



Analisis Sentimen Ulasan Pengunjung Terhadap Tempat Wisata Religi Walisongo Menggunakan Metode Supervised Learning

Pandu Rizki Maulidiah

Universitas Pembangunan Nasional “Veteran” Jawa Timur

Amalia Anjani Arifiyanti

Universitas Pembangunan Nasional “Veteran” Jawa Timur

Dhian Satria Yudha Kartika

Universitas Pembangunan Nasional “Veteran” Jawa Timur

Jl. Raya Rungkut Madya, Gunung Anyar, Surabaya, Jawa Timur (60294)

Korespondensi penulis: 19082010046@student.upnjatim.ac.id

Abstract: *The rapid development of information technology has made visitors of Walisongo Tourisms able to provide reviews through the map application that often used by public, Google Maps. With the reviews given by other users, it can help others who want to visit the place and become a source of data material for the local government in optimizing the services, facilities, and infrastructure of Walisongo Tourisms. To help simplify the process of processing reviews, a sentiment analysis process is needed which is assisted by an algorithm that will be in charge of classifying these reviews. This study aims to carry out a sentiment analysis of visitor ratings of the Walisongo Tourisms using supervised learning, such as Decision Trees (CART), K-Nearest Neighbor, Multinomial Naïve Bayes, and Support Vector Machine kernel RBF. The result of this study indicate that the model who has the highest accuracy value is SVM-RBF model classification with accuracy value 87,12% and f1-score value of negative class 89%, 68% for neutral class, and 91% for positive class.*

Keywords: *Classification, Reviews, Sentiment, Supervised Learning*

Abstrak: Pesatnya perkembangan teknologi informasi membuat pengunjung tempat wisata religi Walisongo dapat memberikan ulasan melalui aplikasi peta yang sering digunakan oleh masyarakat yakni Google Maps. Dengan ulasan yang diberikan pengguna lain, dapat membantu sesama pengguna yang ingin mengunjungi tempat tersebut serta menjadi sumber masukan bagi pemerintah setempat dalam mengoptimalkan pelayanan, sarana, dan prasarana wisata religi Walisongo. Untuk membantu mempermudah proses pengolahan ulasan, dibutuhkan proses analisis sentimen terkait ulasan tersebut yang dibantu dengan suatu algoritma yang akan bertugas untuk mengklasifikasikan ulasan tersebut. Penelitian ini bertujuan untuk melakukan analisis sentimen penilaian pengunjung terhadap tempat wisata religi Walisongo dengan metode *supervised learning*, yakni *Decision Trees (CART)*, *K-Nearest Neighbor*, *Multinomial Naïve Bayes*, dan *Support Vector Machine – kernel RBF*. Penelitian ini bertujuan untuk membuktikan algoritma terbaik dalam klasifikasi sentimen. Hasil penelitian ini menunjukkan bahwa model yang memiliki nilai akurasi tertinggi yakni model dari klasifikasi SVM-RBF dengan nilai akurasi tertinggi dari model lainnya, yakni sebesar 87,12% dengan nilai *f1-score* sebesar 89% untuk nilai negatif, 68% untuk nilai netral, dan 91% untuk nilai positif.

Kata kunci: Klasifikasi, Sentimen, *Supervised Learning*, Ulasan.

LATAR BELAKANG

Berwisata adalah suatu kegiatan yang tak bisa dipisahkan dalam bermasyarakat. Dengan berwisata, seseorang dapat merasakan suasana serta kesenangan baru yang belum pernah didapat selama beraktivitas sehari-hari (Imron, 2019). Dengan tersebarnya berbagai tempat wisata di Indonesia, kegiatan berwisata harus dimanfaatkan dan dioptimalkan dari segi pelayanan maupun fasilitas. Tempat wisata memiliki beberapa jenis, salah satunya adalah tempat wisata religi.

Tempat wisata religi adalah tempat wisata dengan kegiatan yang berkaitan dengan agama terkait (Marsono, 2016). Salah satu wisata religi yang terkenal adalah Makam Walisongo yang berlokasi di beberapa kota di pulau Jawa. Walisongo adalah sebutan untuk kesembilan wali yang menyebarkan serta menyiarkan agama Islam di pulau Jawa (Amarudin, 2020). Kesembilan tempat wisata religi tersebut ramai dikunjungi oleh pengunjung, baik dari dalam maupun luar kota. Dengan intensitas pengunjung yang cukup tinggi, pemerintah setempat harus menyediakan pelayanan, sarana, serta prasarana yang memadai agar pengunjung dapat berwisata dengan nyaman.

