

Sentimen Analisis Rumah Makan Rawon Sakinah Dengan Metode *Naïve Bayes Classifier*

M.Ridwan¹⁾, Rudi Hariyanto²⁾, Muslim Alamsyah³⁾

^{1,2,3)} Program Studi Informatika Universitas Merdeka Pasuruan

Article Info

Article history:

Received Juli 26, 2024
Revised Agustus 14, 2024
Accepted November 11, 2024

Keywords:

*Analysis,
Sentimen,
Metode naïve bayes
Classifier*

ABSTRACT

Rumah Makan Rawon Sakinah merupakan rumah makan yang terkenal di kota Pasuruan. Jika anda sedang mencari kuliner khas Kota Pasuruan, warga sekitar dengan suara bulat akan menjawab Rawon Sakinah. Rumah makan ini sudah berdiri selama 30 tahun. beralamat di Jalan Bangilan No. 80 Kota Pasuruan. Pada penelitian ini akan dianalisis sentimen pengunjung Rumah Makan Rawon Sakinah berdasarkan review Google. Untuk mengetahui review pengunjung Rumah Makan Rawon Sakinah apakah review sentimennya positif atau negatif. Tahapan analisis data yaitu pengolahan teks untuk membersihkan data, pembobotan kata, pelabelan data ke dalam kelas positif dan negatif, pengklasifikasian, dan visualisasi data dengan wordcloud. sentimen positif berjumlah 237 dan sentimen negatif berjumlah 13. Pada penelitian ini menggunakan metode Naïve Bayes. Nilai akurasi yang diperoleh untuk metode Naïve Bayes sebesar 92% dengan sentimen positif. Dari hasil penelitian sentimen review didapatkan bahwa mayoritas pengunjung memiliki sentimen positif.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Muslim Alamsyah

Program Studi Teknik Informatika, Universitas Merdeka Pasuruan, Indonesia

Email: muslim@unmerpas.ac.id

1. PENDAHULUAN

Kemajuan yang pesat di bidang teknologi dan komunikasi bisa sangat dirasakan pada saat ini, tak luput dengan perkembangan teknologi internet yang semakin maju dan meluas, kini persebaran internet sudah bisa dirasakan hingga kepenjuru negeri.

Bahkan untuk orang yang sekedar ingin mencari referensi untuk rencana kuliner, kini sudah dapat memanfaatkan banyak media sosial, untuk mencari foto-foto hingga ulasan atau review tentang kuliner tersebut dari para pengunjung sebelumnya. Media sosial yang sering digunakan untuk acuan kuliner adalah google, google adalah situs web online terbesar didunia, situs ini dapat membantu para kuliner domestic atau mancanegara dalam perencanaan perjalanan kuliner mereka.

Rumah Makan Rawon Sakinah

Rumah Makan Rawon Sakinah merupakan rumah makan terkenal di Kota Pasuruan dengan ciri khas Sup Rawonnya. *Rawon* adalah masakan Indonesia berupa sup daging berkuah hitam dengan campuran bumbu khas yang menggunakan kluwek.

Outscraper

Outscraper adalah layanan penyedia data (pengikis data) yang mudah digunakan. Platform ini menyediakan beberapa fitur-fitur canggih yaitu pengikisan (scraper) data google maps, google maps ulasan, ekstraktor lalulintas & arahan google maps, produk Amazon, ulasan Amazon, email dan kontak. Outscraper memiliki dua layanan yaitu layanan berbayar dan gratis. Layanan berbayar memiliki fitur data yang lengkap dan layanan gratis hanya beberapa fitur data.

Text Mining

Menurut (Suyanto 2019) data mining ditunjukkan untuk mengekstrak(mengambil inti sari) pengetahuan dari kesimpulan data sehingga didapatkan struktur yang dapat dimengerti manusia, serta meliputi basis data dan manajemen data, prapemrosesan data, pertimbangan model, dan inferensi, ukuran keterikatan, pertimbangan kompleksitas pasca pemrosesan terhadap struktur yang ditemukan, visualisasi, dan online updating.

Pembobotan kata TF-IDF

Pembobotan kata adalah proses pemberian bobot term pada dokumen(term weighting). Pembobotan kata dasar berdasarkan dengan menghitung frekuensi term yang sering muncul pada document. Frekuensi kemunculan(*term frequency*) adalah untuk mengetahui seberapa term yang mewakili *document*.

