

Contents list available at www.jurnal.unimed.ac.id

CESS
(Journal of Computing Engineering, System and Science)

journal homepage: <https://jurnal.unimed.ac.id/2012/index.php/cess>



Perbandingan KNN, Decision Tree Dan Naïve Bayes Untuk Analisis Sentimen Marketplace Bukalapak

Comparasion KNN, Decision Tree and Naïve Bayes for Sentimen Analysis Marketplace Bukalapak

Elisa Nathania Halim¹, Baenil Huda², Anggi Elanda^{3*}

^{1,2} Program Studi Sistem Informasi, Falkutas Ilmu Komputer, Universitas Buana Perjuangan Karawang
Jl. Ronggo Waluyo Sirnabaya, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361

³ Program Studi Teknik Informatika, STMIK Rosma

Jl. Parahiyangan, Adiarsa Bar., Kec. Karawang Bar., Karawang, Jawa Barat 41311

email: ¹si19.elisahalim@mhs.ubpkarawang.ac.id, ²baenil88@ubpkarawang.ac.id, ³anggi@rosma.ac.id

ABSTRAK

Banyaknya platform saat ini dapat memudahkan pengguna dapat memberikan ulasan, salah satunya pada aplikasi Bukalapak yang terdapat pada *Google Play Store*. Dimana ulasan negatif maupun positif dapat mempengaruhi nilai aplikasi tersebut, pengguna juga dapat terpengaruh oleh ulasan sentimen aplikasi. Oleh karena itu perlu diadakan analisa sentimen untuk mengelompokan ulasan negatif maupun positif, penelitian ini menggunakan data ulasan sebanyak 1000 ulasan kemudian diklasifikasi menggunakan aplikasi RapidMiner menggunakan tiga metode yaitu KNN, Naive Bayes, dan juga Decision Tree. Hasil dari metode KNN memperoleh nilai *accuracy* 85,03%, *precision* 84.98%, dan *recall* 100.00%, lalu untuk metode Naive Bayes memperoleh nilai *accuracy* 73,95%, *precision* 100.00%, dan juga *recall* 69.26%, dan untuk metode Decision Tree memperoleh nilai *accuracy* 89.12%, *precision* 88.62%, dan *recall* 100.00%. hal ini dapat membuktikan bahwa metode Decision Tree lebih unggul dari pada metode KNN dan juga metode Naive Bayes.

Kata Kunci: *RapidMiner, KNN, Naive Bayes, Decision Tree, Analisis Sentimen.*

ABSTRACT

The number of platforms currently makes it easier for users to provide reviews, one of which is the Bukalapak application on the *Google Play Store*. While both negative and positive reviews can influence the value of the app, users can also be affected by the app's sentiment reviews. Therefore, it is necessary to carry out sentiment analysis to classify negative and positive reviews. This research uses review data of 1000 reviews and then classifies them using the RapidMiner application using three methods, namely KNN, Naive Bayes, and also the Decision Tree. The results of the KNN method obtained accuracy values of 85.03%, precision of 84.98%, and recall of 100.00%, then for the Naive Bayes method obtained accuracy values

*Penulis Korespondensi:
email: anggi@rosma.ac.id

of 73.95%, precision of 100.00%, and recall of 69.26%, and for the Decision Tree method obtained 89.12% accuracy value, 88.62% precision, and 100.00% recall. this can prove that the Decision Tree method is superior to the KNN method and also the Naive Bayes method.

Keywords: *RapidMiner, KNN, Naive Bayes, Decision Tree, Sentimen Analysis.*

1. PENDAHULUAN

Belanja online sekarang ini sangat digemari oleh banyak kalangan mulai dari anak remaja hingga orang tua. Belanja online sangat mudah dan tidak perlu membuang tenaga kesana kemari untuk mencari barang yang kita inginkan, di Indonesia ada banyak sekali tempat belanja online yang sering kita temukan. Akan tetapi saat ingin belanja online pengguna akan melihat seberapa bagus aplikasi tersebut melalui opini pengguna yang sudah menggunakan aplikasi tersebut lebih dulu, opini juga dapat berpotensi besar bagi perusahaan yang ingin mengetahui umpan balik masyarakat terhadap aplikasi yang sudah dibuat. Hasil opini dari para pengguna aplikasi dapat diolah menjadi sebuah informasi dimana dapat melihat seberapa banyak opini positif maupun negatif dari sebuah aplikasi belanja online tersebut.

