

BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Tabel 2. 1 Tabel Penelitian Terdahulu

Judul Jurnal	Nama Jurnal	Penulis / Tahun	Metode	Hasil	Kesimpulan
Text Mining Untuk Analisis Sentimen Pelanggan Terhadap Layanan Uang Elektronik Menggunakan Algoritma Support Vector Machine [8]	SISFOKOM	Fajar Romadoni, Yyuyun Umaidah, Betha Nurina Sari (2020)	Text Mining, Support Vector Machine, Knowledge Discovery in Databases (KDD), Confusion Matrix, Web Scraping	Algoritma Support Vector Machine dapat mencapai tingkat akurasi sebesar 98.7%, hal tersebut menunjukkan bahwa algoritma Support Vector Machine memiliki performa yang sangat baik untuk mengklasifikasikan data sentimen positif maupun negatif. Proses preprocessing merupakan proses yang menentukan tingkat akurasi dimana bahwa semakin bersih data yang digunakan maka data	Algoritma Support Vector Machine membutuhkan data numerik untuk memproses data, makadari itu dilakukan transformasi data menggunakan teknik term frequency. Nilai akurasi yang didapat yaitu sebesar 98.7%.

Judul Jurnal	Nama Jurnal	Penulis / Tahun	Metode	Hasil	Kesimpulan
				yang diolah oleh Vector Machine akan semakin bagus.	
Analisis Sentimen Pengguna E-Wallet Dana dan Gopay Pada Twitter Menggunakan Metode Support Vector Machine (SVM) [10]	TEKNIKA	Muhamad Kabirur Rifa, Moch. Hafid Totohendarto, M. Rafi Muttaqin (2023)	Analisis Sentimen, Support Vector Machine (SVM)	Model yang dibentuk dengan Support Vector Machine (SVM) menghasilkan akurasi 92% untuk DANA dan 90% untuk Gopay.	Algoritma Support Vector Machine terbukti dapat digunakan untuk pembuatan model analisis sentimen aplikasi E-Wallet. Hasil yang didapatkan sudah diatas 70% sehingga dapat digunakan.
Analisis Sentimen Pengguna Dompot Digital DANA Pada Kolom Komentar Google Play Store Dengan Metode Klasifikasi Support Vector Machine [11]	JATI	Wahyu Eko Saputro, Haris Yuana, Wahyu Dwi Puspitasari (2023)	Analisis Sentimen, Klasifikasi, Support Vector Machine	Akurasi yang didapatkan dengan menggunakan algoritma Support Vector Machine untuk aplikasi DANA yaitu diatas 80%. Hasil evaluasi menunjukkan 617 <i>true negative</i> dan 263 <i>true positive</i> .	Hasil evaluasi menunjukkan bahwa lebih banyak data dengan sentimen negatif untuk aplikasi DANA.

Judul Jurnal	Nama Jurnal	Penulis / Tahun	Metode	Hasil	Kesimpulan
Analisis Pemasaran Media Sosial Untuk Menentukan Influencer dan Topik Pembicaraan (Studi Kasus: Go-Pay dan OVO) [12]	JMM Online	Assyifa Reska Sisilia, Refi Rifaldi Windya Giri (2020)	Text Mining, User Generated Content, Social Network Analysis, Topic Modeling	Pengolahan data menunjukkan bahwa akun yang memiliki pengaruh pada jaringan sosial GO-Pay yaitu akun @ptinkad dan @alfamart. Untuk akun yang berpengaruh kepada OVO yaitu akun @ptinka dan @bareksacom.	Akun-akun yang menjadi influencer dan berpengaruh pada media sosial untuk Go-Pay yaitu @ptinka dan @alfamart. Untuk akun yang berpengaruh kepada OVO yaitu akun @ptinka dan @bareksacom. Ketiga akun tersebut memiliki interaksi antar akun, dan penyebaran informasi yang tinggi dalam jaringan sosial Go-Pay dan OVO.
Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest [9]	JPTIHK	Fanka Angelina Larasati, Dian Eka Ratnawati, Buce Trias Hanggara (2022)	Random Forest, API Google-play-scraper	Algoritma random forest digunakan untuk melakukan analisis sentimen, dan menggunakan Accuracy.	Penggunaan library API Google-play-scraper memudahkan peneliti untuk mengambil data ulasan aplikasi melalui kolom review setiap aplikasi di Google Play Store. Nilai akurasi yang didapatkan dalam menggunakan algoritma

