

IMPLEMENTASI TEXT MINING PADA ULASAN APLIKASI FLIP DENGAN KLASIFIKASI SUPPORT VECTOR MACHINE

Aqiela Fasya Kamila¹, Corrina Angelina Putri², Hanifa Adinda Putri³

Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Brawijaya, Malang, Indonesia

Abstrak

Beberapa tahun terakhir, terdapat aplikasi yang membantu transaksi keuangan tanpa dikenakan biaya admin, salah satunya adalah aplikasi Flip. Tetapi, ketika mencapai batas maksimal transaksi, pengguna Flip akan tetap dikenakan biaya admin kembali. Atas hal tersebut, para pengguna Flip menuliskan ulasan terkait aplikasi pada *google play* guna meningkatkan pelayanan dalam bertransaksi. Data ulasan diambil dari *google play* dengan metode *crawling data* sebanyak 100 ulasan terbaru sejak tanggal 19 Maret 2023 hingga 21 Maret 2023. Kemudian, dilakukan tiga tahap *preprocessing data*, yaitu tahap *data tokenization* yang berfungsi untuk menghilangkan tanda baca, angka, karakter, mengubah huruf yang tidak baku menjadi baku, dan mengubah huruf besar menjadi huruf kecil, tahap *stemming* yang berfungsi untuk menghapus kata imbuhan diawal dan diakhir kata sehingga menghasilkan kata dasar, dan tahap *stopword removal* yang berfungsi untuk menghapus kata hubung dan kata yang tidak relevan. Setelah melalui proses *preprocessing data*, data ulasan tersebut dibentuk label guna mengetahui variabel x dan y . Hasil dari pelabelan tersebut akan digunakan sebagai *input* untuk proses klasifikasi model SVM. Hasil model SVM menunjukkan akurasi 0.70, artinya model SVM dapat mengklasifikasikan data dengan benar sebanyak 70%.

Kata kunci: Flip, Klasifikasi, *Text Mining*, *Support Vector Machine*

PENDAHULUAN

Latar Belakang

Penggunaan *mobile banking* atau e-wallet sebagai wadah transaksi secara online menjadi pilihan yang tepat karena mudah, cepat, efisien, dan efektif. Namun, penggunaan *mobile banking* atau e-wallet memiliki biaya admin yang harus dikeluarkan jika transaksi yang dilakukan tidak dengan bank yang sama. Sehingga, masyarakat memilih untuk menggunakan aplikasi tambahan, yaitu aplikasi Flip. Aplikasi Flip digunakan sebagai jembatan transaksi antar bank cukup mudah, hal yang perlu dilakukan hanya melakukan transfer terlebih dahulu ke rekening Flip sesuai dengan bank yang dimiliki, lalu Flip akan membantu meneruskan transaksi ke rekening tujuan.

Namun, aplikasi ini masih memiliki keterbatasan dalam transaksi bebas biaya admin. Ketika mencapai batas maksimal transaksi, pengguna Flip akan tetap dikenakan biaya admin kembali. Selain itu, tidak sedikit pengguna yang mengalami kendala eror pada aplikasi ini. Atas hal tersebut, para pengguna Flip menuliskan ulasan terkait aplikasi pada *google play* guna meningkatkan pelayanan dalam bertransaksi.

Text mining adalah salah satu penambang informasi dari data yang berupa tulisan, dokumen atau text dalam bentuk klasifikasi maupun *clustering*. Tujuan dilakukan penelitian ini adalah untuk menganalisis ulasan positif dan negatif dari aplikasi flip yang kemudian diklasifikasikan menggunakan *Support Vector Machine* (SVM). Batasan masalah pada penelitian ini adalah data yang diambil

sebanyak 100 ulasan terbaru melalui google play dengan pengguna yang berada di Indonesia dan analisis menggunakan bahasa pemrograman Python.

Text Mining

Menurut Sunoto & Wasito (2014), *text mining* adalah salah satu teknik penambang informasi dari data yang berupa tulisan, dokumen atau text dalam bentuk klasifikasi maupun *clustering*. *Text mining* merupakan bagian dari data *mining* dimana akan memproses kata-kata serta dokumen-dokumen yang bisa jadi dalam jumlah sangat besar. Untuk memproses data yang sangat besar akan memakan sumber daya yang tidak sedikit kaitannya dengan pengolahan data tersebut. Disinilah diperlukannya sebuah pemrosesan awal atau *preprocessing* data *text* tersebut sebelum data tersebut dilakukan proses *text mining* sesuai algoritma yang akan diterapkan.

