

IMPLEMENTASI *TEXT MINING* UNTUK ANALISIS PREFERENSI MASYARAKAT TERHADAP TEMPAT WISATA DI INDONESIA

Yohanes Hans Kristian¹⁾, Kestrilia Rega Prilianti²⁾, Paulus Lucky Tirma Irawan³⁾

^{1,2,3} Prodi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Ma Chung Jl. Villa Puncak Tidar N-01, Tidar, Malang
E-mail : ¹311510027@student.machung.ac.id, ²kestrilia.rega@machung.ac.id, ³paulus.lucky@machung.ac.id

ABSTRAK

Pariwisata merupakan salah satu sektor unggulan Indonesia yang perlu dijaga dan dikembangkan untuk meningkatkan daya tariknya. Untuk itu diperlukan informasi preferensi masyarakat terkait suatu tempat wisata tertentu yang bisa didapat salah satunya dari media sosial Twitter menggunakan *text mining*. Pada penelitian ini telah dibuat aplikasi untuk melakukan analisis preferensi masyarakat terhadap tempat wisata di Indonesia dengan menerapkan *text mining* menggunakan analisis sentimen dan analisis faktor dengan studi kasus candi Borobudur dan candi Prambanan. Algoritma *Naive Bayes Classifier* (NBC) digunakan pada analisis sentimen, dan metode *Principal Component Analysis* (PCA) digunakan pada analisis faktor. Dari penelitian yang sudah dilakukan didapatkan hasil akurasi klasifikasi sebesar 100% untuk topik candi Borobudur, 80.59% untuk topik candi Prambanan, dan 90.48% untuk akurasi rata-rata menggunakan algoritma NBC. Metode PCA menghasilkan 10 faktor positif dan 7 faktor negatif untuk topik candi Borobudur, sedangkan untuk topik candi Prambanan didapatkan 8 faktor positif dan 3 faktor negatif. Seluruh faktor yang terbentuk telah divalidasi dan diinterpretasikan oleh pakar. Dapat disimpulkan jika aplikasi yang dibuat dapat digunakan untuk mengetahui informasi preferensi masyarakat terhadap tempat wisata di Indonesia.

Kata Kunci: analisis faktor, analisis sentimen, NBC, PCA, *text mining*.

ABSTRACT

Tourism is one of Indonesia's leading sectors that needs to be maintained and developed to increase its attractiveness. For this reason, public preference information regarding a particular tourist spot that can be obtained is one of them from Twitter social media using text mining. In this study an application has been made to analyze people's preferences for tourist attractions in Indonesia by applying text mining using sentiment analysis and factor analysis with a case study of Borobudur temple and Prambanan temple. The *Naive Bayes Classifier* (NBC) algorithm is used in sentiment analysis, and the *Principal Component Analysis* (PCA) method is used in factor analysis. From the research that has been done, the results of classification are 100% for the topics of Borobudur temple, 80.59% for the topic of Prambanan temple, and 90.48% for the average accuracy using the NBC algorithm. The PCA method produced 10 positive factors and 7 negative factors for the topic of Borobudur temple, while for the topic of Prambanan temple there were 8 positive factors and 3 negative factors. All factors formed have been validated and interpreted by experts. It can be concluded if the application made can be used to find out information on people's preferences for tourist attractions in Indonesia.

Keywords: faktor analysis, NBC, PCA, sentiment analysis, *text mining*.

PENDAHULUAN

Pariwisata merupakan salah satu sektor unggulan Indonesia. Kekuatan pariwisata Indonesia menempati urutan ke-9 dari 10 besar duni, dan menjadi yang terbaik di Asia Tenggara. Selain itu, Indonesia juga tercatat sebagai salah satu dari 20 negara dengan pertumbuhan sektor pariwisata tercepat. Menurut Arief Yahya, dalam beberapa tahun ke depan, sebagian ekonomi Indonesia akan didukung oleh sektor pariwisata. Hal ini menunjukkan bahwa sektor pariwisata di Indonesia memegang peranan yang cukup penting, salah satunya adalah penerimaan devisa negara. Penerimaan devisa negara pada tahun 2017 dari sektor pariwisata adalah sebesar US\$ 15 miliar. Oleh karena itu, sektor pariwisata Indonesia perlu dijaga dan dikembangkan, salah satunya dengan meningkatkan daya tarik wisatawan. Informasi terkait preferensi masyarakat terhadap tempat wisata dibutuhkan untuk meningkatkan daya tarik tempat wisata.

Masalah yang dihadapi saat ini adalah banyaknya data terkait preferensi masyarakat terhadap tempat wisata terutama di media sosial, sehingga untuk mengolahnya secara

manual membutuhkan waktu, tenaga, bahkan uang yang tidak sedikit.

