

Sentiment Analysis of Brompt Application Reviews on the Google Play Store Using the Support Vector Machine Algorithm

[Analisis Sentimen Ulasan Aplikasi Brompt Di Google Play Store Menggunakan Algoritma Support Vector Machine]

Rifqi Yusril Muslikhin¹⁾, Ika Ratna Indra Astutik²⁾

¹⁾Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

²⁾Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

Email Penulis Korespondensi: 181080200200@umsida.ac.id¹⁾, ikaratna@umsida.ac.id²⁾

Abstract. *Brompt is an android-based application that can be downloaded on the Google Play Store. In this application there are several services that can facilitate customers to carry out various activities related to Honda motorbikes. These activities include: Booking Service, Motorbike Info Catalogue, Honda Accessories, etc. Sentiment analysis is a technique to classify a document into several sentiments using text analysis methods. This research uses the dataset contained in the brompt application to classify the sentiment of the application reviews. The dataset will be grouped into two sentiments, namely positive and negative. In this case the researcher uses the Support Vector Machine algorithm with a dataset consisting of 3359 review data. By using the Support Vector Machine algorithm, the highest classification result of 87,6% was obtained using the k-fold cross validation method which was obtained from the 8th fold iteration using a RBF kernel and a ratio of training data and test data of 90:10. This proves that the application of the Support Vector Machine algorithm in document classification is able to produce high enough accuracy.*

Keywords - brompt, google play store, sentiment analysis, support vector machine

Abstrak. *Brompt adalah sebuah aplikasi berbasis android yang bisa di download di Google Play Store. Didalam aplikasi ini terdapat beberapa layanan yang dapat mempermudah pelanggan untuk melakukan berbagai kegiatan terkait dengan sepeda motor Honda. Kegiatan tersebut antara lain: Booking Service, Katalog Info Motor, Aksesoris Honda, dll. Analisis sentimen merupakan teknik untuk mengelompokkan suatu dokumen ke dalam beberapa sentimen dengan menggunakan metode analisis teks. Penelitian ini menggunakan dataset yang terdapat pada aplikasi brompt untuk mengklasifikasikan sentimen dari ulasan aplikasi tersebut. Dataset akan dikelompokkan kedalam dua sentimen yaitu positif dan negatif. Dalam hal ini peneliti menggunakan algoritma Support Vector Machine dengan dataset yang terdiri dari 3359 data ulasan. Dengan menggunakan algoritma Support Vector Machine, didapatkan hasil klasifikasi tertinggi sebesar 87,6% dengan menggunakan metode k-fold cross validation yang di dapat dari iterasi fold ke-8 dengan menggunakan kernel RBF dan perbandingan data training dan data uji 90:10. Hal ini membuktikan bahwa penerapan algoritma Support Vector Machine dalam klasifikasi dokumen mampu menghasilkan akurasi yang cukup tinggi.*

Kata Kunci - analisis sentimen, brompt, google play store, support vector machine

I. PENDAHULUAN

Servis sepeda motor merupakan kegiatan untuk melakukan perbaikan maupun perubahan motor agar dapat berjalan dengan baik seperti aslinya[1]. Pelanggan yang sebelumnya untuk melakukan servis motor harus datang ke tempat, jika tempat servis ramai maka harus antri terlebih dahulu. Dengan berkembangnya teknologi yang semakin pesat, kini banyak tersedia aplikasi booking servis motor. Pelanggan hanya perlu mengisi form yang ada di aplikasi dan menentukan tanggal dan jam servis tanpa harus datang ke tempat servis. Saat ini tersedia beberapa aplikasi booking servis motor salah satunya adalah Brompt yang bisa di download di Google Play Store.

