat pelayanan dari masing-masing aplikasi uang elektronik tersebut. Penelitian ini akan membandingkan dua algoritma yang digunakan dalam pembuatan model analisis sentimen yaitu *Support Vector Machine (SVM)* dan *Random Forest*.



Gambar 3. 1 Grafik Popularitas E-Wallet 2022

Gambar 3.1 merupakan grafik popularitas uang elektronik di Indonesia, dihasilkan oleh riset yang dilakukan oleh *Daily Social* [10]. Grafik tersebut menunjukkan bahwa ketiga algoritma yang dipilih dalam penelitian ini yaitu berdasarkan 3 peringkat popularitas aplikasi tertinggi yaitu DANA, Gopay dan OVO. Fokus dari penelitian ini yaitu untuk membandingkan hasil dari kedua algoritma yang digunakan sehingga dapat menemukan algoritma terbaik untuk

digunakan dalam analisis sentimen aplikasi uang elektronik. Algoritma *Support Vector Machine* (SVM) dan *Random Forest* digunakan dalam penelitian ini karena telah digunakan dalam penelitian sebelumnya dan terbukti dapat menghasilkan akurasi yang tinggi. Selain dari itu, penelitian ini juga diharapkan dapat memberikan informasi mengenai sisi positif dan negatif yang dimiliki oleh ketiga aplikasi tersebut, sehingga dapat memberikan evaluasi.

3.2 Data Collection

Pengumpulan data dilakukan dengan cara mengambil ulasan pengguna aplikasi OVO, Gopay dan DANA di Google Play Store. Metode pengambilan data yang digunakan yaitu *scraping* data dengan menggunakan library *google-play-scraper* pada program Python. Periode data yang diambil dan digunakan untuk penelitian ini yaitu mulai dari Januari 2023 sampai Februari 2024. Hasil dari *scraping* data tersebut kemudian di ubah ke dalam format CSV agar dapat di analisis melalui Google Colab menggunakan bahasa pemrograman Python. Adapun struktur data yang digunakan dalam penelitian ini terdapat pada Tabel 3.1.

Tabel 3. 1 Tabel Variabel

Variabel	Keterangan
<i>Content</i>	Variabel yang memuat ulasan pengguna terkait dengan aplikasi uang elektronik
Sentiment	Variabel yang memuat hasil kategorisasi ulasan (<i>Positive</i> atau <i>Negative</i>)

Berdasarkan struktur data pada Tabel 3.1, variabel yang digunakan dalam penelitian ini terbagi menjadi dua variabel yaitu sebagai berikut.

3.2.1 Variabel Independen

Variabel independen merupakan variabel bebas yang dapat mempengaruhi perubahan terhadap variabel dependen. Berdasarkan dataset yang digunakan dalam penelitian yang dilakukan akan menggunakan variabel *content* sebagai variabel independen.

3.2.2 Variabel Dependen

Variabel dependen merupakan variabel yang dipengaruhi oleh variabel bebas (variabel independen). Berdasarkan dataset yang digunakan dalam penelitian yang dilakukan akan menggunakan variabel *sentiment* sebagai variabel dependen.

3.3 Data Labeling

Langkah yang dilakukan setelah data didapatkan yaitu melakukan pelabelan terhadap data yang akan di analisis. Data akan di labelkan menjadi dua label yaitu positif dan negatif berdasarkan ulasan yang diberikan oleh pengguna. Proses data labeling dilakukan dengan menggunakan *library transformers* pada Python. *Library transformers* merupakan model *deep learning* yang sudah di latih sebelumnya agar dapat mengidentifikasi kalimat positif dan negatif.

3.4 Metode Penelitian

3.4.1 Metode Penyelesaian

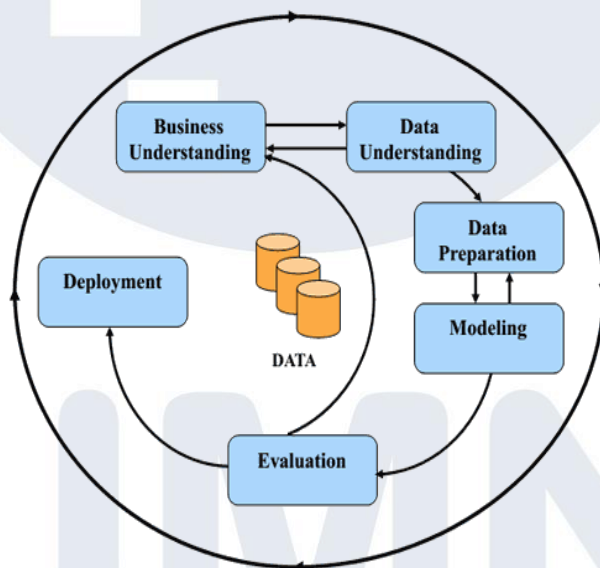
Penelitian ini akan menggunakan metode *machine learning* dengan dua algoritma yang akan dikomparasikan untuk menentukan aplikasi perusahaan uang elektronik (e-wallet) yang paling bagus diantara tiga merek yang telah disebutkan diatas. Kedua algoritma tersebut adalah *Support Vector Machine* (SVM) dan *Random Forest* karena sudah terbukti dapat digunakan dan cocok untuk analisis sentimen dalam penelitian terdahulu yang telah tertulis dalam BAB I dan BAB II. Kedua algoritma tersebut nantinya akan dikomparasikan dan digunakan untuk mencari aplikasi yang terbaik dari ketiga perusahaan aplikasi uang elektronik (*e-money*) terbesar.