Solusi dari masalah tersebut adalah menerapkan *text mining* untuk mengolah data terkait preferensi masyarakat terhadap tempat wisata di Indonesia dengan sumber data yang bisa diperoleh dari media sosial Twitter menggunakan Twitter API. Data yang diperoleh kemudian dilakukan analisis sentimen menggunakan algoritma *Naive Bayes Classifier* (NBC), dan dilanjutkan dengan analisis faktor menggunakan metode *Principal Component Analysis* (PCA).

Pada penelitian sebelumnya (2018) telah dilakukan penerapan *text mining* menggunakan sumber data dari Google Maps [1] dan TripAdvisor [2] sampai pada tahap analisis sentimen menggunakan algoritma NBC. Dalam penelitian ini akan dilakukan

implementasi *text mining* sampai pada tahap analisis faktor.

Tujuan dari penelitian ini adalah mengembangkan aplikasi sistem analisis preferensi masyarakat terhadap tempat wisata yang ada di Indonesia dengan mengimplementasikan *text mining* berdasarkan data dari Twitter menggunakan algoritma NBC dan metode PCA untuk mendapatkan informasi terkait tempat wisata tertentu.

Adapun batasan masalah dalam penelitian ini adalah menggunakan data dari Twitter berupa *tweets* pengguna Twitter dalam bahasa Indonesia dengan studi kasus candi Borobudur dan candi prambanan yang dianalisis menggunakan Python dan menggunakan web untuk menampilkan hasil analisis.

METODE

Penelitian ini dilakukan dalam 6 tahapan utama yaitu, analisis masalah, pengumpulan data, desain sistem, pengembangan sistem, pengujian, dan simpulan analisis faktor. Berikut adalah penjelasan untuk masing-masing tahap.

1. Analisis masalah

Pada tahapan ini ditemukan bahwa aliran informasi di media sosial cepat dan banyak disebabkan oleh banyaknya pengguna media sosial di Indonesia yaitu 49% dari total penduduk Indonesia. Aliran informasi yang cepat dan banyak ini akan menyulitkan jika diolah secara manual, oleh karena itu diterapkan *text mining* yang dilanjutkan dengan analisis sentimen dan analisis faktor.

2. Pengumpulan data

Data yang dikumpulkan diperoleh dari media sosial Twitter menggunakan Twitter API. Dari data yang terkumpul, dilakukan klasifikasi secara manual ke dalam kelas sentimen positif dan kelas sentimen negatif, dan dilanjutkan dengan pembagian yaitu sebesar 80% untuk data latih dan 20% untuk data uji. Data latih digunakan pada analisis sentimen menggunakan algoritma NBC untuk melakukan klasifikasi sentimen.

3. Desain sistem

Sistem yang dibuat pada penelitian ini menggunakan Python untuk melakukan *text mining*, analisis sentimen, dan analisis faktor. Antarmuka pengguna dibuat menggunakan web dengan bantuan framework flask.

Text mining terdiri dari dua tahap yaitu pra-pemrosesan teks dan seleksi fitur. Tahap pra-pemrosesan teks berfungsi untuk mengolah data *tweet* mentah yang belum terstruktur menjadi data *tweet* yang telah terstruktur dengan cara merubah semua huruf menjadi huruf kecil, menghapus link dan mention, dan menghapus karakter selain huruf (case folding), memisah kalimat menjadi masing-masing kata penyusunnya (tokenizing), menghapus kata-kata yang tidak penting berdasarkan data stopword (filtering), dan mengubah setiap kata menjadi bentuk kata dasar (stemming). Hasil dari pra-pemrosesan teks berupa Document Term Matrix (DTM), yaitu matriks yang menggambarkan frekuensi untuk masing-masing kata dengan banyaknya baris yang menunjukkan jumlah data, dan banyaknya kolom yang menunjukkan jumlah variasi kata. Contoh DTM dapat dilihat pada tabel 1.

Tabel 1. DTM

Tweet	Kata 1	Kata 2	Kata ...	Kata M
Tweet 1	1	0	...	3
Tweet 2	0	0	...	0
Tweet
Tweet N	2	1	...	0

Setelah terbentuk DTM dari hasil pra-pemrosesan teks, dilanjutkan dengan seleksi fitur, yaitu penghapusan kata yang memiliki frekuensi kurang dari batas yang ditentukan, yaitu 1. Hasil dari seleksi fitur berupa tabel frekuensi seperti pada tabel 2.