Judul Jurnal	Nama Jurnal	Penulis / Tahun	Metode	Hasil	Kesimpulan
					Random Forest dengan data test 20% yaitu sebesar 84%
Analisis Sentimen Aplikasi DANA Menggunakan Naïve Bayes Classifier dan Support Vector Machine [13]	JUTISI	Gabriel Genaro Warow, Henry Pandia (2024)	Analisis Sentimen, Support Vector Machine, Naïve Bayes	Support Vector Machine menghasilkan akurasi yang lebih besar untuk analisis sentimen dibandingkan algoritma Naïve Bayes. Hasil analisis sentimen menunjukkan bahwa 84.7% pengguna DANA memberikan sentimen negatif.	Algoritma Support Vector Machine terbukti lebih efektif untuk klasifikasi analisis sentiment dibandingkan algoritma Naïve Bayes. DANA memiliki sentimen negatif.
Implementasi Scraping Data Untuk Sentiment Analysis Pengguna Dompot Digital Dengan Menggunakan Algoritma Machine Learning [14]	JANITRA	Eka Yuniar, Dwi Safiroh, Dian Wahyuningsih (2022)	Scraping Data, Analisis Sentimen	Hasil penelitian menunjukkan bahwa Gopay memiliki sentiment positif sebesar 79,6%.	Gopay memiliki sentimen positif yang sangat baik.
Analisis Sentimen Kualitas Layanan Teknologi Pembayaran	JEISBI	Enggarbela Ogi Intan Pratiwi, Wiyli	Analisis sentimen, Support Vector Machine (SVM), Naïve Bayes dan	Hasil dari sentimen yang telah dilakukan terhadap kualitas layanan OVO dan Dana yaitu DANA	Kualitas layanan aplikasi DANA sebagai pembayaran elektronik memiliki sentimen yang negatif.

Judul Jurnal	Nama Jurnal	Penulis / Tahun	Metode	Hasil	Kesimpulan
Elektronik pada Twitter (Studi Kasus Ovo dan Dana) [15]		Yustanti (2021)	K-Nearest Neighbour	memiliki sentimen negatif yang lebih banyak	
Analisis Sentimen E-Wallet Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization [16]	BITS	Vina Vamilina, Rice Novita (2023)	Text Mining, Support Vector Machine (SVM), Analisis Sentimen	Hasil dari sentimen yang telah dilakukan terhadap kualitas layanan DANA memiliki sentimen negatif yang lebih banyak daripada sentimen positif. Hasil Sentimen OVO seimbang antara sentimen positif dan negatif.	Hasil sentimen terhadap kualitas layanan OVO memiliki hasil yang cukup seimbang antara sentimen positif dan negatif. DANA memiliki lebih banyak sentimen negatif.
Penerapan Metode Naïve Bayes Classifier Pada Analisis Sentimen Analisis Gopay [17]	JATI	Riskawati, Fatihanursari, Ilin, Ade Rizki Rinaldi (2024)	Text Mining, Analisis Sentimen, Naïve Bayes	Hasil dari klasifikasi menggunakan algoritma Naïve Bayes, Gopay memiliki sentimen yang sangat positif.	Gopay memiliki kualitas aplikasi yang sangat baik dengan sentimen dan ulasan yang mendukung.