Dengan pesatnya perkembangan teknologi informasi, pengunjung dapat memberikan tanggapan, komentar, maupun ulasan melalui aplikasi peta yang sering digunakan oleh masyarakat yakni Google Maps. Ulasan yang diberikan pengunjung dapat dijadikan sumber masukan bagi pemerintah setempat dalam mengoptimalkan pelayanan, sarana, dan prasarana wisata religi Walisongo. Untuk membantu mempermudah proses pengolahan ulasan, dibutuhkan proses analisis sentimen terkait ulasan tersebut yang dibantu dengan suatu algoritma yang akan bertugas untuk mengklasifikasikan ulasan tersebut.

Penelitian ini bertujuan untuk melakukan analisis sentimen penilaian pengunjung terhadap tempat wisata religi Walisongo dengan metode *supervised learning*. *Dataset* yang akan digunakan yakni data ulasan pengguna Google Maps yang telah mengunjungi Makam Walisongo dalam rentang waktu 2020 – 2022 dengan teknik *web scraping*. Data yang diperoleh dari *web scraping* akan melalui proses *text mining*, *text preprocessing*, dan *term weighting*. Kemudian dari hasil pembobotan tersebut akan diklasifikasikan menggunakan empat algoritma dari *supervised learning*, yakni *Decision Trees (CART)*, *K-Nearest Neighbor*, *Multinomial Naïve Bayes*, dan *Support Vector Machine – kernel RBF*. Dari hasil penelitian ini diharapkan dapat mengetahui algoritma yang terbaik untuk klasifikasi sentimen.

KAJIAN TEORITIS

Dasar Teori

1. Analisis Sentimen

Analisis sentimen adalah bidang pengolahan bahasa alami dan komputasi linguistik yang bertujuan untuk menganalisis pendapat, evaluasi, sikap, dan emosi seseorang apakah orang tersebut berkenaan dengan suatu topik (Khofifah, Rahayu, & Yusuf, 2022).

2. *Decision Tree Classifier (CART)*

Decision Tree adalah salah satu dari metode *supervised learning* yang merepresentasikan sebuah pohon dengan atribut sebagai *node* (Somantri & Dairoh, 2019). Salah satu dari *decision tree* adalah Algoritma CART yang melakukan pendekatan non parametrik tanpa memerlukan

asumsi distribusi (Pahmi, Saepudin, Maesarah, Solehudin, & Wulandari, 2018). Untuk menentukan atribut sebagai akar dari *Decision Tree* diperlukan rumus untuk menghitung *entropy* dan *gain*. Berikut rumus dari *entropy* dan *gain*.

a. Rumus *Gain*

$$Gain(A) = Entropy(S); \sum_{i=1}^n P_i \times \log_2(P_i)$$

Dengan keterangan :

- S = Himpunan Kasus
 A = Atribut
 n = Jumlah Sampel
 |S_i| = Jumlah Kasus pada Partisi ke-i
 |S| = Jumlah Kasus dalam S

b. Rumus *Entropy*

$$Entropy(S_1, S_2, S_3, \dots, S_n) = \sum_{i=1}^n P_i \times \log_2$$

Dengan keterangan :

- S = Himpunan Kasus
 n = Jumlah Sampel
 P_i = Proporsi Kelas

3. *K-Nearest Neighbor Classifier*

K-Nearest Neighbor (selanjutnya disingkat menjadi k-NN) adalah salah satu metode klasifikasi dengan menghitung jarak antara dua titik pada *data training* dan titik pada *data testing* yang didefinisikan dengan rumus Euclidean sebagai berikut (Surohman, Aji, Rousyati, & Wati, 2020).

$$d = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2}$$

Dengan keterangan :

- d = Jarak Euclidean
 x_{2i} = Nilai pada *data testing* ke-i
 x_{1i} = Nilai pada *data training* ke-i
 p = Banyaknya atribut

4. *Multinomial Naïve Bayes Classifier*

Multinomial Naïve Bayes yang merupakan algoritma klasifikasi dengan menentukan kategori berdasarkan kata yang muncul dan jumlah kemunculan (Wardani, Prahutama, & Kartikasari, 2020). *Multinomial Naïve Bayes* memiliki persamaan yang menyatakan bahwa probabilitas dokumen d yang terletak pada kategori c sebagai berikut.

$$P(c|d) \propto P(c) \prod_{1 \leq k \leq n_d} P(t_k | c)$$

Dengan keterangan :

