

Seleksi Fitur *Information Gain* pada Analisis Sentimen Terhadap Ulasan Aplikasi Flip dengan Algoritma *Support Vector Machine*

Isma'il Muhammad Falih¹, Nur Hafifah Matondang², Nurul Chamidah³

Program Studi Informatika / Fakultas Ilmu Komputer

Universitas Pembangunan Nasional Veteran Jakarta

Jl. RS. Fatmawati, Pondok Labu, Jakarta Selatan, DKI Jakarta, 12450, Indonesia

imfalih@gmail.com¹, nurhafifahmatondang@upnvj.ac.id², nurul.chamidah@upnvj.ac.id³

Abstrak. Flip merupakan aplikasi untuk melakukan transfer antar bank berbeda tanpa dikenakan biaya serta menyediakan jasa pembelian pulsa serta paket data yang umumnya bekerja sebagai jembatan transaksi antar bank berbeda. Ulasan yang diberikan pengguna aplikasi Flip banyak berisi opini yang membangun maupun mengkritik yang dapat dijadikan masukan bagi pengembang aplikasi Flip. Tujuan dilakukannya penelitian ini adalah untuk membangun model klasifikasi sentimen menggunakan metode Support Vector Machine dan seleksi fitur Information Gain terhadap ulasan aplikasi Flip pada layanan Google Play. Dalam penelitian ini, ulasan akan dibagi menjadi dua kategori yakni positif dan negatif berdasarkan pelabelan manual oleh 3 penilai, yang kemudian dilakukan preprocessing, seleksi fitur, serta pembagian data sebesar 80% data train dan 20% data test sebelum pembuatan model. Terdapat dua model yakni model tanpa seleksi fitur (model SVM) dan model dengan seleksi fitur (model SVM-IG). Hasil evaluasi menunjukkan akurasi sebesar 91.97%, presisi sebesar 95.53%, recall sebesar 91.45%, dan AUC sebesar 0.9215 yang tergolong sangat baik untuk model SVM, sedangkan untuk model SVM-IG yakni akurasi sebesar 96.25%, presisi sebesar 99.10%, recall sebesar 94.87%, dan AUC sebesar 0.9672 yang tergolong sangat baik.

Kata Kunci: Analisis Sentimen, *Support Vector Machine*, Seleksi Fitur, *Information Gain*, Flip.

1 Pendahuluan

Tidak dapat dipungkiri bahwa transaksi nontunai ramai dilakukan akhir-akhir ini. Salah satu jenis transaksi nontunai yang ramai dilakukan adalah transfer antar bank menggunakan m-banking melalui smartphone, namun dalam setiap transfer antar bank melalui m-banking terdapat biaya administrasi yang dikenakan kepada pengirim. Terdapat sebuah aplikasi di bidang jasa keuangan yakni Flip yang dapat mengatasi permasalahan biaya administrasi tersebut. Solusi yang diberikan meningkatkan minat masyarakat dalam menggunakan aplikasi Flip.

Flip merupakan aplikasi untuk melakukan transfer antar bank berbeda tanpa dikenakan biaya. Aplikasi Flip juga menyediakan jasa pembelian pulsa serta paket data. Umumnya, Flip bekerja sebagai jembatan transaksi antar bank berbeda [1]. Pihak Flip mengungkapkan per bulan September 2021 mereka telah melayani sekitar 6 juta pengguna dengan jumlah transaksi yang tercatat mencapai hingga triliunan rupiah setiap bulannya [2]. Pada tanggal 23 April 2021, Forbes merilis daftar 30 Under 30 Asia 2021 atau daftar 30 anak muda yang memiliki terobosan di bawah 30 tahun yang berasal dari Asia untuk tahun 2021. Dari daftar tersebut, Flip masuk dalam kategori finance & venture capital atau keuangan dan modal ventura. Flip didirikan pada tahun 2015 oleh Rafi Putra Arriyan yang menjabat sebagai CEO bersama dengan kedua temannya, Ginanjar Ibnu Solikhin, dan Luqman Sungkar [3].