Penelitian pertama menjelaskan bahwa penggunaan Algoritma Support Vector Machine dapat mengklasifikasikan data sentimen positif dan data sentimen negatif, penelitian tersebut dilakukan pada objek OVO. Proses preprocessing menunjukkan bahwa semakin akurasi yang dihasilkan akan menjadi semakin besar apabila data yang akan diolah adalah data yang bersih [8]. Penelitian kedua menjelaskan mengenai persentase perusahaan uang digital yang paling banyak diketahui oleh masyarakat yaitu peringkat pertama diduduki oleh DANA, kemudian disusul oleh Gopay dan OVO. Penelitian tersebut membuktikan bahwa algoritma Support Vector Machine (SVM) dapat digunakan untuk analisis sentimen aplikasi uang elektronik dengan hasil yang baik yaitu diatas 70% [10]. Penelitian ketiga melakukan analisis sentimen terhadap aplikasi uang elektronik DANA. Hasil evaluasi menunjukan bahwa aplikasi uang elektronik DANA memiliki sentimen negatif yang lebih banyak. Akurasi yang didapatkan dengan menggunakan algoritma Support Vector Machine yaitu sebesar 80,81% [11]. Penelitian keempat menganalisa aplikasi uang elektronik yang sangat berpengaruh bagi Go-Pay, OVO dan DANA melalui interaksi yang dilakukan. Akun-akun yang memiliki interaksi antar akun, dan penyebaran informasi yang tinggi dalam jaringan sosial Go-Pay, OVO dan DANA merupakan kriteria yang menentukan apakah akun sosial media tersebut sangat berpengaruh bagi perusahaan uang digital. Didapatkan bahwa Gopay, OVO dan DANA merupakan aplikasi dengan interaksi yang terbaik [12]. Penelitian kelima menjelaskan bahwa terdapat library API bernama google-play-scraper yang berfungsi untuk mengambil ulasan suatu aplikasi di Google Play Store secara langsung. Analisis sentimen dilakukan dengan menggunakan algoritma Random Forest dengan pembagian data sebesar 20% data testing. Teruji bahwa pembagian data sebesar 20% yang dapat menghasilkan akurasi paling bagus yaitu sebesar 84% [9]. Penelitian keenam membuktikan bahwa algoritma yang lebih baik untuk klasifikasi analisis sentiment yaitu algoritma Support Vector Machine dibandingkan dengan Naïve Bayes. Hasil analisis sentimen pada penelitian tersebut menunjukan bahwa mayoritas pengguna aplikasi uang elektronik memberikan sentimen negatif. Data menunjukan bahwa DANA memiliki 84.7% sentimen negatif, sehingga dapat disimpulkan bahwa DANA memiliki sentimen yang sangat

negatif [13]. Penelitian ketujuh membuktikan bahwa aplikasi uang elektronik Gopay memiliki sentimen positif yang sangat tinggi. Data dalam penelitian tersebut menunjukkan bahwa Gopay memiliki sentimen positif sebesar 79,6%, sedangkan sentimen negatif yang dimiliki hanya sebesar 20,4% [14]. Penelitian kedelapan menyimpulkan bahwa kualitas aplikasi pembayaran elektronik DANA memiliki lebih banyak sentimen negatif dibandingkan dengan OVO. Jurnal tersebut menggunakan data dengan periode waktu 2 bulan dan DANA memiliki sentimen negatif sebesar 82% dari keseluruhan data tersebut [15]. Penelitian kesembilan menyimpulkan bahwa kualitas aplikasi OVO memiliki sentimen yang positif. Walaupun demikian, angka dari hasil sentimen OVO cukup seimbang antara positif dan negatif, dari 1000 data yang disebutkan dalam jurnal tersebut, 579 ulasan memiliki sentimen positif dan 421 ulasan memiliki sentimen negatif, sedangkan DANA memiliki lebih banyak sentimen negatif [16]. Penelitian kesepuluh menyimpulkan bahwa kualitas aplikasi Gopay memiliki sentimen yang sangat positif. Jurnal tersebut menggunakan data sebanyak 1000 data, dan dari data tersebut terdapat 712 sentimen positif dan 288 sentimen negatif. Dapat disimpulkan bahwa Gopay memiliki 72% sentimen positif dan tergolong memiliki kualitas aplikasi yang baik [17].