$P(t_k|c)$ = Conditional Probability dari t_k

$P(c)$ = Prior Probability

Kategori terbaik dari *Multinomial* adalah kategori yang memiliki nilai *Maximum a Posteriori* (MAP) class C_{map} :

$$P(t_k|c) = \frac{N_k + 1}{|V| + N'}$$

Dengan keterangan :

$|V|$ = Jumlah seluruh *term* pada *data training*

N_k = Jumlah kemunculan t_k

N' = Jumlah total *term* pada dokumen

5. Support Vector Machine – Kernel RBF

Support Vector Machine (selanjutnya disingkat dengan SVM) adalah salah satu algoritma yang mengklasifikasikan data berdasarkan *hyperplane* (garis pembatas) yang memisahkan antara suatu kelas dengan kelas lain (Fitri, Yuliani, Rosyida, & Gata, 2020). Pada penelitian ini akan menggunakan *kernel RBF* dikarenakan *kernel* ini paling banyak digunakan untuk memecahkan masalah klasifikasi data. Berikut merupakan rumus persamaan dari *kernel RBF* :

$$K(x_i, x_j) = \exp\{-\gamma \|x_i - x_j\|^2\}, \gamma > 0$$

Penelitian Terdahulu

Penelitian yang pertama yakni dilakukan oleh (Ginantra, Yanti, Prasetya, Sarasvananda, & Wiguna, 2022) dengan judul Analisis Sentimen Ulasan Villa di Ubud Menggunakan Metode *Naïve Bayes*, *Decision Tree*, dan K-NN. Penelitian tersebut dilakukan dengan tujuan untuk mengetahui opini wisatawan terkait pengalaman berlibur di Villa daerah Ubud sehingga pengelola villa dapat menjaga kualitas dan citra positif di dunia pariwisata. Metode dari penelitian ini menggunakan *Naïve Bayes*, *Decision Tree*, dan K-NN. Adapun hasil dari penelitian ini adalah nilai akurasi terbesar diraih oleh K-NN sebesar 91,26%, kemudian nilai akurasi *decision tree* sebesar 87,78%, dan *naïve bayes* mendapat nilai akurasi sebesar 86,66%.

Penelitian kedua yakni dilakukan oleh (Ilmawan & Mude, 2020) dengan judul Perbandingan Metode Klasifikasi *Support Vector Machine* dan *Naïve Bayes* untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store. Penelitian ini bertujuan untuk mengukur kinerja SVM jika dibandingkan dengan metode klasifikasi lain selain *Maximum Entropy*.

Metode dari penelitian ini menggunakan *Support Vector Machine* dan *Naïve Bayes*. Adapun hasil dari penelitian ini adalah hasil akurasi SVM lebih tinggi dibandingkan dengan hasil akurasi dari *Naïve Bayes* dengan perolehan nilai akurasi SVM sebesar 81,46% dan nilai akurasi *Naïve Bayes* sebesar 75,41%.

Penelitian ketiga yakni dilakukan oleh (Fitri, Yuliani, Rosyida, & Gata, 2020) dengan judul Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma *Naïve Bayes*, *Random Forest*, dan *Support Vector Machine*. Penelitian ini bertujuan untuk mengumpulkan ulasan-ulasan dari penilaian aplikasi yang akan diolah menjadi sebuah kesimpulan aplikasi belajar *online* manakah yang memiliki sentimen terbaik. Metode dari penelitian ini menggunakan *Naïve Bayes*, *Random Forest*, dan *Support Vector Machine*. Adapun hasil dari penelitian ini yakni dari hasil *preprocessing* didapati bahwa komentar positif lebih banyak daripada komentar negatif. Kemudian untuk pengujian didapati hasil bahwa nilai akurasi tertinggi diraih oleh *Random Forest* sebesar 97,16% dan nilai AUC (*Area Under Curve*) sebesar 0,996. Kemudian disusul SVM sebesar 96,01% dan AUC 0,543, serta nilai akurasi terendah diraih *Naïve Bayes* sebesar 94,16% dan AUC sebesar 0,999.

METODE PENELITIAN

Data ulasan yang telah *discraping* akan diberi kelas label positif, negatif, dan netral secara manual. Kemudian data ulasan tersebut akan melalui tahap *text preprocessing* dan *term weighting* agar dapat diproses pada saat tahap perancangan model. Data yang telah melalui *text preprocessing* dan *term weighting* akan dimasukkan dalam perancangan model dari setiap algoritma, yakni algoritma *Decision Tree Classifier*, *K-Nearest Neighbor*, *Multinomial Naïve Bayes*, dan *Support Vector Machine* kernel RBF. Hasil dari perancangan model tersebut akan diuji performanya menggunakan *confusion matrix* dan *classification report*.

HASIL DAN PEMBAHASAN

Data ulasan yang diperoleh dari kesembilan tempat wisata religi Walisongo dengan rentang waktu 2020 – 2022 berjumlah 4031 data. Sebelum data ini melalui tahap *text preprocessing*, terlebih dahulu akan diberi label kelas sentimen positif, negatif, dan netral secara manual dengan total ulasan per kelas sentimen sejumlah 2699 untuk kelas positif, 716 untuk kelas negatif, dan 616 untuk kelas netral. Dikarenakan selisih per kelas cukup jauh, data ulasan akan melalui teknik *sampling random oversampling*. Berikut contoh dari data ulasan yang telah diberi label sentimen.

Tabel 1. Pelabelan Data

Review	Label
Masih kental perpaduan budaya jawa dan islam.. Kalau dari runding text makam buka jam 9 pagi dan tutup jam 10 malam	Netral
Destinasi wisata Religi Makam wali songo yang rame dikunjungi wisatawan dari berbagai daerah rame dikunjungi ditiap hari libur tanggal merah maupun pada hari Minggu untuk warga lokal sendiri biasa mengunjungi makam ditiap hari Kamis malam Jumat	Netral
bagus makamnya	Positif
Makin bagus penataan	Positif
Bintang 1 untuk pihak pengelola yang membiarkan pengemis berkaliaran bebas, kami yang dari jauh sangat risih perihal tersebut, padahal di sunan-sunan yang lain udah tertib dan walaupun ada tidak separah di sini (silahkan baca bintang 1 lainnya perihal masalah tsb)	Negatif

Text Preprocessing dan Term Weighting

Data ulasan yang telah diberi label akan melalui serangkaian tahap dari *text preprocessing*, yakni tahap *cleansing* yang berfungsi untuk mengatasi emoji maupun kata yang tidak diperlukan, *translating* yang berfungsi untuk menerjemahkan ulasan yang menggunakan bahasa Inggris ke bahasa Indonesia, *case folding* berfungsi untuk mengubah kata yang mengandung huruf kapital untuk dirubah kedalam huruf kecil, *tokenizing* berfungsi untuk memisahkan teks berdasarkan panjang teks dan menghilangkan tanda baca, *filtering stopwords* berfungsi untuk menghilangkan kata yang tidak berhubungan dengan analisis sentimen, dan *stemming* berfungsi untuk merubah kata berimbuhan menjadi kata dasar.

Data ulasan yang telah melalui *text preprocessing* akan dilanjutkan ke tahap *term weighting* yang merupakan tahap untuk memberikan nilai bobot tiap kata yang terdapat pada data dengan menghitung jumlah kemunculannya (Parasati, Bachtiar, & Setiawan, 2020). *Term Weighting* yang akan digunakan adalah pembobotan menggunakan metode *Term Frequency – Inverse Document Frequency* (selanjutnya disingkat menjadi TF – IDF).

Perancangan Model

Perancangan model dilakukan dengan data ulasan yang telah melalui *random oversampling*. Perancangan model dilakukan dengan bantuan dari *library sklearn* pada bahasa pemrograman Python.

Uji Validasi

Berikut hasil perbandingan akurasi, *precision*, *recall*, dan *f1-score* dari masing-masing model yang telah dibangun.

Tabel 2. Perbandingan Hasil Uji Validasi

Percobaan Skenario		Akurasi	Precision	Recall	F1-Score	Support
CART (Gini)	Negatif	83,65%	90%	87%	89%	143
	Netral		55%	77%	64%	136
	Positif		93%	84%	89%	528
	<i>Macro Average</i>		79%	83%	80%	807
	<i>Weighted Average</i>		86%	84%	84%	807
CART (Entropy)	Negatif	83,03%	87%	87%	87%	143
	Netral		54%	73%	62%	136
	Positif		93%	85%	89%	528
	<i>Macro Average</i>		78%	81%	79%	807
	<i>Weighted Average</i>		85%	83%	84%	807
K-Nearest Neighbor	Negatif	78,20%	72%	76%	74%	143
	Netral		55%	69%	61%	136
	Positif		88%	81%	85%	528
	<i>Macro Average</i>		72%	75%	73%	807
	<i>Weighted Average</i>		80%	78%	79%	807
Multinomial Naïve Bayes	Negatif	83,03%	78%	88%	83%	143
	Netral		61%	74%	67%	136
	Positif		92%	84%	88%	528
	<i>Macro Average</i>		77%	82%	79%	807
	<i>Weighted Average</i>		84%	83%	83%	807
SVM-RBF	Negatif	87,12%	94%	85%	89%	143
	Netral		72%	65%	68%	136
	Positif		89%	93%	91%	528
	<i>Macro Average</i>		85%	81%	83%	807
	<i>Weighted Average</i>		87%	87%	87%	807

Berdasarkan tabel di atas, dapat disimpulkan bahwa model yang memiliki nilai akurasi tertinggi yakni model dari klasifikasi SVM-RBF yang menggunakan data *random oversampling*. Model klasifikasi ini memiliki nilai akurasi tertinggi dari model lainnya, yakni sebesar 87,12% dengan nilai *f1-score* sebesar 89% untuk nilai negatif, 68% untuk nilai netral, dan 91% untuk nilai positif.

KESIMPULAN DAN SARAN

Kesimpulan yang didapatkan dalam penelitian ini adalah algoritma *Support Vector Machine* kernel RBF memiliki tingkat akurasi yang lebih tinggi dibanding dengan algoritma lain, yakni sebesar 87,12%. Dalam pengujian model ini didapati bahwa ada kecenderungan untuk mendeteksi kelas sentimen netral sebagai kelas sentimen positif. Adapun saran untuk penelitian selanjutnya adalah memperbanyak data latih untuk perancangan model agar tidak terjadi suatu kecenderungan prediksi pada salah satu kelas sentimen.

DAFTAR REFERENSI

Amarudin, A. A. (2020). Walisongo: Halal Tourism dan Perekonomian Masyarakat. *Al-Tsaman: Jurnal Ekonomi dan Keuangan Islam*, 2(2), 101-113.

- Fitri, E., Yuliani, Y., Rosyida, S., & Gata, W. (2020). Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest, dan Support Vector Machine. *TRANSFORMTIKA*, 18(1), 71-80.
- Ginantra, N. L., Yanti, C. P., Prasetya, G. D., Sarasvananda, I. B., & Wiguna, I. K. (2022). Analisis Sentimen Ulasan Villa di Ubud Menggunakan Metode Naive Bayes, Decision Tree, dan K-NN. *Jurnal Nasional Pendidikan Teknik Informatika : JANAPATI*, 11(3), 205-216.
- Ilmawan, L. B., & Mude, M. A. (2020). Perbandingan Metode Klasifikasi Support Vector Machine dan Naive Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store. *ILKOM Jurnal Ilmiah*, 12(2), 154-161.
- Imron, A. (2019). *Analisis Sentimen Terhadap Tempat Wisata di Kabupaten Rembang Menggunakan Metode Naive Bayes Classifier*. Yogyakarta: Universitas Islam Indonesia.
- Khofifah, W., Rahayu, D. N., & Yusuf, A. M. (2022). Analisis Sentimen Menggunakan Naive Bayes untuk Melihat Review Masyarakat terhadap Tempat Wisata Pantai di Kabupaten Karawang pada Ulasan Google Maps. *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi*, 16(4), 171-180.
- Marsono. (2016). *Dampak Pariwisata Religi Kawasan Masjid Sunan Kudus Terhadap Ekonomi, Lingkungan, dan Sosial-Budaya*. Yogyakarta: Gadjah Mada University Press.
- Pahmi, S., Saepudin, S., Maesarah, N., Solehudin, U. I., & Wulandari. (2018). Implementation of CART (Classification and Regression Trees) Algorithm for Determining Factors Affecting Employee Performance. *2018 International Conference on Computing, Engineering, and Design (ICCED)*, 57-62.
- Parasati, W., Bachtiar, F. A., & Setiawan, N. Y. (2020). Analisis Sentimen Berbasis Aspek pada Ulasan Pelanggan Restoran Bakso President Malang dengan Metode Naive Bayes Classifier. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 4(4), 1090-1099.
- Somantri, O., & Dairoh. (2019). Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining. *Jurnal Edukasi dan Penelitian Informatika*, 5(2), 191-196.
- Surohman, Aji, S., Rousyati, & Wati, F. F. (2020). Analisa Sentimen Terhadap Review Fintech dengan Metode Naive Bayes Classifier dan K-Nearest Neighbor. *Evolusi: Jurnal Sains dan Manajemen*, 8(1), 93-105.
- Wardani, N. S., Prahutama, A., & Kartikasari, P. (2020). Analisis Sentimen Pemindahan Ibu Kota Negara dengan Klasifikasi Naive Bayes untuk Model Bernoulli dan Multinomial. *Jurnal Gaussian*, 9(3), 237-246.