3.4.2 Alur Penelitian

Tabel 3. 2 Tabel Perbandingan Alur Penelitian

CRISP-DM	KDD	SEMMA
<i>Business Understanding</i>	<i>Selection</i>	<i>Sample</i>
<i>Data Understanding</i>	<i>Pre Processing</i>	<i>Explore</i>
<i>Data Preparation</i>	<i>Transformation</i>	<i>Modify</i>
<i>Modeling</i>	<i>Data Mining</i>	<i>Model</i>
<i>Evaluation</i>	<i>Interpretation/Evaluation</i>	<i>Assessment</i>
<i>Deployment</i>	-	-

Tabel 3.2 merupakan tabel perbandingan antara tiga framework yaitu *Cross-Industry Standard Process for Data Mining* (CRISP-DM), KDD dan SEMMA.



Gambar 3. 2 Flowchart CRISP-DM [33]

Alur penelitian yang digunakan pada penelitian ini yaitu *Cross-Industry Standard Process for Data Mining* (CRISP-DM) untuk mengukur tingkat kualitas aplikasi perusahaan keuangan elektronik (*e-wallet*). CRISP-DM memiliki 6 siklus dalam prosesnya. Alasan penelitian ini menggunakan metode CRISP-DM karena sudah populer dan terbukti berhasil dalam penelitian analisis sentimen. Terdapat 6 proses yang akan diimplementasikan kedalam penelitian ini, yaitu sebagai berikut [33]:

1. *Business Understanding*

Penggunaan uang elektronik (e-wallet) yang meningkat menyebabkan munculnya banyak merek perusahaan aplikasi uang elektronik (e-wallet) di Indonesia. Penelitian ini bertujuan untuk membandingkan ketiga perusahaan aplikasi uang elektronik yaitu OVO, Gopay dan DANA sehingga dapat menemukan kualitas aplikasi dari masing-masing uang elektronik tersebut. Analisis sentimen dilakukan dengan harapan agar masyarakat Indonesia dapat memilih untuk menggunakan aplikasi uang elektronik dengan masing-masing kekurangan dan kelebihan. Masing-masing perusahaan uang elektronik diharapkan juga dapat mengambil kesimpulan dari penelitian ini mengenai kelebihan dan kekurangan yang didapatkan berdasarkan pendapat pengguna agar dapat meningkatkan kualitas aplikasi perusahaan.

2. *Data Understanding*

Sumber data yang digunakan untuk melakukan penelitian ini, yaitu menggunakan metode scraping data melalui platform Google Play Store dari masing-masing aplikasi OVO, Gopay dan DANA. Sumber data diambil dari Google Play Store dengan tujuan untuk mendapatkan perbandingan terhadap masing-masing kualitas dari ketiga aplikasi uang elektronik tersebut. Google Play Store memiliki ulasan yang dituliskan oleh pengguna masing-masing aplikasi uang elektronik tersebut, sehingga data cocok digunakan untuk analisis sentimen berdasarkan pendapat pengguna.

3. *Data Preparation*

Persiapan data yang diambil menggunakan text mining akan melalui proses *text preprocessing* yang terdapat lima tahap yaitu *remove duplicates*, *transform case*, *tokenization*, *normalization* dan *filter stopword*.

- *Remove duplicates* : Proses penghapusan data yang redundan atau sama.
- *Transform case* : Proses perubahan teks konten dari huruf besar menjadi huruf kecil, penghapusan simbol-simbol dan emoji.
- *Tokenization* : Proses pemisahan teks dengan cara memotong setiap kata menjadi bagian yang lebih kecil.
- *Normalization* : Proses penyesuaian dalam pengejaan kata dan singkatan yang ada di dalam kata tersebut.
- *Filter Stopword* : Penghapusan kata ganti, kata sambung serta tanda baca lainnya seperti titik dan koma.