2.2 Landasan Teori

2.2.1 Dompet Elektronik (*E-Wallet*)

Dalam Pasal 1 angka 7 Peraturan Bank Indonesia Nomor 18/40/PBI/2016 pada Tahun 2016 mengenai Penyelenggaraan Pemrosesan Transaksi Pembayaran, dompet elektronik (*e-wallet*) didefinisikan sebagai layanan elektronik yang digunakan untuk menyimpan data pembayaran antara lain alat pembayaran dengan menggunakan teknologi uang elektronik kartu dengan teknologi *chip* yang dapat menampung dana untuk pembayaran. Bank Indonesia sangat memperhatikan penggunaan dompet elektronik (*e-wallet*) untuk diterapkan terhadap masyarakat Indonesia. Gerakan Nasional Non Tunai (GNNT) secara resmi dijalankan pada tanggal 14 Agustus 2014 oleh Gubernur Bank Indonesia untuk menyadarkan masyarakat terhadap penggunaan pembayaran nontunai (*cashless*), sehingga dapat diterapkan dalam adat dan

budaya masyarakat dalam transaksi sehari-hari [10]. Terdapat beberapa *e-wallet* yang sudah beredar dan banyak digunakan di Indonesia seperti GoPay, OVO dan DANA.

2.2.1.1 GoPay

GoPay merupakan dompet elektronik (*e-wallet*) yang dikeluarkan oleh PT Dompet Anak Bangsa dan telah terdaftar dan diawasi oleh Bank Indonesia. GoPay memiliki fungsi yang sama dengan uang tunai yang digunakan sebagai alat transaksi yang sah. Penggunaan GoPay dilakukan dengan cara menyetor terlebih dahulu uang tunai kedalam rekening GoPay. Pengguna harus terlebih dahulu membuat akun yang didaftarkan dalam aplikasi GoJek untuk menggunakan fitur dompet elektronik (*e-wallet*) GoPay. GoPay memiliki beberapa fitur yaitu *top-up*, *payment*, dan *transfer of funds*. *Top-up* merupakan fitur yang berfungsi untuk user melakukan pemindahan uang tunai ke rekening GoPay, *top-up* dapat dilakukan melalui bank atau driver secara langsung. *Payment* merupakan fitur yang berfungsi untuk melakukan pembayaran. *Transfer of funds* merupakan fitur yang berfungsi untuk memindahkan dana ke akun GoPay lain [12].

2.2.1.2 OVO

OVO merupakan dompet elektronik yang dikeluarkan oleh PT Visionet Internasional. OVO adalah *e-wallet* yang dapat digunakan sebagai media transaksi keuangan. OVO dibentuk dengan misi sosial untuk mendukung program pemerintah terkait dengan Gerakan Nasional Non Tunai. Sama seperti *e-wallet* lainnya, OVO merupakan dompet elektronik yang berbentuk uang elektronik dan dapat diakses melalui aplikasi *e-wallet* OVO. Penggunaan OVO dilakukan dengan cara memasukan uang tunai kedalam rekening OVO. OVO telah bekerja sama dengan 500 ribu merchant dalam 300 kota di Indonesia [12].

2.2.1.3 DANA

DANA merupakan dompet elektronik yang dikeluarkan oleh PT Espay Debit Indonesia pada tahun 2018. DANA dapat digunakan melalui aplikasi DANA yang dapat diunduh melalui Google Play Store dan Apple Store. DANA dapat digunakan sama seperti dompet elektronik lainnya yaitu sebagai media transaksi keuangan. DANA menyediakan layanan transaksi mulai dari membayar tagihan listrik, pembelian pulsa, investasi dan transaksi antar bank. DANA merupakan salah satu pencetus dompet elektronik yang memiliki review terbanyak yaitu sebanyak 5,5 juta review dalam platform Google Play Store [18].

2.2.2 Text Mining

Text Mining merupakan salah satu metode yang digunakan untuk melakukan proses ekstraksi informasi melalui dokumen-dokumen teks yang tidak terstruktur. *Text mining* bekerja dengan cara menambang data berupa teks yang berasal dari dokumen dengan tujuan untuk mencari kata-kata yang dapat mewakili seluruh isi dokumen tersebut sehingga dapat dilakukan penelitian yang lebih lanjut. *Text mining* biasanya digunakan untuk penelitian yang bertujuan untuk menemukan pola-pola menarik dari data yang berukuran besar [19].