Brompt adalah sebuah aplikasi berbasis android yang bisa di download di Google Play Store. Didalam aplikasi ini terdapat beberapa layanan yang dapat mempermudah pelanggan untuk melakukan berbagai kegiatan terkait dengan sepeda motor Honda. Kegiatan tersebut antara lain : Booking Service, Katalog Info Motor, Aksesoris Honda[2]. Di Google Play Store sampai Desember 2021 aplikasi Brompt telah di download lebih dari 500 ribu kali dan mendapatkan rating 4.5[3].

Google Play Store merupakan tempat untuk mengunduh berbagai aplikasi berbasis android sehingga pengguna memiliki banyak pilihan aplikasi untuk digunakan. Google Play Store memiliki fitur yang memberikan rating aplikasi seperti user rating dan fitur review. Secara umum, pengguna ingin mengetahui aplikasi mana yang akan diinstal, sehingga pentingnya fungsi rating dan ulasan sering digunakan untuk menilai aplikasi[4], namun sebagian besar dari ulasan yang diberikan pengguna tidak sesuai peringkat, sehingga sulit bagi pengguna untuk menentukan klasifikasi

aplikasi. Dengan menggunakan text mining, ulasan pengguna Google Play Store dapat diklasifikasikan menjadi ulasan positif dan negatif, yang kemudian dapat dibuat opsional untuk pengguna baru yang menggunakan aplikasi.

Text mining adalah ilmu mengubah teks menjadi informasi dengan memprediksi pola dan tren menggunakan model statistik[5]. Dalam proses text mining terdapat bobot kata yang tujuannya adalah untuk memberi nilai/bobot terhadap term yang terkandung dalam sebuah dokumen atau komentar[6]. Metode pembobotan satu kata adalah TF-IDF (Term Frequency Inverse Document Frequency).

Pembobotan TF-IDF adalah metode yang menghitung bobot setiap kata yang diekstraksi. Metode TF-IDF menggabungkan dua konsep yaitu jumlah term dalam satu dokumen dan penghitungan term yang muncul dalam dokumen berbeda yang dianggap tidak berhubungan[7]. Setelah proses pembobotan kata selesai, proses selanjutnya adalah klasifikasi. Salah satu metode klasifikasi adalah SVM (Support Vector Machine). SVM adalah teknik prediksi yang baik dalam klasifikasi dan regresi. Tujuan dari algoritma SVM adalah untuk mengklasifikasikan teks menggunakan fitur berbobot indeks sebanyak term[8]. Pada penelitian ini klasifikasi dilakukan dengan menggunakan metode support vector machine.

Text Mining

Text mining adalah bidang data mining di mana data mining dalam hal ini didefinisikan dalam format teks dan sumber data biasanya diambil dari dokumen. Tujuannya adalah menemukan kata-kata yang bisa mewakili isi dari data mining. Membuat dokumen dengan cara yang memungkinkan kita melakukan analisis keterhubungan antar dokumen[5]. Text mining menganalisis dan mengklasifikasikan dokumen berdasarkan kata yang dikandungnya, dan menentukan kesamaan antar dokumen variabel lain[9].

Text Preprocessing

Text preprocessing merupakan tahapan dari pengolahan teks awal hingga penyusunan teks menjadi data yang siap diolah. Dokumen tersebut harus dipisahkan menjadi beberapa tingkatan. Dokumen tersebut dapat dibagi menjadi bab, subbab, paragraf, kalimat, dan terakhir potongan kata/token[10]. Tahapan text processing sebagai berikut:

- 1) Cleaning adalah langkah yang bertujuan untuk membersihkan teks dari tanda baca, angka dan lain-lain yang tidak diperlukan dalam klasifikasi[11].
- 2) Case Folding adalah langkah yang bertujuan untuk membuat semua teks menjadi huruf kecil[12].
- 3) Tokenizing adalah langkah yang bertujuan untuk mengubah sebuah kalimat menjadi susunan kata[13].
- 4) Normalisasi adalah pengembalian kata tidak baku ke dalam kamus bahasa Indonesia yang baku[14].
- 5) Filtering adalah langkah yang bertujuan untuk menghapus kata yang tidak penting dari hasil tokenizing [13].
- 6) Stemming merupakan tahapan dalam mengubah dan membuang semua token yang berimbuhan menjadi kata. Tujuan tahapan Stemming adalah untuk menghilangkan imbuhan pada awalan dan akhiran sehingga menjadi token kata dasar yang sama[13].