Tabel 2. Tabel frekuensi

Kata	Frekuensi
Kata 1	3
Kata
Kata M	3

Pada bagian analisis sentimen, digunakan algoritma NBC untuk mengklasifikasikan *tweet* ke dalam sentimen positif maupun sentimen negatif. Untuk dapat melakukan klasifikasi algoritma NBC memerlukan data latih yang didapat dari tabel frekuensi dan dihitung nilai ln probabilitas masing-masing kata dengan rumus pada persamaan (1).

$$\ln p(w | C_k) = \ln \frac{f + 1}{\sum_w (f_w + 1)} \quad (1)$$

Dimana :

- C_k : kelas positif atau negatif
- W : kata
- F : frekuensi kemunculan kata
- f_w : frekuensi kata pada data latih

Selanjutnya dilakukan klasifikasi dengan cara menghitung nilai ln probabilitas positif maupun negatif dari suatu *tweet* untuk menentukan kelas sentimennya menggunakan rumus pada persamaan (2).

$$\ln p(C_k | t) = \sum_y \ln P_y + \sum_x f_x * \ln \frac{f + 1}{\sum_w (f_w + 1)} \quad (2)$$

Dimana :

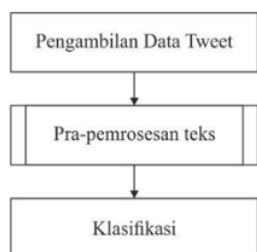
- C_k : kelas positif atau negatif
- t : *tweet*
- w : kata pada data latih
- y : kata pada data uji yang terdapat pada data latih
- x : kata pada data uji yang tidak terdapat pada data latih
- $\ln P_y$: Ln probabilitas dari kata yang terdapat pada data latih

Pada bagian analisis faktor, digunakan metode PCA untuk mengelompokkan kata-kata yang memiliki korelasi tinggi ke dalam satu faktor. Jumlah faktor ditentukan oleh

pengguna sampai didapatkan hasil pengelompokan yang sesuai.

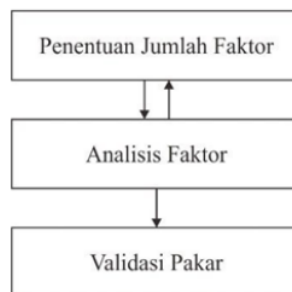
4. Pengembangan sistem

Pada tahapan ini, dikembangkan sistem yang terdiri dari analisis sentimen dan analisis faktor. Pada bagian analisis sentimen, pengguna dapat mengambil data *tweet* dari Twitter melalui Twitter API dengan memasukkan pilihan nama tempat wisata beserta jumlah data yang ingin diambil. Setelah itu sistem akan mengambil data sesuai dengan masukan pengguna, melakukan pra- pemrosesan teks, dan mengklasifikasikan ke dalam sentimen positif dan negatif. Hasil klasifikasi akan ditampilkan kepada pengguna. Diagram analisis sentimen dapat dilihat pada gambar 1.



Gambar 1. Diagram analisis sentiment

Pada bagian analisis faktor, pengguna perlu memilih objek tempat wisata yaitu candi Borobudur atau Prambanan, dan memasukkan jumlah faktor untuk masing-masing kelas sentimen. Setelah jumlah faktor ditentukan oleh pengguna, sistem akan melakukan analisis faktor dan menampilkan hasilnya. Jika hasil pengelompokan faktor masih belum sesuai, pengguna perlu mengulang proses dengan menentukan jumlah faktor sampai mendapatkan hasil analisis faktor berupa pengelompokan faktor yang sesuai. Langkah selanjutnya adalah validasi yang dilakukan oleh pakar untuk melakukan interpretasi faktor yang telah terbentuk. Diagram analisis faktor dapat dilihat pada gambar 2.



Gambar 2. Diagram analisis factor

5. Pengujian

Pengujian dilakukan untuk menguji algoritma berupa akurasi dari algoritma NBC dengan rumus seperti pada persamaan (3).

$$A = \frac{x}{n} \times 100 \quad (3)$$

Dimana :

A : akurasi

x : jumlah *tweet* yang diklasifikasi dengan benar

n : total data uji

Selain itu juga dilakukan pengujian rata-rata waktu komputasi dari masing-masing proses yang terdapat pada sistem dengan rumus pada persamaan (4).

$$\bar{X} = \frac{t}{n} \quad (4)$$

Dimana :

\bar{X} : rata-rata

t : total waktu

n : jumlah percobaan

6. Simpulan analisis faktor

Simpulan analisis faktor didapat dari hasil perhitungan menggunakan metode PCA berupa faktor-faktor yang kemudian dilakukan validasi dan interpretasi oleh pakar.

HASIL DAN PEMBAHASAN

Adapun beberapa penelitian terdahulu yang serupa dengan penelitian ini adalah sebagai berikut. Pemanfaatan analisis sentimen untuk pemeringkatan populasi tujuan wisata berdasarkan data dari media sosial Facebook dan Instagram menggunakan algoritma *Naive Bayes Classifier* [3]. Hasil dari penelitian tersebut berupa pemeringkatan tujuan wisata yang paling populer.