Text Mining memiliki proses dan langkah-langkah yaitu sebagai berikut:

- *Text Preprocessing*

Pada proses ini, text akan di proses agar dapat digunakan untuk analisis sentimen. Langkah-langkah yang dilakukan untuk proses ini yaitu:

- a. *To lower case*, merupakan metode dimana semua karakter huruf akan diubah menjadi karakter huruf kecil.
- b. *Tokenizing*, merupakan proses yang merubah kalimat utuh menjadi kata-kata.
- c. *Remove number*, yaitu proses yang digunakan untuk menghilangkan angka dalam kata.

d. *Remove url*, yaitu proses yang digunakan untuk menghilangkan url yang terdapat dalam kata.

e. *Remove punctuation*, yaitu menghilangkan tanda baca dan delimiter.

- ***Feature Selection***

a. *Stopword*, merupakan kosakata yang tidak sesuai dengan ciri dari suatu dokumen atau teks.

b. *Stemming*, merupakan proses pemetaan dan penguraian dari suatu kata menjadi bentuk kata dasar.

Pada penelitian ini, pengumpulan data dengan metode text mining akan dilakukan melalui *platform* Google Play Store.

2.2.3 Analisis Sentimen

Analisis sentimen merupakan salah satu metode yang digunakan untuk mendeteksi sikap-sikap terhadap objek atau orang. Tujuan dari analisis sentimen yaitu untuk mendapatkan persentase sentimen positif dan sentimen negatif terhadap objek penelitian. Sentimen analisis dilakukan untuk melakukan penilaian terhadap emosi, sikap, pendapat dan evaluasi yang disampaikan oleh seseorang terhadap objek penelitian. Analisis sentimen dapat dilakukan dengan menggunakan pendekatan *machine learning* dengan menggunakan algoritma untuk mengklasifikasikan teks berdasarkan dataset yang telah diberi label sentimen. Analisis sentimen dapat memberikan manfaat terhadap objek yang diteliti agar dapat mendapatkan informasi mengenai sentimen positif atau negatif yang dimiliki, sehingga dapat membantu dalam melakukan evaluasi dan perbaikan. Analisis sentimen memiliki beberapa langkah yaitu sebagai berikut [20]:

- *Level 1*

Level kalimat: Pada level kalimat, proses yang dilakukan yaitu mencari sentimen netral dan negatif serta positif untuk setiap baris kalimat.

- *Level 2*

Level dokumen: Mencari sentimen netral, negatif atau positif untuk seluruh dokumen

- *Level 3*

Level aspek: Pada level aspek, proses analisis sentimen melakukan pengelompokan terhadap hasil sentimen yang sama.

- *Level 4*

Level pengguna: Analisis sentimen dimanfaatkan dengan visualisasi data untuk diimplementasikan kepada user.

2.3 Framework dan Algoritma

2.3.1 CRISP-DM

Cross-Industry Standard Process for Data Mining yang disingkat menjadi CRISP-DM dikembangkan pertama kali pada tahun 1996 oleh lima perusahaan. CRISP-DM merupakan standar yang ditujukan untuk melakukan proses analisis pengambil keputusan dan pemecahan masalah dari suatu bisnis atau penelitian. CRISP-DM memiliki 6 tahapan yaitu sebagai berikut [21]:

1. *Business Understanding*

Pada tahap ini tujuan, objektif, dan target bisnis yang ingin dicapai harus ditentukan. Tahap ini juga dilakukan untuk menilai kondisi awal bisnis sehingga memiliki gambaran dari sumber daya yang akan diperlukan.

2. *Data Understanding*

Pada tahap ini pengumpulan dan eksplorasi data dilakukan untuk memastikan bahwa kualitas data dapat digunakan untuk keperluan bisnis. Exploratory Data Analysis (EDA) juga akan dilakukan pada tahap ini.

3. *Data Preparation*

Pada tahap ini data yang sudah di eksplorasi akan dilakukan proses preprocessing dimana data akan dibersihkan dan menyesuaikan format sehingga cocok untuk keperluan analisis.

4. Data Modelling

Pada tahap ini data akan diaplikasikan kepada model yang akan digunakan. Pembuatan model dapat dibuat lebih dari satu model untuk dilakukan komparasi.