Support Vector Machine

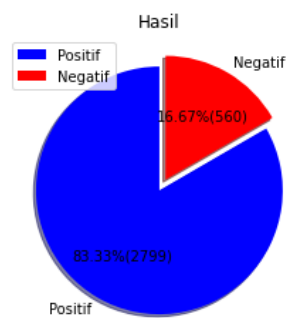
Support vector machine adalah metode klasifikasi yang bekerja dengan memisahkan kelas menggunakan bidang pemisah linier yang disebut hyperplane. Setiap kelas disebut support vector. Kita dapat menemukan hyperplane yang optimal dengan mengukur tepi hyperplane dan menemukan titik maksimumnya[15].

Untuk menyelesaikan problem non linear, SVM dimodifikasi dengan memasukkan fungsi Kernel. Kernel trick memberikan berbagai kemudahan untuk menentukan support vector, kita hanya cukup mengetahui fungsi kernel yang dipakai, dan tidak perlu mengetahui wujud dari fungsi non linear. Kernel yang umum digunakan dalam SVM adalah kernel linear, kernel sigmoid, kernel RBF, kernel polynomial[15].

II. METODE

Pengumpulan data ulasan dilakukan dengan menggunakan google play scrapper. Data pada penelitian ini diambil dari ulasan aplikasi brompit yang ada di google play store. Dataset kemudian dikelompokkan menjadi 2 kelas yaitu kelas positif dan negatif. Pelabelan data untuk menentukan ulasan positif dan negatif dilakukan menggunakan library vader sentiment dalam bahasa pemrograman python.

Data yang didapatkan berjumlah 3359 data. Dataset disimpan dalam format .csv yang nantinya akan diinputkan ke dalam sistem. Presentase jumlah data positif dan negatif dapat dilihat pada gambar 1.



Gambar 1. Jumlah ulasan positif dan negatif

Dari total data yang berjumlah 3359 data, kemudian penulis membagi dataset menjadi data latih serta data uji dengan beberapa perbandingan. Hal ini dikarenakan banyaknya data training mempengaruhi tingkat akurasi.

III. HASIL DAN PEMBAHASAN

Proses pertama adalah text preprocessing. Text preprocessing berguna untuk mengolah data sehingga data akan siap digunakan untuk proses klasifikasi.

	content	cleansing
0	pertama kali service (1000km), ada beberapa ke...	pertama kali service km ada beberapa keluhan s...
1	Aplikasi ini sangat membantu. 1. Kita bisa pes...	Aplikasi ini sangat membantu Kita bisa pesan w...
2	Saya lebih suka pakai buku servis yang manual ...	Saya lebih suka pakai buku servis yang manual ...
3	Pelayanan sudah bagus tapi saya ada masalah. S...	Pelayanan sudah bagus tapi saya ada masalah Sa...
4	Mantul,ngk ush antre pnjang,karena udh boking ...	Mantulngk ush antre pnjangkarena udh boking le...
5	Aplikasi ini memberikan kemudahan dan nyaman...	Aplikasi ini memberikan kemudahan dan nyaman...