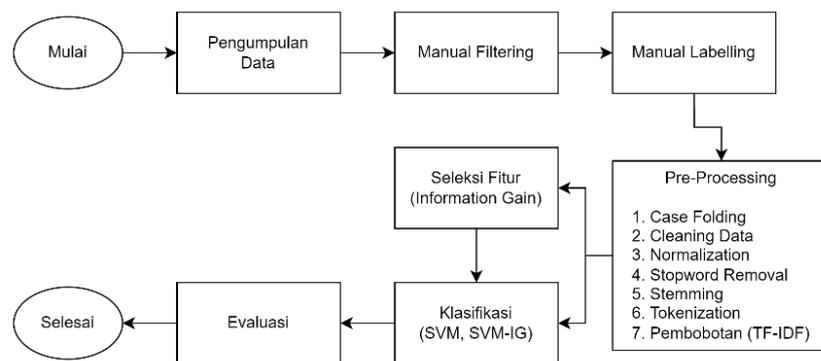
Umumnya ulasan aplikasi pada layanan Google Play yang diberikan pengguna aplikasi terkait setelah pengguna tersebut mengunduh dan menggunakan aplikasi pada ponsel pintar mereka [4]. Ulasan terhadap aplikasi yang diberikan pengguna dapat dijadikan masukan untuk pengembang dalam pengambilan keputusan, baik ulasan yang positif maupun negatif.

Terdapat berbagai metode yang dapat digunakan pada analisis sentimen, metode Support Vector Machine (SVM) dipilih karena pada penelitian yang telah dilakukan oleh Rokhman yang membandingkan algoritma SVM dengan Decision Tree pada analisis sentimen ulasan aplikasi Gojek menunjukkan bahwa akurasi SVM lebih tinggi dibandingkan Decision Tree yakni 90,2% dan 89,8% [5]. Kemudian pada tahun 2019, telah dilakukan penelitian oleh Sari yang melakukan implementasi seleksi fitur Information Gain (IG) pada klasifikasi ulasan pengguna aplikasi Mandiri Online menggunakan Naive Bayes (NB), menunjukkan bahwa IG dapat meningkatkan akurasi yang semula 93,33% pada klasifikasi NB (K-Fold = 9) menjadi 95,33% pada klasifikasi NB-IG (K-Fold = 8) dengan peningkatan sebesar 2% [6].

Pada penelitian ini, digunakan metode Support Vector Machine dengan menggunakan seleksi fitur Information Gain untuk menghindari Curse of Dimensionality pada SVM. Kedua metode yang digunakan telah dibuktikan pada penelitian sebelumnya memiliki hasil akurasi yang baik. Pada penelitian ini, dilakukan analisis sentimen terhadap ulasan aplikasi Flip yang terdapat pada layanan Google Play.

Penelitian ini bertujuan untuk membangun model klasifikasi sentimen menggunakan metode Support Vector Machine dan seleksi fitur Information Gain terhadap ulasan aplikasi Flip pada layanan Google Play dan mengetahui pengaruh penggunaan seleksi fitur Information Gain dan performa dari model yang dibuat.

2 Metodologi Penelitian



Gambar. 1. Penelitian ini terbagi menjadi beberapa tahapan yakni, pengumpulan data, *manual filtering*, *manual labelling*, *pre-processing*, seleksi fitur (*Information Gain*), klasifikasi (*Support Vector Machine*), dan evaluasi.

2.1 Pengumpulan Data

Pengumpulan data yang dilakukan dengan cara *scraping*. *Scraping data* dilakukan dengan bahasa pemrograman *python* dan memanfaatkan *library google-play-scraper* yang telah tersedia pada *python*. Data ulasan hasil proses *scraping* disimpan dengan format TSV yang kemudian dikonversi menjadi format CSV dengan berisi fitur ulasan.

2.2 Manual Filtering

Manual filtering dilakukan dengan menghapus data-data yang tidak relevan (tidak memiliki sentimen, tidak memiliki makna, dan tidak berhubungan dengan objek penelitian) serta menyalurkan bahasa menjadi bahasa Indonesia secara manual menggunakan Google Translate.

2.3 Manual Labelling

Pelabelan data ulasan dilakukan oleh 3 orang penilai secara manual berdasarkan sudut pandang masing-masing penilai dalam memaknai kalimat ulasan kedalam 2 kategori yakni positif (1) dan negatif (2). Penentuan label kelas positif atau negatif ditentukan pada pemungutan suara tertinggi atau mayoritas dari ketiga penilai.