5. *Evaluation*

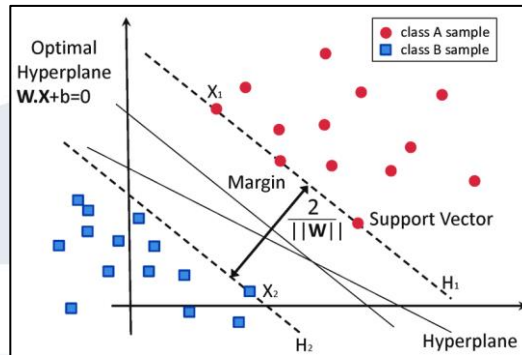
Pada tahap ini performa data dan model akan di evaluasi dan dibandingkan sehingga tidak ada tahapan yang terlewatkan dan mendapatkan model yang paling cocok untuk keperluan bisnis.

6. *Deployment*

Deployment merupakan proses implementasi model dan menggunakannya dalam keperluan bisnis.

2.3.2 *Support Vector Machine (SVM)*

Support Vector Machine (SVM) merupakan salah satu algoritma yang digunakan untuk klasifikasi. *Support Vector Machine (SVM)* dikenalkan pertama kali pada tahun 1992 oleh Vapnik yang bergerak dalam bidang *pattern recognition*. *Support Vector Machine (SVM)* membutuhkan training set positif dan negatif untuk mengklasifikasikan dua kelas. *Support Vector Machine (SVM)* bekerja untuk menemukan pemisah yang terbaik untuk mengklasifikasikan kedalam dua kelas dan memaksimalkan margin antara dua kelas tersebut. Pemisah antar kelas tersebut disebut dengan *hyperplane*.



Gambar 2. 1 Support Vector Machine (SVM)

Gambar 2.1 merupakan gambaran dari cara kerja algoritma *Support Vector Machine (SVM)*. *Hyperplane* tersebut ditentukan berdasarkan titik yang berada di antara setiap kelas dengan jarak terdekat dengan kelas lainnya, sehingga disebut dengan *support vector*. *Support Vector Machine (SVM)* bekerja dengan menggunakan formula, salah satu yang paling sederhana yaitu formula yang linier [22].

$$w \cdot x + b = 0$$

Rumus 2. 1 Formula Sederhana SVM Linear

Keterangan :

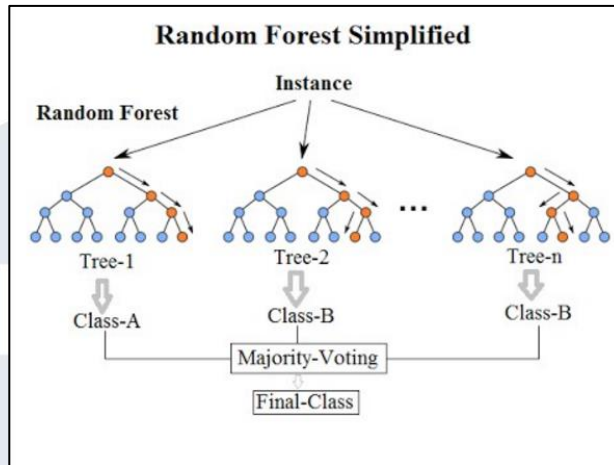
w = parameter hyperplane (garis tegak lurus antara garis *hyper plane* dan titik support vector)

x = titik data masukan Support Vector Machine

b = parameter *hyperlane* yang dicari (nilai bias)

2.3.3 *Random Forest*

Random Forest diperkenalkan pertama kali pada tahun 2001 oleh Breiman. Breiman memperkenalkan algoritma *Random Forest* dengan menunjukkan beberapa kelebihan yang dimiliki seperti nilai error yang lebih rendah, performa yang tinggi dalam klasifikasi dan memiliki metode efektif untuk mengestimasi data *missing* [23]. *Random Forest* merupakan salah satu algoritma machine learning yang digunakan untuk melakukan klasifikasi dan regresi.