4. *Modeling*

Pembuatan model akan dibuat menggunakan algoritma *Support Vector Machine (SVM)* dan *Random Forest*. Kedua algoritma tersebut akan dibentuk *data training* dan *data testing*. Algoritma *Support Vector Machine (SVM)* dan *Random Forest* akan menggunakan rasio 80:20 karena sudah terbukti menghasilkan akurasi yang tinggi dalam salah satu penelitian terdahulu. Kedua model tersebut akan dilakukan komparasi untuk mengetahui model yang mendapatkan hasil dengan akurasi lebih tinggi dan paling cocok untuk mengurutkan tingkat kualitas aplikasi perusahaan uang elektronik (*e-wallet*).

5. *Evaluation*

Pada evaluasi model, akurasi yang dihasilkan dari setiap model yang dibentuk akan diperhatikan sehingga dapat mengetahui model yang paling cocok. Metode evaluasi yang digunakan adalah *confusion matrix* untuk melihat perhitungan accuracy, recall, f1-score dan hasil *testing* dari setiap algoritma. Berdasarkan kesimpulan evaluasi tersebut, maka tingginya akurasi dari model tersebut dapat mengurutkan tingkat kualitas aplikasi perusahaan uang elektronik (*e-wallet*).

6. *Deployment*

Pada penelitian ini, tahap deployment akan dilakukan dengan menggunakan data yang telah dikumpulkan kedalam model analisis sentimen untuk menghasilkan akurasi prediksi.

3.5 Teknik Pengumpulan Data

Data utama yang digunakan pada penelitian ini berasal dari platform Google Play Store melalui ketiga uang elektronik (*e-wallet*) tersebut yaitu OVO, Gopay dan DANA. Pengumpulan data tersebut dilakukan dengan metode *text mining* menggunakan library API google-play-scraper. Data yang diambil merupakan ulasan yang ditulis oleh pengguna aplikasi ketiga uang elektronik tersebut. Hasil yang di dapatkan yaitu berupa format .csv. Periode data yang diambil dan digunakan yaitu dari Januari 2023 sampai dengan Februari 2024.

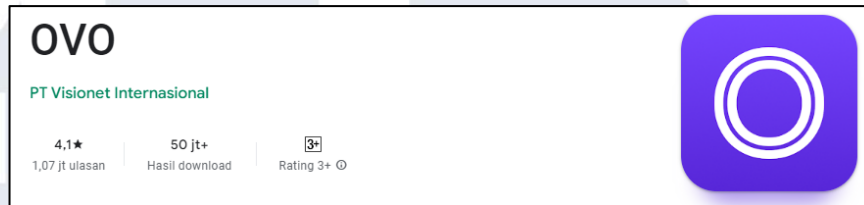
3.6 Teknik Pengambilan Sampel

Penelitian ini mengambil sampel dari user yang sudah pernah menggunakan salah satu aplikasi uang elektronik (*e-wallet*). User merupakan seseorang atau kelompok tertentu yang telah memiliki pengalaman dalam menggunakan aplikasi uang elektronik (*e-wallet*) tersebut sehingga dapat memberikan *feedback*.

Sampel merupakan sekelompok elemen yang dipilih dari kelompok yang lebih besar dengan harapan mempelajari kelompok yang lebih kecil akan dapat mengungkapkan informasi penting tentang kelompok yang lebih besar. Pengambilan sampel pada penelitian ini menggunakan rumus Slovin dengan tingkat margin error sebesar 5% untuk mengetahui perhitungan minimal jumlah sampel yang dibutuhkan dari total jumlah populasi yang ada dalam setiap ulasan masing-masing aplikasi uang elektronik tersebut [25]. Data sampel yang sudah didapatkan kemudian dibagi menjadi sebuah kelompok, dalam penelitian ini kelompok yang dibagi adalah user yang memberikan pendapat positif dan negatif. Berikut merupakan hitungan dari sampel yang dibutuhkan dalam populasi setiap aplikasi uang elektronik yang ada.

3.6.1 OVO

Populasi ulasan yang terdapat pada aplikasi uang elektronik OVO yaitu sebanyak 1,07 juta ulasan, dapat dilihat pada Gambar 3.2. Dengan demikian, perhitungan sampel yang dibutuhkan dari jumlah populasi tersebut yaitu sebagai berikut.