Gambar 2. Hasil Proses Cleaning

Gambar 2 merupakan hasil proses cleaning. Tahapan ini dilakukan untuk membersihkan ulasan dari tanda baca, angka, simbol, dan sebagainya.

	cleansing	case_folding
0	pertama kali service km ada beberapa keluhan s...	pertama kali service km ada beberapa keluhan s...
1	Aplikasi ini sangat membantu Kita bisa pesan w...	aplikasi ini sangat membantu kita bisa pesan w...
2	Saya lebih suka pakai buku servis yang manual ...	saya lebih suka pakai buku servis yang manual ...
3	Pelayanan sudah bagus tapi saya ada masalah Sa...	pelayanan sudah bagus tapi saya ada masalah sa...
4	Mantulngk ush antre pnjangkarena udh boking le...	mantulngk ush antre pnjangkarena udh boking le...
5	Aplikasi ini memberikan kemudahan dan nyaman...	aplikasi ini memberikan kemudahan dan nyaman...

Gambar 3. Hasil Proses Case Folding

Gambar 3 merupakan hasil proses case folding. Tahapan ini dilakukan untuk membuat data ulasan menjadi lowercase atau menggunakan huruf kecil.

	case_folding	tokenize
0	pertama kali service km ada beberapa keluhan s...	[pertama, kali, service, km, ada, beberapa, ke...
1	aplikasi ini sangat membantu kita bisa pesan w...	[aplikasi, ini, sangat, membantu, kita, bisa, ...
2	saya lebih suka pakai buku servis yang manual ...	[saya, lebih, suka, pakai, buku, servis, yang,...
3	pelayanan sudah bagus tapi saya ada masalah sa...	[pelayanan, sudah, bagus, tapi, saya, ada, mas...
4	mantulngk ush antre pnjangkarena udh boking le...	[mantulngk, ush, antre, pnjangkarena, udh, bok...
5	aplikasi ini memberikan kemudahan dan nyaman...	[aplikasi, ini, memberikan, kemudahan, dan, ke...

Gambar 4. Hasil Proses Tokenizing

Gambar 4 merupakan hasil proses tokenizing. Tahapan ini dilakukan untuk mengubah kalimat pada data ulasan menjadi susunan kata.

	tokenize	normalisasi
0	[pertama, kali, service, km, ada, beberapa, ke...	[pertama, kali, servis, kilometer, ada, bebera...
1	[aplikasi, ini, sangat, membantu, kita, bisa, ...	[aplikasi, ini, sangat, membantu, kita, bisa, ...
2	[saya, lebih, suka, pakai, buku, servis, yang,...	[saya, lebih, suka, pakai, buku, servis, yang,...
3	[pelayanan, sudah, bagus, tapi, saya, ada, mas...	[pelayanan, sudah, bagus, tapi, saya, ada, mas...
4	[mantulngk, ush, antre, pnjangkarena, udh, bok...	[bagus tidak, perlu, antre, pnjangkarena, suda...
5	[aplikasi, ini, memberikan, kemudahan, dan, ke...	[aplikasi, ini, memberikan, kemudahan, dan, ke...

Gambar 5. Hasil Proses Normalisasi

Gambar 5 merupakan hasil proses normalisasi. Tahapan ini dilakukan untuk mengubah kata yang tidak baku, bahasa gaul, kata yang disingkat menjadi kata yang baku.

	normalisasi	stopwords
0	[pertama, kali, servis, kilometer, ada, bebera...	[kali, servis, kilometer, keluhan, aplikasi, s...
1	[aplikasi, ini, sangat, membantu, kita, bisa, ...	[aplikasi, membantu, pesan, bengkel, repot, am...
2	[saya, lebih, suka, pakai, buku, servis, yang,...	[suka, pakai, buku, servis, manual, banding, b...
3	[pelayanan, sudah, bagus, tapi, saya, ada, mas...	[pelayanan, bagus, coba, order, pesanan, baran...
4	[bagus tidak, perlu, antre, pnjangkarena, suda...	[bagus tidak, antre, pnjangkarena, booking, ap...
5	[aplikasi, ini, memberikan, kemudahan, dan, ke...	[aplikasi, kemudahan, kenyamanan, servis, moto...