Penerapan *text mining* untuk menganalisis pelayanan dari ojek daring menggunakan algoritma NBC menggunakan data dari Twitter [4]. Hasil klasifikasi dibagi ke dalam tiga kelas yaitu sentimen positif, sentimen negatif, dan sentimen netral, dengan hasil akurasi algoritma NBC sebesar 80%.

Penerapan *text mining* untuk melakukan pengelompokan data *tweet* Shopee Indonesia [5]. Pengelompokan dilakukan menggunakan K-means. Penelitian tersebut menghasilkan jenis konten pada kelompok yang memiliki jumlah *retweet* yang tinggi.

Penerapan *text mining* untuk klasifikasi konten berita [6]. Penelitian tersebut menggunakan algoritma *Naive Bayes Classifier* untuk melakukan klasifikasi. Hasil dari penelitian tersebut berupa sistem klasifikasi berita berbasis web.

Penerapan *text mining* untuk klasifikasi kategori cerita pendek [7]. Penelitian tersebut menggunakan algoritma *Naive Bayes* untuk melakukan klasifikasi. Hasil dari penelitian tersebut adalah berupa akurasi sebesar 64.36% untuk algoritma *Naive Bayes*.

Penerapan *text mining* untuk analisis sentimen perguruan tinggi swasta, yang menggunakan data dari Twitter dan diolah menggunakan bahasa pemrograman R [8]. Klasifikasi sentimen dilakukan menggunakan algoritma *Multinomial Naive Bayes* dan dilanjutkan dengan analisis faktor. Hasil dari penelitian tersebut adalah akurasi sebesar 87.3% untuk klasifikasi sentimen.

Penerapan *text mining* untuk

penentuan kategori tugas akhir mahasiswa berdasarkan strak [9]. Penelitian tersebut menggunakan algoritma *Naive Bayes* untuk melakukan penentuan kategori. Hasil dari penelitian tersebut adalah akurasi sebesar 87% untuk algoritma *Naive Bayes*.

Penelitian lain menggunakan analisis sentimen masyarakat terhadap produk maupun layanan suatu perusahaan menggunakan metode *Deep Belief Network* (DBN) dengan sumber data dari Twitter [10]. Penelitian tersebut mengklasifikasikan sentimen menjadi tiga kelas yaitu sentimen positif, sentimen negatif, dan sentimen netral. Hasil dari penelitian tersebut adalah akurasi untuk metode DBN sebesar 93.31%.

Penerapan lain dari *text mining* juga dilakukan pada penelitian untuk menganalisis sentimen masyarakat terhadap sebuah film [11]. Penelitian tersebut menggunakan algoritma *Support Vector Machine* (SVM) untuk membagi sentimen menjadi tiga kelas yaitu sentimen positif, sentimen negatif, dan sentimen netral menggunakan sumber data dari Twitter. Hasil dari penelitian tersebut adalah rata-rata akurasi algoritma SVM menggunakan 60, 70, 80, dan 90 persen data latih berturut-turut adalah sebesar 70.06%, 76.83%, 81.07%, dan 83.3%. Selain itu juga terdapat penelitian analisis sentimen masyarakat terhadap hotel menggunakan algoritma SVM berbasis *Particle Swarm Optimization* (PSO) [12]. Penelitian tersebut mengklasifikasikan sentimen menjadi kelas positif dan negatif. Hasil dari penelitian tersebut adalah akurasi algoritma SVM sebesar 96.94% menggunakan PSO.

Penerapan *text mining* juga dapat dilakukan untuk menganalisis opini masyarakat terhadap tokoh politik [13]. Penelitian tersebut menggunakan data dari Twitter dan algoritma NBC untuk mengklasifikasikan sentimen menjadi tiga kelas yaitu positif, negatif, dan netral. Hasil dari penelitian tersebut adalah akurasi sebesar 61% untuk algoritma NBC dengan data latih untuk

masing-masing sentimen sebanyak 100, dan akurasi sebesar 91% untuk algoritma NBC dengan data latih untuk masing-masing sentimen sebanyak 700.