Dalam menganalisa hasil opini diperlukan metode untuk melihat hasil akurasi seberapa bagus aplikasi tersebut. Banyak sekali metode yang dapat digunakan, dan metode yang sering digunakan untuk menganalisis opini pengguna aplikasi belanja online yaitu ada K-Nearest Neighbor(KNN), Decision Tree, dan Naive Bayes. Metode ini digunakan untuk mengkalsifikasi hasil opini pengguna aplikasi tersebut menjadi positif maupun negatif serta mengeluarkan hasil akurasi.

Jumlah pengguna aktif dalam komunikasi online memiliki jumlah data yang sangat banyak sehingga memunculkan data besar, munculnya kumpulan data dalam jumlah sangat besar, rumit, dan tidak terstruktur menjadi salah satu sumber daya yang sangat dapat diolah untuk memperoleh hasil sentimen yang lebih akurat[1], salah satu tempat adanya big data yaitu *play store*. *Google* memiliki layanan yang menyediakan konten-konten, aplikasi, musik, dan masih banyak lagi. Dan salah satu fitur yang terdapat pada *google* yaitu fitur rating yang dimana semua pengguna layanan tersebut dapat memberikan penilaian terhadap layanan dari *google* tersebut. Aplikasi yang terdapat pada *play store* adalah aplikasi belanja *online* salah satu aplikasi tersebut adalah Bukalapak. Bukalapak merupakan aplikasi belanja *online* di Indonesia yang sangat besar, aplikasi ini digunakan untuk melakukan proses belanja secara *online* tanpa harus bertemu dengan pemilik toko tersebut.

Pada penelitian sebelumnya dilakukan penelitian mengjategorikan dan menganalisis pandangan pengguna Bukalapak untuk memanfaatkan ulasan-ulasan pengguna yang diambil melalui komentar aplikasi Bukalapak dengan menggunakan metode naive bayes mendapatkan hasil accuracy sebesar 83%, recall 80.33%, F1-Score 80.66%, dan precision 82%. Lalu ulasan tersebut setelah dianalisis sampai evaluasi akan divisualisasikan menggunakan *wordcloud*[2].

Ulasan-ulasan yang diberikan oleh pengguna aplikasi Bukalapak mengharapkan akan mendapatkan *feedback* bagi pengembang agar ditingkatkan kembali layanan, namun dengan banyaknya ulasan yang masuk maka dari itu sistem *e-commerce* tidak mencukupi hal-hal apa saja yang berkaitan untuk segera diperbaiki. Tidak adanya metode untuk mengelompokkan ulasan negatif dan positif. Beserta adanya pengelompokan ulasan pengguna berdasarkan jenis akan memudahkan pengembang untuk mendapatkan informasi ulasan sentimen pengguna aplikasi Bukalapak.

Penelitian ini memiliki tujuan untuk mengetahui sentimen pengguna aplikasi Bukalapak yang dapat menghasilkan saran untuk pihak pengembang agar dapat memperbaiki fitur-fitur pelayanan dan fitur-fitur lainnya untuk mendapatkan hasil yang bagus dari pengguna. Proses klasifikasi ini menggunakan *machine learning* dengan menggunakan tiga metode yaitu algoritma KNN, Decision Tree, dan Naive Bayes untuk melihat hasil pengelompokan sentimen, dan untuk mengetahui hasil dari perbandingan dari *precision, accuracy, F1 Score*, dan juga *recall*. Hasil dari penelitian ini diharapkan dapat memberikan hasil pengelompokan teks dengan baik dan akurat agar mendapatkan informasi dan saran yang sangat bermanfaat bagi pihak yang membutuhkan.

2. DASAR/TINJAUAN TEORI

Penelitian ini memiliki beberapa istilah untuk digunakan sebagai pendukung dalam penelitian ini yaitu :

2.1. Analisis sentiment

Analisis sentimen merupakan salah satu cara yang dapat dipakai untuk membuat opini yang diberikan dari pengguna aplikasi, sentimen analisis juga mewujudkan salah satu proses yang digunakan untuk mengekstrak, membuat data tekstual dan menangkap data. Ada tiga jenis opini yang sering ditemui yaitu netral, positif, dan negatif, dengan opini tersebut perusahaan dapat mempertimbangkan atau menganalisis kembali produk yang mereka buat[3].

Sentimen analisis juga merupakan bidang studi yang menganalisa penilaian, sikap, emosi, dan pendapat terhadap suatu produk. Pentingnya menganalisis sentimen berkaitan dengan pesatnya opini di banyak media seperti play store, media sosial, dan lain-lain[4].

2.2. Ecommerce

E-commerce yaitu proses penjualan dan pembelian antar pihak pertama dan pihak kedua dan terjadi disuatu indusrti melalui media intern, e-commerce juga merupakan kegiatan bisnis yang menyangkut konsumen, pihak ketiga, servis provider, dan manufaktur dengan menggunakan jaringan komputer yaitu internet[2].

2.3. Naive Bayes

Naive bayes merupakan sebuah metode melalui klasifikasi yang sesuai dengan teori teorma bayes. Pada metode ini pengklasifikasian berdasarkan probabilitas yang bisa terjadi dimasa depan[5].

2.4. Decision Tree

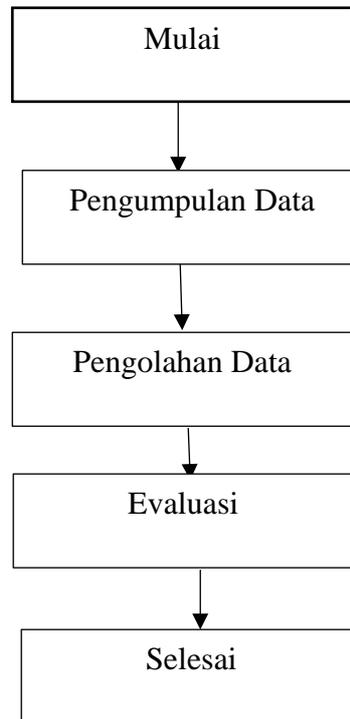
Decision Tree merupakan teknik pengelompokan dari suatu pohon keputusan, dimana sebuah data dapat dibidang sebagai simpul. Simpul yang paling utama dapat dikatakan sebagai pohon keputusan disebut sebagai root. *Decision tree* sangat bergantung pada keputusan aturan *if-then* tapi tidak memerlukan matrik dan juga parameter[6].

2.5. KNN

KNN adalah algoritma paling sederhana dan sering digunakan untuk menyelesaikan suatu masalah pengelompokan. KNN berulang kali dipakai sebagai pengelompokan data dan teks. Teknik pengelompokan kepada suatu obyek pada data yang memiliki jarak yang sangat dekat[7].

3. METODE

Penelitian ini dilakukan untuk menganalisa opini pengguna aplikasi bukalapak di platform *Play Store*.



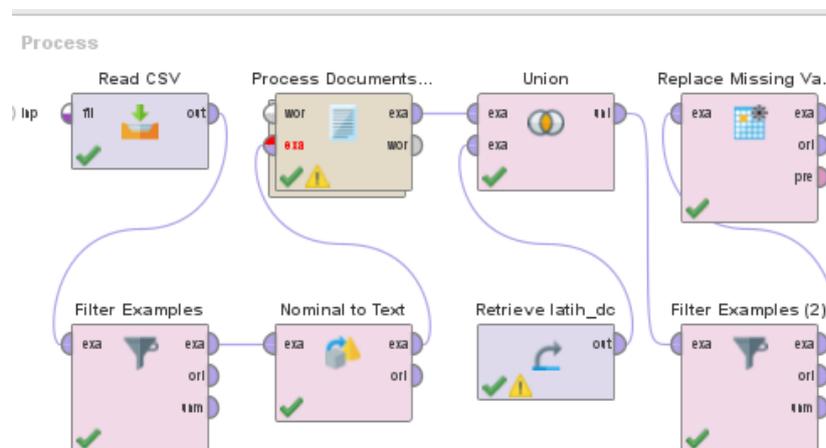
Gambar 1. Metode

3.1. Pengumpulan Data

Pengumpulan data dilakukan menggunakan python dengan menggunakan *library google-scrape*, pengambilan data ini dilakukan dengan mengumpulkan opini yang terbaru. Data yang diambil sebanyak 1000 data

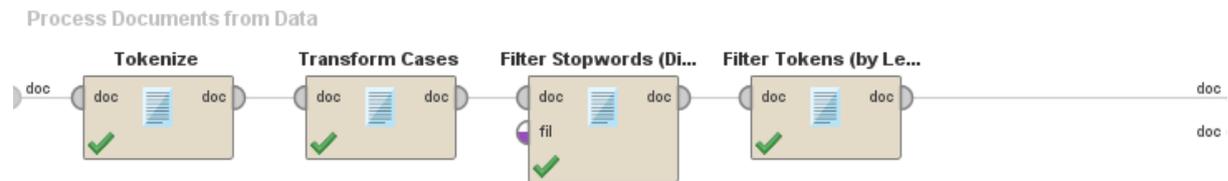
3.2. Pengolahan Data

Data yang sudah di kumpulkan dalam bentuk csv akan diolah di Rapidminer dengan dilakukannya proses seperti gambar dibawah ini.



Gambar 2. Pengolahan Data

Data berupa csv di lakukan *convert Nominal to Text* untuk mengubah nilai yang bersifat nominal ke teks[8]. Kemudian data akan melalui tahap *Process Documents from data* untuk *tokenize, filter stopwords, transform cases, filter tokens*.



Gambar 3. Process Dokumen

Tokenize adalah pemotongan input berdasarkan setiap kata yang menyusun[9], Transform cases merupakan Langkah persiapan data yang mengubah semua huruf menjadi huruf kecil[10], lalu ada tahap stopwords dilakukan untuk menghilangkan kata atau kalimat yang tidak penting, dan terakhir ada filter tokens yaitu menghapus kata-kata dengan panjang huruf tertentu.

3.3. Evaluasi

Tahap ini adalah tahap evaluasi klasifikasi dengan mengukur performa menggunakan confusion matrix terhadap algoritma KNN, naïve Bayes, dan Decision Tree.

4. HASIL DAN PEMBAHASAN

Penelitian yang telah dilakukan akan dijelaskan seperti dalam metode sebelumnya, berikut hasil dan pembahasan dari metode yang digambarkan.

4.1. Pengumpulan Data

Pengumpulan data dilakukan untuk membuat dataset berupa opini-opini pengguna aplikasi Bukalapak, setelah data opini-opini pengguna Bukalapak sudah dikumpulkan akan dijadikan sebuah file berupa csv. Pengumpulan data diambil di sebuah aplikasi yaitu *Play Store*. Kemudian data berupa csv tadi dimasukkan kedalam *Rapidminer* dan dilakukan *convert Nominat to Text*.

4.2. Pengolahan Data

Setelah tahap pengumpulan data dilanjutkan dengan tahap pengolahan data, ditahap ini dilakukan beberapa proses yaitu *tokenize, cases folding, stopwords, dan filter tokens*

4.2.1. tokenize

Tahap ini dilakukan untuk memecahkan kalimat menjadi per kata, menghilangkan tanda baca, dan menghilangkan angka.

Tabel 1. Tokenize

bgus cepat dan tdk lambat

4.2.2. cases folding

Tahap ini adalah dimana tahap mengubah semua huruf yang ada di data menjadi huruf kecil semua. Tabel dibawah adalah contoh kalimat yang sudah di *transform cases*.

Table 2. Transform Cases

Akun bukalapak saya tidak bisa login karena no hp sudah tidak aktif bagaimana solusinya...

4.2.3. Stopword

Tahap ini merupakan tahap dimana akan dihilangkan suatu kata yang tidak penting pada kalimat, table dibawah adalah salah satu kalimat yang sudah di stopwords.

Table 3. stopwords

alhamdulillah sangat lancer dalam transaksi dibukalapak

4.2.4. filter tokens

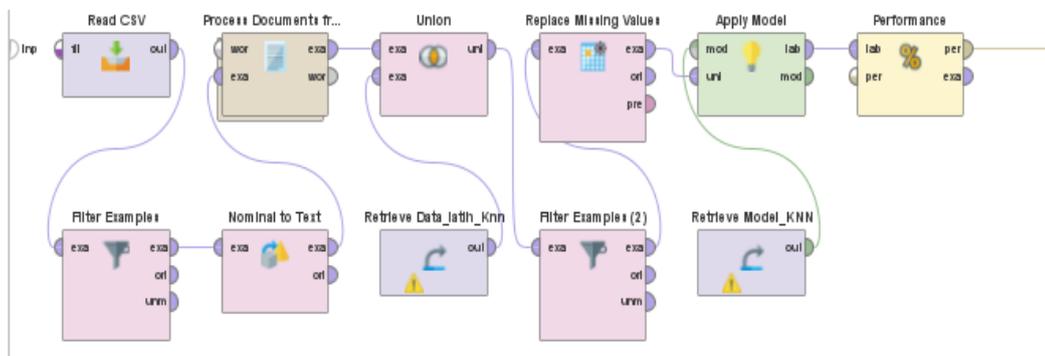
Tahap ini dilakukan untuk menghapus kata-kata dengan panjang yang sudah ditentukan, dalam penelitian ini dengan minimal 4 karakter dan maksimal 25 karakter.

4.3. Evaluasi

Penelitian ini akan melakukan pengujian menggunakan 3 algoritma yaitu KNN, Naive Bayes, dan juga Decision Tree dan membandingkan dari ketiga algoritma itu dimana yang paling cocok untuk menganalisis sentiment aplikasi Bukalapak.

4.3.1 Pengelompokan KNN

Pengelompokan menggunakan KNN mendapatkan hasil *accuracy* 85.03%, *precision* 84.98%, *recall* 100.00%. Pada gambar 4 menunjukkan proses pengelompokan menggunakan Rapidminer dengan metode KNN. Yang dimana data sudah dilatih dan di uji menggunakan metode KNN dan menghasilkan data model KNN.



Gambar 4. proses klasifikasi KNN menggunakan Rapidminer

Kemudian hasil *confusion matrix* menggunakan metode KNN seperti pada gambar 5

accuracy: 85.03%

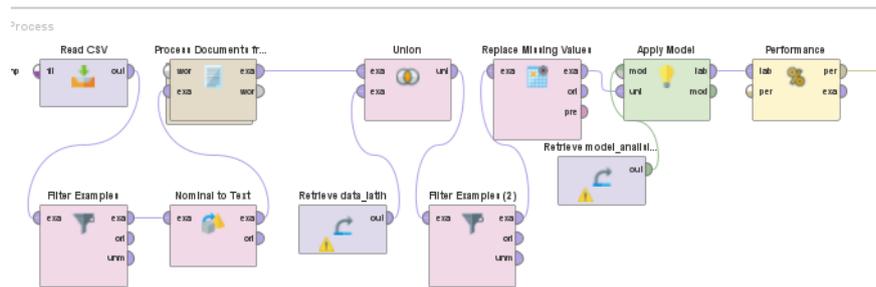
	true Positif	true Negatif	class precision
pred. Positif	1698	300	84.98%
pred. Negatif	0	6	100.00%
class recall	100.00%	1.96%	

Gambar 5. Confusion Matrix KNN

Pada gambar 5 menunjukkan hasil *confusion matrix* yang menggunakan metode KNN yaitu menghasilkan ulasan positif yang diprediksi benar sebanyak 1698 ulasan dan yang salah sebanyak 300 ulasan, lalu untuk hasil ulasan negatif yang diprediksi benar sebanyak 6 ulasan dan yang salah sebanyak 0 ulasan.

4.3.2 Pengelompokan Naive Bayes

Pengelompokan menggunakan Naive Bayes mendapatkan hasil *accuracy* 73.95%, *precision* 100.00%, *recall* 69.26%. Pada gambar 6 menunjukkan proses pengelompokan menggunakan Rapidminer dengan metode Naive Bayes. Yang dimana data sudah dilatih dan di uji menggunakan metode Naive Bayes dan menghasilkan data model Naive Bayes .



Gambar 6. Proses Klasifikasi Naive Bayes Menggunakan RapidMiner

Kemudian hasil *confusion matrix* menggunakan metode Naive Bayes seperti gambar 7

accuracy: 73.95%

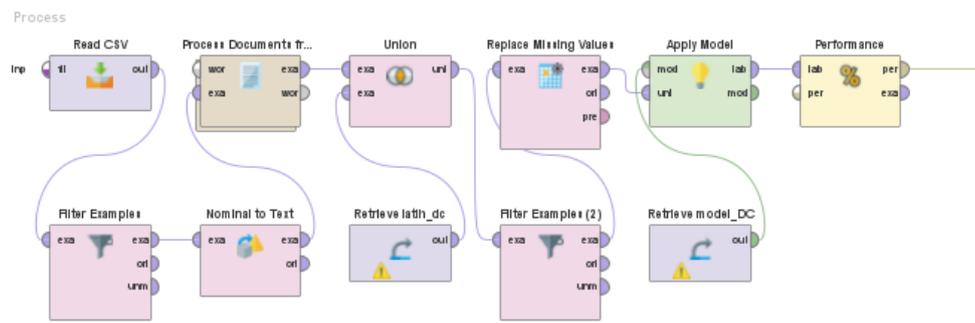
	true Positif	true Negatif	class precision
pred. Positif	1176	0	100.00%
pred. Negatif	522	306	36.96%
class recall	69.26%	100.00%	

Gambar 7. Confusion Matrix Naive Bayes

Pada gambar 7 menunjukkan hasil *confusion matrix* yang menggunakan metode Naive Bayes yaitu menghasilkan ulasan positif yang diprediksi benar sebanyak 1176 ulasan dan yang salah sebanyak 0 ulasan, lalu untuk hasil ulasan negatif yang diprediksi benar sebanyak 306 ulasan dan yang salah sebanyak 522 ulasan.

4.3.3 Pengelompokan Decision Tree

Pengelompokan menggunakan Decision Tree mendapatkan hasil *accuracy* 89.12%, *precision* 88.62%, *recall* 100.00%. Pada gambar 8 menunjukkan proses pengelompokan menggunakan Rapidminer dengan metode Decision Tree. Yang dimana data sudah dilatih dan di uji menggunakan metode Decision Tree dan menghasilkan data model Decision Tree .



Gambar 8. Proses Klasifikasi Decision Tree Menggunakan RapidMiner
 Kemudian hasil *confusion matrix* menggunakan metode Decision Tree seperti gambar 9

accuracy: 89.12%

	true Positif	true Negatif	class precision
pred. Positif	1698	218	88.62%
pred. Negatif	0	88	100.00%
class recall	100.00%	28.76%	

Gambar 9. Confusion Matrix Decision Tree

Pada gambar 9 menunjukkan hasil *confusion matrix* yang menggunakan metode Decision Tree yaitu menghasilkan ulasan positif yang diprediksi benar sebanyak 1698 ulasan dan yang salah sebanyak 218 ulasan, lalu untuk hasil ulasan negatif yang diprediksi benar sebanyak 88 ulasan dan yang salah sebanyak 0 ulasan

Berdasarkan hasil pengelompokan menggunakan tiga metode yaitu KNN, Naive Bayes, dan juga Decision Tree dapat disimpulkan bahwa kinerja metode yang lebih unggul yaitu Decision Tree karena nilai *accuracy*, *precision*, dan *recall* lebih tinggi dan lebih bagus dari pada metode KNN dan Naive Bayes. Seperti yang ditunjukkan pada tabel 4.

Table 4. Perbandingan Evaluasi Metode KNN, Naive Bayes, dan Decision Tree

Pengelompokan	Accuracy	Precision	Recall
KNN	85.03%	84.98%	100.00%
Naive Bayes	73.95%	100.00%	69.26%
Decision Tree	89.12%	88.62%	100.00%

5. KESIMPULAN

Berdasarkan hasil dari pengujian 1000 data ulasan menggunakan aplikasi RapidMiner dengan menggunakan metode KNN, Naive Bayes, dan juga Decision Tree dapat disimpulkan bahwa metode Decision Tree lebih unggul mulai dari nilai *accuracy*, *precision*, dan juga *recall*. Lalu hasil *confusion matrix* metode Naive Bayes lebih unggul dari pada KNN dan Decision Tree untuk memprediksi ulasan positif dengan hasil pengelompokan benar sebanyak 1176 ulasan dan yang salah 0. Sedangkan dalam memprediksi ulasan negative metode KNN, untuk memprediksi ulasan negative dengan hasil sebanyak pengelompokan benar sebanyak 88 ulasan dan yang salah 0 ulasan.

Hasil penelitian ini mendapatkan informasi mengenai sentiment positif maupun negative, dan pihak pengembang dapat melakukan perbaikan untuk ulasan negative menjadi lebih baik.

UCAPAN TERIMA KASIH

Penulis berterima kasih kepada pembimbing dalam merevisi penulisan ini, kemudian rekan-rekan yang sudah memberikan masukan, dan juga yang paling utama kepada Tuhan karena bisa sampai ini diberi kemudahan dalam penulisan artikel ini.

REFERENSI

- [1] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, p. 113, 2018, doi: 10.26418/jp.v4i2.27526.
- [2] I. Di Estika, I. Darmawan, and O. N. Pratiwi, "Analisis Sentimen Ulasan Pengguna Untuk Peningkatan Layanan Menggunakan Algoritma Naïve Bayes (Studi Kasus : Bukalapak)," *e-Proceeding Eng.*, vol. 8, no. 2, pp. 2735–2745, 2021.
- [3] N. Tri Romadloni, I. Santoso, S. Budilaksono, and M. Ilmu Komputer STMIK Nusa Mandiri Jakarta, "Perbandingan Metode Naive Bayes, Knn Dan Decision Tree Terhadap Analisis Sentimen Transportasi Krl Commuter Line," *J. IKRA-ITH Inform.*, vol. 3, no. 2, pp. 1–9, 2019.
- [4] Raksaka Indra Alhaqq, I Made Kurniawan Putra, and Yova Ruldeviyani, "Analisis Sentimen terhadap Penggunaan Aplikasi MySAPK BKN di Google Play Store," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 11, no. 2, pp. 105–113, 2022, doi: 10.22146/jnteti.v11i2.3528.
- [5] S. M. Salsabila, A. Alim Murtopo, and N. Fadhilah, "Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier," *J. Minfo Polgan*, vol. 11, no. 2, pp. 30–35, 2022, doi: 10.33395/jmp.v11i2.11640.
- [6] O. Somantri and D. Dairoh, "Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 2, p. 191, 2019, doi: 10.26418/jp.v5i2.32661.
- [7] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.
- [8] S. Chohan, A. Nugroho, A. M. B. Aji, and W. Gata, "Analisis Sentimen Pengguna Aplikasi Duolingo Menggunakan Metode Naïve Bayes dan Synthetic Minority Over Sampling Technique," *Paradig. - J. Komput. dan Inform.*, vol. 22, no. 2, pp. 139–144, 2020, doi: 10.31294/p.v22i2.8251.
- [9] M. Afdal and L. Waroka, "Shopee Application Review Classification Using Probabilistic Neural Network Algorithm And K-Nearest Neighbor Klasifikasi Ulasan Aplikasi Shopee Menggunakan Algoritma Probabilistic Neural Network Dan K-Nearest Neighbor," vol. 2, no. 1, pp. 49–58, 2022.
- [10] S. I. Nurhafida, F. Sembiring, J. Raya, and C. No, "Analisis Sentimen Aplikasi Novel Online Di Google Play Store Menggunakan Algoritma Support Vector Machine (SVM)," vol. 6, pp. 317–327, 2022.