2.4 Preprocessing

Preprocessing adalah tahap dilakukannya pembersihan data dan konversi teks agar memiliki standar sesuai dengan yang dibutuhkan [7]. Setelah data diberi label, harus dilakukan *preprocessing* atau praproses data untuk menyiapkan data yang masih mentah dan banyak *noise* menjadi data yang terstruktur dan bersih. *Preprocessing* terdiri dari 7 langkah yang secara terurut yakni *Case Folding*, *Cleaning Data*, *Normalization*, *Stopword Removal*, *Stemming*, *Tokenization*, dan Pembobotan (TF-IDF).

a) Case Folding

Case folding merupakan tahapan untuk menyamaratakan semua huruf pada data menjadi huruf kecil dengan tujuan untuk menghindari ketidakkonsistenan seperti *case sensitive* yang diakibatkan penggunaan kapitalisasi huruf yang tidak seragam [6].

b) Cleaning Data

Cleaning data atau pembersihan data dilakukan dengan membersihkan data ulasan dari hal-hal yang tidak berguna dan mengganggu. Hal-hal tersebut seperti angka, tanda baca, emoji, tagar, dan lain-lain guna mengurangi *noise* dan mencegah terjadinya masalah kualitas data. Data *noise* dapat berupa data tidak lengkap, data duplikat, data *outlier*, dan data tidak valid [8].

c) Normalization

Pada tahapan ini, dilakukan pembakuan kata sesuai dengan Kamus Besar Bahasa Indonesia (KBBI). Setiap kata pada data ulasan dinormalisasi dengan mengubah kata tidak baku, singkatan, serta *typo* (*typographical error* atau kesalahan ketik) menjadi kata baku.

d) Stopword Removal

Tahap *stopword removal* menghapus kata-kata gramatikal yang tidak relevan dari token seperti preposisi, konjungsi, dan lain-lain. Hal ini dikarenakan kata-kata tersebut dianggap tidak berpengaruh dan tidak penting serta akan mengganggu proses klasifikasi. Korpus yang digunakan merupakan kamus dari *library NLTk* yang ditambah dengan beberapa *stopword* tambahan.

e) Stemming

Proses *stemming* dilakukan agar kata pada data ulasan diubah menjadi kata dasar memanfaatkan *library sastrawi*. Proses *stemming* menghapus imbuhan atau afiks seperti me-, ber-, di-, -kan, -an, dan lain-lain.

f) Tokenization

Tokenization atau tokenisasi merupakan proses segmentasi teks dengan memanfaatkan spasi atau tanda baca. Masukan pada proses tokenisasi adalah data teks dan keluaran yang dikeluarkan adalah token-token kata [9].

g) Pembobotan (TF-IDF)

Pembobotan kata merupakan proses penentuan bobot suatu kata berdasarkan tingkat kepentingannya. TF-IDF (*Term Frequency-Inverse Document Frequency*) membobotkan kata dalam suatu teks sebanding dengan kemunculannya pada teks tersebut, namun berbanding terbalik dengan teks yang lain. Berikut adalah rumus dasar dari TF-IDF [9].

$$IDF_i = \log \frac{N}{DF_i} \tag{1}$$

$$w_{ij} = TF_{ij} \cdot IDF_i \tag{2}$$

Keterangan:

w_{ij} = Bobot TF-IDF kata i pada teks/ dokumen j .

N = Jumlah total teks/ dokumen.

DF_i = Jumlah teks/ dokumen dengan kata i didalamnya.

IDF_i = Rasio jumlah total teks/ dokumen dengan jumlah teks/ dokumen dengan kata i didalamnya.

TF_i = Jumlah kemunculan kata i pada teks/ dokumen j .

2.5 Seleksi Fitur *Information Gain*

Seleksi fitur merupakan suatu proses pemilihan fitur yang paling penting, signifikan, relevan, berguna atau berkontribusi tinggi dari kumpulan data dengan tujuan mereduksi dimensi. *Information Gain* merupakan teknik pengukuran perubahan nilai entropi setelah dilakukan segmentasi pada dataset berdasarkan fitur. *Information Gain* menghitung seberapa banyak informasi relevan yang disediakan oleh fitur mengenai kelas atau label pada suatu dataset. Entropi sendiri ialah sebuah metrik pengukuran yang menentukan ketidakteraturan dalam suatu data dengan mengukur ketidakhomogenitas pada fitur tertentu [8]. Berikut adalah rumus-rumus yang digunakan pada seleksi fitur *Information Gain* [6].

- a) Rumus entropi record dan total.

$$Ent(S) = \sum_{i=1}^m -p_i \cdot \log_2 p_i \quad (3)$$

- b) Rumus entropi fitur.

$$Ent_F(S) = \sum_{j=1}^n -p_j \cdot Ent(S_j) \quad (4)$$

- c) Rumus Information Gain.

$$Gain(S, F) = Ent(S) - |Ent_F(F)| \quad (5)$$

Keterangan:

S = Sampel (record atau total).

F = Fitur.

m = Jumlah maksimum fitur.

n = Jumlah nilai pada kelas.

p_i = Rasio jumlah sampel di kelas i terhadap total sampel.

p_j = Rasio total sampel j terhadap total sampel pada fitur.

$Ent(S_j)$ = Nilai entropi untuk sampel j .

2.6 Klasifikasi *Support Vector Machine (SVM)*

Support Vector Machine (SVM) merupakan metode klasifikasi yang dapat digunakan untuk klasifikasi data linier maupun non linier dengan asumsi bahwa apabila dua kelas data tidak dapat dipisahkan oleh *hyperplane* secara linier, maka *hyperplane* yang optimal didapatkan setelah data dipetakan pada ruang dimensi yang lebih tinggi [10]. Pada SVM terdapat fungsi kernel, fungsi ini bertujuan untuk memetakan ruang vektor yang pada dasarnya tidak dapat dipisahkan secara linier menjadi ruang vektor yang dapat dipisahkan secara linier [9]. Berdasarkan uraian tersebut, *Support Vector machine (SVM)* adalah algoritma klasifikasi yang memanfaatkan pembatas antar dua kelas atau *hyperplane*. Untuk menemukan *hyperplane* pada data non linier diperlukan penggunaan fungsi kernel pada saat memetakan ruang vektor yang lebih tinggi atau biasa disebut *kernel trick*. Fungsi kernel yang umum digunakan yakni kernel *linear* (dasar), kernel *polynomial*, kernel *radial basis function (RBF)*, dan kernel *sigmoid*.

2.7 Evaluasi

Evaluasi atau pengujian dilakukan pada data *testing* menggunakan model klasifikasi yang telah dibuat dan direpresentasikan sebagai *confusion matrix*. Performa model klasifikasi yang dibuat dapat dilihat dari nilai akurasi, nilai presisi, nilai *recall* hingga nilai AUC (*Area Under Curve*) [11]. Terdapat pedoman dalam menggunakan nilai AUC yang dibagi menjadi 5 kategori yakni, 0.50-0.60 sebagai gagal, 0.60-0.70 sebagai buruk, 0.70-0.80 sebagai cukup, 0.80-0.90 sebagai baik, 0.90-1.00 sebagai sangat baik [12].

3 Hasil dan Pembahasan

3.1 Pengumpulan Data

| | A |
|----|--|
| 1 | Aplikasinya sangat membantu...sukses selalu |
| 2 | Saya sangat suka membantu sekali buat yg wirausaha seperti saya |
| 3 | aplikasi transfer uang tanpa biaya admin. |
| 4 | Tlong dong uang saya di kembalikan, dana refundnya ðŸŽµ masa cara kerja smpe 1x24 jam .. aplikasi tidak sesuai sama realita di iklan.. Hoaks |
| 5 | Lebih bagus lagi kalau batas transaksi per hari nya ditambah ya. Jgn cuma 5 juta. Alias bs tetap gratis walau transaksi di atas 5 juta. |
| 6 | Sejauh ini masih bagus aplikasinya |
| 7 | Auto hapus, saya sudah transfer ke Flip buat kirim ke Dana tapi gagal padahal di SeaBank sudah kepotong... Mending pake aplikasi Ovo atau Dana aja... Ga jelas aplikasinya kecewa ðŸ”E |
| 8 | Maaf utk situasi saat ini agak lumayan kecewa karena fitur transfer via website sudah gabisa musti unduh aplikasi sepertinya sudah bnyak untung |
| 9 | nambah untung dari aplikasi |
| 10 | Keren, bisa hemat banyak |
| 10 | 22a we a12 we we 12 we 22aaaaaa . Ccx |

Gambar. 2. Data ulasan diambil pada tanggal 14 Oktober 2021 dan berjumlah 1000 data terbaru pada aplikasi Flip versi 1.19.29 pada layanan Google Play. Data ulasan yang disimpan hanya berisikan kalimat ulasan atau *content*.

3.2 Manual Filtering

Manual filtering berhasil menyeleksi sebanyak 65 data yang tidak relevan dari 1000 data ulasan dan menghasilkan 935 data ulasan bersih. Proses dari manual filtering dapat dilihat pada Tabel 1.

Tabel 1. Proses *Manual Filtering*

| Sebelum | Sesudah |
|---------------------------------------|---------------------------------------|
| AðŸ”EðŸ”EðŸ”EðŸ”EðŸ”E | (data dihapus) |
| 22a we a12 we we 12 we 22aaaaaa . Ccx | (data dihapus) |
| Aplikasi yang luar biasa. Good job | Aplikasi yang luar biasa. Kerja bagus |

3.3 Manual Labelling

Manual labelling menghasilkan sebanyak 585 data ulasan positif dan 350 data ulasan negatif dari total 935 data. Penentuan label kelas positif atau negatif ditentukan melalui pemungutan suara tertinggi atau mayoritas dari ketiga penilai. Sentimen pengguna terhadap aplikasi Flip sebesar 63% sentimen positif dan 37% sentimen negatif. Proses dari *manual labelling* dapat dilihat pada Tabel 2.

Tabel 2. Proses *Manual Labelling*

| Data Ulasan | Penilai 1 | Penilai 2 | Penilai 3 | Hasil Akhir |
|--|-----------|-----------|-----------|-------------|
| Sangat membantu dlm pengiriman beda rekening.... #suksesselalu | Positif | Positif | Positif | Positif |
| Membantu dalam hal transfer uang antar bank | Positif | Positif | Positif | Positif |
| lambat prosesnya! | Negatif | Negatif | Negatif | Negatif |
| Sangat membantu sekali dalam hal tranfers antar bank. | Positif | Positif | Positif | Positif |
| Lama transfernya ribet | Negatif | Negatif | Negatif | Negatif |

3.4 Preprocessing

Preprocessing terdiri dari 7 langkah yang secara terurut yakni *Case Folding*, *Cleaning Data*, *Normalization*, *Stopword Removal*, *Stemming*, *Tokenization*, dan Pembobotan (TF-IDF).

a) **Case Folding**

Tahap *case folding* merubah kapitalisasi seluruh data ulasan menjadi huruf kecil (*lowercase*) sepenuhnya untuk menghindari terjadinya *case sensitive*. Proses *case folding* ditunjukkan pada Tabel 3.

Tabel 3. Proses *Case Folding*

| Data | Masukan | Keluaran |
|------|--|--|
| D1 | Sangat membantu dlm pengiriman beda rekening.... #suksesselalu | sangat membantu dlm pengiriman beda rekening.... #suksesselalu |
| D2 | Membantu dalam hal transfer uang antar bank | membantu dalam hal transfer uang antar bank |
| D3 | lambat prosesnya! | lambat prosesnya! |
| D4 | Sangat membantu sekali dalam hal tranfers antar bank. | sangat membantu sekali dalam hal tranfers antar bank. |
| D5 | Lama transfernya ribet | lama transfernya ribet |

b) **Cleaning Data**

Pada tahap ini dilakukan pembersihan data ulasan dari hal-hal yang tidak berguna dan mengganggu seperti angka, tanda baca, emoji, tagar (#), dan lain-lain. Hal ini dilakukan untuk mengurangi *noise* dan mencegah terjadinya masalah kualitas data. Proses *cleaning data* ditunjukkan pada Tabel 4.

Tabel 4. Proses *Cleaning Data*

| Data | Masukan | Keluaran |
|------|--|--|
| D1 | sangat membantu dlm pengiriman beda rekening.... #suksesselalu | sangat membantu dlm pengiriman beda rekening |
| D2 | membantu dalam hal transfer uang antar bank | membantu dalam hal transfer uang antar bank |
| D3 | lambat prosesnya! | lambat prosesnya |
| D4 | sangat membantu sekali dalam hal tranfers antar bank. | sangat membantu sekali dalam hal tranfers antar bank |
| D5 | lama transfernya ribet | lama transfernya ribet |

c) **Normalization**

Pembakuan kata dilakukan pada tahap ini agar sesuai dengan Kamus Besar Bahasa Indonesia (KBBI), kata yang diubah meliputi kata tidak baku, singkatan, dan *typo* (*typographical error* atau kesalahan ketik). Proses *Normalization* ditunjukkan pada Tabel 5.

Tabel 5. Proses *Normalization*

| Data | Masukan | Keluaran |
|------|--|--|
| D1 | sangat membantu dlm pengiriman beda rekening | sangat membantu dalam pengiriman beda rekening |
| D2 | membantu dalam hal transfer uang antar bank | membantu dalam hal transfer uang antar bank |
| D3 | lambat prosesnya | lambat prosesnya |
| D4 | sangat membantu sekali dalam hal tranfers antar bank | sangat membantu sekali dalam hal transfer antar bank |
| D5 | lama transfernya ribet | lama transfernya susah |

d) **Stopword Removal**

Proses *Stopword Removal* merupakan tahapan yang bekerja menghapus kata-kata gramatikal yang tidak relevan dari data ulasan seperti preposisi, konjungsi, dan lain-lain. Proses *Stopword Removal* ditunjukkan pada Tabel 6.

Tabel 6. Proses *Stopword Removal*

| Data | Masukan | Keluaran |
|------|---------|----------|
|------|---------|----------|

| | | |
|----|--|-----------------------------------|
| D1 | sangat membantu dalam pengiriman beda rekening | membantu pengiriman beda rekening |
| D2 | membantu dalam hal transfer uang antar bank | membantu transfer uang bank |
| D3 | lambat prosesnya | lambat prosesnya |
| D4 | sangat membantu sekali dalam hal transfer antar bank | membantu transfer bank |
| D5 | lama transfernya susah | transfernya susah |

e) *Stemming*

Proses ini mengubah kata pada data ulasan menjadi kata dasar dengan menghapus imbuhan atau afiks (me-, ber-, -kan, -an, dan lain-lain). Proses *Stemming* ditunjukkan pada Tabel 7.

Tabel 7. Proses *Stemming*

| Data | Masukan | Keluaran |
|------|-----------------------------------|---------------------------|
| D1 | membantu pengiriman beda rekening | bantu kirim beda rekening |
| D2 | membantu transfer uang bank | bantu transfer uang bank |
| D3 | lambat prosesnya | lambat proses |
| D4 | membantu transfer bank | bantu transfer bank |
| D5 | transferya susah | transfer susah |

f) *Tokenization*

Proses ini merupakan proses pemisahan kata dari kalimat sehingga menghasilkan token-token kata. Proses *Tokenization* ditunjukkan pada Tabel 8.

Tabel 8. Proses *Tokenization*

| Data | Masukan | Keluaran |
|------|---------------------------|--|
| D1 | bantu kirim beda rekening | ['bantu', 'kirim', 'beda', 'rekening'] |
| D2 | bantu transfer uang bank | ['bantu', 'transfer', 'uang', 'bank'] |
| D3 | lambat proses | ['lambat', 'proses'] |
| D4 | bantu transfer bank | ['bantu', 'transfer', 'bank'] |
| D5 | transfer susah | ['transfer', 'susah'] |

g) *Pembobotan (TF-IDF)*

Pembobotan dilakukan terhadap 805 kata dari jumlah dokumen sebanyak 935 ulasan menggunakan persamaan (1) dan (2) memanfaatkan *library TfidfVectorizer* pada *python*, sehingga seluruh kata memiliki bobot agar dapat digunakan pada proses seleksi fitur dan klasifikasi.

3.5 Seleksi Fitur *Information Gain*

Seleksi fitur *information gain* dilakukan untuk model yang menggunakan seleksi fitur dengan menggunakan persamaan (3), (4), (5). Terdapat 3 model yang menggunakan seleksi fitur yakni model SVM-IG >0.001, model SVM-IG >0.002, dan model SVM-IG >0.003 yang masing-masing memiliki jumlah kata 698 kata untuk batas '>0.001', 313 kata untuk batas '>0.002', dan 237 kata untuk batas '>0.003' yang nantinya akan digunakan salah satu yang memiliki performa tertinggi untuk dibandingkan dengan model tanpa seleksi fitur.

3.6 Klasifikasi *Support Vector Machine*

Sebelum proses klasifikasi dilakukan pembagian data sebesar 80% data latih dan 20% data uji dari total sebanyak 935 data ulasan. Pemodelan dengan SVM diterapkan terhadap 748 data latih dan menghasilkan sebanyak 4 model yakni model SVM, model SVM-IG >0.001, model SVM-IG >0.002, dan model SVM-IG >0.003 dan untuk menentukan parameter yang digunakan masing-masing model memanfaatkan *library GridSearchCV* pada *python*. Parameter setiap model ditunjukkan pada Tabel 9.

Tabel 9. Parameter Setiap Model

| Model | | Jumlah Fitur | Parameter SVM yang digunakan | | |
|--------|--------|--------------|------------------------------|-----|-------|
| | | | Kernel | C | Gamma |
| SVM | - | 805 | RBF | 5 | 1 |
| SVM-IG | >0.001 | 798 | RBF | 5 | 1 |
| | >0.002 | 313 | RBF | 0.5 | 1 |
| | >0.003 | 237 | RBF | 5 | 0.1 |

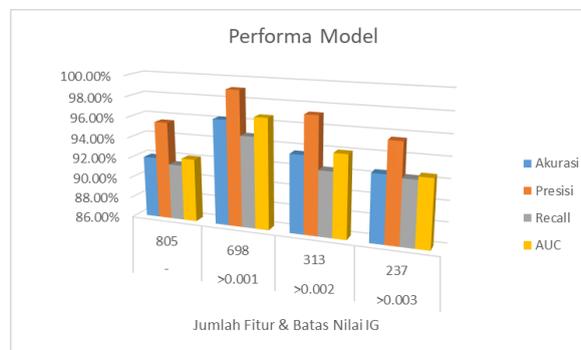
3.7 Evaluasi

Evaluasi dilakukan terhadap 187 data uji menggunakan hasil pemodelan pada proses klasifikasi. Evaluasi setiap model ditunjukkan pada Tabel 10.