Gambar 3. 3 Jumlah Ulasan OVO

$$n = \frac{N}{1 + N(e)^2}$$

$$n = \frac{1070000}{1 + 1070000(0.05)^2}$$

$$n = \frac{1070000}{1 + 1070000(0.0025)}$$

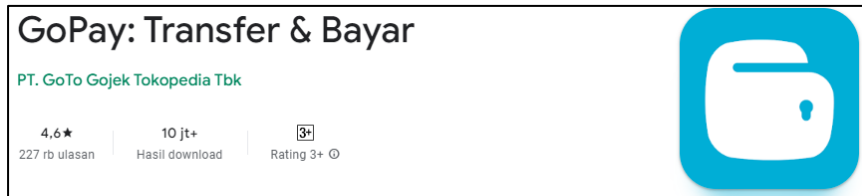
$$n = \frac{1070000}{2676}$$

$$n = 399.851 \sim 400 \text{ (dibulatkan menjadi 400)}$$

Berdasarkan perhitungan rumus slovin diatas terhadap populasi ulasan aplikasi uang elektronik OVO, maka jumlah minimal sampel yang harus digunakan untuk penelitian adalah 400 ulasan.

3.6.2 Gopay

Populasi ulasan yang terdapat pada aplikasi uang elektronik gopay yaitu sebanyak 227 ribu ulasan, dapat dilihat pada Gambar 3.3. Dengan demikian, perhitungan sampel yang dibutuhkan dari jumlah populasi tersebut yaitu sebagai berikut.



Gambar 3. 4 Jumlah Ulasan Gopay

$$n = \frac{N}{1 + N(e)^2}$$

$$n = \frac{227000}{1 + 227000(0.05)^2}$$

$$n = \frac{227000}{1 + 227000(0.0025)}$$

$$n = \frac{227000}{1 + 227000(0.0025)}$$

$$n = \frac{227000}{568.5}$$

$$n = 399.296 \sim 400 \text{ (dibulatkan menjadi 400)}$$

Berdasarkan perhitungan rumus slovin diatas terhadap populasi ulasan aplikasi uang elektronik Gopay, maka jumlah minimal sampel yang harus digunakan untuk penelitian adalah 400 ulasan.

3.6.3 DANA

Populasi ulasan yang terdapat pada aplikasi uang elektronik DANA yaitu sebanyak 5,86 juta ulasan, dapat dilihat pada Gambar 3.4. Dengan demikian, perhitungan sampel yang dibutuhkan dari jumlah populasi tersebut yaitu sebagai berikut.

UNIVERSITAS
MULTIMEDIA
NUSANTARA



Gambar 3. 5 Jumlah Ulasan DANA

$$n = \frac{N}{1 + N(e)^2}$$

$$n = \frac{5860000}{1 + 5860000(0.05)^2}$$

$$n = \frac{5860000}{1 + 5860000(0.0025)}$$

$$n = \frac{5860000}{1 + 5860000(0.0025)}$$

$$n = \frac{5860000}{14651}$$

$$n = 399.972 \sim\sim 400 \text{ (dibulatkan menjadi 400)}$$

Berdasarkan perhitungan rumus slovin diatas terhadap populasi ulasan aplikasi uang elektronik DANA, maka jumlah minimal sampel yang harus digunakan untuk penelitian adalah 400 ulasan.

3.7 Teknik Analisis Data

Tabel 3. 3 Tabel Perbandingan Tools

Tools	Kelebihan	Kekurangan
Rstudio	Open Source, memiliki package dan function yang lengkap, multifungsi dapat digunakan dalam proses cleansing sampai pembuatan model.	Cukup kompleks, UI dan GUI tidak terlalu user friendly, eksekusi coding membutuhkan memori yang besar.

Tools	Kelebihan	Kekurangan
Jupyter Notebook	Memiliki UI dan GUI yang user friendly, mudah untuk diaplikasikan, membutuhkan memori yang kecil	Sulit untuk mengidentifikasi kesalahan, kurang terorganisir, sulit untuk melakukan <i>debuging</i> .
Rapid Minner	Memiliki fasilitas lengkap untuk melakukan proses evaluasi, memiliki banyak metode integrasi data	Harus memiliki pengalaman dalam menggunakan database, kurang untuk bereksperimen

Tools yang akan digunakan dalam penelitian ini adalah Jupyter Notebook dengan menggunakan platform Google Colab. Jupyter Notebook menggunakan bahasa pemrograman Python yang memiliki banyak library untuk mendukung penelitian ini. Selain itu, Jupyter Notebook mampu untuk memproses data yang berukuran besar dan tidak memiliki limitasi ukuran data. Pembuatan data testing dan data training juga dapat dilakukan menggunakan Jupyter Notebook. Dengan menggunakan Jupyter Notebook maka seluruh proses dari penelitian ini dapat diaplikasikan.