Selain itu juga terdapat penelitian yang melakukan implementasi *text mining* untuk menganalisis sentimen masyarakat terhadap tokoh publik menggunakan data dari Twitter dan algoritma NBC dikombinasikan dengan *term frequency* dan *Term Frequency-Inverse Document Frequency (TF-IDF)*. [14]. Hasil dari penelitian tersebut adalah akurasi sebesar 79.91% untuk algoritma NBC dengan fitur *term frequency*, dan 79.68% untuk algoritma NBC dengan fitur TF-IDF. Implementasi *text mining* yang serupa juga terdapat pada penelitian untuk menganalisis sentimen masyarakat terhadap calon presiden Indonesia 2014 menggunakan data dari Twitter dan algoritma NBC [15]. Hasil dari penelitian tersebut adalah 47.7% sentimen positif, 26.4% sentimen negatif, dan 25.9% sentimen netral untuk calon pertama. Sedangkan untuk calon kedua adalah 37.6% sentimen positif, 34.4% sentimen negatif, dan 27.9% sentimen netral.

Dari penelitian yang telah dilakukan, didapatkan beberapa hasil. Hasil yang pertama adalah hasil pengumpulan data. Pada pengumpulan data diperoleh data-data yang relevan untuk topik candi Borobudur, sedangkan untuk topik candi Prambanan diperoleh data-data yang kurang relevan. Data yang relevan dalam penelitian ini berarti data yang memiliki sentimen positif dan sentimen negatif terkait topik tempat wisata. Data yang berhasil dikumpulkan sebanyak 3739 *tweet* untuk topik candi Borobudur dan 2752 *tweet* untuk topik candi Prambanan. Dari data yang telah diperoleh tersebut, dilakukan klasifikasi manual untuk membagi data ke dalam dua kelas yaitu sentimen positif dan sentimen negatif. Hasil klasifikasi manual berupa 95 *tweet* sentimen positif dan 95 *tweet* sentimen negatif untuk topik candi Borobudur, 95 *tweet* sentimen positif dan 10 *tweet* sentimen negatif untuk topik candi Prambanan. Dari hasil klasifikasi manual dapat dilihat

jika terdapat masalah berupa kekurangan data untuk topik candi Prambanan dengan sentimen negatif. Masalah seperti ini dapat diatasi dengan mencari sumber data lain selain Twitter.

Hasil kedua berupa hasil analisis sentimen yang terdiri dari proses *text mining* dan klasifikasi *tweet* menggunakan algoritma NBC.

1. *Text mining*

Proses *text mining* terdiri dari pengambilan data dan pra-pemrosesan teks. Pada tahap pengambilan data, sistem berhasil mengambil data sesuai dengan masukan pengguna, yang kemudian dilakukan pra-pemrosesan teks. Tahap pra-pemrosesan teks dapat berjalan dengan baik terlihat dari semua huruf yang sudah berubah menjadi huruf kecil, tidak terdapat karakter lain selain huruf, hilangnya *mention* dan *link*, hilangnya kata yang terdapat pada stopword, setiap *tweet* terpisah menjadi masing-masing kata penyusunnya, dan semua kata baku telah berubah ke dalam bentuk kata dasar.

2. Hasil klasifikasi

Dari hasil klasifikasi, terdapat beberapa *tweet* yang seharusnya memiliki sentimen positif namun diklasifikasikan ke dalam sentimen negatif, oleh karena itu diterapkan nilai *threshold* terhadap sentimen negatif. Artinya, untuk dapat diklasifikasikan ke dalam sentimen negatif, nilai ln probabilitas negatif suatu *tweet* harus lebih besar minimal sebanyak nilai *threshold* dari nilai ln probabilitas positif. Nilai *threshold* ditentukan dari percobaan klasifikasi terhadap *tweet* data uji dengan melihat akurasi tertinggi yang dihasilkan oleh setiap nilai *threshold*.

Hasil ketiga berupa hasil analisis faktor yang telah divalidasi oleh pakar. Topik candi Borobudur menghasilkan 10 faktor positif dan 7 faktor negatif, sedangkan topik candi Prambanan menghasilkan 8 faktor positif dan 3 faktor negatif. Berikut adalah hasil analisis faktor terhadap topik candi Borobudur dan candi Prambanan.

1. Faktor positif candi Borobudur yaitu wisata keagamaan, keajaiban dunia, atraksi budaya, wisata alam, tujuan

wisata, wisata budaya, pariwisata Indonesia, traveling, daya tarik, dan kawasan wisata.

2. Faktor negatif candi Borobudur yaitu atraksi, pesaing, biaya, perawatan, kenyamanan, kesan yang buruk, dan aktivitas.
3. Faktor positif candi Prambanan yaitu wisata budaya, keindahan, acara, pemandangan, spot foto, objek wisata, atraksi, dan wisata jogja.
4. Faktor negatif candi Prambanan yaitu sarana, tiket masuk, dan fasilitas publik.

Hasil keempat berupa hasil uji coba yang terdiri dari uji akurasi klasifikasi, uji coba analisis faktor, dan uji waktu komputasi.