Crawling Data (Pengambilan Data)

Crawling data merupakan sebuah metode untuk mengumpulkan data informasi yang diinginkan secara otomatis yang bersumber dari *website* (Udapure *et al.*, 2014). Dengan menggunakan data hasil *crawling*, data bisa diproses sesuai kebutuhan.

Preprocessing

Preprocessing adalah tahap sebelum memasuki analisis lebih lanjut. Tujuannya adalah untuk menghilangkan teks dari informasi yang tidak diperlukan atau *noise*. Pada analisis *text mining*, terdapat tiga tahap dalam *preprocessing text* yaitu (Sunoto & Wasito, 2014):

1. Tokenization

Proses *Tokenization* ini terbagi menjadi 3 tahapan, yaitu:

1. Cleansing

Proses ini berguna untuk menghilangkan data ulasan dari kata atau kalimat yang tidak diperlukan seperti tanda baca, angka, dan karakter pada ulasan. Contoh pada *cleansing*, ketika ada kata “makan2” akan berubah menjadi “makan”, kata “bagaimana?” berubah menjadi “bagaimana”, dan kata “bagus 🍑” berubah menjadi “bagus”.

2. Normalization

Proses ini berguna untuk memperbaiki ejaan kata sesuai dengan kamus slang yang telah dibuat serta menerjemahkan bahasa asing menjadi Bahasa Indonesia. Perbaikan ini membantu untuk mengubah kata yang tidak baku menjadi kata baku. Contoh pada *normalization*, Ketika terdapat kata tidak baku “pake” akan diubah menggunakan *slang* menjadi “pakai”, kata asing “download” akan diubah menggunakan slang menjadi “unduh”.

3. Case Folding

Proses ini berguna untuk mengubah huruf kapital menjadi huruf kecil. Jika tidak dilakukan *case folding*, kata “BAGUS”, “Bagus”, dan “bagus” akan dianggap menjadi tiga

kata yang berbeda. Contoh pada tahap *case folding*, Ketika ada kata “SEMANGAT” akan diubah menjadi “semangat”.

2. *Stemming*

Stem memiliki arti yaitu akar kata, sehingga proses *stemming* berguna untuk menghapus kata imbuhan dari setiap kata, baik kata imbuhan yang berada di depan maupun di belakang. Proses ini membantu untuk menyelaraskan kata, jika terdapat kata “peranan” dan “berperan”, akan menjadi dua kata yang berbeda sehingga diperlukan proses *stemming* untuk mengubah dua kata tersebut menjadi hanya satu kata, yaitu “peran”. Contoh kata imbuhan adalah me-, pe-, ber-, -in, -eh, -an, -nya, -kah, per-an, ber-an, me-kan, dan masih banyak lagi.

3. *Stopword Removal*

Proses *stopword removal* berguna untuk menghapus kata *stopword* atau kata penghubung seperti yang, dan, atau, tetapi, dan sebagainya. Selain kata penghubung, kata *stopword* juga termasuk pada kata ganti waktu, *key*, kata ganti orang, kata kerja, jumlah. *Stopword* adalah kata-kata dengan tata bahasa yang tidak relevan dengan konteks teks, sehingga perlu dihapus agar lebih efisien. *Key* merupakan topik dari data yang diunduh, perlu dihapus karena *key* akan menjadi jumlah kata yang paling sering disebut sehingga tahap klasifikasi tidak akan merepresentasikan data asli. Contoh pada proses *stopword removal*, ketika terdapat kata “saya” “sangat” “sudah”, maka semua kata tersebut dihapus.

Pelabelan

Pelabelan adalah tahap memberikan label pada data yaitu ulasan di Flip. Proses ini digunakan untuk memisahkan antara variabel *x* dan *y*. Jenis data komentar yang diambil akan diubah menjadi dataset yang kemudian akan dilakukan pelabelan pada setiap baris ulasan data tersebut apakah kalimat itu positif, negatif atau netral. Cara kerja pelabelan sendiri dengan memahami makna kalimat berdasarkan konteks yang dibicarakan bukan penilaian kata per kata.

Klasifikasi

Klasifikasi digunakan untuk memasukkan objek ke dalam kelas maupun kategori yang sudah ditentukan (Parapat, *et al.*, 2018), Terdapat empat jenis klasifikasi yang dapat digunakan untuk pengujian *text mining*, yaitu sebagai berikut:

1. Naïve Bayes
2. *K-Nearest Neighbor* (KNN)
3. *Support Vector Machine* (SVM)
4. *Neural Network* (NN)

Support Vector Machine (SVM)

Support Vector Machine merupakan salah satu teknik untuk klasifikasi maupun regresi yang berada dalam satu kelas dengan KNN, yaitu dalam kelas *supervised learning*. Metode ini melakukan pemetaan nonlinear yang berfungsi untuk mengubah data ke dimensi yang lebih tinggi, dengan

mengimplementasikan *learning bias*. Tingkat akurasi pada model yang dihasilkan SVM bergantung dengan fungsi kernel dan parameter yang digunakan (Parapat, *et al.*, 2018).

Matriks Konfusi

Pengujian ketelitian hasil klasifikasi bertujuan untuk melihat kesalahan klasifikasi sehingga dapat diketahui persentase ketepatannya (akurasi) hasil klasifikasi (Bashit *et al.*, 2019). Matriks konfusi (*confussion matrix*) merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Pada dasarnya matriks konfusi mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya (Solichin, 2017). Uji ketelitian ini dilakukan agar dapat diketahui tingkat kepercayaan terhadap pemakaian hasil klasifikasi untuk analisis dan keperluan berikutnya (Bashit *et al.*, 2019).

METODE

Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa data sekunder, yaitu data ulasan aplikasi Flip yang diperoleh dari *google play*. Data yang diambil sebanyak 100 ulasan terbaru berbahasa Indonesia, sejak tanggal 19 maret 2023 hingga 21 maret 2023. Data ulasan tersebut memiliki atribut nama, tanggal mengunggah ulasan, ulasan, dan nilai skor.

Analisis dan Penyelesaian Masalah

Tahap-tahap yang dilakukan dalam penelitian ini sebagai berikut;

1. Melakukan tahap pertama *preprocessing data*, yaitu *data tokenization*.
2. Melakukan tahap kedua *preprocessing data*, yaitu *stemming*.
3. Melakukan tahap ketiga *preprocessing data*, yaitu *stopword removal*.
4. Mengklasifikasikan setiap kata masuk ke dalam sentimen positif, negatif, atau netral.
5. Melakukan klasifikasi menggunakan *Support Vector Machine* (SVM).

HASIL DAN PEMBAHASAN

Crawling Data (Pengambilan Data)

Berikut ini merupakan hasil pengambilan 100 ulasan teratas sejak tanggal 19 maret 2023 hingga 21 Maret 2023 dari aplikasi Flip:

Tabel 1. Crawling Data

No	Nama	Tanggal	Ulasan	Skor
1	dja rum	2023-03-21 05:04:35	mau masuk aja susah,cuma stuck dologo	1
2	nur qoma	2023-03-21 05:03:39	aplikasi transaksi TERBAIKK 👍	5
3	Atik 1	2023-03-21 05:01:08	Mantappp trnsfer jd lbh mudah dan gratis admin kerennnn tingkatkan pelayanan flip 😊😊😊👍👍👍👍	5
⋮	⋮	⋮	⋮	⋮

No	Nama	Tanggal	Ulasan	Skor
98	Ridho Fadlipradana	2023-03-19 17:00:16	Bagus!, Tolong di tingkatkan lagi kecepatan transfer nya, terutama di Tanggal" Tertentu 😊	5
99	JOVAN RV	2023-03-19 16:21:21	Kalau ada kendala transfer, jgn menunggu 1 - 2 hari kerja, terlalu lama. urusan org jadi terhenti, karena transfer belum terkirim, kembalikan dana ke saldo flip si pengirim Secepat mungkin. Flip ini hanya sarana, bukan utama.	1
100	Dede Wahyu	2023-03-19 15:24:30	Apk nya bagus mudah di gunakan	5

Berdasarkan Tabel 1, kolom ulasan masih memiliki kalimat yang kurang efektif, sehingga perlu dilakukan *preprocessing text*.

Tokenization

Tokenization merupakan tahap awal dalam melakukan *preprocessing data*. Pada tahap ini, kata-kata dalam kalimat ulasan akan dibersihkan dari angka, tanda baca, serta karakter yang tidak diperlukan, ejaan kata akan dibenarkan, dan huruf kapital akan diubah ke huruf kecil. Berikut ini disajikan Tabel 2. yang berisi ulasan setelah melalui tahap *Tokenization*.

Tabel 2. Tokenization

No	Ulasan	Tokenization
1	mau masuk aja susah,cuma stuck dilogo	ingin masuk saja susah hanya macet di logo
2	aplikasi transaksi TERBAIKK 👍	aplikasi transaksi terbaik
3	Mantappp trnsfer jd lbh mudah dan gratis admin kerennnn tingkatkan pelayanan flip 😊😊😊👍👍👍👍	mantap transfer jadi lebih mudah dan gratis admin keren tingkatkan pelayanan flip
:	:	:
98	Kalau ada kendala transfer, jgn menunggu 1 - 2 hari kerja, terlalu lama. urusan org jadi terhenti, karena transfer belum terkirim, kembalikan dana ke saldo flip si pengirim Secepat mungkin. Flip ini hanya sarana, bukan utama.	kalau ada kendala transfer jangan menunggu hari kerja terlalu lama urusan orang jadi terhenti karena transfer belum terkirim kembalikan dana ke saldo flip si pengirim secepat mungkin flip ini hanya sarana bukan utama
99	Apk nya bagus mudah di gunakan	aplikasinya bagus mudah digunakan
100	Untuk pengiriman dari luar negara ke indonesia apakah bisa	untuk pengiriman dari luar negara ke indonesia apakah bisa

Berdasarkan Tabel 2, kalimat ulasan pada kolom *tokenization* sudah tidak terdapat angka, tanda baca, karakter dan huruf kapital, sehingga *preprocessing data* dapat dilakukan ke tahap *stemming*. Pada tahap *stemming*, setiap kata akan direduksi menjadi bentuk kata dasar. Hal ini dilakukan untuk mengenali kata-kata yang memiliki makna serupa, sehingga analisis yang dihasilkan akan konsisten.

Stemming

Stemming merupakan proses untuk merubah kata-kata berimbuhan menjadi kata dasar. Tujuan dari proses *stemming* adalah mengurangi variasi kata yang memiliki makna yang sama. Berikut ini disajikan Tabel 3. Yang berisi ulasan setelah melalui tahap *stemming*.

Tabel 3. Stemming

No	Tokenization	Stemming
1	ingin masuk saja susah hanya macet di logo	ingin masuk saja susah hanya macet di logo
2	aplikasi transaksi terbaik	aplikasi transaksi baik
3	mantap transfer jadi lebih mudah dan gratis admin keren tingkatkan pelayanan flip	mantap transfer jadi lebih mudah dan gratis admin keren tingkat layan flip
⋮	⋮	⋮
98	kalau ada kendala transfer jangan menunggu hari kerja terlalu lama urusan orang jadi terhenti karena transfer belum terkirim kembalikan dana ke saldo flip si pengirim secepat mungkin flip ini hanya sarana bukan utama	kalau ada kendala transfer jangan tunggu hari kerja terlalu lama urus orang jadi henti karena transfer belum kirim kembali dana ke saldo flip si kirim cepat mungkin flip ini hanya sarana bukan utama
99	aplikasinya bagus mudah digunakan	aplikasi bagus mudah guna
100	untuk pengiriman dari luar negara ke indonesia apakah bisa	untuk kirim dari luar negara ke indonesia apakah bisa

Berdasarkan Tabel 3, kata yang memiliki imbuhan telah dihilangkan sehingga setiap kata sudah terbentuk ke kata dasar. Tahap *preprocessing data* selanjutnya adalah *stopword removal*, dimana kata-kata yang tidak memberikan informasi penting akan dihapus, seperti ‘dan’, ‘atau’, dan ‘di’.

Stopword Removal

Stopword removal merupakan proses untuk mengeliminasi kata-kata yang tidak memberikan informasi penting dalam analisis. Tahap ini dilakukan untuk berfokus pada kata-kata yang lebih relevan dalam analisis sehingga meningkatkan kualitas dalam analisis teks. Berikut ini disajikan Tabel 4. Yang berisi ulasan setelah melalui tahap *stopword removal*.

Tabel 4. Stopword Removal

No	Stemming	Stopword Removal
1	ingin masuk saja susah hanya macet di logo	masuk susah macet logo
2	aplikasi transaksi baik	aplikasi transaksi
3	mantap transfer jadi lebih mudah dan gratis admin keren tingkat layan flip	mantap transfer mudah gratis admin keren tingkat layan flip
⋮	⋮	⋮
98	kalau ada kendala transfer jangan tunggu hari kerja terlalu lama urus orang jadi henti karena transfer belum kirim kembali dana ke saldo flip si kirim cepat mungkin flip ini hanya sarana bukan utama	kendala transfer tunggu kerja urus orang henti transfer kirim dana saldo flip si kirim cepat flip sarana utama
99	aplikasi bagus mudah guna	aplikasi bagus mudah
100	untuk kirim dari luar negara ke indonesia apakah bisa	kirim negara indonesia

Pelabelan

Proses pelabelan merupakan tahap pemisahan antara variabel y dan variabel x dari ulasan setelah melalui tahap *stopword removal* serta menyimpulkan ulasan tersebut apakah termasuk dalam

klasifikasi positif, negatif atau netral. Penyajian proses pelabelan dari ulasan aplikasi Flip terdapat di Tabel 5, sebagai berikut:

Tabel 5. Pelabelan

No	Stopword Removal	Sentiment
1	masuk susah macet logo	negatif
2	aplikasi transaksi	netral
3	mantap transfer mudah gratis admin keren tingkat layan flip	positif
⋮	⋮	⋮
98	kendala transfer tunggu kerja urus orang henti transfer kirim dana saldo flip si kirim cepat flip sarana utama	negatif
99	aplikasi bagus mudah	positif
100	kirim negara indonesia	netral

Klasifikasi SVM

Sebelum melakukan pengklasifikasian, data ulasan dipartisi ke dalam bentuk *training* dan *testing* yang dipilih secara random. Partisi yang dilakukan memiliki rasio 80:20, dimana data *training* berjumlah 80 dan data *testing* berjumlah 20. Setelah dilakukan partisi data, barulah proses pengklasifikasian dilakukan. Hasil pengklasifikasian menggunakan SVM sebagai berikut:

Tabel 6. Klasifikasi SVM

No	No Ulasan	Stopword Removal	Aktual	Prediksi SVM
1	0	masuk susah macet logo	negatif	negatif
2	4	flip hasil ubah ribet mudah praktis kembang inovasi layan flip bintang tekan flip hebat kerja bagus moga sukses	positif	positif
3	10	coba bagus	positif	positif
⋮	⋮	⋮	⋮	⋮
18	80	mantap	positif	positif
19	83	mudah	positif	positif
20	90	unduh tipu isi saldo masuk koin saldo flip bayar listrik opsi bayar koin flip kedaluwarsa laku koin flip tipu kejam	negatif	netral

Matriks Konfusi

Setelah diketahui hasil prediksi dari klasifikasi SVM, hasil tersebut dimasukkan ke dalam matriks konfusi bersamaan dengan data *testing* aktual. Hasil matriks konfusi pada klasifikasi ulasan aplikasi Flip adalah sebagai berikut:

Tabel 7. Matriks Konfusi

Prediksi		Aktual			
		Positif	Netral	Negatif	Total
	Positif	9	1	1	11
	Netral	2	3	2	7
	Negatif	0	0	2	2
	Total	11	4	5	20

Berdasarkan matriks konfusi yang terbentuk, akurasi dari klasifikasi SVM adalah sebagai berikut:

$$\begin{aligned} akurasi &= \frac{n_{11} + n_{22} + n_{33}}{T} \\ &= \frac{9 + 3 + 2}{20} = \frac{14}{20} \\ &= 0.70 \end{aligned}$$

KESIMPULAN

Berdasarkan model SVM, nilai akurasi yang didapatkan untuk mengklasifikasikan ulasan Aplikasi Flip menjadi ulasan positif, netral dan negatif adalah 0.70. Artinya, model SVM dapat mengklasifikasikan data dengan benar sebanyak 70% dari total data. Tetapi pada penelitian ini belum dapat dikatakan bahwa model SVM merupakan model terbaik untuk proses pengklasifikasian karena pemilihan model terbaik memerlukan model lain yang digunakan sebagai pembandingan.

DAFTAR PUSTAKA

- Parapat, I. M., Furqon, M. T., Sutrisno. Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. 2(10): 3165.
- Sunoto, Y., Wasito, B. (2014). Analisis Testimonial Wisatawan Menggunakan *Text Mining* Dengan Metode *Naïve Bayes* dan *Decision Tree*, Studi Kasus Pada Hotel-Hotel di Jakarta. *Jurnal Informatika dan Bisnis*, 3(2): 42-43.
- Udapure, T.V., Kale, R. D., Dharmik, R. (2014). Study of Web Crawler and its Different Types. *Journal of Computer Engineering*. 16(1): 2.