Term Frequency – *Inverse Document frequency*(TF-IDF)

TF-IDF adalah bobot *term* dalam terhadap *document*. Berikut rumus TF-IDF.

$$w_{ij} = tf_{ij} \times idf_j$$

Sentiment

Analisis sentiment merupakan pendapat, perasaan, sikap, dan emosi orang-orang sehubungan dengan produk, organisasi, layanan, film, individu, politik, dan topic tertentu lainnya. Untuk menganalisa pendapat dalam teks itu sangat umum untuk menilai mereka dengan polaritas dan dengan demikian dapat mengkalsifikasikan kata atau frasa di antara tiga kategori : positif, negative atau netral.(Guevara,2018).

Naïve Bayes Classifier

Naïve bayes classifier merupakan teknik klasifikasi berdasarkan teorema bayes dengan asumsi independensi di antara para predictor. Klasifikasi Naïve Bayes meprediksi peluang masa di depan berdasarkan pengalaman di masa sebelumnya (Imron2019).

$$P(A|B) = \frac{(P(B|A) * P(A))}{P(B)}$$

$P(A/B)$ = data dengan kelas tidak dikenal

$P(A)$ = h

ipoteis data X adalah kelas Khusus

$P(B)$ = probabilitas hipotesis H didasarkan pada kondisi

$P(B|A)$ = probabilitas hipotesis X didasarkan pada kondisi H

$P(X)$ = probabilitas X

2. METODE

Penelitian yang akan dilakukan dalam pengerjaan Skripsi ini dari awal hingga akhir.

Pengumpulan data ulasan

Pengambilan data atau crawling google ulasan dengan cara mendownload data ulasan melalui web Outscraper.com dengan menyalin alamat web google ulasan.

populasi dan data sampel

Populasi dalam penelitian ini adalah semua data google ulasan Rumah Makan Rawon Sakinah yang terdapat pada google ulasan dengan cara mendownload ulasan melalui web Outscraper.com sebanyak 250 ulasan pengunjung.

variable penelitian

Dalam penelitian ini terdapat beberapa variabel diantaranya seperti berikut.

1. Penilaian ulasan atau rating bintang : pengunjung kuliner Rumah Makan Rawon Sakinah , penilaian rating ini bernilai 1 sampai 5 bintang.
2. Nama : pengunjung kuliner Rumah Makan Rawon Sakinah yang menulis ulasan atau mereview di google ulasan.
3. Review ulasan : pengunjung memberi deskripsi isi ulasan pengunjung Rumah Makan Rawon Sakinah.

Analisis ulasan data

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

1. Deskriptif adalah untuk mengetahui gambaran umum Rumah Makan Rawon Sakinah tentang ulasan pengunjung .
2. Processing data adalah bertujuan untuk membersihkan data ulasan google agar bisa diolah.

3. Klasifikasi data adalah bertujuan untuk mengetahui kelas sentiment positif dan negative .
4. Word cloud adalah untuk perhitungan frekuensi kata topic apa yang paling banyak muncul dalam google ulasan pengunjung Rumah Makan Rawon Sakinah.

3. HASIL DAN PEMBAHASAN

Tahapan untuk menguji hasil dari penggunaan data di Rumah Makan Rawon Sakinah sebanyak 250 Ulasan.

Pengambilan ulasan data (crawling data)

Ulasan data penelitian diperoleh dari google ulasan dengan cara web *scraping* menggunakan Outscraper.com.

Pemrosesan Data

Pemrosesan data melalui analisi deskriptif berfungsi untuk mendeskripsikan, atau memberi gambaran secara umum tentang sentiment atau penilaian tentang Rumah Makan Rawon Sakinah berdasarkan pada data google ulasan.

Text Preprocessing

Data text perlu di bersihkan sebelum masuk ke model machine learning, proses pembersihan ini disebut dengan text preprocessing. Tahapan text preprocessing diantaranya yaitu cleasing, tokenizing , filtering dan lain lain.

a. Case Folding

case folding adalah tahapan text preprocessing yang berfungsi untuk mengubah huruf besar menjadi huruf kecil atau lowertext.

b. Filtering

filtering adalah pembersihan text stopword pada tahapan text preprosesing. Stopword adalah kata-kata yang tidak memiliki makna. Contoh stopwords dalam bahasa inggris “if”, “the”, “but”, “and”, “or”.

c. Cleansing

Cleansing adalah tahapan text preprocessing bertujuan pembersihan text dari tab, new line, back slice, mention, link, hastag, URL.

d. Remove Number

Remove number adalah tahapan text preprocessing bertujuan pembersihan angka

e. Remove Punctuation

Remove punctuation adalah pada tahapan text preprocessing bertujuan pembersihan text dari tanda baca .

f. Remove Short Words

Remove short words adalah tahapan text preprocessing bertujuan untuk pembersihan text dari kata – kata pendek kurang dari 3 huruf yg tidak penting misalnya “ex”, “ur”, “go”.

g. Tokenizing

Tokenizing adalah tahapan text preprocessing berfungsi sebagai pemisahan text menjadi potongan per-kata.

h. Remove Extend Stopwords

Remove extends stopwords adalah tahapan text preprocessing bertujuan sebagai pembersih text dari stopwords yang sebelumnya tidak terhapus pada filtering.

Pembobotan TF-IDF

Melewati tahapan text preprocessing selanjutnya data google ulasan harus dikonversikan terlebih dahulu dalam bentuk angka. mengkonversikan data google ulasan menjadi angka dengan menggunakan pembobotan kata TF-IDF.

Tabel 1. hasil TF-IDF

Term	Weight
Good	0.046526
Rawon	0.045862
Delicious	0.038802
Food	0.029185
Taste	0.028722

Analisis Sentiment

Hasil dari pembobotan kata maka proses selanjutnya pelabelan kelas sentiment. Menurut sumber datanya analisis sentiment dibedakan menjadi dua jenis, yaitu jenis pada level document dan analisis pada level kalimat. Dalam penelitian ini data akan dilakukan analisis sentiment untuk tiap ulasan, maka dalam penelitian ini berjenis pada level kalimat (fine grained sentiment analysis).

Tiga cara pelabelan dalam analisis sentiment google ulasan, yaitu berdasarkan nilai rating, kamus lexicon dan pemberian label secara manual satu-persatu. berikut penjelasan kelebihan dan kekurangan dari masing-masing pelabelan.

a. Pelabelan Berdasarkan Rating

Pelabelan ini biasanya melabelkan data menjadi kelas sentiment positif dan kelas sentiment negative, contoh data bersentiment positif jika nilai rating 5 bintang dan nilai negative jika nilai rating kurang dari >5. Kelebihan dari metode ini adalah mudah diterapkan, hanya merubah nilai rating dengan rumus if. akan tetapi kelemahan dari metode ini adalah nilai rating tidak menunjukkan bahwa ulasan dengan rating yang baik bersentiment positif, misalnya dalam kalimat ulasan tersebut mengandung kalimat negasi.

b. Pelabelan Berdasarkan Lexicon

Data yang dilebelkan dengan cara memilah satu ulasan berdasarkan kata kedalam kata positif atau negative, jika kata tidak terdapat pada kata positif atau negative maka data tersebut masuk kategori netral. Jumlah kata kemudian dilakukan kalkulasi dan diketahui ulasan tersebut masuk ke kategori sentiment positif, netral dan negative. Kelebihan proses ini adalah lebih akurat dibandingkan cara pelabelan berdasarkan rating, cara ini memeriksa ulasan berdasarkan kata yang terkandung. Kelemahan cara ini jika terdapat kata dalam kamus lexicon yang tidak sesuai dengan keadaan pada kenyataannya.

c. Pelabelan Manual Dengan Membaca Ulasan

Cara pelabelan ini dilakukan dengan membaca ulasan satu persatu lalu ditentukan ulasan tersebut masuk kedalam sentiment positif atau negative. kelebihan cara ini dari hasil yang didapat lebih akurat atau sesuai dengan kenyataan dan tidak perlu proses komputasi. Namun kelemahan dari cara ini dikhawatirkan adanya kecenderungan opini seseorang dan cara ini sulit diterapkan pada data skala besar.

Proses pelabelan kelas sentiment pada google ulasan menggunakan library vader lexicon secara otomatis dan untuk memanggil fungsi Sentiment Intensity Analyzer dari package NLTK. Data dilabelkan menjadi dua (2) yaitu kelas sentiment positif dan kelas sentiment negative. Data diproses dengan menerapkan algoritma sentiment analisis vader. Algoritma vader menghitung skor dari persentiment ulasan (kalimat). skor yang didapatkan dari proses tersebut berupa positif, negative dan netral. Hasil skor digabungkan sehingga menghasilkan nilai compound. Nilai compound merupakan matriks yang menghitung skor yang telah dinormalisasikan antara -1 dan +1. Nilai compound diatas ≥ 0 dinyatakan kelas sentiment

positif sedangkan nilai compound dibawah <0 dinyatakan kelas sentiment negative. Dapat juga ditentukan dari besar skor presentase kata positif dan negative dalam satu kalimat ulasan tersebut. jika kalimat ulasan presentase skor positif lebih tinggi dari kata negative maka dinyatakan kelas sentiment positif dan sebaliknya jika nilai presentase skor negative lebih tinggi dari positif maka dinyatakan kelas sentiment negative. Script proses pelebelan data dan hasil sebagai berikut.

Data Latih Dan Data Uji

Data latih digunakan untuk klasifikasi yang berisi pengetahuan digunakan untuk prediksi kelas sentiment baru, sebelum klasifikasi data ulasan terlebih dahulu perlu dibagi menjadi data latih dan data uji. Semakin banyak data yang dilatih maka semakin baik algoritma dalam memahami data.

Data uji untuk melihat presentase algoritma klasifikasi dalam keberhasilan melakukan klasifikasi dengan benar. Analisis ini dibutuhkan data 80% pada data latih dan data uji 20% di masing masing data. Berikut proses partisi data latih dan data uji dengan script dibawah ini.

Tabel 2. Hasil data latih dan data uji

DATA	ULASAN
LATIH	200
UJI	50

Klasifikasi

Klasifikasi pada data uji dengan mempelajari pengetahuan yang di dapat pada data latih. Data latih terdapat kelas sentiment positif dan kelas sentiment negative yang dipelajari cirri-ciri kata yang terdapat pada kelas sentiment masing-masing. Berikut script metode Naïve Bayes.

Table 3. Coufision matrix

True table	Predict table	
	Negative	Positif
Negative	0	1
Positif	2	55

$$\text{perhitungan akurasi } 80\% = \frac{0+55}{2+55+1} 100\% = \frac{55}{58} 100\% = 95\%$$

Table 4. Coufision matrix

True table	Predict table	
	Negative	Positif
Negative	0	4
Positif	0	46

$$\text{perhitungan akurasi } 70\% = \frac{0+71}{3+71+1} 100\% = \frac{71}{75} 100\% = 95\%$$

Table 5. Coufision matrix

True table	Predict table	
	Negative	Positif
Negative	0	1
Positif	1	23

$$\text{perhitungan akurasi } 90\% = \frac{0+21}{1+23+1} 100\% = \frac{23}{25} 100\% = 92\%$$

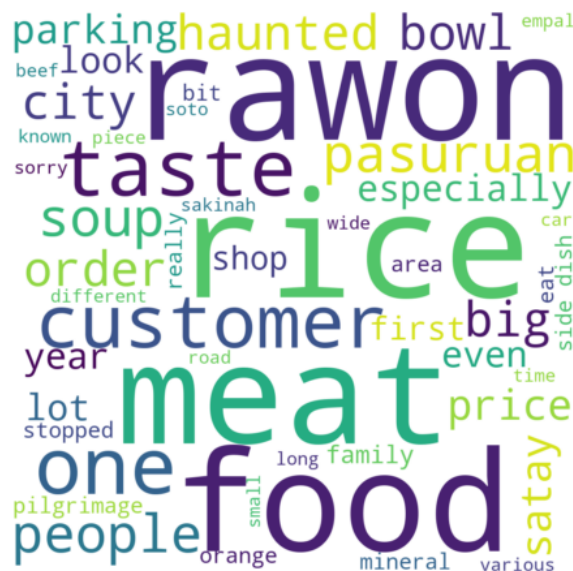
Visualisai

Visualisai bertujuan untuk mengekstrak informasi berupa topic yang paling banyak dibicarakan atau sering diulas oleh pengunjung Rumah Makan Rawon Sakinah. sehingga banyak teks ulasan yang ada, untuk diambil informasi yang paling penting. Untuk visualisai hasil klasifikasi sentiment menggunakan Word Cloud.

Wordcloud adalah representasi dari suatu data ulasan yang menampilkan kata yang sering di ulas dan kata-kata penting yang sering muncul terdapat pada data ulasan tersebut. Semakin tinggi presentase kemunculan kata maka semakin tinggi pula frekuensi kemunculan kata tersebut.

Data Sentiment Positif

Data ulasan kelas sentiment positif di indefikasikan berdasarkan banyak ulasan positif yang berdasarkan banyak nya frekuensi kata dalam ulasan. Berikut hasil visualisasi ulasan positif dari hasil ekstraksi informasi yang di dapat pada ulasan pengunjung Rumah Makan Rawon Sakinah . script proses untuk melihat kata yang sering muncul pada ulasan positif berikut dibawah ini.



Data Sentiment Negative

- Matematika*, vol. 5, pp. 808–813, 2022, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [6] I. Rahmawati and T. R. Fitriani, “Analisis Sentimen Menggunakan Algoritma Logistic Regression Pada Penerbangan Lion Air berdasarkan Ulasan Pengguna Platform Online,” *Jejaring Penelitian dan Pengabdian Masyarakat (JPPM)*, vol. 1, no. 1, pp. 3024–8167, 2023.
 - [7] J. A. Zulqornain, Indriati, and P. P. Adikara, “Analisis Sentimen Tanggapan Masyarakat Aplikasi Tiktok Menggunakan Metode Naïve Bayes dan Categorical Proportional Difference (CPD),” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 7, pp. 2886–2890, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>
 - [8] A. Pradhan, M. R. Senapati, and P. K. Sahu, “Improving sentiment analysis with learning concepts from concept, patterns lexicons and negations: Improving sentiment analysis with learning concepts,” *Ain Shams Engineering Journal*, vol. 13, no. 2, Mar. 2022, doi: 10.1016/j.asej.2021.08.004.
 - [9] B. K. Hananto, A. Pinandito, and A. P. Kharisma, “Penerapan Maximum TF-IDF Normalization Terhadap Metode KNN Untuk Klasifikasi Dataset Multiclass Panichella Pada Review Aplikasi Mobile,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 12, pp. 6812–6823, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>
 - [10] P. Bafna, D. Pramod, and A. Vaidya, “Document Clustering: TF-IDF Approach,” in *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, 2016, pp. 61–66.
 - [11] B. Das and S. Chakraborty, “An Improved Text Sentiment Classification Model Using TF-IDF and Next Word Negation.”
 - [12] P. Sudhir and V. D. Suresh, “Comparative study of various approaches, applications and classifiers for sentiment analysis,” *Global Transitions Proceedings*, vol. 2, no. 2, pp. 205–211, Nov. 2021, doi: 10.1016/j.gltp.2021.08.004.
 - [13] H. Rahab, A. Zitouni, and M. Djoudi, “SANA: Sentiment analysis on newspapers comments in Algeria,” *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 7, pp. 899–907, Sep. 2021, doi: 10.1016/j.jksuci.2019.04.012.
 - [14] F. F. Irfani, M. Triyanto, A. D. Hartanto, and Kusnawi, “Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine,” *Jurnal Bisnis, Manajemen dan Informatika*, vol. 16, no. 3, pp. 258–266, 2020.
 - [15] Sulistiyanto, S. (2022). Perancangan Software The Smart Tourism Bali Dengan Android. Joutica, 7(1), 551-555.
 - [16] I. Novitasari, T. B. Kurniawan, D. A. Dewi, and Misinem, “Analisis Sentimen Masyarakat Terhadap Tweet Ruang Guru Menggunakan Algoritma Naive Bayes Classifier (NBC),” *Jurnal Mantik*, vol. 6, no. 3, pp. 2685–4236, 2022.
 - [17] D. Ardiansyah, A. Saepudin, R. Aryanti, E. Fitriani, and Royadi, “Analisis Sentimen Review Pada Aplikasi Media Sosial Tiktok Menggunakan Algoritma K-NN dan SVM Berbasis PSO,” *Jurnal Informatika Kaputama (JIK)*, vol. 7, no. 2, pp. 233–241, 2023.