Gambar 6. Hasil Proses Filtering Stopwords

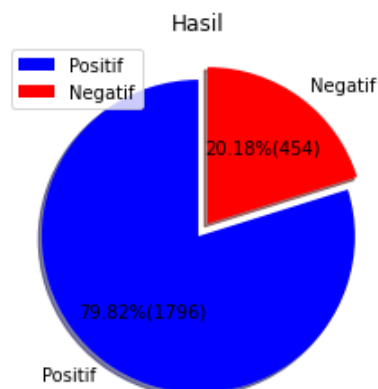
Gambar 6 merupakan hasil proses filtering stopwords. Tahapan ini dilakukan untuk menghapus kata yang tidak diperlukan.

	stopwords	stemming
0	[kali, servis, kilometer, keluhan, aplikasi, s...	[kali, servis, kilometer, keluh, aplikasi, sat...
1	[aplikasi, membantu, pesan, bengkel, repot, am...	[aplikasi, bantu, pesan, bengkel, repot, ambil...
2	[suka, pakai, buku, servis, manual, banding, b...	[suka, pakai, buku, servis, manual, banding, b...
3	[pelayanan, bagus, coba, order, pesanan, baran...	[layan, bagus, coba, order, pesan, barang, bro...
4	[bagus tidak, antre, pnjangkarena, booking, ap...	[bagus tidak, antre, pnjangkarena, booking, ap...
5	[aplikasi, kemudahan, kenyamanan, servis, moto...	[aplikasi, mudah, nyaman, servis, motor, antre...

Gambar 7. Hasil Proses Stemming

Gambar 7. Merupakan hasil proses stemming. Tahapan ini dilakukan untuk mengubah sebuah kata kedalam bentuk kata dasarnya.

Setelah proses text preprocessing, data yang sebelumnya berjumlah 3359 data berkurang menjadi 2250 data. Presentase data ulasan setelah proses text preprocessing dapat dilihat pada gambar 8.



Gambar 8. Jumlah Ulasan Positif dan Negatif Setelah Proses Text Preprocessing

Penerapan Algoritma Support Vector Machine

Beberapa kernel dapat digunakan dalam algoritma SVM ini antara lain kernel linier, kernel polynomial, kernel radial basis function (RBF), dan kernel sigmoid. Dalam penelitian ini, peneliti menguji akurasi terbaik semua kernel dengan membandingkan tiga set data, yaitu 90:10, 80:20, dan 70:30.

Tabel 1. Hasil Pengujian Menggunakan Kernel Linear

Perbandingan	Hasil Pengujian
80% : 20%	82,2%
90% : 10%	80%
70% : 30%	81,4%

Tabel 2. Hasil Pengujian Menggunakan Kernel Polynomial

Perbandingan	Hasil Pengujian
80% : 20%	82,4%
90% : 10%	83,1%
70% : 30%	83,8%

Tabel 3. Hasil Pengujian Menggunakan Kernel RBF

Perbandingan	Hasil Pengujian
80% : 20%	82%
90% : 10%	84%
70% : 30%	83,7%

Tabel 4. Hasil Pengujian Menggunakan Kernel Sigmoid

Perbandingan	Hasil Pengujian
80% : 20%	82,4%
90% : 10%	83,1%
70% : 30%	83,8%

Dari hasil pengujian yang dapat dilihat pada tabel 1, tabel 2, tabel 3, dan tabel 4 menghasilkan akurasi tertinggi sebesar 84% yang didapatkan menggunakan kernel RBF dengan perbandingan data 90:10.

Pengujian Menggunakan Metode K-Fold Cross Validation

Metode K-fold Cross Validation memungkinkan semua data mendapatkan kesempatan untuk menjadi data latih dan data uji. Pada penelitian ini dilakukan pengujian sebanyak 10 fold dengan perbandingan data training dan data uji 90:10 dengan menggunakan kernel RBF.