1. Uji akurasi klasifikasi

Uji akurasi dilakukan terhadap 59 data uji dengan 38 data uji topik candi Borobudur dan 21 data uji topik candi Prambanan. Data uji merupakan data yang sudah diklasifikasikan secara manual ke dalam kelas sentimen positif dan kelas sentimen negatif. Pada penelitian ini, diterapkan nilai *threshold* untuk meningkatkan akurasi klasifikasi. Penentuan nilai *threshold* untuk topik candi Borobudur didasarkan dari percobaan pada tabel 3.

Tabel 3. Percobaan nilai *threshold* topik candi Borobudur

Threshold	Akurasi (%)
0	97.37
0.5	97.37
0.75	100
1	97.37

Dari hasil percobaan nilai *threshold* terhadap topik candi Borobudur, didapatkan nilai 0.75 sebagai nilai *threshold* terbaik dengan akurasi sebesar 100%. Sedangkan percobaan nilai *threshold* untuk topik candi Prambanan dapat dilihat pada tabel 4.

Tabel 4. Percobaan nilai *threshold* topik candi Prambanan

Threshold	Akurasi (%)
0	28.57
...	...
14	76.19
15	80.95
16	80.95
17	80.95
18	76.19

Dari hasil percobaan nilai *threshold* terhadap topik candi Prambanan, didapatkan nilai 15, 16, dan 17 sebagai nilai *threshold* terbaik dengan akurasi sebesar 80.95%, untuk selanjutnya digunakan nilai 15 sebagai nilai *threshold*.

Sesuai dengan percobaan penentuan nilai *threshold*, dari pengujian klasifikasi topik candi Borobudur menggunakan nilai *threshold* sebesar 0.75, didapatkan hasil 38 *tweet* yang dapat diklasifikasi dengan benar dari total 38 *tweet* yang diuji, sehingga menghasilkan akurasi sebesar 100% dengan perhitungan sebagai berikut.

$$Akurasi = \frac{38}{38} \times 100 = 100\% \quad (5)$$

Sedangkan hasil pengujian klasifikasi topik candi Prambanan, didapatkan hasil 17 *tweet* yang dapat diklasifikasi dengan benar dari total 21 *tweet* yang diuji menggunakan nilai *threshold* sebesar 15, sehingga menghasilkan akurasi sebesar 80.95% dengan perhitungan sebagai berikut.

$$Akurasi = \frac{17}{21} \times 100 = 80.95\% \quad (6)$$

Sedangkan perhitungan akurasi rata-rata dari kedua topik sebesar 90.48% dengan perhitungan sebagai berikut.

$$Akurasi = \frac{100 + 80.95}{2} = 90.48\% \quad (7)$$

2. Uji waktu komputasi
Pengujian waktu komputasi dilakukan pada masing-masing proses di dalam aplikasi analisis preferensi. Setiap proses dilakukan pengujian waktu komputasi sebanyak lima kali dan kemudian dihitung reratanya. Data yang digunakan untuk melakukan pengujian adalah 100 *tweet* untuk pengambilan data, 38 *tweet* untuk pra-pemrosesan teks, 38 *tweet* untuk klasifikasi, 152 *tweet* untuk analisis faktor topik candi Borobudur, dan 105 *tweet* untuk topik candi Prambanan. Selain itu pengujian juga dilakukan dengan tidak menghapus cache dan menghapus cache. Hasil perbandingan rata-rata waktu komputasi dapat dilihat pada tabel 5.

Tabel 5. Uji waktu komputasi

No	Proses	Waktu (detik) ±	
		Tanpa cache	Dengan cache
1.	Pengambilan <i>tweet</i>	4.54	4.62
2.	Pra-pemrosesan teks	9.41	0.05
3.	Klasifikasi <i>tweet</i>	0.1	0.05
4.	Analisis faktor topik candi Borobudur	0.12	0.12
5.	Analisis faktor topik candi Prambanan	0.1	0.1

SIMPULAN

Aplikasi analisis preferensi masyarakat terhadap tempat wisata di Indonesia menggunakan data dari Twitter telah berhasil dibuat.
5.

