

Metode *Fuzzy-KNN* Untuk Analisis Sentimen Objek Wisata Di Sulawesi Utara Pada Trip Advisor

Fuzzy-KNN Method For Sentiment Analysis Of Tourist Objects In North Sulawesi On Trip Advisor

Aprilia Viktoria Gorahe¹, Irene R. H. Tangkawarouw², Vivi Peggie Rantung³

^{1,2,3} Teknik Informatika, Fakultas Teknik, Universitas Negeri Manado

Article Info	ABSTRAK
Article history: Received: Aug 20, 2024 Revised: Sep 18, 2024 Accepted: Sep 28, 2024	Sulawesi Utara dikenal dengan keindahan alamnya yang memikat serta kekayaan budayanya yang unik. Seiring dengan perkembangan teknologi dan meningkatnya penggunaan media sosial serta platform ulasan wisata, seperti Trip Advisor, banyak informasi berharga tentang pengalaman wisatawan yang dapat dieksplorasi untuk mendukung pengembangan sektor pariwisata. Metode <i>Fuzzy-KNN</i> digunakan untuk mengatasi kompleksitas analisis sentimen pada dataset yang tidak terstruktur dan beragam. Sulawesi Utara dipilih sebagai studi kasus karena kekayaan objek wisatanya yang beragam dan meningkatnya penggunaan platform Trip Advisor sebagai sumber ulasan wisata. Penelitian ini berfokus pada pengembangan model analisis sentimen yang dapat mengklasifikasikan ulasan pengguna menjadi kategori positif, negatif, atau netral dengan menggunakan pendekatan <i>Fuzzy-KNN</i> . Metode ini menggabungkan keunggulan dari <i>K-Nearest Neighbors (KNN)</i> dengan logika fuzzy, yang memungkinkan penanganan lebih baik terhadap ketidakpastian dan variasi dalam data ulasan wisata. Metode <i>Fuzzy-KNN (Fuzzy K-Nearest Neighbors)</i> dipilih dalam penelitian ini karena kemampuannya yang efektif dalam menangani data yang tidak pasti dan bervariasi. Dengan menggunakan metode ini, analisis sentimen dapat dilakukan dengan lebih akurat dan menyeluruh, sehingga memberikan gambaran yang jelas mengenai persepsi wisatawan terhadap objek wisata di Sulawesi Utara.
Kata kunci Sulawesi Utara, Metode KNN, Analisis Online Review Tripadvisor, Wisata TripAdvisor	
Keywords North Sulawesi, KNN method, Online Analysis of Tripadvisor Reviews, TripAdvisor Travel	ABSTRACT <i>North Sulawesi is known for its enchanting natural beauty and unique cultural richness. As technology develops and the increasing use of social media and travel review platforms, such as Trip Advisor, a lot of valuable information about tourists' experiences can be explored to support the development of the tourism sector. The Fuzzy-KNN method is used to overcome the complexity of sentiment analysis on unstructured and diverse datasets. North Sulawesi was chosen as a case study because of its rich diversity of tourist attractions and the increasing use of the Trip Advisor platform as a source of tourist reviews. This research focuses on developing a sentiment analysis model that can classify user reviews into positive, negative or neutral categories using the</i>

Fuzzy-KNN approach. This method combines the advantages of K-Nearest Neighbors (KNN) with fuzzy logic, which allows better handling of uncertainty and variation in travel review data.

Corresponding Author:

Irene R. H. Tangkawarouw

Teknik Informatika,

Universitas Negeri Manado,

Jl. Kampus Unima, Kelurahan Tonsaru, Kecamatan Tondano Selatan.

Email:irene.tangkawarow@unima.ac.id

PENDAHULUAN

Obyek wisata yang ada di Sulawesi Utara ini merupakan salah satu tujuan wisata yang terkenal baik bagi wisatawan domestic maupun. Pengembangan dan peningkatan layanan bagi wisatawan terus dilakukan oleh penyelenggara tempat wisata. Informasi yang didapatkan dari timbal balik wisatawan yang mengunjungi obyek wisata tersebut dapat menjadi masukan bagi peningkatan layanan dan pengembangan obyek wisata. Informasi ini dapat didapatkan melalui ulasan-ulasan yang diunggah oleh para wisatawan melalui platform TripAdvisor mengenai obyek wisata Sulawesi Utara.

Ulasan pada TripAdvisor ini dapat memberikan informasi berharga karena platform ini merupakan platform yang populer secara global dalam memberikan ulasan dan rekomendasi mengenai kepariwisataan. Penggalan informasi mengenai sentimen wisatawan terhadap objek-objek wisata di Sulawesi Utara dilakukan dengan metode analisis sentimen. Analisis sentimen ini akan menggali sentimen positif maupun negatif.

Banyaknya review mengenai tempat wisata yang telah dikunjungi sangat membantu calon pengunjung lainnya. Ulasan-ulasan ini memberikan informasi yang mendetail tentang keadaan dan fasilitas di tempat tersebut. Selain itu, review dari pengunjung sebelumnya bisa menjadi referensi berharga bagi mereka yang ingin mendapatkan gambaran nyata tentang pengalaman di sana. Hal ini memudahkan calon wisatawan dalam memutuskan apakah tempat tersebut sesuai dengan harapan dan kebutuhan mereka. Dengan demikian, review tersebut tidak hanya informatif, tetapi juga memberikan masukan penting untuk menentukan apakah tempat wisata tersebut layak untuk dikunjungi atau tidak. Dunia digital ini memudahkan semua orang untuk mengakses data di media sosial, seperti yang penulis lakukan untuk penelitian ini, yaitu adalah tentang opini publik di beberapa objek wisata di Sulawesi Utara yang terus meningkat peminatnya dan banyak yang memberi ulasan di berbagai tempat wisata pada media sosial TripAdvisor.

TripAdvisor adalah situs perjalanan terbesar di dunia dengan lebih dari 661 juta ulasan dan opini publik. Banyak ulasan tersebut berasal dari berbagai bisnis objek wisata di Sulawesi Utara. Di media sosial, pengunjung sering berbagi pendapat mereka tentang destinasi wisata yang telah mereka kunjungi. Mereka memberikan opini berdasarkan tempat, spot-spot wisata, dan suasana yang mereka alami. Ulasan ini mencakup berbagai perspektif, baik positif maupun negatif, yang dapat membantu wisatawan lain dalam membuat keputusan. Terdapat penelitian mengenai analisis sentimen sebelumnya menggunakan algoritma K-Nearest Neighbor (K-NN). Penelitian sebelumnya yaitu penerapan algoritma K-Nearest Neighbor pada sentimen review agen,

Analisis sentimen kurikulum 2013 pada Twitter menggunakan Ensemble Feature dan metode K-Nearest Neighbor.

Penelitian ini dilakukan untuk mengurangi ambiguitas, makna yang terkandung dalam ulasan pelanggan di beberapa objek wisata menggunakan Analisis sentiment atau penelitian ini berfokus pada aplikasi spesifik yang dapat melakukan peninjauan klasifikasi untuk melihat pola kata atau kalimat yang positif atau negatif. Penelitian ini mengusulkan kombinasi sentimen analisis dengan pembelajaran mesin dan FK-NN untuk memprediksi sentimen ulasan positif dan negatif. Metode ini bertujuan untuk meningkatkan akurasi dalam klasifikasi sentimen dari berbagai ulasan yang tersedia. Sentimen analisis digunakan untuk memahami opini dan perasaan dari teks ulasan. Pembelajaran mesin kemudian diterapkan untuk mempelajari pola-pola dalam data ulasan tersebut. FK-NN (Fuzzy K-Nearest Neighbors) digunakan untuk memperbaiki prediksi dengan mempertimbangkan faktor-faktor ketidakpastian dalam data ulasan.

Dataset adalah ulasan yang diambil dari situs web yang dipilih, yaitu TripAdvisor. Penelitian ini melakukan langkah-langkah pra-pemrosesan, seperti stemming, stopwords, dan tokenizer. Lalu, sentimen Analisis dilakukan dengan mengklasifikasikan data dengan Fuzzy K-Metode Nearest Neighbor (FK-NN). Hasil dari penelitian ini adalah hasil analisis berdasarkan sentimen tergolong positif dan negatif. Analisis sentimen atau “penambahan opini” berfokus pada aplikasi spesifik yang mengklasifikasikan data dengan Fuzzy K-Metode Nearest Neighbor (FK-NN).

METODE PENELITIAN

Penelitian ini berfokus pada pengembangan model analisis sentimen yang dapat mengklasifikasikan ulasan pengguna menjadi kategori positif, negatif, atau netral dengan menggunakan pendekatan Fuzzy-KNN. Metode ini menggabungkan keunggulan dari K-Nearest Neighbors (KNN) dengan logika fuzzy, yang memungkinkan penanganan lebih baik terhadap ketidakpastian dan variasi dalam data ulasan wisata.

Teknik Pengumpulan Data

Pengumpulan data dilakukan dengan cara mengambil data ulasan pengunjung objek wisata yang ada di Sulawesi Utara dari kolom komentar situs TripAdvisor dengan cara web scraping menggunakan Web Scraper yang merupakan salah satu ekstensi dari chrome. Dengan menggunakan ekstensi Chrome Web Scraper, bisa dengan mudah melakukan scraping data dari TripAdvisor tanpa harus menulis kode. Data yang diperoleh terkait ulasan objek wisata di Sulawesi Utara dapat dilihat pada tabel 3.1.

Tabel 3. 1 Data Ulasan Berbahasa Indonesia yang ada di situs TripAdvisor yang berhasil di scrap

No	Penilaian	Nama	Ulasan
1.	5,0	Michael R French	Rame dengan pengunjung, letaknya dekat kemana - mana.

2.	5,0	163KTY	Cocok untuk wisata edukasi sejarah. Kalau mengunjungi tempat sejarah bersama anak2 sebaiknya di siang hari
3.	5,0	Jade S	Sangat bagus untuk pembelajaran dan wisata. kurang leluasa dalam mengeksplor seluruh ruangan. Kurang tersedia tempat rehat setelah berkeliling.
4.	5,0	Lintu247	Secara keseluruhan gedung ini masih bagus dan terawat.
...	...		
250.	5,0	Nita	Tempatnya adem dan pemandangannya indah sekali

Metode K-Nearest Neighbor

Metode K-Nearest Neighbor (k-NN) pertama kali di perkenalkan awal tahun 1960. Metode ini bekerja intensif ketika diberikan data training yang besar dan tidak populer sampai dengan tahun 1960. Ketika tekin komputasi sedang berkembang (Baharuddin et al., 2019).

Metode

Metode yang digunakan pada penelitian ini merupakan metode penelitian eksperimen, dengan tahapan sebagai berikut:

a. Pengumpulan Data

Penulis mengambil data dari situs review *www.tripadvisor.com*. Data yang diambil terdiri dari 50 review positif dan 50 review negative.

b. Pengolahan Data

Preprocessing pada dataset yang dilakukan yaitu :

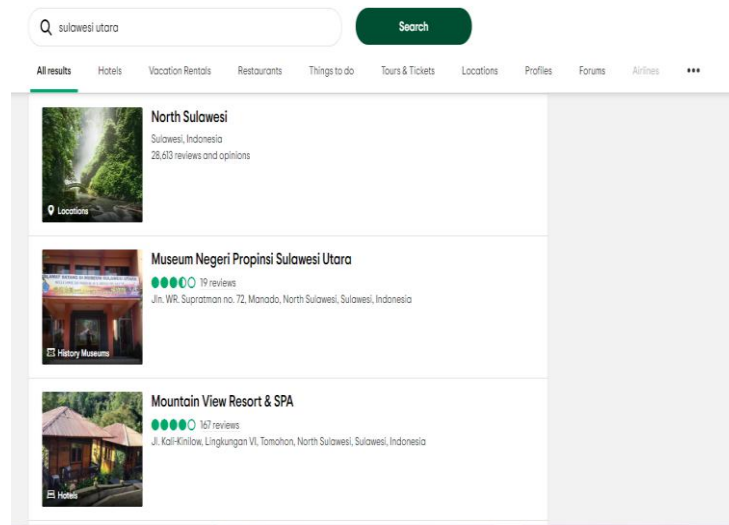
1) Analisis Data

Metode analisis yang digunakan dalam penelitian ini antara lain sebagai berikut.

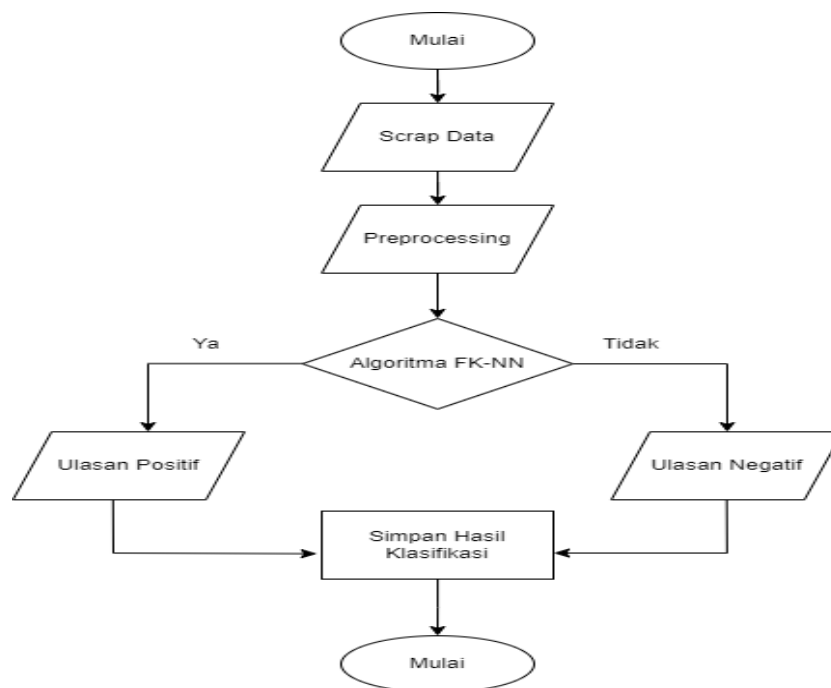
2) Data

Data dalam penelitian ini diperoleh dari kolom komentar situs TripAdvisor dengan cara web scraping menggunakan bahasa pemrograman Python. Data ulasan Sulawesi Utara berbahasa Inggris dapat diperoleh dari situs *web.www.tripadvisor.com* dengan nama "Sulawesi Utara", data yang akan diambil adalah data rating, name dan review. Data ulasan berbahasa Indonesia dapat diperoleh dari situs *web www.tripadvisor.co.id* dengan nama "Sulawesi Utara", data yang akan diambil adalah data penilaian, nama dan ulasan. Jumlah data yang

diperolehdari proses web scraping, untuk data ulasan berbahasa Indonesia sebanyak 250 ulasan.



Gambar 3. 1 Tampilan ulasan dari beberapa destinasi wisata di Sulawesi utara



Gambar 3. 2 Proses analisis sentiment menggunakan metode FK-NN

HASIL DAN PEMBAHASAN

Hasil Penelitian

1. Text Preprocessing

Text processing merupakan suatu proses untuk menyeleksi data text agar menjadi lebih terstruktur lagi, sehingga data menjadi siap untuk dilakukan analisis. Pada proses text processing ini terdapat beberapa tahapan diantaranya:

Tokenizing dan Stopwords Removal

Tahapan tokenizing adalah mengidentifikasi kata dalam teks menjadi beberapa urutan yang terpotong oleh spasi atau karakter special dan menghilangkan tanda baca dan angka pada teks. Berikut contoh hasil dari tokenizing. Stopwords Removal berfungsi untuk membuang kata-kata yang sering muncul dan bersifat umum, kurang menunjukkan relevansinya dengan teks. Kata-kata yang akan dibuang didefinisikan dalam stopword list.

Tabel 4. 1 Hasil Tokenizing dan Stopwords Removal

Kode	Case Folding	Hasil Tokenizing & Stopwords Removal
U1	Danau Linaw terkenal dengan danau tiga warnanya. Dengan pemandangan yang menakjubkan dan menenangkan, danau ini menawarkan ketenangan bagi para pengunjungnya. Dengan tiket seharga Rp 25 ribu, Anda bisa mendapatkan kopi atau teh gratis. Pasti akan lebih baik jika tidak hujan dan ada musik akustik lokal yang mengiringi.	danau linaw terkenal dengan danau tiga warnanya. dengan pemandangan yang menakjubkan dan menenangkan, danau ini menawarkan ketenangan bagi para pengunjungnya. dengan tiket seharga rp 25 ribu, anda bisa mendapatkan kopi atau teh gratis. tentunya akan lebih baik jika tidak hujan dan ada musik akustik lokal yang mengiringi.
U2	Danau dengan suasana yang cozy, dengan udaranya yang sangat sejuk, nyaman dikunjungi menjelang sore hari, kendalanya hanya jalanan menuju kesana masih relatif sempit	danau dengan suasana yang cozy dengan udaranya yang sangat sejuk nyaman dikunjungi menjelang sore hari kendalanya hanya jalanan menuju kesana masih relatif sempit
U3	Dingin pastinya. Pemandangannya sejuk, staffnya ada beberapa yg terlalu flat. Overall menarik untuk dikunjungi	dingin pastinya pemandangannya sejuk staffnya ada beberapa yg terlalu flat overall menarik untuk dikunjungi
U4	Pemandangan bagus, ada resto yg cukup besar, makanan lumayan dari segi harga & rasa, hanya bau belerang yang agak menyengat, cocok untuk santai rame2, klo hari libur padat & ramai	pemandangan bagus ada resto yg cukup besar makanan lumayan dari segi harga rasa hanya bau belerang yang agak menyengat cocok untuk santai klo hari libur padat ramai
U5	Bagi yang berminat, ini adalah tempat yang wajib dikunjungi karena ada banyak fenomena vulkanik yang berbeda untuk dilihat. Pastikan juga untuk mengunjungi Lahendong. Tempat ini terletak sedikit lebih jauh dan memberikan gambaran yang menarik dan lengkap tentang vulkanisme.	bagi yang berminat, ini adalah tempat yang wajib dikunjungi karena terdapat banyak fenomena vulkanik yang berbeda untuk dilihat. pastikan juga untuk mengunjungi lahendong. tempat ini terletak sedikit lebih jauh dan memberikan gambaran yang menarik dan lengkap tentang vulkanisme.

Sentimen dengan F-KNN

Setelah data diproses melalui text preprocessing, langkah selanjutnya adalah melakukan sentiment dengan menggunakan metode Fuzzy K-Nearest Neighbor. Namun sebelum mengimplementasikan sentimen menggunakan F-Knn pastikan bahwa data yang akan di analisis telah menjadi numerik dan menjadi tipe data float.

Mengonversi Kolom Rating Menjadi Numerik

python

Copy code

```
df['Rating'] = df['Rating'].apply(pd.to_numeric, errors='coerce')
```

- **Tujuan:** Memastikan bahwa semua nilai dalam kolom Rating adalah numerik (tipe float atau int).
- **Fungsi `pd.to_numeric`:** Fungsi ini mengonversi nilai dalam kolom menjadi tipe data numerik. Jika ada nilai yang tidak bisa dikonversi (misalnya, string yang tidak berupa angka), nilai tersebut akan diubah menjadi NaN (Not a Number).
- **Parameter `errors='coerce'`:** Memastikan bahwa jika ada nilai yang tidak bisa dikonversi menjadi numerik, nilai tersebut akan diganti dengan NaN daripada menimbulkan error.

Memeriksa Nilai NaN Setelah Konversi

python

Copy code

```
print(df['Rating'].isna().sum())
```

- **Tujuan:** Menghitung dan menampilkan jumlah nilai NaN dalam kolom Rating setelah proses konversi.
- **Fungsi `isna()`:** Menghasilkan seri boolean di mana True menunjukkan posisi nilai NaN.
- **Fungsi `sum()`:** Menjumlahkan nilai True untuk memberikan jumlah total nilai NaN dalam kolom Rating.

Menghapus Baris dengan Nilai NaN dalam Kolom Rating

- **Tujuan:** Menghapus baris dalam DataFrame df yang memiliki nilai NaN di kolom Rating.
- **Fungsi `dropna(subset=['Rating'])`:** Menghapus semua baris yang memiliki NaN dalam kolom Rating. Argumen `subset=['Rating']` menentukan bahwa penghapusan hanya didasarkan pada kolom Rating.

Tabel 4. 2 Tabel Sentimen dengan F-KNN

Kode	Case Folding	Sentimen
U1	Danau Linaw terkenal dengan danau tiga warnanya. Dengan pemandangan yang menakjubkan dan menenangkan, danau ini	Positif

	menawarkan ketenangan bagi para pengunjungnya. Dengan tiket seharga Rp 25 ribu, Anda bisa mendapatkan kopi atau teh gratis. Pastinya akan lebih baik jika tidak hujan dan ada musik akustik lokal yang mengiringi.	
U2	Danau dengan suasana yang cozy, dengan udaranya yang sangat sejuk, nyaman dikunjungi menjelang sore hari, kendalanya hanya jalanan menuju kesana masih relatif sempit	Positif
U3	Dingin pastinya. Pemandangannya sejuk, staffnya ada beberapa yg terlalu flat. Overall menarik untuk dikunjungi	Positif
U4	Pemandangan bagus, ada resto yg cukup besar, makanan lumayan dari segi harga & rasa, hanya bau belerang yang agak menyengat, cocok untuk santai rame2, klo hari libur padat & ramai	Positif
U5	Bagi yang berminat, ini adalah tempat yang wajib dikunjungi karena ada banyak fenomena vulkanik yang berbeda untuk dilihat. Pastikan juga untuk mengunjungi Lahendong. Tempat ini terletak sedikit lebih jauh dan memberikan gambaran yang menarik dan lengkap tentang vulkanisme.	Negatif

B. TF-IDF

TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan langkah dalam menghitung bobot kata (term) dari suatu data. Frekuensi istilah menghitung jumlah kemunculan suatu istilah dalam dokumen. di sisi lain, document frequency adalah jumlah dokumen di mana istilah tersebut muncul.

Berdasarkan analisis dan pengujian yang sudah dilakukan maka kesimpulan penelitian ini adalah algoritma k-NN mampu melakukan klasifikasi ulasan secara otomatis, hasil pengujian yang didapatkan dengan menggunakan total kurang lebih 250 data sampel dengan pembagian 80% *dataset* dan 20% data *test* mendapatkan rata-rata nilai akurasi sebesar 40%, nilai rata-rata presisi sebesar 40%, nilai rata-rata recall sebesar 100%. Dengan hasil pengujian tersebut bahwa algoritma k-NN cukup baik jika diimplementasikan dalam analisis sentimen. Proses klasifikasi semakin akurat jika *dataset*

yang digunakan dalam penelitian berjumlah banyak. Nilai akurasi, presisi, dan *recall* bisa lebih ditingkatkan jika data sampel yang digunakan lebih banyak.

KESIMPULAN

Kesimpulan dari penelitian ini adalah bahwa penggunaan algoritma K-Nearest Neighbor (KNN) dalam analisis sentimen terhadap ulasan atau review destinasi atau tempat wisata menunjukkan hasil yang cukup memuaskan. Dengan menggunakan metode tersebut, penelitian berhasil mengimplementasikan langkah-langkah text preprocessing, TF-IDF, dan penggunaan F-KNN secara sistematis untuk mendapatkan prediksi sentimen yang akurat dan dapat diandalkan.

SARAN

1. Penggunaan dataset yang lebih besar dan representatif dapat membantu meningkatkan akurasi dan generalisasi model. Dengan demikian, disarankan untuk mengumpulkan lebih banyak data ulasan dari berbagai sumber dan destinasi wisata.
2. Pemilihan parameter *k* dalam algoritma KNN perlu diperhatikan dengan cermat. Pengujian lebih lanjut dengan variasi nilai *k* dapat membantu menemukan nilai optimal yang sesuai dengan karakteristik data.

DAFTAR PUSTAKA

- Fikri Baihaqi, G., Ratnawati, D. E., & Hanggara, B. T. (2022). Analisis Sentimen Wisata Alun-Alun Kota Batu menggunakan Algoritma Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 6(12), 6010–6018. <http://j-ptiik.ub.ac.id>
- Korompot, M. A., Rantung, V. P., Kembuan, O., Informatika, T., Teknik, F., & Manado, U. N. (2024). *Aplikasi Evaluasi Analisis Sentimen Terhadap Layanan Akademik Menggunakan Algoritma K - Nearest Neighbor Berbasis Web*. 18–26.
- Nurlaili, F., & Irawan, R. H. (2021). Penerapan Metode KNN Dalam Menentukan Titik Lokasi Wisata Di Kecamatan Ngancar. *Prosiding SEMNAS INOTEK ...*, 272–276. <https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/1143%0Ahttps://proceeding.unpkediri.ac.id/index.php/inotek/article/download/1143/742>
- Rizky, R., Tigusti, W., Ratnawati, D. E., & Anam, S. (2018). Implementasi Fuzzy K-Nearest Neighbor (FK-NN) Untuk Mengklasifikasi Fungsi Senyawa Berdasarkan Simplified Molecular Input Line Entry System (SMILES). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(12), 6331–6338.
- L. Wati Evelina, “Komunitas adalah Pesan: Studi Netnografi Virtual di Situs Wisata TripAdvisor,” *War. ISKI*, vol. 01, no. 02, pp. 65-74[1] L. Wati Evelina, “Komunitas adalah Pesan:”, 2018.
- D. Tabuni, G. H. M. Kapantouw, and L. R. Rengkung, “Strategi Pengembangan Kawasan Wisata Danau Linouw Di Kota Tomohon Provinsi Sulawesi Utara,”

Agri-Sosioekonomi, vol. 11, no. 3A, p. 143, 2015, doi:
10.35791/agrsosek.11.3a.2015.10624.

- R. Sarudin, “Analisis Online Review Tripadvisor . Com Terhadap Minat Pembelian Produk Jasa Akomodasi Di Hotel Manhattan Tripadvisor . Com Review Online Analysis on the Interest of Buying Accommodation Services in Hotel,” *J. Hosp. dan Pariwisata*, vol. 7, pp. 33–46, 2021
- .