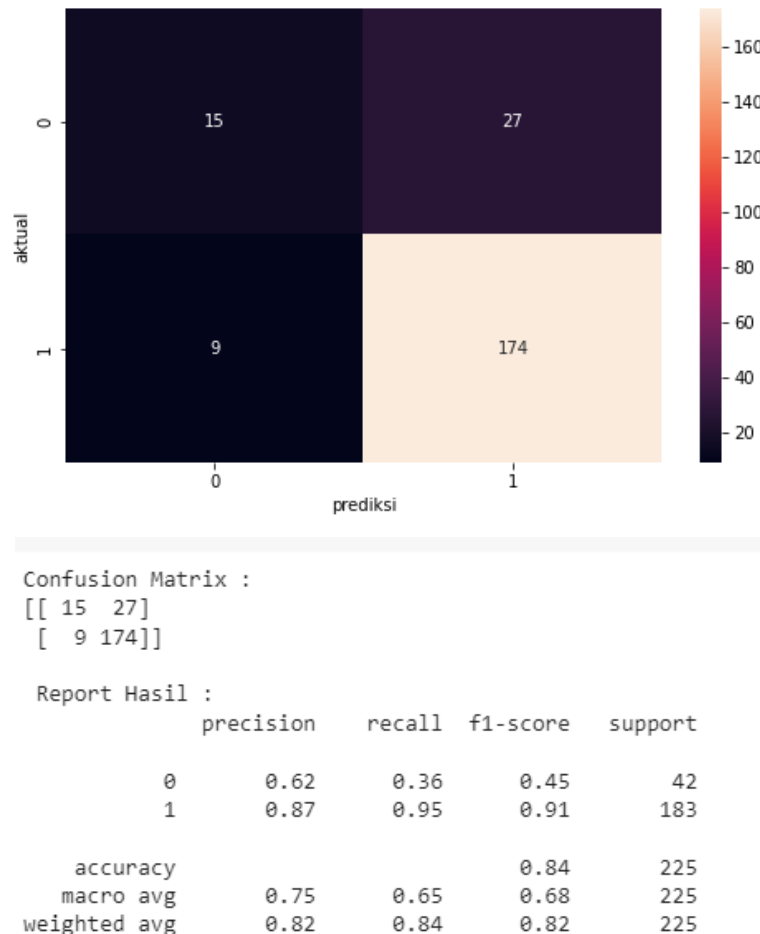
Tabel 5. Hasil Pengujian Menggunakan K-fold Cross Validation

Iterasi ke-	Hasil Pengujian
1	79,8%
2	86,2%
3	83,7%
4	83,7%
5	86,2%
6	83,1%
7	84,6%
8	87,6%
9	85,1%
10	87,1%

Pada tabel 5 dapat dilihat akurasi yang didapatkan menggunakan metode k-fold cross validation dimana akurasi tertinggi didapatkan pada iterasi ke-8 dengan akurasi sebesar 87,6%. Sedangkan untuk akurasi paling rendah didapatkan pada iterasi ke-1 dengan akurasi sebesar 79,8%.

Evaluasi Menggunakan Confusion Matriks

Evaluasi dilakukan untuk menarik kesimpulan mengenai kinerja dari algoritma support vector machine. Tahapan ini dilakukan untuk memperoleh besarnya akurasi, presisi, recall, dan f1-score. Hasil confusion matriks dapat dilihat pada gambar 9.



Gambar 9. Hasil Confusion Matriks

Berdasarkan hasil evaluasi menggunakan confusion matrix, didapatkan nilai presisi untuk sentimen negatif(0) dan positif(1) masing-masing 62% dan 87%. Hasil recall sentimen negatif(0) dan positif(1) sebesar 36% dan 95%. Sedangkan untuk nilai dari f1-score untuk sentimen negatif(0) dan positif(1) sebesar 45% dan 91%. Akurasi algoritma support vector machine sebesar 84%.