16
Berdasarkan hasil penelitian dan pengujian yang telah dilakukan terhadap aplikasi tersebut, dapat disimpulkan beberapa hal sebagai berikut,

1. Analisis sentimen dapat diimplementasikan terhadap tempat wisata yang memiliki data latih yang cukup. Pada penelitian ini dapat dilihat jika topik candi Borobudur memiliki data yang cukup, maka menghasilkan akurasi klasifikasi yang lebih tinggi dibandingkan akurasi klasifikasi topik candi Prambanan yang memiliki data lebih sedikit.
2. Tahap pra-pemrosesan teks dapat berjalan dengan baik pada *tweet* yang memiliki kata baku. Jika *tweet* mengandung kata tidak baku, kata tidak baku tersebut dapat mengganggu proses klasifikasi serta analisis faktor.
3. Algoritma *Naive Bayes Classifier* yang digunakan pada tahap analisis sentimen dapat melakukan klasifikasi *tweet* dengan baik menggunakan data latih yang cukup dan seimbang antara sentimen positif dan sentimen negatif dengan akurasi sebesar 100% untuk topik candi Borobudur menggunakan nilai *threshold* sebesar 0.75, akurasi sebesar 80.95% untuk topik candi Prambanan menggunakan nilai *threshold* sebesar 15, dan akurasi sebesar 93.22% untuk keseluruhan topik.
4. Analisis faktor yang dilakukan pada aplikasi ini dapat mengelompokkan setiap kata dengan baik ke dalam faktor-faktor yang sesuai dengan jumlah faktor sebanyak 10 untuk sentimen positif dan 7 untuk sentimen negatif topik candi Borobudur, dan jumlah faktor sebanyak 8 untuk sentimen positif dan 3 untuk sentimen negatif topik candi Prambanan.

SARAN

Aplikasi ini diharapkan dapat digunakan untuk mengetahui informasi preferensi masyarakat yang lebih luas terhadap tempat wisata di Indonesia

DAFTAR PUSTAKA

- [1] Wilianto, L., Pudjiantoro, T.H. dan Umbara, F.R. 2017, 'Analisis Sentimen terhadap Tempat Wisata dari Komentar Pengunjung dengan Menggunakan Metode Naive Bayes Classifier Studi Kasus Jawa Barat', *prosiding SNATIF Ke -4*, Cimahi, hal. 439-448.
- [2] Fanissa, S., Fauzi, M.A. dan Adinugroho, S. 2018, Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2:2766-2770.
- [3] Murnawan dan Sinaga, A. 2017, Pemanfaatan Analisis Sentimen untuk Pemingkatan Popularitas Tujuan Wisata, *Jurnal Penelitian Pos dan Informatika*, 2:109-120.
- [4] Nugroho, D.G., Chrisnanto, Y.H. dan Wahana, A. 2016, 'Analisis Sentimen Pada Jasa Ojek Online Menggunakan Metode Naive Bayes', *prosiding SNST ke-7*, Semarang, hal. 156-161.
- [5] Indraloka, D.S. dan Santosa, B. 2017, Penerapan Text Mining untuk Melakukan Clustering Data *Tweet* Shopee Indonesia, *Jurnal Sains dan Seni ITS*, 6:A51-A56.
- [6] Kurniawan, B., Effendi, S., dan Sitompul, O.S. 2012, Klasifikasi Konten Berita dengan Metode Text Mining, *Jurnal Dunia Teknologi Informasi*, 1:14-19.
- [7] Somantri, O. 2017, Text Mining untuk Klasifikasi Kategori Cerita Pendek Menggunakan Naive Bayes (NB), *Jurnal Telematika*, 12:1.
- [8] Hartono, B. 2017, 'Implementasi Text Mining untuk Analisis Sentimen Perguruan Tinggi Swasta Melalui Media Sosial', *Skripsi*, Universitas Ma Chung, Malang.
- [9] Hermanto, A. 2016, Implementasi Text Mining Menggunakan Naive Bayes untuk Penentuan Kategori *Tugas Akhir* Mahasiswa Berdasarkan Abstraksinya, *Konvergensi*, 12:1-10.
- [10] Zulfa, I. dan Winarko, E. 2017, Sentimen Analisis Tweet Berbahasa Indonesia dengan Deep Belief Network, *IJCCS*, 11:187-198.
- [11] Rahutomo, F., Saputra, P.Y. dan Fidyawan, M.A. 2018, Implementasi Twitter Sentiment Analysis untuk Review Film Menggunakan Algoritma Support Vector Machine, *Jurnal Informatika Polinema*, 4:93-100.
- [12] Indrayuni, E. 2016, Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization, *Jurnal Evolusi*, 4:20-27.
- [13] Pramono, A., Indriati, R. dan Nugroho A. 2017, 'Sentiment Analysis Tokoh Politik pada Twitter', dipresentasikan pada *Seminar Nasional Inovasi Teknologi di Kediri*, 22 Februari.
- [14] Hidayatullah, A.F. dan Azhari, S.N. 2014, 'Analisis Sentimen dan Klasifikasi Kategori Terhadap Tokoh Publik pada Twitter', dipresentasikan pada *Seminar Nasional Informatika* di Yogyakarta, 12 Agustus.
- [15] Nurhuda, F., Sihwi, S.W. dan Doewes, A. 2013, Analisis

Sentimen Masyarakat terhadap
Calon Presiden 2014 berdasarkan
Opini dari Twitter Menggunakan

Metode Naive Bayes Classifier,
Jurnal ITSMART, 2:35-42.