Tabel 10. Parameter Setiap Model

| Model | | Jumlah Fitur | Pengukuran Performa | | | |
|--------|--------|--------------|---------------------|---------|--------|--------|
| | | | Akurasi | Presisi | Recall | AUC |
| SVM | - | 805 | 91.97% | 95.53% | 91.45% | 0.9215 |
| SVM-IG | >0.001 | 798 | 96.25% | 99.10% | 94.87% | 0.9672 |
| | >0.002 | 313 | 93.58% | 97.29% | 92.30% | 0.9401 |
| | >0.003 | 237 | 92.51% | 95.57% | 92.30% | 0.9258 |

Berdasarkan Tabel 10 dapat dilihat bahwa pada kasus analisis sentimen ulasan aplikasi Flip, model dengan seleksi fitur yang memiliki performa tertinggi adalah model SVM-IG >0.001, maka dari itu model SVM-IG >0.001 digunakan sebagai pembandingan dengan model tanpa seleksi fitur (model SVM).



Gambar 3. Gambaran performa keempat model yang telah dibuat dalam bentuk grafik.

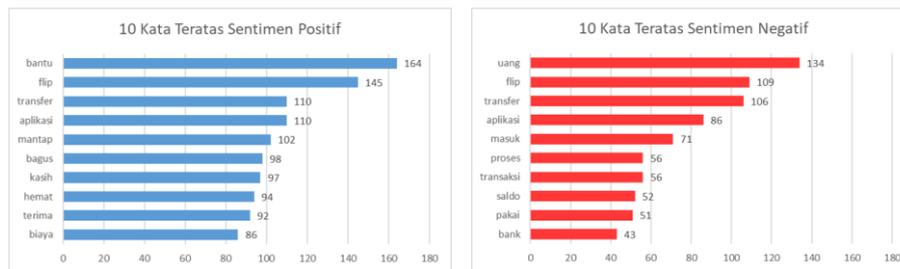
3.8 Wordcloud



Gambar 4. Hasil visualisasi *wordcloud* untuk sentimen positif (biru) dan negatif (merah) dari data ulasan aplikasi Flip.

Kata-kata pada ulasan positif ('bantu', 'mantap', 'terima', 'kasih', dan lain-lain) menyiratkan bahwa aplikasi Flip sangat berguna dan memberikan kepuasan terhadap penggunaannya. Sedangkan kata-kata pada ulasan negatif ('uang', 'transfer', 'transaksi', 'masuk', dan lain-lain) menyiratkan bahwa aplikasi Flip memberikan kekecewaan dan dalam proses transfer/ transaksinya dikeluhkan oleh beberapa penggunaannya.

3.9 Frekuensi Kata



Gambar. 5. Hasil visualisasi Frekuensi kata untuk sentimen positif (biru) dan negatif (merah) dari data ulasan aplikasi Flip.

Kata yang paling sering muncul pada sentimen positif ulasan aplikasi Flip adalah ‘bantu’ sebanyak 164 kali kemunculan, sedangkan kata yang paling sering muncul pada sentimen positif ulasan aplikasi Flip adalah ‘uang’ sebanyak 134 kali kemunculan.

4 Kesimpulan dan Saran

4.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, termasuk hasil analisis dan pembahasan, berikut merupakan beberapa hal yang dapat disimpulkan:

1. Model klasifikasi dibangun setelah data ulasan aplikasi Flip versi 1.19.29 yang diperoleh pada tanggal 14 Oktober 2021 melalui tahap manual filtering untuk menghilangkan noise, manual labelling untuk memberi label pada data ulasan, preprocessing untuk membersihkan data dan memberi bobot pada data ulasan, serta melalui tahapan seleksi fitur Information Gain (IG) bagi model SVM-IG. Pembagian data dibagi menjadi 80% data train dan 20% data test yang kemudian diklasifikasikan menggunakan kernel Radial Basis Function (RBF) pada Support Vector Machine (SVM).
2. Hasil evaluasi menunjukkan performa untuk model tanpa seleksi fitur (model SVM) yang optimal diperoleh pada klasifikasi SVM dengan kernel RBF, nilai $C = 5$, nilai $\Gamma = 1$, dengan jumlah fitur sebanyak 805 fitur. Performa dari model SVM memperoleh akurasi sebesar 91.97%, presisi sebesar 95.53%, recall sebesar 91.45%, dan nilai AUC sebesar 0.9215 yang tergolong sangat baik. Sedangkan untuk hasil evaluasi performa model dengan seleksi fitur (model SVM-IG >0.001) yang optimal diperoleh pada klasifikasi SVM-IG dengan kernel RBF, nilai $C = 5$, nilai $\Gamma = 1$, dengan jumlah fitur sebanyak 698 fitur. Performa dari model SVM memperoleh akurasi sebesar 96.25%, presisi sebesar 99.10%, recall sebesar 94.87%, dan nilai AUC sebesar 0.9672 yang tergolong sangat baik.
3. Penggunaan seleksi fitur Information Gain dengan parameter batas nilai $IG >0.001$ pada model yang telah dibuat mengurangi fitur sebanyak 107 fitur dan menunjukkan peningkatan performa yang ditandai dengan meningkatnya akurasi sebesar 4.28%, presisi sebesar 3.57%, recall sebesar 3.42% dan nilai AUC sebesar 0.0457.

4.2 Saran

Berdasarkan penelitian yang telah dilakukan, termasuk hasil analisis dan pembahasan, dapat diberikan saran yakni menggunakan metode klasifikasi dan seleksi fitur lain, serta menggunakan dataset aplikasi Flip versi terbaru atau aplikasi sejenis.

Referensi

- [1] Flip.id, “Transfer Dana Tanpa Biaya.” <https://flip.id/> (accessed Oct. 13, 2021).
- [2] L. D. Jatmiko, “Flip Klaim Raih 6 Juta Pengguna hingga September 2021,” Sep. 28, 2021. <https://teknologi.bisnis.com/read/20210928/266/1448009/flip-klaim-raih-6-juta-pengguna-hingga-september-2021> (accessed Oct. 13, 2021).
- [3] V. Ratriani, “Mengenal Flip, aplikasi transfer uang gratis antar bank di Forbes 30 under 30 Asia,” Apr. 26, 2021. <https://keuangan.kontan.co.id/news/mengenal-flip-aplikasi-transfer-uang-gratis-antar-bank-di-forbes-30-under-30->

- asia (accessed Oct. 13, 2021).
- [4] S. Watmah, Suryanto, and Martias, “Komparasi Metode K-NN, Support Vector Machine, Dan Random Forest Pada E-Commerce Shopee,” *Insa - J. Inov. dan Sains Tek. Elektro*, vol. 2, no. 1, pp. 15–21, 2021.
 - [5] K. A. Rokhman, Berlilana, and P. Arsi, “Perbandingan Metode Support Vector Machine Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online,” *JOISM J. Inf. Syst. Manag.*, vol. 2, no. 2, pp. 1–7, 2021.
 - [6] A. E. Sari, S. Widowati, and K. M. Lhaksana, “Klasifikasi Ulasan Pengguna Aplikasi Mandiri Online di Google Play Store dengan Menggunakan Metode Information Gain dan Naive Bayes Classifier,” 2019, vol. 6, pp. 9143–9157.
 - [7] F. A. Nugraha, N. Harani, and R. Habibi, *Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning*. Bandung: Kreatif Industri Nusantara, 2020. [Online]. Available: https://books.google.co.id/books?id=f738DwAAQBAJ&hl=id&source=gbs_navlinks_s
 - [8] P. Gupta and N. K. Sehgal, *Introduction to Machine Learning in the Cloud with Python: Concepts and Practices*. Springer International Publishing, 2021. doi: 10.1007/978-3-030-71270-9.
 - [9] T. Jo, *Text Mining: Concepts, Implementation, and Big Data Challenge*. Berlin: Springer International Publishing, 2019. doi: <https://doi.org/10.1007/978-3-319-91815-0>.
 - [10] B. Makhabel, P. Mishra, N. Danneman, and R. Heimann, *R: Mining Spatial, Text, Web, and Social Media Data*. Birmingham: Packt Publishing, 2017. [Online]. Available: https://www.google.co.id/books/edition/R_Mining_spatial_text_web_and_social_med/HHg5DwAAQBAJ?hl=en&gbpv=0
 - [11] Riyanto and A. Azis, “Application of the Vector Machine Support Method in Twitter Social Media Sentiment Analysis Regarding the Covid-19 Vaccine Issue in Indonesia,” *J. Appl. Data Sci.*, vol. 2, no. 3, pp. 102–108, 2021.
 - [12] C. Steven and Wella, “The Right Sentiment Analysis Method of Indonesian Tourism in Social Media Twitter Case Study: The City of Bali,” *IJNMT (International J. New Media Technol.)*, vol. 7, no. 2, pp. 102–110, 2020, doi: <https://doi.org/10.31937/ijnmt.v7i2.1732>.