IV. KESIMPULAN

Berdasarkan hasil pengujian yang dilakukan dapat disimpulkan bahwa algoritma support vector machine terbukti dapat digunakan untuk klasifikasi data. Kernel RBF mendapatkan akurasi tertinggi dibanding kernel lainnya dengan besar akurasi 84%. Hasil confusion matriks menunjukkan bahwa akurasi dari algoritma support vector machine dalam penelitian ini sebesar 84%. Akurasi tertinggi dalam penelitian ini didapatkan menggunakan metode k-fold cross validation dengan besar akurasi 87,6% yang didapatkan dari iterasi fold ke-8 dari pengujian k-fold cross validation.

UCAPAN TERIMA KASIH

Penulis mengucapkan terimakasih yang sebesar besarnya kepada pihak-pihak yang berkontribusi dalam penelitian ini. Terimakasih kami sampaikan kepada :

- 1) Laboratorium Informatika Fakultas Sains dan Teknologi Universitas Muhammadiyah Sidoarjo.
- 2) Rekan mahasiswa Universitas Muhammadiyah Sidoarjo

REFERENSI

- [1] Priyanto, S. A. (2019). Aplikasi Pemesanan Servis Sepeda Motor Online Berbasis Android (Doctoral dissertation, Universitas Panca Marga Probolinggo).
- [2] Mpmhondajatim. (2019). Diakses dari <https://www.mpmhondajatim.com/brompit>.
- [3] Google Play Store. (2021). Diakses dari <https://play.google.com/store/apps/details?id=com.mpm.brompit>.
- [4] Herlinawati, Nuraeni, et al. (2020) "Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine." CESS (Journal of Computer Engineering, System and Science) 5.2: 293-298.
- [5] Raymond J Mooney, C. (2006). Machine Learning Text Categorization. Machine Learning Text Categorization, 1–6.
- [6] Deolika, A., Kusriani, K., & Luthfi, E. T. (2019). Analisis Pembobotan Kata Pada Klasifikasi Text Mining. (JurTI) Jurnal Teknologi Informasi, 3(2), 179-184.
- [7] Luqyana, W. A. (2018). Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine (Doctoral dissertation, Universitas Brawijaya).
- [8] Arifin, N., Enri, U., & Sulistiyowati, N. (2021). Penerapan Algoritma Support Vector Machine (SVM) dengan TF-IDF N-Gram untuk Text Classification. STRING (Satuan Tulisan Riset dan Inovasi Teknologi), 6(2), 129-136.
- [9] Findawati, Y., & Rosid, M. A. (2020). Buku Ajar Text Mining. Umsida Press, 1-123.
- [10] Fauzi, A. (2022). Penerapan Algoritma Text Mining dan Lexrank dalam Meringkas Teks Secara Otomatis. Bulletin of Data Science, 1(2), 65-72.
- [11] Retnawiyati, E., Fatoni, M. M., & Negara, E. S. (2015). Analisis Sentimen pada Data Twitter dengan Menggunakan Text Mining terhadap Suatu Produk. Departemen Teknik Informatika Universitas Bina Darma.
- [12] Pradikdo, A. C., & Ristyawan, A. (2018). Model klasifikasi abstrak skripsi menggunakan text mining untuk pengkategorian skripsi sesuai bidang kajian. Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer, 9(2), 1091-1098.
- [13] Ling, J., Kencana, I. P. E. N., & Oka, T. B. (2014). Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square. E-Jurnal Matematika, 3(3), 92-99.
- [14] Lina Auliya, Z. (2019). KLASIFIKASI SENTIMEN TERHADAP APLIKASI SHOPEE DENGAN METODE NAÏVE BAYES CLASSIFIER BERDASARKAN ULASAN DI GOOGLE PLAY (Doctoral dissertation, Universitas Islam Negeri Sultan Syarif Kasim Riau).
- [15] Nugroho, A. S., Witarto, A. B., & Handoko, D. (2003). Support vector machine. Proceeding Indones. Sci. Meeting Cent. Japan.

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.