

## Analisis Sentimen Ulasan Aplikasi Dana dengan Metode *Random Forest*

Fanka Angelina Larasati<sup>1</sup>, Dian Eka Ratnawati<sup>2</sup>, Buce Trias Hanggara<sup>3</sup>

Program Studi Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Brawijaya  
Email: <sup>1</sup>fankanglna@student.ub.ac.id, <sup>2</sup>dian\_ilkom@ub.ac.id, <sup>3</sup>buce\_trias@ub.ac.id

### Abstrak

Dompot Digital adalah aplikasi elektronik yang dapat digunakan untuk melakukan transaksi keuangan secara *online*, tanpa uang fisik dan tanpa kartu, pengguna hanya menggunakan *smartphone* dalam melakukan transaksi keuangannya. Aplikasi dompet digital yang populer disaat pandemi ini yaitu Dana. Aplikasi Dana memiliki banyak pengguna, sehingga sering kali terdapat ulasan positif, negatif dan netral yang tidak relevan dengan *rating* yang diberikan di *Google Play Store*. Data ulasan aplikasi Dana akan diperoleh menggunakan teknik *Web Scraping* menggunakan *API Google-Play-Scraper*. Data hasil *scraping* selanjutnya akan dilakukan *text preprocessing* untuk membersihkan teks agar data dapat dieksekusi. Analisis sentimen dapat mendeteksi apakah suatu teks mengandung opini positif, negatif maupun netral dari ulasan pengguna yang tidak sesuai dengan *rating*. *Random Forest* adalah metode dalam analisis yang terdiri dari beberapa pohon keputusan sebagai *classifier*. Pada Penelitian ini, digunakan metode *random forest* dengan membagi tiga kelas sentimen yaitu positif, negatif dan netral juga mengambil indikator evaluasi yaitu akurasi, *recall*, presisi dan *f-measure*. Pengujian dilakukan berdasarkan jumlah *tree* dan kedalaman *tree* terhadap 1354 data dengan membagi data 250 data per kelas nya. Berdasarkan hasil pengujian dan analisis yang sudah dilakukan, dengan perbandingan data latih dan data uji 80%:20% diperoleh nilai *precision* 84%, *recall* 84%, *F1-Score* 84% dan *accuracy* sebesar 84% dengan kedalaman *tree* 65 dan jumlah *tree* 400.

**Kata kunci:** *Dompot Digital, ulasan aplikasi Dana, Web Scraping, text preprocessing, analisis sentimen, Random Forest*

### Abstract

*Digital Wallet is an electronic application that can be used to make financial transactions online, without physical money and without cards, users only use smartphones in conducting financial transactions. A popular digital wallet application during this pandemic is Dana. The Dana app has many users, so there are often positive, negative and neutral reviews that are irrelevant to the rating given on the Google Play Store. The Dana application review data will be obtained using the Web Scraping technique using the Google-Play-Scraper API. The data from the scraping will then be done by text preprocessing to clean the text so that the data can be executed. Sentiment analysis can detect whether a text contains positive, negative or neutral opinions from user reviews that do not match the rating. Random Forest is a method in analysis that consists of several decision trees as a classifier. In this study, the random forest method was used by dividing three classes of sentiment, namely positive, negative and neutral. Also taking evaluation indicators, namely accuracy, recall, precision and f-measure. The test is carried out based on the number of trees and the depth of the tree to 1354 data by dividing the data by 250 data per class. Based on the results of tests and analyzes that have been carried out, with a comparison of training data and test data of 80%: 20%, namely precision 84%, recall 84%, F1-Score 84% and accuracy of 84% with a tree depth of 65 and a tree number of 400.*

**Keywords:** *Digital Wallet, Dana application review, Web Scraping, text preprocessing, sentiment analysis, Random Forest*

## 1. PENDAHULUAN

Dompot Digital adalah aplikasi yang digunakan untuk melakukan transaksi keuangan

secara daring, pengguna cukup menggunakan *smartphone* dalam melakukan transaksi (Bimo, 2021). Di masa pandemi COVID – 19 dompet digital menjadi alat pembayaran yang dianjurkan

karena meminimalkan kontak secara langsung saat bertransaksi. Bank Indonesia mencatat bahwa transaksi uang elektronik bulan Februari 2021 naik 26,42% dari tahun sebelumnya mencapai Rp 19,2 Triliun dibandingkan tahun sebelumnya (Richard, 2021).

Aplikasi dompet digital yang populer disaat pandemi setelah Gopay dan OVO yaitu Dana (Rahardyan, 2021). Aplikasi Dana diluncurkan pada tahun 2018 memungkinkan pengguna melakukan transaksi keuangan nontunai mulai dari pembayaran kebutuhan sehari-hari hingga *top up* saldo investasi. Berdasarkan survei YouGov, Dana adalah dompet digital tercepat dalam peningkatan pengguna dengan pertumbuhan tercepat sekitar 4% di masa COVID-19 (Liputan 6, 2021). Aplikasi Dana memiliki banyak pengguna, sehingga sering kali terdapat ulasan positif, negatif dan netral yang tidak relevan dengan *rating* yang diberikan di *Google Play Store*.

*Google Play Store* adalah toko aplikasi resmi yang dikembangkan oleh Google yang memungkinkan pengguna dalam mencari dan mengunduh aplikasi untuk sistem *Android* (Wikipedia, 2022). Data ulasan aplikasi Dana akan diperoleh menggunakan teknik *Web Scraping* yaitu proses mengekstraksi data dari halaman website. Data ulasan aplikasi Dana pada *Google Play Store* akan dilakukan *scraping* menggunakan *API Google-Play-Scraper*. *Google-Play-Scraper* adalah *API* untuk mengekstraksi data informasi aplikasi dan ulasan aplikasi dari *Google play store* dengan lebih mudah tanpa ketergantungan eksternal (Ginting, 2021). Data hasil *scraping* selanjutnya akan dilakukan *text preprocessing*.

*Text Preprocessing* merupakan suatu proses membersihkan teks yang dilakukan agar data dapat memenuhi persyaratan untuk dieksekusi (Sanjaya & Absar, 2015). *Text preprocessing* penting dilakukan dikarenakan keadaan teks mempengaruhi hasil akurasi (Baskoro, Susanto, & Khomsah, 2021). Data ulasan aplikasi akan dilakukan *Case folding*, *tokenizing*, *stopword removal* dan *stemming* (Fitri, Yuliani, & Gata, 2020). Ulasan yang telah dilakukan *text preprocessing* dapat menjadi sumber data untuk dianalisa dan diklasifikasi menggunakan analisis sentimen.

Analisis sentimen adalah teknik pemrosesan data untuk mendeteksi opini dalam bentuk teks (Adminlp2m, 2022). Analisis sentimen digunakan untuk mendeteksi teks yang berisi ulasan pelanggan yang bernada positif,

negatif maupun netral dari ulasan pengguna yang tidak sesuai dengan *rating* yang diberikan. Metode yang digunakan dalam penelitian ini yaitu *Random Forest*. Metode *Random Forest* berisi gabungan dari *Decision Tree* untuk melakukan klasifikasi, *Random Forest* merupakan algoritma *ensemble* dimana untuk mendapatkan keputusan akhir dilakukan *voting majority* dari semua model *Decision Tree* (Rohman, Purwanto, & Santoso, 2018). Algoritma *random forest* lebih akurat dibandingkan *Support Vector Machine* dan *Naïve Bayes* dengan peningkatan akurasi 7,16%. (Fitri, Yuliani, & Gata, 2020).

Penelitian analisis sentimen sebelumnya dilakukan oleh Boma Bayu Baskoro, Irwan Susanto dan Siti Khomsah. Penelitiannya menggunakan data komentar pelanggan hotel di Purwokerto yang diunduh dari situs *tripadvisor.co.id*. Metode yang digunakan *Random Forest Classifier* dan TF-IDF. Pada *preprocessing* dilakukan normalisasi, *stemming*, dan *stopword* dengan menambahkan kata *stopword* selain dalam *library sastrawi*. Hasil penelitian didapatkan bahwa akurasi model mencapai 87,23% dan jika tanpa proses *stemming*, akurasi model hanya 87,01%.

Berdasarkan uraian diatas, penulis akan meneliti bagaimana analisis sentimen ulasan aplikasi Dana menggunakan metode *Random Forest* dengan membagi kelas sentimen menjadi positif, negatif dan netral. Evaluasi yang digunakan yaitu akurasi, *recall*, presisi dan *f-measure*. Data diambil dari ulasan pengguna aplikasi Dana pada *Google Play Store* dengan *scraping* menggunakan *API Google-Play-Scraper* dan data teks dilakukan *preprocessing* sebelum diklasifikasi. Tujuan penelitian ini diharapkan dapat mengetahui hasil analisis sentimen pada kelas positif, negatif dan netral serta menjadi evaluasi untuk meningkatkan kualitas pelayanan pada aplikasi Dana.

## 2. LANDASAN KEPUSTAKAAN

### 2.1. Data Mining

*Data Mining* adalah cara atau proses yang dilakukan untuk mendapatkan informasi dari data yang besar. Hasilnya dapat digunakan untuk pengambilan keputusan jika diperlukan untuk masa mendatang (Buulolo, 2020). Tahapan dari proses *data mining* yaitu (Wahyudi, Indrajit, & Muhammad, 2017) :

a) *Selection*, yaitu proses pemilihan data dari

serangkaian data.

- b) *Processing*, yaitu proses membersihkan data seperti menghapus data duplikat, pengecekan data tidak konsisten, dan perbaikan kesalahan di data.
- c) *Transformation*, yaitu proses pengubahan format data dengan menggunakan coding.
- d) Proses *data mining*, yaitu proses mencari pola pada data menggunakan metode atau teknik tertentu.
- e) *Interpretation/Evaluation*, yaitu proses memahami pola atau informasi didapatkan untuk dilakukan *iterasi*.

## 2.2. Google Play Store

*Google Play Store* yaitu aplikasi khusus yang hanya terdapat di sistem android dikembangkan oleh google agar pengguna dapat melihat dan mengunduh aplikasi (Wikipedia, 2022). Pada penelitian *ini google play store* menjadi sumber data untuk mengumpulkan ulasan pengguna aplikasi Dana.

## 2.3. Dana

Dana adalah aplikasi dompet digital berasal dari Indonesia yang terdapat di *Google Play Store* digunakan untuk pembayaran digital *cashless* dan *cardless* baik *online* maupun *offline*. Dana menyediakan fitur mulai dari pembelian pulsa, tagihan listrik, hingga *top-up* saldo investasi. Tujuannya adalah agar mempermudah pengguna dalam bertransaksi dengan cepat, nyaman dan tetap aman (Dana, 2022).

## 2.4. Text Mining

*Text Mining* adalah proses untuk mengekstrak informasi dari data sumber untuk dilakukan analisis, mengelompokkan informasi berdasarkan kata-kata, agar dapat mengetahui hubungan dari data sumber lainnya (Findawati & Rosid, 2020).

*Text Mining* bertugas khusus untuk melakukan pengkategorian teks (*text categorization*) dan pengelompokan teks (*text clustering*). Teks yang sudah dikategorikan dan dikelompokkan selanjutnya dilakukan pemecahan kalimat menjadi kata (tokenisasi), lalu huruf besar dirubah menjadi kecil (kapitalisasi), dilakukan perubahan kata yang berimbuhan menjadi kata dasar (*stemming*), perhitungan dan pengelompokan kata (*filtering*) (Putri & Setiadi, 2014). Penerapan *text mining* misalnya analisis sentimen, preferensi pelanggan, penyaringan

spam, dan lainnya.

## 2.5. Analisis Sentimen

Analisis sentimen merupakan terapan dari *text mining*. Data yang sudah dikumpulkan akan dilakukan analisis agar mendapatkan pandangan sehingga kesimpulan dapat dibuat. Hasilnya yaitu sentiment positif, negatif maupun netral (Bahrawi, 2019). Analisis sentimen memiliki 4 jenis yaitu:

- 1) Analisis sentimen bertingkat, dimana digunakan untuk mengartikan pemberian bintang dalam ulasan. 5 bintang sangat positif dan 1 bintang sangat negatif
- 2) Deteksi Emosi, dimana digunakan untuk mendeteksi emosi seperti rasa Bahagia, marah, sedih, dan frustrasi.
- 3) Berbasis aspek, dimana digunakan untuk mengetahui aspek atau fitur yang memiliki sentiment positif, negatif atau netral
- 4) Multibahasa, dimana mendeteksi bahasa di dalam teks sesuai dengan bahasa yang dipilih

## 2.6. Web Scraping

*Web Scraping* adalah metode pengambilan informasi secara otomatis dari situs *web* dan data diambil dengan cara ekstraksi (Ayani, Pratiwi, & Hafiz, 2019). Tahapan umum *web scraping* yaitu (Flores, Permatasari, & Jasa, 2020):

- 1) Meminta HTTP dalam memperoleh sumber daya yang ditargetkan
- 2) Permintaan akan diproses oleh server dengan format URL berisi permintaan GET/HTTP yang berisi POST
- 3) Setelah diterima, sumber daya yang diminta akan diambil dan dikirim kembali dengan format yang beragam

Pada penelitian ini menggunakan *API Google-Play-Scraper* dalam pengambilan data.

### 2.6.1 API Google-Play-Scraper

*API Google-Play-Scraper* adalah metode untuk mengambil data dari *Google Play Store* tanpa dependensi eksternal menggunakan bahasa pemrograman python. Data yang diambil dapat berupa informasi aplikasi seperti judul aplikasi, *developer url*, kategori aplikasi, keseluruhan *rating* dan *review*, deskripsi, *thumbnail*, *rating* konten dan *screenshot* aplikasi. Selain informasi dari aplikasi *API google-play-scraper* juga dapat mengambil data ulasan pengguna seperti nama, foto, *rating*, tanggal, *comment likes*, dan *comment* (Zub, 2022). Berikut ini langkah-langkah dalam

mengambil data ulasan di *google play store* menggunakan *API Google-Play-Scraper*:

- 1) Melakukan install API Google-Play-Scraper dengan menjalankan ‘pip install google-play-scraper’
- 2) Melakukan import packages ‘from google\_play\_scraper import app’ ‘import pandas as pd’ ‘import numpy as np’
- 3) Mengambil aplikasi ID yang terdapat di url seperti ‘id.dana’
- 4) Melakukan scraping review
- 5) Memasukan hasil scraping ke dalam data frame

## 2.7. Random Forest

*Random Forest* adalah metode *machine learning* yang terdiri dari gabungan *Decision Tree* untuk dilakukan klasifikasi dimana untuk memperoleh keputusan akhir dilakukan *voting majority* (Rohman, Purwanto, & Santoso, 2018). Berikut ini langkah-langkah dalam membuat *Decision tree* atau pohon keputusan:

- 1) Menghitung nilai *gini impurity* pada setiap fitur
- 2) Menghitung nilai *gini impurity* dari setiap kondisi *split*
- 3) Menghitung nilai *average gini impurity* dari setiap kondisi *split*
- 4) Menghitung nilai *information gain* dari setiap kondisi *split*
- 5) Mencari nilai *information gain* tertinggi untuk dijadikan *root*
- 6) Ulangi Langkah 2-4 sampai k tree

## 2.8. Text Preprocessing

*Text Preprocessing* adalah proses membersihkan *text* sehingga menjadi berkualitas dan memenuhi persyaratan untuk dieksekusi pada sebuah algoritma (Sanjaya & Absar, 2015). langkah-langkah dalam *text preprocessing* yaitu:

- 1) *Case Folding*, dilakukan untuk mengkonverensi teks dalam dokumen data menjadi *lower case*.
- 2) *Tokenizing*, dilakukan untuk memecah teks menjadi per kata.
- 3) *Stopword removal*, dilakukan penghilangan kata tidak begitu penting.
- 4) *Stemming*, dilakukan penghapusan imbuhan dan akhiran.

## 2.9. Pembobotan Kata

Pembobotan kata atau *Term Frequency-Inverse Document Frequency* (TF-IDF) adalah tahapan untuk memberikan nilai pada setiap kata

yang ada bertujuan untuk menemukan *frequency* kemunculan *term* dan dikalikan dengan *inverse* dari *frequency* dokumen (Rohman, Purwanto, & Santoso, 2018). Berikut ini rumus IDF pada persamaan (1)

$$idf = \log \left( \frac{N}{df_t} \right) \quad (1)$$

Berikut ini rumus TF-IDF pada persamaan (2)

$$tf - idf_{td} \times idf_t \quad (2)$$

## 2.10. Confusion Matrix

*Confusion Matrix* yaitu model dengan nilai fungsi untuk apakah klasifikasi pada metode yang dinilai memiliki label baik, buruk atau netral (Han, 2011). Berikut ini *confusion matrix* dengan tiga kelas sentimen seperti pada Tabel 1.

Tabel 1. *Confusion Matrix*

Kelas	Prediksi Positif	Prediksi Negatif	Prediksi Netral
Aktual Positif	True Positive (TP)	False Negative1 (FN1)	False Netral1 (FNt1)
Aktual Negatif	False Positive1 (FP1)	True Negative (TN)	False Netral2 (FNt2)
Aktual Netral	False Positive2 (FP2)	False Negative 2 (FN2)	True Netral (TNt)

## 2.10. K-Fold Cross Validation

*K-Fold Cross Validation* merupakan model validasi silang yang digunakan untuk memprediksi model dan memprediksi keakuratan model. Tujuan dilakukannya *K-Fold Cross Validation* yaitu untuk menghilangkan data yang bias (Tempola, Muhammad, & Khairan, 2018). Berikut ini contoh model 3 *K-Fold Cross Validation* pada Tabel 2.

Tabel 2. Model 3 *K-Fold Cross Validation*

1	2	3
1	2	3
1	2	3

Pada Tabel 2 terdapat kotak berwarna biru dimana merupakan data *testing* dan kotak putih merupakan data *training*. Data yang digunakan pada model *3-Fold Cross Validation* akan dieksekusi 3 kali dan setiap data subset memiliki kesempatan menjadi data testing dan data

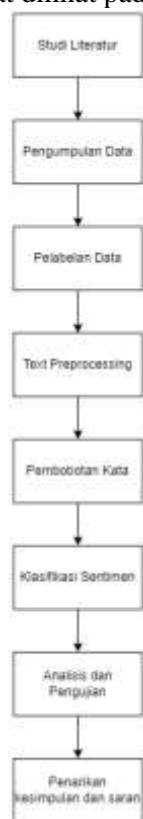


training.

### 3. METODOLOGI PENELITIAN

#### 3.1. Metodologi Penelitian

Pada penelitian ini menggunakan algoritma *random forest* dalam melakukan analisis sentimen terhadap ulasan aplikasi, Tahapan yang dilakukan yaitu studi literatur, pengumpulan data, pelabelan data, *text preprocessing*, pembobotan kata, klasifikasi sentimen, analisis dan pengujian, dan penarikan dan kesimpulan. Diagram alir dapat dilihat pada Gambar 1.



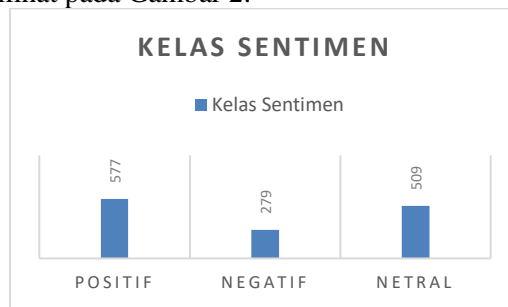
Gambar 1. Diagram Alir penelitian

### 4. ANALISIS DAN PENGUJIAN

#### 4.1. Analisis Deskriptif

Pengujian ini menggunakan ulasan aplikasi Dana dari *Google Play Store*. Hasil *Scraping* terdiri dari *rating* dan ulasan aplikasi dengan jumlah 1354 data. Penelitian ini menggunakan 3 kelas sentimen yaitu positif, negatif dan netral. Kelas sentimen positif berisi pujian, rasa puas, senang dan menyukai dengan jumlah 577 ulasan. Kelas sentimen negatif berisi keluhan terhadap aplikasi, protes, dan penilaian buruk dengan jumlah 279 ulasan. Kelas sentimen netral berisi saran, masukan dan pertanyaan dengan jumlah

509 ulasan. Hasil Klasifikasi sentimen dapat dilihat pada Gambar 2.



Gambar 2. Hasil Klasifikasi Sentimen

#### 4.2. Pengujian dan Analisis Parameter Terbaik

Pada pengujian ini menggunakan kedalaman *tree* dan jumlah *tree* dengan hasil terbaik, yaitu kedalaman 65 karena memiliki nilai *precision* 79%, *recall* 71%, dan *F1 score* 73% yang dimana lebih unggul. Pada Jumlah *Tree* akan diambil jumlah 400 *tree* karena memiliki nilai lebih unggul yaitu *precision* 79%, *recall* 71%, *F1-Score* 71%. Pengujian dilakukan dengan *K-Fold Cross Validation* dengan jumlah 5 *fold*. Hasil Pengujian parameter terbaik dapat dilihat pada Tabel 3.

Tabel 3. Hasil Pengujian Parameter Terbaik

Precision	Recall	F1-Score	Akurasi
0.77	0.68	0.67	0.76

Dari hasil pengujian parameter terbaik didapatkan nilai rata-rata *precision* yaitu 77%, rata-rata *recall* 68% rata-rata *F1-score* 67% dan rata-rata *accuracy* 76%. Berikut ini hasil pengujian parameter terbaik perkelasnya Tabel 4.

Tabel 4. Hasil Pengujian Parameter Terbaik Perkelas

Class	Precision	Recall	F1-Score	Akurasi
-1	0.83	0.32	0.46	0.77
0	0.62	0.83	0.71	
1	0.90	0.95	0.93	
Macro Average	0.78	0.70	0.70	

Berdasarkan Tabel 4 dapat dilihat jumlah *tree* 400 dan kedalaman 65 menghasilkan nilai *precision* pada kelas negatif yaitu 83%, kelas netral 62% dan kelas positif 90% hal ini berarti bahwa, tingkat kemampuan *system* dalam mencari ketepatan informasi pada proposisi label prediksi dibandingkan keseluruhan hasil yang diprediksi cukup baik.

Nilai *recall* pada kelas negatif mendapatkan

nilai dibawah 50% yaitu 32% berarti tingkat keberhasilan model dalam menemukan data yang benar *negative* rendah, sedangkan kelas netral mendapatkan nilai 83%, dan kelas positif mendapatkan nilai 95% yang berarti tingkat keberhasilan sistem dalam menemukan kembali data yang benar netral dan benar positif tinggi.

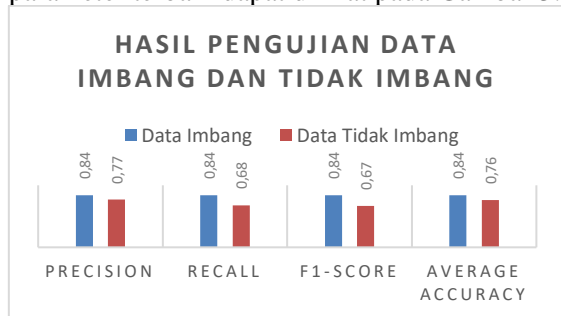
*F1-Score* merupakan perbandingan rata rata dari *presision* dan *recall* yang dibobotkan. Nilai *F1-Score* pada kelas negatif yaitu 46% termasuk dalam golongan rendah dikarenakan kelas negatif mendapatkan nilai *recall* yang rendah yaitu 32%, sedangkan kelas netral dan positif mendapatkan hasil yang cukup tinggi yaitu netral 71%, dan kelas positif 93%.

Nilai *accuracy* yang didapatkan dari hasil pengujian parameter terbaik yaitu kelas negatif, netral, dan positif bernilai 77% dimana merupakan rasio prediksi benar dari keseluruhan data.

Berdasarkan Tabel 4 didapatkan bahwa pada kelas *negative*, nilai *recall* dan *F1-Score* rendah yaitu 32% dan 41%. Ini dikarenakan data ulasan pada kelas *negative* lebih sedikit dibandingkan data kelas positif maupun netral yaitu 279 kelas *negative*, 577 kelas positif dan 509 kelas netral.

#### 4.3. Hasil Pengujian Data Imbang

Pada pengujian sebelumnya didapatkan nilai *recall* dan *F1-Score* pada kelas *negative* rendah yaitu 32% dan 41%. Berdasarkan nilai rendah yang didapatkan dilakukan pengujian dengan menyeimbangkan data kelas positif, *negative*, dan netral pada parameter terbaik yang didapatkan yaitu *tree* dengan kedalaman 65 dan jumlah *tree* 400 menjadi masing-masing berjumlah 250 data. Visualisasi Pengujian data imbang dan tidak imbang pada pengujian parameter terbaik dapat dilihat pada Gambar 3.



Gambar 3. Data Imbang dan Tidak Imbang

Berdasarkan Gambar 6.4 hasil pengujian data imbang dan tidak imbang didapatkan bahwa nilai *precision* mengalami kenaikan pada data

imbang menjadi 84% atau kenaikan sebesar 7%. Selanjutnya, Pada nilai *recall* data imbang mengalami kenaikan menjadi 84% atau kenaikan sebesar 16%. Nilai *F1-Score* juga mengalami kenaikan menjadi 84% atau kenaikan sebesar 17%. Terakhir, Pada nilai *accuracy* mengalami kenaikan sebesar 8% menjadi 84%. Hal ini menunjukkan bahwa kemampuan sistem mengidentifikasi data lebih baik saat data imbang dibandingkan dengan data yang tidak imbang yaitu dengan data positif, negatif, dan netral berjumlah 250 data.

#### 4.4. Analisis Sentimen Aplikasi Dana

Pada pengujian ini menggunakan ulasan aplikasi Dana yang telah dikelompokkan menjadi 3 kelas sentimen yaitu positif, *negative* dan netral bertujuan untuk mencari kata yang sering muncul. Dari 1354 data ulasan, ulasan yang mengandung sentiment positif yaitu 577 ulasan, sentimen negatif yaitu 279 ulasan dan sentimen netral yaitu 509 ulasan. Analisis ini dilakukan menggunakan *wordcloud* agar dapat menampilkan kata yang sering muncul pada ulasan. Pada kelas positif kata yang sering muncul dapat dilihat pada Gambar 4.



Gambar 4. Kata Positif yang Sering Muncul

Pada Gambar 4 didapatkan 10 kata yang sering muncul yaitu “bagus”, “mantap”, “sangat membantu”, “sangat bagus”, “mudah”, “membantu”, “transaksi”, “good”, “cepat”, dan “Dana”. Kata ini merupakan kata yang sering digunakan dalam ulasan pengguna aplikasi Dana di *google play store*. Hal ini mengartikan bahwa pengguna merasa puas terhadap pelayanan aplikasi Dana dikarenakan memberikan kemudahan, cepat dan membantu pengguna dalam bertransaksi.

Selanjutnya, kata yang sering muncul pada kelas netral dapat dilihat pada Gambar 5.



Gambar 5. Kata Netral yang Sering Muncul

Pada Gambar 5 terdapat 10 kata yang sering muncul yaitu “saldo”, “update”, “kenapa”, “tolong”, “masuk”, “transaksi”, “download”, “akun”, “uang”, “transfer”. Hal ini menunjukkan bahwa kata “saldo”, “update”, “transaksi”, “download”, “akun”, “uang”, “masuk” dan “transfer” menunjukkan bahwa pengguna sering bertanya atau memberikan saran dan masukan terkait 8 kata tersebut. Sedangkan kata “kenapa” merepresentasikan bahwa masih ada pengguna yang bertanya di kolom ulasan dan kata “tolong” mempresentasikan bahwa pengguna meminta bantuan pada pihak Dana untuk membantu menyelesaikan permasalahan.

Lalu, kata ulasan yang sering muncul pada kelas *Negative* dapat dilihat pada Gambar 6.



Gambar 6. Kata Negatif yang Sering Muncul

Pada Gambar 6 didapatkan 10 kata yang sering muncul yaitu “lemot”, “lama”, “saldo”, “update”, “susah”, “topup”, “transaksi”, “email”, “cs” dan “respon”. Kata “lemot”, “susah” dan “lama” menunjukkan bahwa pengguna merasa kesusahan dalam mengakses aplikasi Dana, kata “saldo”, “topup”, “transaksi” menunjukkan bahwa pengguna merasa kesusahan dalam mengisi saldo, top up dan melakukan transaksi, kata “email”, “cs”, dan “respon” menunjukkan bahwa kekecewaan pengguna karena lamanya respon dari customer service, kata “update” menunjukkan bahwa pengguna mengalami kesulitan dalam melakukan update aplikasi.

Berdasarkan analisis sentimen dapat disimpulkan bahwa Dana perlu meningkatkan

pelayanan *customer service* agar lebih tanggap dalam menindak lanjuti keluhan dari pelanggan. Dana juga perlu meningkatkan kemampuan aplikasi dalam melakukan topup dan transfer karena banyak mengalami gangguan. Selain itu, selalu melakukan upgrade aplikasi agar tidak “lemot” saat diakses.

## 5. PENUTUP

### 5.1. Kesimpulan

Berdasarkan Hasil penelitian mengenai analisis sentimen ulasan aplikasi Dana dengan metode *random forest*, dapat disimpulkan bahwa:

- 1) Pengambilan data terhadap ulasan aplikasi Dana diambil dari *Google Play Store* menggunakan teknik *web scraping* dan *tools jupyter notebook* dengan API *Google-Play-Scraper*. Total data yang diperoleh sebanyak 1354 data. Data yang dilakukan *scraping* yaitu *rating* dan ulasan pengguna aplikasi Dana.
- 2) *Text* yang dilakukan *preprocessing* pertama dilakukan pelabelan sentimen positif, negatif dan netral. Data yang sudah diberi label selanjutnya dilakukan *text preprocessing* untuk membersihkan data agar siap diteliti. Tools yang digunakan yaitu *google collab* dengan tahapan *case folding*, *remove number*, *remove punctuation*, *remove single char*, *remove whitespace*, *tokenization*, *stopwords*, *normalization*, dan *stemming*.
- 3) Nilai akurasi, *recall*, presisi, dan *f-measure* pada klasifikasi tiga kelas sentimen yaitu positif, negatif dan netral terhadap ulasan aplikasi Dana menggunakan metode *Random Forest* yaitu menghasilkan nilai yang berbeda disetiap pengujian kedalaman dan jumlah *tree*. Namun, dalam parameter terbaik dengan data yang diseimbangkan menjadi 250 data per kelasnya nilai yang didapatkan dari rasio data latih 80% dan data uji 20% yaitu *precision* 84%, *recall* 84%, *F1-Score* 84%, dan *accuracy* 84%. Hal ini menunjukkan, sistem dapat mengidentifikasi data dengan kedalaman *tree* 65 dan jumlah *tree* 400 dengan baik.
- 4) Hasil analisis sentimen pada kelas positif, negatif dan netral terhadap ulasan aplikasi Dana yaitu menggunakan metode *wordcloud* untuk mencari kata yang sering muncul dari 3 kelas sentimen yaitu positif, negatif dan netral. Pada kelas sentimen positif berjumlah

577 berisi pujian, rasa menyukai, senang dan puas, label negatif berjumlah 279 berisi keluhan, protes dan penilaian buruk dan label netral berjumlah 509 berisi saran, masukan, dan pertanyaan. Pada Label positif kata yang paling banyak muncul yaitu “bagus”, “mantap”, dan “sangat membantu”. Sedangkan pada Label negatif kata yang paling banyak keluar yaitu “lemot”, “topup”, dan “cs”. Kata pada label Netral yang paling banyak keluar yaitu “saldo”, “update”, dan “kenapa”. Dari kata yang sering muncul, dapat ditarik kesimpulan bahwa Dana perlu meningkatkan pelayanan *customer service* agar lebih tanggap dalam menindak lanjuti keluhan dari pelanggan. Dana juga perlu meningkatkan kemampuan aplikasi dalam melakukan *topup* dan transfer karena banyak mengalami gangguan. Selain itu, selalu melakukan upgrade aplikasi agar tidak “lemot” saat diakses.

## 5.2. Saran

Adapun saran penulis untuk penelitian selanjutnya yaitu menggunakan dataset yang lebih banyak agar hasil penelitian maksimal.

## 6. DAFTAR PUSTAKA

- Adminlp2m. (2022, Februari). Analisis Sentimen (Sentiment Analysis) : Definisi, Tipe dan Cara Kerjanya.
- Ayani, D. D., Pratiwi, H. S., & H. M. (2019). Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace. *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*.
- Bahrawi. (2019). Sentiment Analysis Using Random Forest Algorithm Online Social Media Base. *Journal of Information Technology and ITS Utilization*.
- Baskoro, B. B., Susanto, I., & Khomsah, S. (2021). Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR). *Journal of Informatics, Information System, Software Engineering and Application*, 021-029.
- Bimo, W. A. (2021). Penilaian Pengguna Dompot Digital Saat Pandemi Covid-19. *Moneter Jurnal Keuangan dan Perbankan*.
- Buulolo, E. (2020). *Data Mining Untuk Perguruan Tinggi*. Yogyakarta: CV Budi Utama.
- Damar, A. M. (2021, November 12). *liputan6.com*. Retrieved from *liputan6.com*: <https://www.liputan6.com/tekno/read/4709582/survei-dana-jadi-dompet-digital-dengan-pertumbuhan-jumlah-pengguna-paling-pesat>
- Dana. (2022). *Dana.id*. Retrieved from *Dana.id*: <https://www.dana.id/about>
- Findawati, Y., & Rosid, M. A. (2020). *Buku Ajar Text Mining*. Sidoarjo: UMSIDA PRESS.
- Fitri, E., Yuliani, Y., & Gata, W. (2020). Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine. *TRANSFORMATIKA*, 71-80.
- Flores, V. A., Permatasari, P. A., & Jasa, L. (2020). Penerapan Web Scraping Sebagai Media Pencarian dan Menyimpan Artikel Ilmiah Secara Otomatis Berdasarkan Keyword. *Majalah Ilmiah Teknologi Elektro*, 19.
- Ginting, S. A. (2021). Analisis Sentimen Dalam Evaluasi Kepuasan Pelanggan Terhadap Pelayanan Jasa Transportasi Online Gojek dan Grab.
- Han, J. K. (2011). *Data Mining Concepts and Techniques Third Edition*. Waltham: Elsevier Inc.
- Putri, E. K., & Setiadi, T. (2014). PENERAPAN TEXT MINING PADA SISTEM KLASIFIKASI EMAIL SPAM MENGGUNAKAN NAIVE BAYES. *Jurnal Sarjana Teknik Informatika*, 2.
- Rahardyan, A. (2021). *Bisnis.com*. Retrieved July 4, 2022, from <https://finansial.bisnis.com/read/20220103/563/1484554/riset-dompet-digital-terpopuler-ovo-dan-gopay-dominan>
- Richard, M. (2021). *Bisnis*. Retrieved January 2022, from <https://finansial.bisnis.com/read/20210319/90/1369582/bank-indonesia-klaim-uang-elektronik-semakin-dicintai>



- Rohman, T. B., Purwanto, D. D., & Santoso, J. (2018). Sentimen Analysis Terhadap Review Rumah Makan di Surabaya Memanfaatkan Algoritma Random Forest. *Seminar Nasional Teknologi Informasi dan Aplikasinya*.
- Sanjaya, S., & Absar, E. A. (2015). Pengelompokan Dokumen Menggunakan Winnowing Fingerprint. *Jurnal CoreIT*, 50-56.
- Tempola, F., Muhammad, M., & Khairan, A. (2018). PERBANDINGAN KLASIFIKASI ANTARA KNN DAN NAIVE BAYES PADA PENENTUAN STATUS GUNUNG BERAPI DENGAN K-FOLD CROSS VALIDATION. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 5, 577-584.
- Wahyudi, T., Indrajit, R. E., & M. F. (2017). PEMANFAATAN STATUS KREDIT NASABAH UNTUK MENGEVALUASI PEMBIAYAAN KPR PADA BANK MUAMALAT INDONESIA MENGGUNAKAN DATA MINING. *Seminar Nasional Sains dan Teknologi 2017*.
- Wikipedia. (2022). *id.wikipedia.org*. Retrieved from [id.wikipedia.org: https://id.wikipedia.org/wiki/Google\\_Play](https://id.wikipedia.org/wiki/Google_Play)
- Zub, D. (2022). *SerpApi*. Retrieved Juli 2022, from <https://serpapi.com/blog/scrape-google-play-store-app-in-python/>