Gambar 2. 2 *Random Forest*

Gambar 2.2 merupakan cara kerja *Random Forest*. *Random Forest* dikembangkan dari metode CART (*Classification and Regression Trees*) dan bekerja dengan menggabungkan banyak *decision tree* sehingga menghasilkan akurasi prediktif yang lebih besar daripada *decision tree* secara tunggal. *Random Forest* menggunakan sistem *voting* untuk memilih pohon keputusan yang dapat menghasilkan akurasi terbaik. *Random Forest* memiliki beberapa keunggulan yaitu dapat menangani berbagai jenis data termasuk *imbalanced data* dan dapat mencegah *overfitting* [9]. *Random Forest* bekerja dengan menggunakan formula *entropy* dan *information gain* yang berfungsi untuk mengukur seberapa informatif sebuah *node* [24].

$$Entropy (S) = \sum_{i=1}^n -p_i \log_2 p_i$$

Rumus 2. 2 Rumus *Entropy Random Forest*

Keterangan :

S = Himpunan dataset

n = Banyaknya jumlah kelas

p_i = Probabilitas kelas ke- i dalam output S

$$\text{Information Gain (A)} = \text{Entropy (S)} = \sum_{i=1}^n \frac{S_i}{S} \times \text{Entropy}(S_i)$$

Rumus 2. 3 Rumus *Information Gain Random Forest*

Keterangan :

A = Atribut

S = Himpunan dataset

$|S_i|$ = Jumlah sampel untuk nilai ke-*i*

$|S|$ = Banyaknya jumlah data

2.4 Tools dan Software

2.4.1 Rumus Slovin

Rumus Slovin digunakan untuk mencari jumlah sampel yang dibutuhkan pada suatu penelitian berdasarkan jumlah populasi yang ada. Untuk mencari jumlah sampel yang dibutuhkan berdasarkan rumus Slovin yaitu sebagai berikut [25].

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

Rumus 2. 4 Rumus Slovin

Keterangan :

n = Ukuran sampel

N = Ukuran populasi

e = Nilai toleransi kesalahan pengambilan sampel

2.4.2 Transformers

Transformers merupakan *library* yang dikembangkan menggunakan bahasa pemrograman python dan memiliki fungsi untuk mengunduh *pre-trained model*. Library transformers umumnya digunakan untuk melakukan NLP (*Natural Language Processing*) seperti melakukan klasifikasi teks untuk analisis sentimen. Salah satu contoh

model untuk melakukan klasifikasi teks yaitu Indonesian-roBERTa, dimana model tersebut telah dilatih untuk mengklasifikasikan teks positif dan negatif dalam bahasa Indonesia [26].

2.4.3 *CountVectorizer*

CountVectorizer merupakan salah satu fitur ekstraksi yang bekerja dengan cara mengubah kumpulan teks menjadi sebuah matrix. *CountVectorizer* akan mengubah teks menjadi vektor yang berisikan frekuensi jumlah kemunculan setiap kata yang ada dalam data set. Vektor akan terbentuk berdasarkan ukuran kata pada data set, jumlah vektor akan bertambah setiap kali kata ditemukan. *CountVectorizer* akan melakukan ekstraksi kata yang memiliki minimal 2 huruf [27].

2.4.4 Google Play Store



Gambar 2. 3 Logo Google Play Store

Google Play Store merupakan sebuah platform aplikasi yang memiliki fungsi sebagai tempat pemasaran aplikasi-aplikasi berbasis Android. Developer Aplikasi dapat memasarkan aplikasi-aplikasi yang telah dibuat melalui Google Play Store. Google Play Store memiliki banyak sekali kategori aplikasi yang dapat di unduh secara berbayar maupun secara gratis. Salah satu kategori yang digunakan untuk penelitian ini yaitu kategori keuangan yang berisikan aplikasi-aplikasi dompet elektronik (*E-Wallet*). Setiap aplikasi di Google Play Store terdapat *reviews* dan *rating* untuk menampung aspirasi serta tanggapan pengguna terhadap aplikasi tersebut [28].

2.4.5 Microsoft Excel



Gambar 2. 4 Logo Microsoft Excel

Microsoft Excel merupakan salah satu perangkat lunak yang memiliki fungsi untuk mengolah data. Microsoft Excel memiliki spreadsheet yang terdiri dari baris dan kolom. Microsoft Excel dapat digunakan untuk berbagai bidang salah satunya yaitu untuk melakukan penambahan, pengurangan, dan perubahan terhadap data. Microsoft Excel menyediakan berbagai macam fitur yang dapat membantu pengguna dalam melakukan analisis terhadap data [29].

2.4.6 Google Colaboratory



Gambar 2. 5 Logo Google Colaboratory

Google Colaboratory (Google Colab) merupakan platform berbasis cloud yang dikembangkan oleh Google untuk mengembangkan dan mengeksekusi kode Python. Google Colaboratory menggunakan basis Jupyter Notebook sebagai lingkungan *notebook* yang interaktif untuk menjalankan kode Python. Google Colaboratory memiliki akses ke dalam *cloud* seperti CPU, GPU dan TPU (*Tensor Processing Units*) dari Google untuk menjalankan kode secara efisien sehingga pengembangan model *machine learning* dapat dilakukan secara efektif. Google

Colaboratory dapat diintegrasikan dengan layanan penyimpanan data berbasis *cloud* yaitu Google Drive sehingga dapat melakukan *backup* secara *real-time* [30].

2.4.7 Python



Gambar 2. 6 Logo Python

Python merupakan salah satu bahasa pemrograman yang dikembangkan pertama kali oleh Guido van Rossum pada awal tahun 1990-an. Python adalah bahasa pemrograman dengan tingkat yang tinggi sehingga sering digunakan untuk mengembangkan aplikasi, web, machine learning, dan artificial intelligence. Python memiliki berbagai macam library dan modul seperti NumPy, Pandas, Pymongo, PyTorch dan masih banyak lagi yang membantu pengguna dalam melakukan pengolahan data menjadi lebih efisien. Python merupakan bahasa pemrograman yang bersifat open-source, sehingga banyak pengguna dalam komunitas yang melakukan pengembangan seperti membuat library dan modul yang dapat memudahkan pengguna lain dalam melakukan analisis data [31].

2.4.8 Confusion Matrix

Confusion Matrix digunakan untuk mengevaluasi kinerja dari algoritma klasifikasi. *Confusion matrix* mengevaluasi kinerja berdasarkan hasil dari data testing yang merepresentasikan performa dari model klasifikasi [32]. *Confusion matrix* memiliki empat parameter yang digunakan untuk mengevaluasi model, dapat dilihat dari Tabel 2.1.

Tabel 2. 2 Confusion Matrix

Class Designation		Actual Class	
		TRUE	FALSE
Predicted Class	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

2.4.8.1 Accuracy

Accuracy merupakan hitungan yang merepresentasikan jumlah dari prediksi yang akurat dari kelas *true positive* maupun kelas *true negative*. Perhitungan akurasi terbaik adalah angka 1.0 dan yang terburuk adalah 0.0. Adapun rumus yang digunakan untuk menghitung *Accuracy* yaitu sebagai berikut [32].

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

Rumus 2. 5 Rumus *Accuracy*

2.4.8.2 Recall

Recall merupakan hitungan yang merepresentasikan jumlah prediksi positif yang akurat (*True positive*). Perhitungan *recall* terbaik adalah angka 1.0 dan yang terburuk adalah 0.0. Adapun rumus yang digunakan untuk menghitung nilai *recall* yaitu sebagai berikut [32].

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

Rumus 2. 6 Rumus *Recall*

2.4.8.3 Precision

Precision merupakan hitungan yang merepresentasikan jumlah true positives yang dibandingkan dengan seluruh data yang memiliki nilai positive. Adapun rumus yang digunakan untuk menghitung nilai *precision* yaitu sebagai berikut [32].

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

2.4.8.4 *F1-Score*

F1-score merupakan hitungan yang merepresentasikan nilai gabungan antara *precision* dan *recall* untuk menghitung nilai rata-rata yang didapatkan dari nilai gabungan tersebut. Adapun rumus yang digunakan untuk menghitung nilai *F1-Score* yaitu sebagai berikut [32].

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

Rumus 2. 8 Rumus *F1-Score*