IMPLEMENTASI TEXT MINING UNTUK ANALISIS PREFERENSI MASYARAKAT TERHADAP TEMPAT WISATA DI INDONESIA

ORIGINALITY REPORT

17%

SIMILARITY INDEX

15%

INTERNET SOURCES

8%

PUBLICATIONS

4%

STUDENT PAPERS

PRIMARY SOURCES

1	repository.widyatama.ac.id Internet Source	1%
2	teknik.trunojoyo.ac.id Internet Source	1%
3	Submitted to Universitas Brawijaya Student Paper	1%
4	www.scribd.com Internet Source	1%
5	jurnal.upnyk.ac.id Internet Source	1%
6	core.ac.uk Internet Source	1%
7	j-ptiik.ub.ac.id Internet Source	1%
8	ecampus.pelitabangsa.ac.id Internet Source	1%
9	Submitted to Universitas Islam Indonesia Student Paper	1%

10	publikasi.dinus.ac.id Internet Source	<1 %
11	Submitted to UIN Sultan Syarif Kasim Riau Student Paper	<1 %
12	etheses.uin-malang.ac.id Internet Source	<1 %
13	repository.upnvj.ac.id Internet Source	<1 %
14	id.scribd.com Internet Source	<1 %
15	jurnal.atmaluhur.ac.id Internet Source	<1 %
16	repository.uksw.edu Internet Source	<1 %
17	Lian Ardiani, Herry Sujaini, Tursina Tursina. "Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak", Jurnal Sistem dan Teknologi Informasi (Justin), 2020 Publication	<1 %
18	Submitted to President University Student Paper	<1 %
19	dspace.uphsurabaya.ac.id:8080 Internet Source	<1 %

20	Fajar Ratnawati. "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter", INOVTEK Polbeng - Seri Informatika, 2018 Publication	<1 %
21	Gamaria Mandar, Gunawan Gunawan. "Peringkasan dokumen berita Bahasa Indonesia menggunakan metode Cross Latent Semantic Analysis", Register: Jurnal Ilmiah Teknologi Sistem Informasi, 2017 Publication	<1 %
22	id.123dok.com Internet Source	<1 %
23	jurnal.ustjogja.ac.id Internet Source	<1 %
24	repository.ipb.ac.id Internet Source	<1 %
25	repository.mercubuana.ac.id Internet Source	<1 %
26	repository.usd.ac.id Internet Source	<1 %
27	dergipark.org.tr Internet Source	<1 %
28	jurnal.untan.ac.id Internet Source	<1 %

29	library.binus.ac.id Internet Source	<1 %
30	www.hindawi.com Internet Source	<1 %
31	Abdurrahman Rahim Thaha, Fauzan Aziz. "Penambangan Teks Pada Tujuan Wisata di Bandung Raya (Studi Kasus: Tangkuban Perahu dan Kawah Putih)", Jurnal Sekretaris & Administrasi Bisnis (JSAB), 2020 Publication	<1 %
32	Emelia Aprodaid Marwa, Ari Budi Kristanto. "Analisis Sentimen Pengungkapan Informasi Manajemen: Text Mining Berbasis Metode VADER", Owner, 2022 Publication	<1 %
33	Putri Suci Rahmadani, Fenny Chintya Tampubolon, Adelia Nurfattul Jannah, Novia Lucky Halen Hutabarat, Allwin M. Simarmata. "Tiktok Social Media Sentiment Analysis Using the Nave Bayes Classifier Algorithm", SinkrOn, 2022 Publication	<1 %
34	alviprofdr.blogspot.com Internet Source	<1 %
35	digilib.uin-suka.ac.id Internet Source	<1 %

36	digilib.unhas.ac.id Internet Source	<1 %
37	e-journal.hamzanwadi.ac.id Internet Source	<1 %
38	e-journal.unair.ac.id Internet Source	<1 %
39	www.neliti.com Internet Source	<1 %
40	openlibrary.telkomuniversity.ac.id Internet Source	<1 %
41	Naisah Marito Putry. "KOMPARASI ALGORITMA KNN DAN NAÏVE BAYES UNTUK KLASIFIKASI DIAGNOSIS PENYAKIT DIABETES MELLITUS", EVOLUSI : Jurnal Sains dan Manajemen, 2022 Publication	<1 %
42	Rakhmi Khalida, Siti Setiawati. "Analisis Sentimen Sistem E-Tilang Menggunakan Algoritma Naive Bayes Dengan Optimalisasi Information Gain", Journal of Informatic and Information Security, 2020 Publication	<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography On