JOURNAL OF INFORMATICS, BUSSINES, EDUCATION, AND INNOVATION TECHNOLOGY

Analisis Sentimen Terhadaap Review Aplikasi Hotel Online Menggunakan Metode K-Nearst Neighbor

Sentiment Analysis of Online Hotel Application Reviews Using the K-Nearst Neighbor Method

Ronaldo Gabriel Pemasi[1*], Vivi P. Rantung [2], Parabelem T. D. Rompas, [3]

[1,2,3] Teknik Informatika, Fakultas Teknik, Universitas Negeri Manado

Article Info

Article history:
Received: Jun 09, 2024
Revised: Jul 10, 2024
Accepted: Jul 25, 2024

Kata kunci
Analisis, Cross, K-Nearest Neighbors, Natural Language Processing, Validation

ABSTRAK

Analisis sentimen terletak di ranah Natural Language Processing (NLP), sebuah cabang ilmu komputer yang bertujuan untuk memahami bahasa manusia secara alami. Evaluasi analisis sentimen dengan algoritma K-Nearest Neighbors (KNN) melibatkan beberapa jenis metrik evalakan yang digunakan untuk menilai kinerja model. Tulisan ini berfokus pada analisis sentimen dan analisis sentimen, yaitu analisis sentimen dengan klik dua kali pada cross validation, otomatis membentuk dua bagian yaitu pelatihan dan pengujian. Dalam tulisan ini, kami akan menyajikan gambaran umum metode analisis sentimen yang digunakan dalam analisis sentimen.

Keywords
Analysis, Cross, K-Nearest Neighbors, Natural Language Processing, Validation

ABSTRACT

Sentiment analysis lies in the realm of Natural Language Processing (NLP), a branch of computer science that aims to understand natural human language. Evaluation of sentiment analysis with the K-Nearest Neighbors (KNN) algorithm involves several types of evaluation metrics that are used to assess model performance. This article focuses on sentiment analysis and sentiment analysis, namely sentiment analysis by double clicking on cross validation, automatically forming two parts, namely training and testing. In this paper, we will present an overview of sentiment analysis methods used in sentiment analysis

Corresponding Author:
Vivi P. Rantung
Prodi Teknik Informatika Fakultas Teknik,
Universitas Negeri Manado,
Jl. Kampus Unima, Tonsaru, Kec. Tondano Selatan, Kab. Minahasa, Sulawesi Utara, 95618.168
Email: vivirantung@unima.ac.id

PENDAHULUAN

Seiring dengan meningkatnya penggunaan internet, konsumen semakin bergantung pada review online untuk membuat keputusan terkait pemesanan hotel. Review ini menjadi sumber informasi utama bagi banyak orang dalam menentukan pilihan tempat menginap. Ketergantungan yang tinggi pada ulasan online ini mencerminkan betapa pentingnya feedback dari pelanggan bagi industri perhotelan.

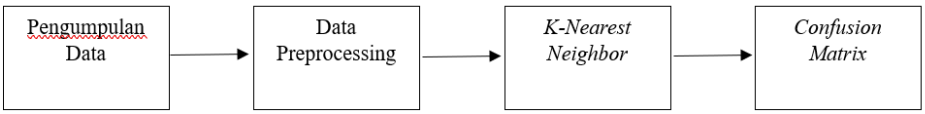
Namun, banyaknya ulasan yang tersedia sering kali membuat sulit bagi manajemen hotel untuk secara manual menyaring and memahami sentimen yang terkandung dalam ulasan tersebut. hal ini dapat menyebabkan beberapa permasalahan utama, antara lain kesulitan dalam memahami sentimen. Volume ulasan yang besar membuat analisis manual menjadi tidak praktis and cenderung tidak akurat. Tanpa alat yang tepat, manajemen mungkin melewatkan tren penting atau salah menafsirkan data, yang dapat berdampak negatif pada strategi bisnis mereka.

Penelitian ini diharapkan dapat memberikan berbagai manfaat bagi pihak hotel untuk memahami sentimen pelanggan secara lebih cepat dan akurat, yang membantu dalam pengambilan keputusan strategis dan operasional.

METODE PENELITIAN

Metodologi yang digunakan untuk melakukan analisis sentimen untuk pemeringkatan popularitas situs pemesanan hotel menggunakan metode k-nearest Neighbor (k-NN) dibagi menjadi lima tahapan yaitu tahap pengumpulan data, text preprocessing, penerapan K-NN dan Melakukan pengujian dengan metode Confusion Matrix.

Berikut adalah tahap urutan penelitian yang akan dilakukan tertera dalam diagram *flow chart* dibawah.



Gambar 1 Tahapan Penelitian

Tahap awal adalah langkah pertama yang dilakukan untuk melakukan mengumpulkan data. Proses yang dilakukan adalah menarik review dari website, pengumpulan data yang dilakukan adalah web scrapping. Setelah data url didapatkan langkah selanjutnya adalah memasukkan url tersebut dalam aplikasi google colabs. Pada penelitian ini adalah Traveloka.com. Setelah semua tahap dilakukan maka langkah terakhir adalah menyimpan dan menjalankan recipe yang telah kita buat sebelumnya.

Untuk tahapan klasifikasi K- Nearest Neighboar klik dua kali pada cross validation yang secara otomatis membentuk dua bagian yaitu training dan testing. Untuk bagian training tarik operator K-NN dan masukkan nilai K. Sedangkan pada bagian testing tarik apply model dan performance. setelah data telah diproses dan diklasifikasikan, kemudian akan dilakukan Pengujian, metode akan dilakukan menggunakan model confusion matrix Alat yang akan digunakan terbagi menjadi dua yaitu Hardwere dan Softwere. Dengan alat penelitian berupa Laptop, aplikasi/software RapidMiner, Google Colab software serta data yang digunakan merupakan data dari aplikasi pemesanan hotel online yaitu Traveloka

HASIL DAN PEMBAHASAN

Pengumpulan data adalah tahap di mana peneliti mengumpulkan data yang diperlukan untuk penelitian. Dalam penelitian ini, teknik pengumpulan data akan difokuskan pada pengumpulan ulasan hotel online yang akan menjadi objek analisis sentimen. Data tersebut dapat diperoleh dari berbagai platform pemesanan hotel online yaitu play.google.com. Pengumpulan data dapat dilakukan dengan menggunakan teknik web scraping atau dengan mengakses google colabs. Penting untuk memastikan bahwa data

yang dikumpulkan representatif dan sesuai dengan tujuan penelitian, serta memperhatikan etika dalam penggunaan data online.

Dengan melakukan studi literatur dan teknik pengumpulan data pada tahap awal, penelitian ini dapat memperoleh pemahaman yang komprehensif tentang topik penelitian dan mempersiapkan langkah-langkah selanjutnya dalam proses analisis sentimen terhadap ulasan hotel online menggunakan metode K-NN.

Data preprocessing merupakan tahap penting dalam analisis data di mana data mentah yang dikumpulkan akan dipersiapkan untuk analisis lebih lanjut. Sub-tahap ini melibatkan langkah-langkah seperti penghapusan karakter khusus, tokenisasi teks menjadi kata-kata individual, menghilangkan stopwords, dan stemming atau lemmatization untuk mengubah kata-kata menjadi bentuk dasar mereka. Tujuan dari data preprocessing adalah untuk membersihkan dan mengubah data menjadi format yang sesuai untuk analisis sentimen.

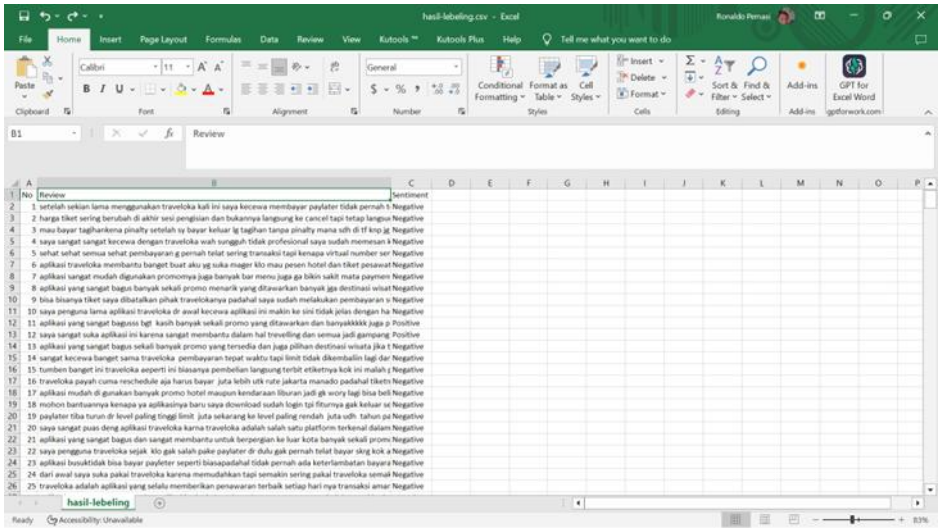
Text Preprocessing

Ditahap selanjutnya setelah data kita kompulkan dilakukanlah tahap cleaning text atau Text Preprocessing. Tahap ini terdiri dari case folding, Tokenizing, filtering, dan Stemming. Berikut ini adalah penjelasan secara rinci mengenai tahapan dalam text preprocessing

1. Case Folding
   Tahap ini merupakan tahapan Dimana teks yang menggunakan huruf capital akan dirubah menjadi huruf kecil. Case folding sendiri merupakan tahapan awal dalam text preprocessing agar memudahkan cleaning text untuk menuju ketahapan selanjutnya.
2. Tokenizing
   Tahap Tokenizing adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya. Tokenisasi secara garis besar memecah sekumpulan karakter dalam suatu teks ke dalam satuan kata, bagaimana membedakan karakter- karakter tertentu yang dapat diperlakukan sebagai pemisah kata atau bukan.
3. Filtering
   Tahap Filtering adalah tahap mengambil kata-kata penting dari hasil token. Bisa menggunakan algoritma stoplist (membuang kata kurang penting) atau wordlist (menyimpan kata penting). Stoplist/stopword adalah kata- kata yang tidak deskriptif yang dapat dibuang dalam pendekatan bag-of-words. Contoh stopwords adalah “yang”, “dan”, “di”, “dari” dan seterusnya.
4. Stemming
   Teknik Stemming diperlukan selain untuk memperkecil jumlah indeks yang berbeda dari suatu dokumen, juga untuk melakukan pengelompokan kata- kata lain yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk atau form yang berbeda karena mendapatkan imbuhan yang berbeda.

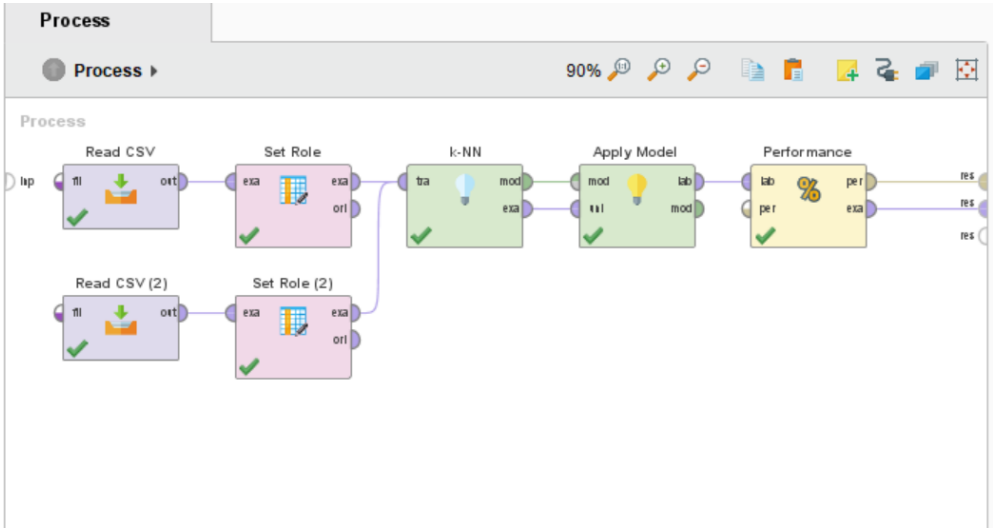
Klasifikasi K-Nearest Neighboar

Tahap ini melibatkan penerapan metode K-Nearest Neighbor (K-NN) untuk mengklasifikasikan sentimen dari setiap ulasan hotel. Dalam K-NN, ulasan yang baru akan diklasifikasikan berdasarkan mayoritas label sentimen dari tetangga terdekatnya dalam ruang fitur. Langkah-langkah ini melibatkan pemilihan nilai K yang optimal, penghitungan jarak antara ulasan yang baru dengan ulasan yang sudah ada, dan penentuan label sentimen berdasarkan mayoritas tetangga terdekat.



Gambar 2 Klasifikasi Data Dalam Tabel Excel

Data ini merupakan data yang telah dilabeli. Setelah data yang dikumpulkan awalnya belum memiliki label. Pemberian label pada setiap dokumen dilakukan berdasarkan nilai prioritas yang digunakan.



Gambar 3 Proses K-NN pada RapidMiner

Pada proses diatas telah dilakukan pemodelan K-NN dengan membandingkan kedua data yang sama, setelah digabungkan maka akang didapatkan hasil sebagai berikut

Open in Turbo Prep Auto Model Filter (397 / 397 examples): all

Row No.	Sentiment	prediction(S...	confidence(...	confidence(...	no	stemming	Positiv
1	Negative	Negative	1	0	2	kecewa nya nyaman paylatter ...	0
2	Negative	Negative	1	0	3	aplikasi traveloka bantu suka ...	0
3	Negative	Negative	1	0	4	terima kasih traveloka tani p...	0
4	Negative	Negative	1	0	5	ui aplikasi udah mantapdan ...	0
5	Negative	Negative	1	0	6	aplikasi bagus bantu bertravel...	0
6	Negative	Negative	1	0	7	tolong konfirmasi cepat mes ...	0
7	Negative	Negative	1	0	8	voucer udah terbit cancel h ...	0.160
8	Negative	Negative	1	0	9	terima kasih refund tiket citili...	0
9	Negative	Negative	1	0	10	aplikasi bagus membantubany...	0.101
10	Positive	Positive	0	1	11	aplikasi bantu banget sih ka...	0.167
11	Negative	Negative	1	0	12	beli tiket pesawat hotel tmpa...	0
12	Negative	Negative	1.000	0	13	bibtang bintang temu bug tra...	0.026
13	Positive	Positive	0	1	14	nih kasih bintang jujur suka ...	0.103
14	Negative	Negative	1	0	15	cari hotel pesawat kereta api...	0

ExampleSet (397 examples, 4 special attributes, 5 regular attributes)

Gambar 3 Hasil Proses K-NN dengan RapidMiner

dari hasil perhitungan pada sistem dalam RapidMiner, dapat didapatkan table berisi kolom prediksi dan confidence yang merupakan hasil dari perhitungan Model K-NN namun ada beberapa perhitungan sentiment yang tidak sesuai seperti review yang dikategorikan negatif namun setelah filter kembali data ternyata merupakan review positif. Maka dilakukan pelabelan manual lagi dengan menetapkan kata positif, dalam hal ini saya mengambil kata “bagus, bantu, muidah” dan setelah difilter kembali ada 64 review positif yang dikategorikan negatif. dan setelah melakukan pengecekan lagi pada sentiment positif ternyata ada beberapa hasil yang tidak sesuai dengan pengambilan kata “sulit, susah, tipu, rugi” ditemukan 4 review negatif yang dikategorikan positif.

mudah pesan tiket jalan promo nya	0	0	0	Negative	FN
makasih banget traveloka bantu saat	0	0	0	Negative	FN
negara indonesia didownload bahasa i	0	0	0	Negative	
beli tiket jamin tiket refund kecewa	0	0	0	Negative	
aplikasi smart mudah ktika jalan lov	0.213	0	0.4019	Negative	FN

Gambar 4 Gambar Sentimen Negatif yang keliru atau False Negative

paylater perlu beli pulsa token andai	0.039	0	0.4421	Negative	
aplikasi cari untung beda sam yg sel	0	0	0	Negative	
aplikasi yg sngat rugi customer proses	0.048	0	0.5815	Positive	FP
jujur ya aplikasi bantu guna paylater	0	0	0	Negative	N
kecewa bayar nga telat nga limit tur	0	0.057	-0.4404	Negative	
tolong cs aplikasi edukasi ya klaim u	0	0	0	Negative	

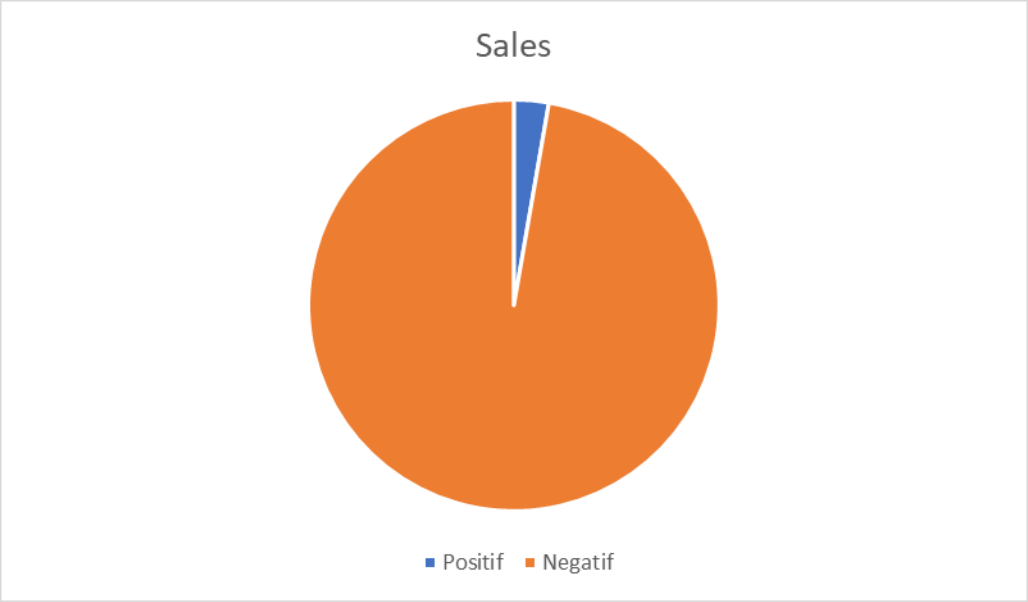
Gambar 5 Gambar Sentimen Positif yang keliru atau False Positif

$$Sentimen\ Positif = \frac{13}{398} \times 100\% = 2{,}26\%$$

$$Sentimen\ Negatif = \frac{320}{398} \times 100\% = 80{,}40\%$$

$$Sentimen\ Positif\ Keliru\ (FP) = \frac{4}{398} \times 100\% = 1{,}00\%$$

$$Sentimen\ Negatif\ Keliru\ (FN) = \frac{64}{398} \times 100\% = 16{,}08\%$$



Gambar 6 Sentimen Positif dan Negatif dengan diagram Pie

Berdasarkan hasil analisis sentiment mengenai review hotel online dapat disimpulkan bahwa sangat banyak tanggapan negatif daripada positif. Pembahasan tentang review hotel online sangatlah penting bagi konsumen dalam menentukan destinasi hotel. Tapi dari sini dapat dilihat perbandingannya sangatlah jauh antara positif dan negatif untuk pemesanan hotel online.

Analisis Validasi Data

Tahap terakhir adalah analisis validasi data, di mana keakuratan dan kinerja model K-NN dievaluasi. Ini melibatkan pembagian data menjadi set pelatihan dan set pengujian, penerapan model K-NN pada set pelatihan, dan evaluasi performa model menggunakan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Tujuan dari analisis validasi data adalah untuk memastikan bahwa model dapat menghasilkan prediksi sentimen yang akurat dan dapat diandalkan.

Dalam penelitian ini, penting untuk menggunakan berbagai metrik evaluasi untuk memastikan model memberikan hasil yang akurat dan dapat diandalkan. langkah-langkah Evaluasi adalah

Membuat Confusion Matrix untuk mendapatkan gambaran umum tentang kesalahan yang dibuat oleh model. Confusion Matrix merupakan tabel perbandingan hasil evaluasi

dari perhitungan model sistem dalam data mining. Dari hasil perhitungan sistem yang didapatkan dari rapid miner adalah sebagai berikut:

accuracy: 100.00%

	true Negative	true Positive	class precision
pred. Negative	384	0	100.00%
pred. Positive	0	13	100.00%

Gambar 7 Hasil Proses *Confusion Matrix* dengan RapidMiner

Bisa dilihat true positif adalah semua data kategori positif yang berhasil terklasifikasi dengan total 13 data. True negatif adalah semua data kategori negatif yang berhasil diklasifikasi dengan total 384 data. Adapun false positif atau pred.positif yang artinya semua data yang berkategori negatif tapi terklasifikasi positif tidak dapat dihasilkan atau nilainya 0. Begitupula dengan false negative atau pred negatif tidak dihasilkan atau 0.
Dari hasil confusion matrix diatas beserta nilai-nilainyac dalam aplikasi RapidMiner sudah otomatis terhitung untuk setiap akurasi, presisi, recall, f1 score, maka sekarang kita akan melakukan perhitungan manual untuk mengetahui hasil dari nilai akurasi, precision, recall dan f1 score.

Akurasi (*Accuracy)*
Mennggambarkan seberapa akurat model dalam mengklasifikasikan dengan benar.

$$Akurasi \frac{TP + TN}{(TP + TN + FP + FN)} = \frac{13 + 384}{(13 + 384 + 4 + 64)} = \frac{398}{465} = 0{,}85\%$$

Presisi (*Precision*)
Menggambarkan akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model.

$$Precision = \frac{TP}{(TP + FP)} = \frac{13}{(13 + 4)} = \frac{13}{17} = 0{,}76\%$$

Recall (*Sensitivity atau True Positive Rate*)
Menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi.

$$Recall = \frac{TP}{(TP + FN)} = \frac{13}{(13 + 64)} = \frac{13}{77} = 0{,}20\%$$

F1 Score
F-1 Score menggambarkan perbandingan rata-rata precision dan recall yang dibobotkan. Accuracy tepat kita gunakan sebagai acuan performansi algoritma jika dataset kita memiliki jumlah data False Negatif dan False Positif yang sangat mendekati (symmetric). Namun jika jumlahnya tidak mendekati, maka sebaiknya kita menggunakan F1 Score sebagai acuan.

$$F1 = 2 \times \frac{Recal \times Presisi}{Recall + Presisi} = 2 \times \frac{0{,}20\% \times 0{,}76\%}{0{,}20\% + 0{,}76\%} = \frac{0{,}152}{0{,}96} = 0{,}31\%$$

KESIMPULAN

Dari penelitian yang sudah dilakukan mengenai analisis sentiment terhadap review aplikasi hotel online yang diambil dari google playstore menggunakan algoritma K-Nearest Neighbor, dapat disimpulkan bahwa:

Klasifikasi system in menggunakan metode K-NN. Pengklasifikasikan sentiment negatif dan positif dengan permasalahan pemesanan hotel online, dengan performa system klasifikasi hamper mencapai 100%. Bisa dilihat dari evaluasi system akurasi 85% , presisi 76%, recall 20%, dan f1 scorenya 31%.

Dari data yang sudah discraping dari website google playstore sebanyak 398 data dalam bentuk review text, setelah dilakukan analisis sentiment didaptakan hasil dengan 13 dokumen positif yang merujuk pada bagus, bantu, mudah. Namun sangat banyak pulan sentiment negative yaitu sebanyak 384 dokumen yang menunjukan bahwa masih ada pengguna yang merasa tidak puas dengan pelayanan aplikasi hotel online.

UCAPAN TERIMAKASIH

Saya ingin menyampaikan terimakasih yang mendalam atas dukungan dan kontribusi luar biasa dalam pelaksanaan penelitian ini. Tanpa kerjasama yang sangat berarti dari semua pihak yang turut serta, penelitian ini tidak akan mencapai kesuksesan seperti yang kita nikmati saat ini. Pertama-tama, saya ingin mengucapkan terima kasih kepada ibu Vivi P. Rantung, ST, MISD selaku Koorprodi Teknik Informatika dan dosen Pembimbing Akademik dan Pembimbing ke-2 Prof. Dr. Parabelem T. D. Rompas, MT yang memberikan arahan, masukan dan bimbingan yang telah mereka berikan. Saya sangat menghargai kesabaran dan waktu yang telah mereka curahkan untuk saya. Selain itu, saya juga ingin menyampaikan rasa terima kasih yang tak terhingga kepada keluarga dan teman-teman yang telah memberikan dukungan moral serta motivasi selama proses penelitian ini. Doa dan semangat yang terus mengalir dari mereka sangat berarti bagi kelancaran penelitian ini. Semoga ungkapan terima kasih ini mencerminkan rasa syukur dan penghargaan saya atas segala bantuan dan dukungan yang telah diberikan. Saya sungguh berterima kasih atas segala hal ini.

DAFTAR PUSTAKA

Rauf, R., Sarbini, D., & Nurdiana. (2017). Physical and Sensory Characteristics of Bread From Composite Wheat and Cassava Flours with Optimum Fermentation and Proportional Water Volume. In *The 3rd International Conference on Science, Technology and Humanity 2017*. Solo, Indonesia, 118–127.

Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research. Applied Sciences, 13(7), 4550.

Bansal, S., & Sharma, R. (2022). Sentiment analysis and opinion mining on educational data: A survey. ScienceDirect.

Khalid, H., & Yadav, S. (2021). Comparative Analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, and Logistic Regression for Sentiment Analysis. ScienceDirect.

Zou, C., Wang, C., & Zhang, Y. (2021). A comprehensive survey on sentiment analysis: Approaches and challenges. MDPI Applied Sciences, 11(3), 1204.

Ningrum, D. R., & Setiawan, N. A. (2020). Sentiment Analysis for Formative Assessment in Higher Education: A Systematic Review. Springer.

Ligthart, A., Catal, C., & Tekinerdogan, B. (2021). Systematic review of sentiment analysis in educational data. ScienceDirect.

Deviyanto, A., & Wahyudi, M. D. R. (2018). Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor.

Ernawati, S., & Wati, R. (2018). Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel.

Mentari, N. D., Fauzi, M. A., & Muflikhah, L. (Tahun tidak disebut). Analisis Sentimen Kurikulum 2013 Pada Sosial Media Twitter Menggunakan Metode K-Nearest Neighbor dan Feature Selection Query Expansion Ranking.

Febrianti, Y. M., Indriati, & Widodo, A. W. (Tahun tidak disebut). Analisis Sentimen Pada Ulasan "Lazada" Berbahasa Indonesia Menggunakan K-Nearest Neighbor (K-NN) Dengan Perbaikan Kata Menggunakan Jaro Winkler Distance.