

Analisis Sentimen Wisatawan Melalui Data Ulasan Candi Borobudur di Tripadvisor Menggunakan Algoritma Naïve Bayes Classifier

Yerik Afrianto Singgalen

Faculty of Business Administration and Communication, Tourism Department, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia

Email: yerik.afrianto@atmajaya.ac.id

Email Penulis Korespondensi: yerik.afrianto@atmajaya.ac.id

Submitted: 04/11/2022; Accepted: 14/12/2022; Published: 30/12/2022

Abstrak—Analisis sentimen pengunjung ke destinasi wisata Candi Borobudur di Indonesia perlu dilakukan untuk mengetahui preferensi produk dan layanan yang diharapkan. Selain itu, analisis sentiment juga bermanfaat bagi pengelola untuk menyesuaikan kebutuhan wisatawan dengan sarana-prasarana yang disediakan di area destinasi wisata. Metode klasifikasi yang digunakan dalam analisis sentimen ialah Naïve Bayes Classifier (NBC) terhadap 3850 ulasan pengunjung di Candi Borobudur. Data ulasan diambil dari halaman website Tripadvisor yang telah disaring berdasarkan bahasa, waktu mengulas, dan karakteristik perjalanan untuk dianalisis secara komprehensif preferensi wisatawan mancanegara. Tahapan penelitian ini terbagi menjadi tiga bagian yakni tahapan persiapan data, pemrosesan data, analisis sentimen dan evaluasi performa algoritma. Adapun, SMOTE Upsampling digunakan untuk menyeimbangkan data. Penelitian ini menunjukkan bahwa hasil implementasi metode klasifikasi Naïve Bayes Classifier (NBC) diperoleh nilai accuracy sebesar 96,36% dan nilai precision sebesar 93,23% serta nilai recall sebesar 100% dengan nilai Area Under Curve (AUC) sebesar 0,714. Selain itu hasil perankingan lima kata yang populer dari data ulasan menunjukkan adanya sorotan kondisi fisik candi, pemandangan dan aktivitas kunjungan wisata di Candi Borobudur dimana empat kata paling populer dalam ulasan pengunjung ialah temple, visit, borobudur, sunrise dan place.

Kata Kunci: Analisis Sentimen; Borobudur; Indonesia, Tripadvisor; Naïve Bayes Classifier

Abstract—Sentiment analysis of visitors to the tourist destinations of Borobudur Temple in Indonesia needs to be done to determine the expected product and service preferences. In addition, sentiment analysis is also helpful for managers to adjust the needs of tourists to the infrastructure provided in the tourist destination area. The classification method used in the sentiment analysis is the Naïve Bayes Classifier (NBC) against 3850 visitor reviews at Borobudur Temple. Review data is pulled from Tripadvisor web pages filtered by language, review time, and travel characteristics to analyze foreign traveler preferences comprehensively. This research stage is divided into three parts: data preparation, data processing, sentiment analysis, and algorithm performance evaluation. In addition, SMOTE Upsampling is used to balance data. The results of implementing the Naïve Bayes Classifier (NBC) classification method obtained an accuracy value of 96.36%, a precision value of 93.23%, and a recall value of 100% with an Area Under Curve (AUC) value of 0.714. In addition, the results of ranking five famous words from the review data show that there are highlights of the physical condition of the temple, scenery, and tourist visit activities at Borobudur Temple, where the four most famous words in visitor reviews are the “temple,” “visit,” “Borobudur,” “sunrise” and “place.”

Keywords: Borobudur; Indonesia; Naïve Bayes Classifier; Sentiment Analysis; Tripadvisor

1. PENDAHULUAN

Analisis sentimen terhadap produk dan jasa di destinasi wisata prioritas Candi Borobudur perlu dikaji secara komprehensif untuk menghasilkan rekomendasi yang tepat dalam menyesuaikan perubahan preferensi pengunjung serta meningkatkan citra pariwisata Indonesia. Terdapat kesenjangan dalam studi tentang analisis sentimen wisatawan yang terbatas pada implementasi metode klasifikasi tanpa menghubungkan hasil klasifikasi untuk menghasilkan sejumlah rekomendasi berdasarkan kata populer dalam kelas positif maupun negatif untuk mengoptimalkan manajemen destinasi wisata. Dalam konteks manajemen destinasi wisata Candi Borobudur, platform digital yang digunakan sebagai sumber data ulasan menjadi salah satu indikator yang berpengaruh terhadap keputusan berkunjung. Disisi lain, studi tentang analisis sentimen di destinasi wisata masih menggunakan sumber data dari media sosial, salah satunya twitter [1]. Sementara itu, terdapat platform digital berbasis website yang dikembangkan khusus untuk memberikan panduan perjalanan wisata melalui informasi akomodasi hotel dan restoran serta jasa transportasi ke lokasi destinasi yang diinginkan wisatawan, salah satunya website Tripadvisor [2]. Hal ini menunjukkan adanya kesenjangan dalam studi terdahulu yaitu belum adanya penjelasan yang secara eksplisit menghubungkan hasil analisis sentimen menggunakan metode klasifikasi untuk menghasilkan sejumlah rekomendasi bagi peningkatan kualitas layanan di suatu destinasi wisata. Adapun, kontribusi dari hasil penelitian ini berupa rekomendasi untuk optimalisasi manajemen destinasi wisata Candi Borobudur. Mempertimbangkan hal tersebut maka penelitian ini bertujuan untuk menganalisis sentimen pengunjung di destinasi wisata super-premium Indonesia yakni Candi Borobudur, menggunakan metode klasifikasi melalui algoritma *Naïve Bayes Classifier* berdasarkan data ulasan pengguna aplikasi website Tripadvisor.

Pendekatan analisis sentimen terhadap citra pariwisata dapat digunakan dalam evaluasi manajemen produk dan layanan destinasi wisata [3]. Dalam perkembangannya, pendekatan *machine learning* digunakan untuk memprediksi sentimen publik berdasarkan ulasan wisatawan di berbagai media digital [4]. Selain itu, Lu & Zheng menunjukkan bahwa data ulasan pengguna media sosial terhadap atraksi wisata dapat diproses menggunakan algoritma metode klasifikasi untuk memperoleh informasi terkait ulasan yang mengandung makna positif, netral dan negatif [5]. Lebih jauh, Luo et al menunjukkan bahwa konsumen yang mengulas produk dan layanan hotel di media digital seperti

website *online hotel booking*, dapat diproses menggunakan algoritma untuk klasifikasi sentimen konsumen [6]. Adapun, Merawati menunjukkan bahwa algoritma *Naïve Bayes Classifier* dapat digunakan untuk mengklasifikasi data ulasan berdasarkan sentimen, baik positif maupun negatif [7]. Hal ini menunjukkan bahwa kajian tentang analisis sentimen destinasi wisata di Indonesia perlu dilakukan untuk menghasilkan rekomendasi peningkatan kualitas produk dan layanan di destinasi wisata.

Studi tentang analisis sentimen terhadap isu-isu kepariwisataan Indonesia menjadi topik populer yang berkontribusi secara empiris maupun teoretis. Putra & Kadyanan menunjukkan bahwa hasil analisis sentimen wisatawan terhadap destinasi wisata menggunakan algoritma *Naïve Bayes Classifier* tidak hanya dalam bentuk visualisasi informasi berdasarkan sentimen positif dan negatif, melainkan juga memantik interpretasi terhadap motivasi wisatawan saat mengunjungi destinasi wisata [8]. Disisi lain, Ratnasari et al. menggunakan algoritma *Naïve Bayes Classifier* dalam penerapan sistem pendukung keputusan yang memudahkan wisatawan untuk memilih objek wisata di aplikasi, dengan demikian wisatawan dapat memenuhi kebutuhan personal sesuai preferensi masing-masing [9]. Hal ini menunjukkan bahwa *Naïve Bayes Classifier* sangat populer digunakan dalam pendekatan klasifikasi untuk mengidentifikasi sentimen wisatawan sebagai pengguna aplikasi digital, baik sistem informasi berbasis *website* maupun *mobile*. Dengan demikian, data ulasan website Tripadvisor dapat digunakan sebagai data untuk menganalisis sentimen pengunjung Candi Borobudur menggunakan metode klasifikasi dan pendekatan *Naïve Bayes Classifier*.

Algoritma *Naïve Bayes Classifier* telah digunakan dalam berbagai kajian analisis sentimen. Salsabila menunjukkan bahwa *Naïve Bayes Classifier* dapat digunakan untuk mengidentifikasi sentimen positif dan negatif pengguna aplikasi, dimana nilai akurasi, *class recall*, *class precision*, serta Area Under Curve (AUC) menunjukkan hasil yang optimal [10]. Adapun, Harliana & Putra menekankan pada akurasi tertinggi yang dihasilkan oleh *Naïve Bayes Classifier* dalam berbagai pengujian data menggunakan *confusion matrix* dan *10-fold cross validation*, terjadi di fold ke-10 [11]. Selain itu, Ibrahim et al. menunjukkan bahwa *Naïve Bayes Classifier* dapat digunakan dalam klasifikasi teks, dimana tahap *pre-processing* diperlukan untuk ekstraksi data berdasarkan *Term Frequency-Inverse Document Frequency* (TF-IDF) agar memperoleh nilai akurasi yang tinggi [12]. Banchhor & Srinivasu menegaskan bahwa tahapan *pre-processing* data diperlukan untuk mempermudah proses kalkulasi, mempertimbangkan kompleksitas *big data* yang memungkinkan adanya format data yang terstruktur dan tidak terstruktur [13]. Rangkaian proses dalam perlu dilakukan dalam tahap *pre-processing* data ialah *data cleaning*, *data integration*, *data selection*, *data transformation*, *mining process*, *pattern evaluation*, and *knowledge presentation*, selanjutnya proses-proses tersebut dapat dilakukan menggunakan berbagai jenis aplikasi seperti RapidMiner [14]. Hal ini menunjukkan bahwa tahapan dalam *preprocessing* data perlu dilakukan untuk memperoleh data yang terstruktur yang dapat dikalkulasikan menggunakan algoritma *Naïve Bayes Classifier* untuk mendapatkan nilai akurasi yang tinggi.

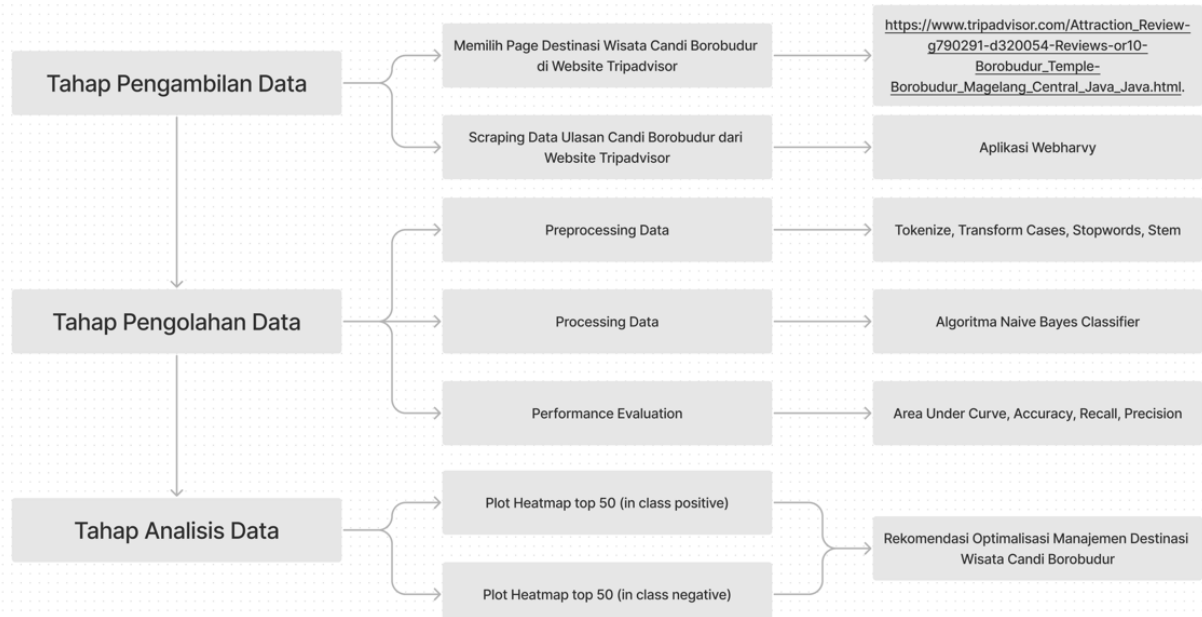
Perkembangan studi tentang implementasi algoritma *Naïve Bayes Classifier* untuk metode klasifikasi, memiliki konteks pembahasan yang beragam. Beberapa studi tentang metode klasifikasi berbasis algoritma *Naïve Bayes Classifier* menunjukkan adanya kebutuhan untuk menghasilkan informasi yang berguna dalam pengambilan keputusan [15]. Bahkan studi tentang *data mining*, dalam hal ini analisis sentimen konsumen terhadap produk dan jasa di menggunakan metode klasifikasi berbasis algoritma *Naïve Bayes Classifier* dinilai efektif dalam menganalisis preferensi konsumen restoran [16]. Analisis sentimen terhadap produk dan jasa dapat dilakukan menggunakan data ulasan konsumen di halaman website yang diproses untuk mendapatkan gambaran tentang perubahan preferensi konsumen hotel [17]. Adapun, beberapa studi menunjukkan bahwa algoritma *Naïve Bayes Classifier* menunjukkan tingkat akurasi yang lebih tinggi, sehingga banyak digunakan dalam berbagai pendekatan analisis sentimen [18], [19]. Mempertimbangkan hal tersebut, maka penelitian ini menggunakan algoritma *Naïve Bayes Classifier* dalam analisis sentimen wisatawan di website Tripadvisor berdasarkan hasil ulasan pengunjung di Candi Borobudur, Provinsi Jawa Tengah, Indonesia.

Perkembangan studi tentang analisis sentimen di sektor pariwisata berkontribusi penting bagi optimalisasi sistem manajemen destinasi pariwisata di Indonesia. Rinaldi et al. menunjukkan bahwa analisis netnografi sentimen pengguna twitter terhadap pembukaan kembali pariwisata di tengah pandemi Covid-19 menjadi populer [1]. Selanjutnya, Godovykh et al. menunjukkan bahwa analisis sentimen terkait pariwisata dan Covid-19 menjadi populer seiring dengan adanya pembatasan mobilitas akses masyarakat di berbagai negara dan daerah [20]. Meningkatnya data digital berupa ulasan produk dan jasa di berbagai macam platform, semakin memperkaya data sehingga dapat dikelola menggunakan berbagai macam pendekatan *big data analysis* untuk mengidentifikasi dan menganalisis sentimen konsumen produk dan jasa bisnis pariwisata, yang dikenal sebagai wisatawan [21]. Disisi lain, perkembangan studi tentang analisis sentimen tidak terbatas pada text dan emojis, melainkan juga gambar yang dilampirkan oleh pengulas berdasarkan konteks dan konten dalam ulasan di website maupun aplikasi digital lainnya [22]. Hal ini menunjukkan bahwa kajian tentang analisis sentimen mengalami perkembangan, demikian juga algoritma dan metode yang diadopsi oleh peneliti untuk meningkatkan akurasi dari hasil analisis sentimen. Dengan demikian dapat diketahui bahwa kajian tentang analisis sentimen penting untuk dilakukan untuk berkontribusi secara empiris dan teoretis.

2. METODOLOGI PENELITIAN

2.1 Tahapan Analisis Sentimen

Proses analisis sentimen wisatawan melalui data ulasan Candi Borobudur di website Tripadvisor menggunakan algoritma Naïve Bayes Classifier terbagi menjadi tiga tahapan sebagai berikut : tahap pertama ialah pengambilan data teks dengan memilih halaman Candi Borobudur di website Tripadvisor kemudian dilakukan proses *scraping* data ulasan menggunakan aplikasi *webharvy*; tahap kedua ialah pengolahan data menggunakan aplikasi Rapidminer dengan memanfaatkan operator untuk *pre-processing* (*tokenize, transform cases, stopwords, stem*), *processing* (*Naïve Bayes Classifier*), dan *performance evaluation* (*Area Under Curve, Accuracy, Recall, Precision*); tahap ketiga ialah tahap analisis data hasil klasifikasi kata (*top 50*) yang divisualisasikan dalam bentuk *plot heatmap* untuk dianalisis sesuai dengan konteks manajemen destinasi wisata di Candi Borobudur, sehingga menghasilkan rekomendasi manajemen destinasi wisata yang relevan dan kontekstual. Adapun, alur dari keseluruhan tahapan dapat dilihat pada gambar 1 berikut ini.



Gambar 1. Alur Proses Analisis Sentimen

Gambar 1 merupakan keseluruhan tahapan penelitian untuk analisis sentimen wisatawan melalui data ulasan Candi Borobudur di website Tripadvisor menggunakan algoritma Naïve Bayes Classifier. Pendekatan analisis sentimen yang digunakan dalam mengidentifikasi dan menganalisis data ulasan konsumen terhadap produk dan jasa bisnis pariwisata yakni wisatawan, diperlukan untuk menghasilkan sejumlah rekomendasi berdasarkan teks yang menjadi sorotan wisatawan dalam bentuk sentimen negatif maupun positif. Setelah mengidentifikasi sentimen positif dan negatif terhadap produk dan jasa, performa algoritma yang digunakan dalam metode klasifikasi analisis sentimen dievaluasi menggunakan *cross validation* untuk mengetahui persentase nilai akurasi, *recall* dan presisi. Dengan demikian, rekomendasi untuk meningkatkan produk dan layanan di suatu destinasi wisata dapat dideskripsikan berdasarkan hasil pengolahan data sentimen. Dalam konteks penelitian ini, proses analisis sentimen sangat relevan dengan konsep penambangan teks (*Text Mining*) atau penambangan data (*Data Mining*) untuk menemukan pengetahuan atau informasi yang memiliki unsur kebaharuan dan manfaat berdasarkan basis data tekstual. Pendekatan penambangan teks memiliki proses yang kompleks tergantung struktur data dalam dokumen. Oleh sebab itu, pemrosesan atau pengorganisasian/pengelompokkan data diperlukan untuk memperoleh hasil yang akurat, terutama karakteristik data dalam jumlah besar (*Big Data*). Hal ini menunjukkan bahwa, ekstraksi dokumen dalam proses pengelompokkan dilakukan karena adanya *noise* pada data dengan struktur teks yang tidak sesuai. Adapun, penyebab struktur teks yang tidak sesuai ialah karena struktur penulisan kalimat masing-masing individu sebagai pengulas, memiliki karakteristik yang beragam. Mempertimbangkan hal tersebut, maka *data pre-processing* meliputi proses *tokenizing, filtering, dan folding* menjadi penting dilakukan untuk menentukan fitur-fitur yang mewakili setiap kata dalam proses klasifikasi menggunakan algoritma *Naïve Bayes Classifier*.

2.2 Algoritma Naïve Bayes Classifier

Naïve Bayes Classifier merupakan metode klasifikasi dengan probabilitas sederhana yang mengaplikasikan teorema bayes dengan asumsi ketidaktergantungan (*independen*) yang tinggi [23]. Penelitian ini didasarkan pada jumlah dataset yang digunakan untuk metode yang mempunyai performansi cepat dan akurat dalam pengklasifikasian. *Naïve Bayes Classifier* hanya membutuhkan data latih (*training data*) yang relatif kecil untuk menentukan estimasi parameter yang dibutuhkan dalam proses klasifikasi. Pada tahap klasifikasi, nilai kategori ditentukan dari data berdasarkan term yang muncul menggunakan persamaan berikut.

$$P(X_k|Y) = \frac{P(Y|X_k)}{\sum_i P(Y|X_i)} \quad (1)$$

Dimana, keadaan posterior (Probabilitas X_k di dalam Y) dapat dihitung dari keadaan prior (Probabilitas Y di dalam X_k) dibagi dengan jumlah dari semua probabilitas Y di dalam semua X_i . Dalam konteks penelitian ini, data teks yang diperoleh dari website Tripadvisor diklasifikasi menggunakan persamaan berikut

$$P(v_1|C=c) = \frac{CountTerms(v_1, docs(c))}{AllTerms(docs(c))} \quad (2)$$

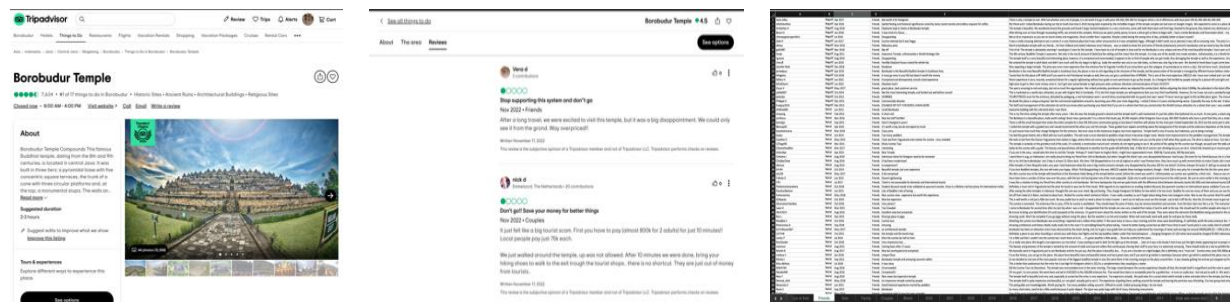
Dimana v_1 merupakan salah satu suku kata yang muncul dalam ulasan pengguna website Tripadvisor terhadap kualitas produk dan layanan di Destinasi Wisata Candi Borobudur. Sedangkan, $CountTerms(v_1, docs(c))$ merujuk pada jumlah kemunculan suatu kata berlabel c (“positif” atau “negatif”). Adapun, $AllTerms(docs(c))$ merujuk pada jumlah semua kata berlabel c yang ada pada dataset. Untuk menghindari adanya nilai nol pada probabilitas maka diimplementasikan laplace smoothing, untuk mengurangi probabilitas dari hasil yang terobservasi, dan jua meningkatkan probabilitas hasil yang belum terobservasi. Dengan demikian, persamaan yang digunakan ialah sebagai berikut :

$$P(v_1|C=c) = \frac{CountTerms(v_1, docs(c))+1}{AllTerms(docs(c))+|V|} \quad (3)$$

Dimana $|V|$ merujuk pada jumlah semua kata dalam data ulasan yang ada di dataset. Dengan demikian, proses klasifikasi data ulasan akan menunjukkan kata dengan nilai tertinggi sebagai representasi perhatian pengulas terhadap produk dan jasa layanan pariwisata di Destinasi Candi Borobudur melalui website Tripadvisor.

2.3 Data Ulasan Candi Borobudur oleh Pengguna Website Tripadvisor

Tripadvisor merupakan platform digital berbasis website yang dapat digunakan sebagai wadah informatif bagi para traveler untuk memberikan rekomendasi perjalanan wisata beserta akomodasi dan transportasi yang dapat digunakan oleh wisatawan ketika bepergian. Selain itu, website Tripadvisor juga menyediakan kolom ulasan bagi para pelancong untuk menceritakan pengalaman hingga mendeskripsikan kesan dan pesan setelah berkunjung. Selain itu, fitur emoji dan rating tersedia dan dapat digunakan oleh pengulas untuk mengekspresikan pengalaman individu maupun berkelompok setelah berkunjung ke suatu destinasi. Dalam konteks penelitian ini, Candi Borobudur menjadi fokus utama dalam pembahasan tentang hasil implementasi algoritma *Naïve Bayes Classifier* dalam analisis sentimen. Berdasarkan data Tripadvisor, terdapat 7.603 ulasan dengan beragam penilaian yaitu *excellent*, *very good*, *average*, *poor*, dan *terrible*. Adapun, data teks yang telah diperoleh ialah sebagai berikut.

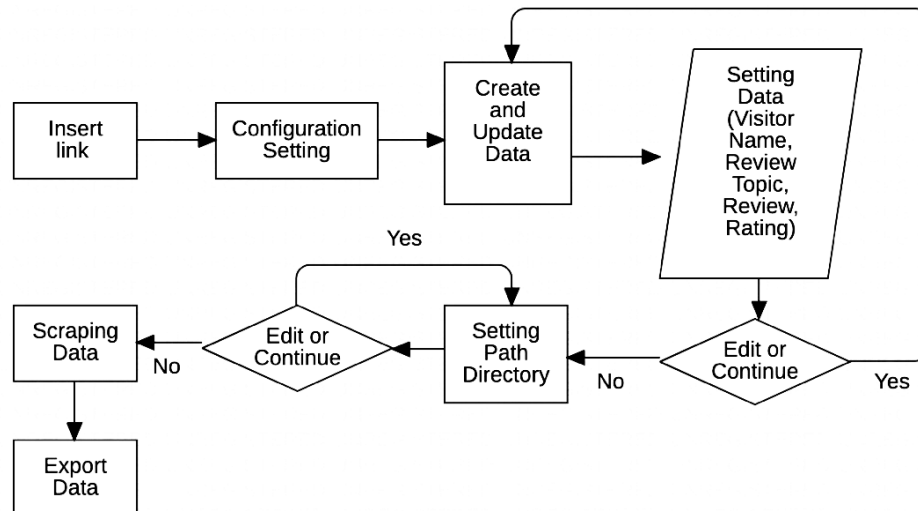


Gambar 2. Data Ulasan Wisatawan di Halaman Website Tripadvisor (Candi Borobudur)

Gambar 2 merupakan *interface* dari Candi Borobudur *page* di website Tripadvisor serta hasil *scraping* data yang diklasifikasikan berdasarkan nama akun, rating, bulan dan tahun postingan, tipe kunjungan, judul ulasan, dan isi ulasan. Ulasan yang tersedia di website Tripadvisor juga diklasifikasikan berdasarkan jenis bahasa yang digunakan oleh pengulas. Hal ini menunjukkan bahwa website Tripadvisor menampung sejumlah data ulasan yang bervariasi dari sisi bahasa, hingga komponen produk dan jasa yang diulas. Meskipun demikian, ulasan pengguna website Tripadvisor masih didominasi oleh data yang tidak terstruktur, sehingga perlu dilakukan proses pengolahan data untuk memperoleh informasi yang valid dan kredibel. Mempertimbangkan hal tersebut, penelitian ini fokus pada data ulasan berbahasa Inggris, dengan mempertimbangkan evaluasi produk dan jasa dalam ulasan wisatawan mancanegara menggunakan bahasa Inggris.

2.4 Pengumpulan Data Menggunakan Webharvy

Pengumpulan data melalui website Tripadvisor dilakukan menggunakan aplikasi Webharvy. Konfigurasi untuk mempermudah proses pengambilan data dilakukan melalui sejumlah tahapan, sebagai berikut : *insert link*; *configuration setting*; *create and update data*; *setting data table (visitor name, review topic, review, rating)*; *setting path directory*; *scraping and export data*. Data yang dihimpun fokus pada ulasan pengguna aplikasi Tripadvisor di destinasi wisata Candi Borobudur, sebagaimana diagram alur dalam gambar berikut.



Gambar 3. Alur Proses Web Scraping Menggunakan Webharvy

Gambar 3 merupakan alur dari proses *scraping* data di website Tripadvisor. Link Candi Borobudur yang digunakan untuk menghimpun jumlah ulasan ialah https://www.tripadvisor.com/Attraction_Review-g790291-d320054-Reviews-or10-Borobudur_Temple-Borobudur_Magelang_Central_Java_Java.html. Adapun proses penyaringan informasi, dipilih berdasarkan identitas akun pengulas (*traveler account identity*), *rating* (1-5), tanggal ulasan (*time of year: January-December*) dan kategori perjalanan (*type of visit: business, couples, family, friends, solo*). Adapun, jenis bahasa yang dipilih ialah bahasa inggris (*english*) dengan pertimbangan bahasa yang paling banyak digunakan oleh pengunjung yang mengulas Candi Borobudur. Berdasarkan data yang ditampilkan dalam website Tripadvisor terdapat 7.603 data ulasan, terdapat 4217 data ulasan dalam bahasa inggris (*english*) yang berhasil dihimpun.

2.5 Data Preprocessing Menggunakan RapidMiner

Dalam tahapan ini, tahap *pre-processing* diperlukan untuk menyiapkan data agar dapat dikalkulasi menggunakan algoritma yang telah ditentukan (*Naïve Bayes Classification*). Preprocessing data sangat berguna untuk menyaring dan menata data agar tidak sama (*duplicate*) atau kosong (*missing data*).

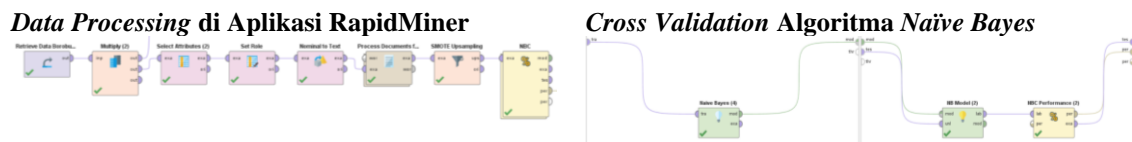


Gambar 4. Tahap Data Preprocessing

Gambar 4 merupakan tahap persiapan data di RapidMiner. Berdasarkan hasil *scraping* di website Tripadvisor terdapat 4217 data ulasan yang perlu disiapkan dalam tahap *preprocessing* sebelum dilakukan proses kalkulasi menggunakan algoritma *Naïve Bayes Classification*, dimana 852 data ulasan tidak memiliki ulasan detail, 722 data yang tidak memiliki tanggal ulasan serta kategori perjalanan. Proses *Data Preprocessing* di aplikasi *RapidMiner* disusun menggunakan beberapa operator-operator sebagai berikut : *select attributes*; *filter example*; dan *process documents form data*. Dalam operator *process documents form data* terdapat operator *tokenize*, *transform cases*, *filter tokens*, *filter stopwords*, dan *stem (snowball)* berdasarkan algoritma TF-IDF (*Term Frequency-Inverse Document Frequency*) dimana Nilai TF (*Term Frequency*) = Jumlah Frekuensi Kata Terpilih/Jumlah Kata. Sedangkan, nilai IDF (*Inverse Document Frequency*) = Log (Jumlah Dokumen/Jumlah Frekuensi Kata Terpilih). Setelah tahap *pre-processing*, diperoleh 3850 data ulasan yang dapat digunakan untuk proses kalkulasi menggunakan *Naïve Bayes Classification* (NBC).

2.6 Data Processing dan Evaluasi Performa Algoritma

Pada tahap *Data processing*, terdapat operator yang digunakan untuk menyelesaikan persoalan *Class Imbalance Problem* (CIP) ialah *Synthetic Minority Oversampling Technique* (SMOTE Upsampling). SMOTE Upsampling sangat efektif digunakan dalam memodifikasi dataset yang tidak seimbang dengan cara membuat sintetik baru dari kelas minoritas agar dapat meningkatkan kinerja metode klasifikasi [24]. Selanjutnya, operator yang digunakan dalam implementasi metode klasifikasi ialah *cross validation* dengan jumlah *fold* 10 dan tipe *sampling* otomatis. Dalam operator *Cross Validation*, terdapat operator *Naïve Bayes*, *Apply Model*, dan *Performance (Binominal Classification)* yang pengaturan parameter ialah *accuracy*, *precision*, *recall*, dan *Area Under Curve* (AUC). Adapun, alur dari data processing dapat dilihat pada gambar 5 berikut ini.



Gambar 5. Data Processing Menggunakan SMOTE Upsampling dan Algoritma Naïve Bayes Classification

Gambar 5 merupakan tahap *data processing* untuk implementasi metode klasifikasi *Naïve Bayes Classification* yang menggunakan operator-operator di aplikasi RapidMiner sebagai berikut : *select attribute (attribute filter type : single; attribute : review detail); set role (attribute name : review detail; target role : regular); nominal to text (attribute filter type : single; attribute : review detail); process document from data (tokenize : regular expression ((?mius)[-!"#\$%&'()*+,-./:;<=>?@\[\]_`{|}~]) dan non-letters; transform cases : lower case; filter tokens by lengths (min 4- max 25), filter stopwords (english), stem : snowballs); SMOTE Upsampling (Number of Neighbours: 5); cross validation (number of fold : 10), Naïve Bayes algorithm, apply model, performance binominal classification (accuracy, precision, recall, AUC)*. Dengan demikian, dapat diperoleh data hasil klasifikasi sentimen positif dan negatif berdasarkan data ulasan pengguna website *Tripadvisor*. Data tersebut dapat diinterpretasikan sebagai ekspresi kepuasan pengunjung ketika berkunjung ke Candi Borobudur, sekaligus mencerminkan kualitas produk dan jasa yang ditawarkan oleh pengelola destinasi. Dengan demikian, uraian penelitian ini akan menghubungkan hasil klasifikasi sentimen pengunjung dengan optimalisasi produk dan layanan pariwisata di Candi Borobudur, Indonesia.

3. HASIL DAN PEMBAHASAN

Kajian tentang implementasi algoritma *Naïve Bayes Classification* untuk menganalisis sentimen pengguna media sosial memiliki kontribusi penting bagi optimalisasi manajemen destinasi pariwisata [7], [25], isu kesehatan [26] dan bisnis [27] serta hal-hal yang berhubungan dengan pemerintahan [28]. Disisi lain, *Andreyestha & Azizah* menunjukkan bahwa *SMOTE Upsampling* digunakan untuk mengatasi persoalan data yang tidak seimbang [29]. Intensitas riset tentang analisis sentimen pengguna media sosial menunjukkan rekomendasi untuk mengoptimalkan sistem layanan, produk, kebijakan, serta pemetaan kondisi sosial di suatu daerah. Hal ini menunjukkan bahwa hasil analisis sentimen merupakan proses pengolahan data yang menunjukkan informasi secara kontekstual sehingga dapat digunakan berdasarkan kepentingan sektoral, salah satunya ialah sektor pariwisata. Sektor pariwisata, memiliki sejumlah aspek kerentanan yang berpengaruh terhadap intensitas kunjungan yaitu citra destinasi wisata [2]. Untuk membentuk citra destinasi yang positif bagi penunjang, pihak manajemen destinasi dapat melakukan studi lanskap destinasi pariwisata perihal dampak pembangunan pariwisata terhadap kondisi lingkungan, sosial dan ekonomi [30]–[34] maupun merancang sistem informasi pariwisata berbasis website yang kontekstual, sebagai strategi pemasaran digital yang menarik para pengunjung [35]–[37]. Dengan demikian, dapat diketahui bahwa luaran dari kajian analisis sentimen pengunjung di destinasi wisata, dapat berkontribusi bagi peningkatan kualitas manajemen destinasi wisata yang kontekstual.

Berdasarkan hasil pengolahan 3850 data teks menggunakan operator *cross validation* untuk mengevaluasi performa algoritma *Naïve Bayes* dalam mengklasifikasi kata ke dalam kelas positif dan kelas negatif di aplikasi *Rapidminer*, dapat diketahui adanya perbedaan nilai *accuracy*, *recall*, *precision*, serta *Area Under Curve (AUC)* sebelum dan setelah menggunakan operator *SMOTE Upsampling*. Sebelum menggunakan operator *SMOTE Upsampling*, diperoleh nilai *accuracy* sebesar 90,52%. Selain itu, nilai *precision* sebesar 5,69% dan nilai *recall* sebesar 14,41%. Adapun, nilai *classification error* sebesar 9,48%. Secara spesifik, terdapat 3469 kata (*pred.positif-true positif*), 95 kata (*pred.positif-true negatif*), 270 kata (*pred.negative-true positif*), dan 16 kata (*pred.negative-true negatif*). Adapun, nilai *Area Under Curve (AUC)* yang diperoleh sebelum menggunakan *SMOTE Upsampling* ialah 0.500 (50%). Setelah menggunakan *SMOTE Upsampling*, nilai *accuracy* meningkat dari 90,52% menjadi 96,36%. Selain itu, nilai *precision* juga meningkat dari 5,69% menjadi 93,23% dan nilai *recall* meningkat dari 14,41% menjadi 100%. Adapun, nilai *classification error* menurun dari 9,48% menjadi 3,64%. Secara spesifik, terdapat 3479 kata (*pred.positif-true positif*), 0 kata (*pred.positif-true negatif*), 272 kata (*pred.negative-true positif*), dan 3739 kata (*pred.negative-true negatif*). Adapun, nilai *Area Under Curve (AUC)* yang diperoleh setelah menggunakan *SMOTE Upsampling* ialah 0.714 (71,4%). Hal ini menunjukkan adanya persoalan data yang tidak seimbang sehingga membutuhkan operator *SMOTE Upsampling* untuk meningkatkan performa dari algoritma *Naïve Bayes Classifier*. Pada tabel 1 dapat dilihat nilai *Confusion Matrix* dari algoritma *Naïve Bayes* sebelum dan setelah menggunakan operator *SMOTE Upsampling*.

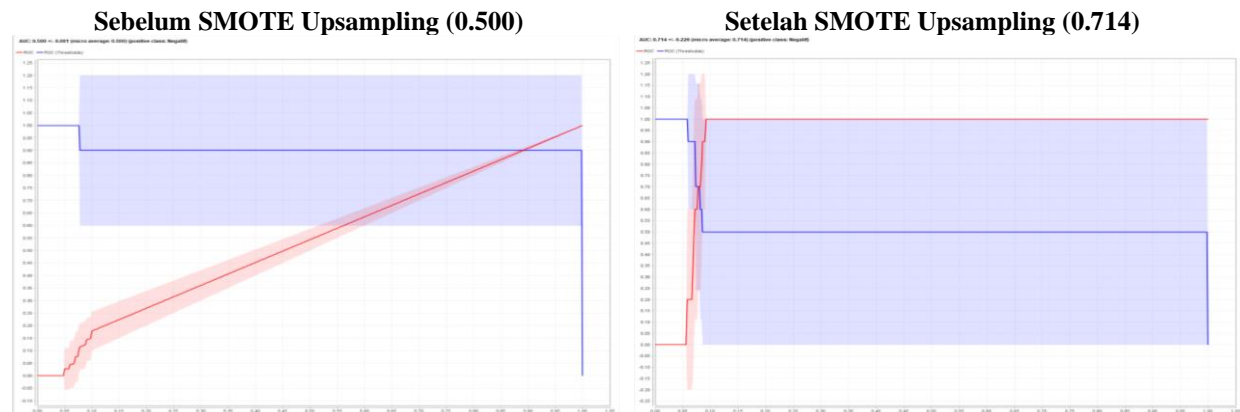
Tabel 1. Evaluasi Confusion Matrix Naïve Bayes sebelum dan setelah menggunakan *SMOTE Upsampling*

Sebelum menggunakan <i>SMOTE Upsampling</i>			
Accuracy	90.52% +/- 1.38% (Micro Average: 90.52%)		
	True Positif	True Negatif	Class Precision
Pred. Positif	3469	95	97.33%
Pred. Negatif	270	16	5.69%
Class Recall	92.78%	14.41%	

Setelah menggunakan SMOTE Upsampling

Accuracy	96.36% +/- 0.58% (Micro Average: 96.36%)		
Pred. Positif	True Positif	True Negatif	Class Precision
Pred. Negatif	3467	0	100%
Class Recall	272	3739	93.22%
	92.73%	100%	

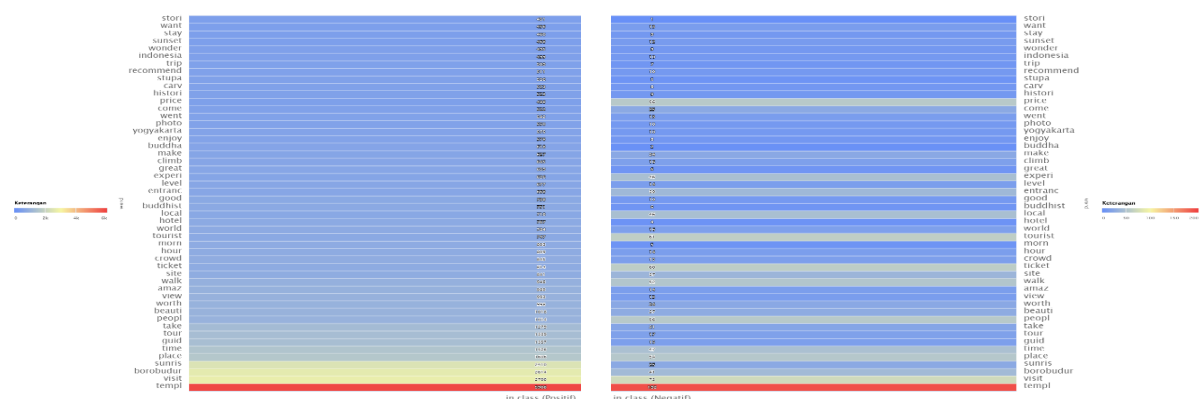
Tabel 1 merupakan hasil dari tahap processing data di aplikasi Rapidminer yang menunjukkan adanya perbedaan nilai *accuracy*, *precision*, dan *recall* sebelum dan setelah menggunakan SMOTE Upsampling. Berdasarkan hasil pengujian nilai *Area Under Curve* (AUC) diketahui bahwa nilai AUC sebelum menggunakan SMOTE Upsampling ialah 0.500 (50%). Setelah menggunakan SMOTE Upsampling, terjadi peningkatan nilai AUC dari dari 0.500 (50%) menjadi 0.714 (71,4%), sebagaimana gambar berikut.



Gambar 4. Area Under Curve (AUC) sebelum dan setelah menggunakan SMOTE Upsampling

Gambar 4 menunjukkan bahwa nilai AUC mengalami peningkatan setelah menggunakan SMOTE Upsampling dari 0,500 menjadi 0,714. Adapun, Kadafi menunjukkan bahwa klasifikasi performance AUC pada nilai 0,50-0,60 (gagal), 0,60-0,70 (rendah), 0,70-0,80 (adil/sama), 0,80-0,90 (baik), 0,90-1,00 (paling baik) [38]. Dengan demikian dapat diketahui bahwa nilai AUC dari hasil kalkulasi data ulasan pengguna website Tripadvisor menggunakan algoritma Naïve Bayes Classifier tergolong baik. Algoritma Naïve Bayes dalam beberapa konteks pengolahan data teks untuk analisis sentimen, mampu menunjukkan performa yang baik [28], [39]. Meskipun, dalam beberapa konteks data teks, hasil perbandingan performa Naïve Bayes dengan beberapa algoritma lainnya seperti *Decision Tree*, *Support Vector Machine*, dan *k-NN* menunjukkan hasil yang sebaliknya [40]. Dengan demikian dapat diketahui bahwa, penggunaan algoritma Naïve Bayes Classifier dalam metode klasifikasi bersifat kontekstual.

Berdasarkan hasil prediksi data teks yang tergolong sentimen positif dan negatif, dapat diketahui 3467 kata dalam ulasan yang mengandung unsur sentimen positif, serta 270 kata dalam ulasan yang mengandung unsur sentimen negatif. Dalam optimalisasi layanan produk dan jasa di destinasi wisata Candi Borobudur, jumlah kata dalam ulasan yang mengandung unsur sentimen positif dapat digunakan sebagai petunjuk untuk mempertahankan hal-hal yang berhubungan atraksi, aksesibilitas, amenitas, dan akomodasi melalui jumlah kata terbanyak yang mencul dalam ulasan. Seperti halnya kata-kata yang berhubungan dengan atraksi ialah “place, borobudur, sunrise, visit, temple” yang mengindikasikan manajemen atraksi pariwisata perlu dioptimalkan dan dipertahankan. Demikian juga sebaliknya, kata dalam ulasan yang mengandung unsur sentimen negatif dapat digunakan sebagai bentuk evaluasi untuk membenahi sistem manajemen destinasi sehingga mampu mempertahankan dan meningkatkan kepuasan pengunjung. Penelitian ini juga menunjukkan 50 kata yang paling sering muncul di setiap ulasan, sebagaimana gambar berikut.



Gambar 5. Plot Heatmap 50 Kata yang Terdeteksi dalam Ulasan di Website Tripadvisor

Gambar 5 menunjukkan bahwa hasil klasifikasi dalam bentuk *heatmap plot*, terdapat 50 kata yang paling sering muncul dalam ulasan pengunjung. Apabila diurutkan berdasarkan 10 kata yang diprediksi termasuk dalam ulasan dengan unsur sentimen positif dan negatif serta paling sering muncul dalam ulasan pengguna *website Tripadvisor* mengenai Candi Borobudur, maka diperoleh kata-kata berikut ini : kata *temple* terdeteksi sebanyak 6100 dari 2673 ulasan; kata *visit* terdeteksi sebanyak 2778 dari 1780 ulasan; kata *borobudur* terdeteksi sebanyak 2655 dari 1513 ulasan; kata *sunrise* terdeteksi sebanyak 2435 dari 1456; kata *place* terdeteksi sebanyak 1660 dari 1218 ulasan; kata *time* terdeteksi sebanyak 1573 dari 1128 ulasan; kata *guide* terdeteksi sebanyak 1372 dari 964 ulasan; kata *tour* terdeteksi sebanyak 1356 dari 900 ulasan; kata *take* terdeteksi sebanyak 1300 dari 957 ulasan; kata *people* terdeteksi sebanyak 1070 dari 812 ulasan. Hal ini mengindikasikan bahwa atraksi menjadi aspek yang dominan dikomentari oleh pengunjung Candi Borobudur. Selain itu, kata “*people*” mengindikasikan kerahaman (*hospitality*) dari masyarakat maupun pengelola destinasi yang meninggalkan kesan yang baik. Studi tentang manajemen destinasi pariwisata menunjukkan bahwa keramahan menjadi salah satu nilai yang sangat penting dalam meningkatkan kepuasan pengunjung [41]. Disisi lain, Hadi & Widyarningsih menunjukkan bahwa daya tarik suatu destinasi dapat berupa nilai-nilai *sapta pesona* yakni keamanan, ketertiban, kebersihan, kesejukan, keindahan, keramahan, dan ketenangan [41]. Hal ini menunjukkan bahwa ulasan pengguna *website Tripadvisor* tentang Candi Borobudur tidak hanya berhubungan dengan atraksi, aksesibilitas, amenitas, dan *ancillary* melainkan juga *sapta pesona*.

Berdasarkan hasil penelitian ini terdapat sejumlah rekomendasi bagi pengelola destinasi wisata Candi Borobudur agar dapat mengoptimalkan produk dan layanan di destinasi wisata, sehingga meningkatkan kepuasan wisatawan. Mempertimbangkan intensitas kata *temple*, *borobudur*, dan *place* yang paling sering muncul menunjukkan kesan terhadap atraksi wisata, maka otentisitas candi perlu dijaga sehingga tidak mengalami degradasi nilai warisan budaya, akibat aktivitas kungkungan wisata yang masif. Selain itu, aspek keramahan perlu dipertahankan sebagai daya tarik yang berhubungan dengan kebudayaan masyarakat lokal dalam berinteraksi maupun berkomunikasi. Mempertimbangkan kata “*people*” yang terdeteksi sebanyak 1070 dari 812 ulasan, maka hal-hal yang berhubungan dengan layanan prima perlu diperhatikan untuk mempertahankan dan meningkatkan kepuasan pengunjung. Meskipun, hasil pengolahan data penelitian ini menunjukkan sentimen positif yang lebih dominan dibandingkan sentimen negatif, citra pariwisata Candi Borobudur perlu dipertahankan agar menjadi daya tarik sekaligus menjadi simbol dari jati diri bangsa Indonesia.

4. KESIMPULAN

Hasil penelitian ini menunjukkan bahwa proses klasifikasi data teks berdasarkan sentimen negatif dan positif menggunakan algoritma *Naïve Bayes Classifier* terhadap data ulasan pengguna *website Tripadvisor* dapat diproses menggunakan aplikasi Rapidminer. Hasil pengolahan data ulasan terhadap destinasi wisata Candi Borobudur menunjukkan adanya performa yang baik. Hal tersebut dapat dilihat dari 3850 data ulasan yang telah diproses menghasilkan nilai akurasi sebesar 96,36%, nilai presisi 93,23% dan recall sebesar 100%. Adapun, nilai *classification error* sebesar 3,64% dengan Area Under Curve (AUC) sebesar 0,714 (71,4%). Disisi lain, hasil klasifikasi sepuluh kata yang terdeteksi paling sering muncul di setiap ulasan pengunjung ialah sebagai berikut: *temple*, *visit*, *borobudur*, *sunrise*, *place*, *time*, *guide*, *tour*, *take*, *people*. Dengan demikian dapat diketahui bahwa hal-hal yang berkesan bagi pengunjung di Candi Borobudur lebih dominan pada aspek atraksi. Dengan demikian, rekomendasi bagi pengelola destinasi wisata Candi Borobudur agar mempertahankan dan meningkatkan performa layanan produk dan jasa yang berhubungan dengan otentikasi warisan budaya serta mengimplementasikan nilai-nilai *sapta pesona* untuk meningkatkan kepuasan pengunjung dan citra destinasi wisata Candi Borobudur sebagai destinasi wisata super prioritas di Indonesia.

REFERENCES

- [1] A. R. Rinaldi, J. Damanik, and D. Mutiarin, “Analisis Netnografi Sentimen Pengguna Twitter Terhadap Pembukaan Kembali Pariwisata Di Tengah Pandemi Covid-19,” *Pariwisata Budaya J. Ilm. Agama Dan Budaya*, vol. 6, no. 1, pp. 27–36, 2021, doi: 10.25078/pba.v6i1.1982.
- [2] Y. A. Singgalen, “Sentiment Analysis on Customer Perception towards Products and Services of Restaurant in Labuan Bajo,” *J. Inf. Syst. Informatics*, vol. 4, no. 3, pp. 511–523, 2022, doi: 10.51519/journalisi.v4i3.276.
- [3] Q. Jiang, C. S. Chan, S. Eichelberger, H. Ma, and B. Pikkemaat, “Sentiment analysis of online destination image of Hong Kong held by mainland Chinese tourists,” *Curr. Issues Tour.*, vol. 24, no. 17, pp. 2501–2522, 2021, doi: 10.1080/13683500.2021.1874312.
- [4] K. Puh and M. B. Babac, “Predicting Sentiment and Rating of Tourist Reviews Using Machine Learning,” *J. Hosp. Tour. Insights*, 2022, doi: 10.1108/JHTI-02-2022-0078.
- [5] Y. Lu and Q. Zheng, “Twitter public sentiment dynamics on cruise tourism during the COVID-19 pandemic,” *Curr. Issues Tour.*, vol. 24, no. 7, pp. 892–898, 2021, doi: 10.1080/13683500.2020.1843607.
- [6] J. Luo, S. Huang, and R. Wang, “A fine-grained sentiment analysis of online guest reviews of economy hotels in China,” *J. Hosp. Mark. Manag.*, vol. 30, no. 1, pp. 71–95, 2021, doi: 10.1080/19368623.2020.1772163.
- [7] N. L. P. Merawati, A. Z. Amrullah, and Ismarmiaty, “Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation,” *Rekayasa Sist. dan Teknol. Inf.*, vol. 1, no. 10, pp. 123–131, 2021, doi: 10.29207/resti.v5i1.2587.
- [8] I. G. N. A. W. Putra and I. G. A. G. A. Kadyanan, “Optimization of Bali Tourism Recommendations Based on Personal

- Motivation of Tourists Using the Naive Bayes Algorithm,” *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 10, no. 1, p. 83, 2021, doi: 10.24843/jlk.2021.v10.i01.p11.
- [9] A. Ratnasari *et al.*, “Penerapan sistem pendukung keputusan untuk pemilihan objek wisata di majalengka menggunakan algoritma naïve bayes,” vol. 9, pp. 54–59, 2019.
- [10] S. M. Salsabila, A. A. Murtopo, and N. Fadhilah, “Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Pada Angket Mahasiswa,” *J. Minfo*, vol. 11, no. 2, pp. 30–35, 2022.
- [11] Harliana and F. N. Putra, “Klasifikasi Tingkat Rumah Tangga Miskin Saat Pandemi Dengan Naïve Bayes Classifier,” *J. Sains dan Inform.*, vol. 7, no. 2, pp. 165–173, 2021, doi: 10.34128/jsi.v7i2.339.
- [12] M. F. Ibrahim, M. A. Alhakeem, and N. A. Fadhil, “Evaluation of Naïve Bayes Classification in Arabic Short Text Classification,” *Al-Mustansiriyah J. Sci.*, vol. 32, no. 4, pp. 42–50, 2021, doi: 10.23851/mjs.v32i4.994.
- [13] C. Banchhor and N. Srinivasu, “Analysis of Bayesian optimization algorithms for big data classification based on Map Reduce framework,” *J. Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00464-4.
- [14] P. S. Dewi, C. K. Sastradipraja, and D. Gustian, “Sistem Pendukung Keputusan Kenaikan Jabatan Menggunakan Metode Algoritma Naïve Bayes Classifier,” *J. Teknol. dan Inf.*, vol. 11, no. 1, pp. 66–80, 2021, doi: 10.34010/jati.v11i1.3593.
- [15] M. Ismail, N. Hassan, and S. S. Bafjaish, “Comparative Analysis of Naive Bayesian Techniques in Health-Related for Classification Task,” *J. Soft Comput. Data Min.*, vol. 1, no. 2, pp. 1–10, 2020, [Online]. Available: https://www.researchgate.net/publication/352312714_Comparative_Analysis_of_Naive_Bayesian_Techniques_in_Health-Related_for_Classification_Task
- [16] V. A. Permadi, “Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura,” *J. Buana Inform.*, vol. 11, no. 2, pp. 141–151, 2020, doi: 10.24002/jbi.v11i2.3769.
- [17] I. E. Tiffani, “Optimization of Naïve Bayes Classifier By Implemented Unigram, Bigram, Trigram for Sentiment Analysis of Hotel Review,” *Jossex*, vol. 1, no. 1, pp. 1–7, 2020.
- [18] A. Firmansyah and Eriswandi, “Analisis Sentimen Terhadap Operator Seluler Telkomsel Menggunakan Algoritma Naive Bayes,” *SIGMA – J. Teknol. Pelita Bangsa*, vol. 10, no. 4, pp. 108–116, 2020.
- [19] L. B. Ilmawan and M. A. Mude, “Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store,” *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 154–161, 2020, doi: 10.33096/ilkom.v12i2.597.154-161.
- [20] M. Godovykh, J. Ridderstaat, C. Baker, and A. Fyall, “COVID-19 and Tourism: Analyzing the Effects of COVID-19 Statistics and Media Coverage on Attitudes toward Tourism,” *Forecasting*, vol. 3, no. 4, pp. 870–883, 2021, doi: 10.3390/forecast3040053.
- [21] A. R. Alaci, S. Becken, and B. Stantic, “Sentiment Analysis in Tourism: Capitalizing on Big Data,” *J. Travel Res.*, vol. 58, no. 2, pp. 175–191, 2019, doi: 10.1177/0047287517747753.
- [22] S. Wei and S. Song, “Sentiment Classification of Tourism Reviews Based on Visual and Textual Multifeature Fusion,” *Wirel. Commun. Mob. Comput.*, vol. 2022, no. 1, pp. 1–10, 2022, doi: 10.1155/2022/9940817.
- [23] F. Nurhuda, S. W. Sihwi, and A. Doewes, “Analisis Sentimen Masyarakat Terhadap Pilpres 2019 Berdasarkan Opini Dari Twitter Menggunakan Metode Naive Bayes Classifier,” *J. ITSMART*, vol. 2, no. 2, pp. 35–42, 2013, doi: 10.51519/journalcisa.v1i3.45.
- [24] C. Cahyaningtyas, Y. Nataliani, and I. R. Widiarsari, “Analisis Sentimen Pada Rating Aplikasi Shopee Menggunakan Metode Decision Tree Berbasis SMOTE,” *Aiti*, vol. 18, no. 2, pp. 173–184, 2021, doi: 10.24246/aiti.v18i2.173-184.
- [25] Y. A. Singgalen, “Analisis Sentimen dan Pemodelan Topik dalam Optimalisasi Pemasaran Destinasi Pariwisata Prioritas di Indonesia,” *J. Inf. Syst. Informatics*, vol. 4, no. 1, pp. 459–470, 2021, [Online]. Available: <http://journal-isi.org/index.php/isi/article/view/171>
- [26] F. F. Mailo and L. Lazuardi, “Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia,” *J. Inf. Syst. Public Heal.*, vol. 4, no. 1, pp. 28–36, 2019, doi: <https://doi.org/10.22146/jisph.44455>.
- [27] N. Legiawati, T. I. Hermanto, and Y. R. Ramadhan, “Analisis Sentimen Opini Pengguna Twitter Terhadap Perusahaan Jasa Ekspedisi Menggunakan Algoritma Naïve Bayes Berbasis PSO,” *J. Ris. Komput.*, vol. 9, no. 4, pp. 930–937, 2022, doi: 10.30865/jurikom.v9i4.4629.
- [28] F. Zamachsari, G. Vangeran Saragih, Susafa’ati, and W. Gata, “Analisis Sentimen Pemindahan Ibu Kota Negara dengan Feature Selection Algoritma Naive Bayes dan Support Vector Machine,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 3, pp. 504–512, 2017, doi: 10.29207/resti.v4i3.1942.
- [29] A. Andreyestha and Q. N. Azizah, “Analisa Sentimen Kicauan Twitter Tokopedia Dengan Optimalisasi Data Tidak Seimbang Menggunakan Algoritma SMOTE,” *Infotek J. Inform. dan Teknol.*, vol. 5, no. 1, pp. 108–116, 2022, doi: 10.29408/jit.v5i1.4581.
- [30] Y. A. Singgalen, “Vegetation Index and Mangrove Forest Utilization through Ecotourism Development in Dodola and Guraping of North Maluku Province,” *J. Manaj. Hutan Trop.*, vol. 28, no. 2, pp. 150–161, 2022, doi: 10.7226/jtfm.28.2.150.
- [31] Y. A. Singgalen, “Tourism Infrastructure Development and Transformation of Vegetation Index in Dodola Island of Morotai Island Regency,” *J. Inf. Syst. Informatics*, vol. 4, no. 1, pp. 130–144, 2022.
- [32] Y. A. Singgalen and D. Manongga, “Monitoring of Mangrove Ecotourism Area using NDVI, NDWI, and CMRI in Dodola Island, Morotai Island Regency, Indonesia,” *J. Ilmu dan Teknol. Kelaut. Trop.*, vol. 14, no. 1, pp. 95–108, 2022.
- [33] Y. A. Singgalen and D. Manongga, “Mangrove-based Ecotourism Sustainability Analysis using NDVI and AHP Approach,” *Indones. J. Comput. Cybern. Syst.*, vol. 16, no. 2, pp. 125–136, 2022, doi: 10.22146/ijccs.68986.
- [34] Y. A. Singgalen, “Priority Analysis of Mangrove Guraping Ecotourism Development Based on Spatial Data Using Process Hierarchy Analysis,” *J. Inf. Syst. Informatics*, vol. 4, no. 1, pp. 1–15, 2022.
- [35] Y. A. Singgalen, “Strategic Planning of Ecotourism Management System Using Ward and Peppard Framework,” *J. Inf. Syst. Informatics*, vol. 4, no. 2, pp. 216–232, 2022, doi: <https://doi.org/10.51519/journalisi.v4i2.245>.
- [36] H. Prasadja and Y. A. Singgalen, “Analysis and Design of Mangrove Ecotourism Management System (SIMANGROVE) of Dodola Island , Morotai Island Regency , Indonesia,” *J. Inf. Syst. Informatics*, vol. 4, no. 2, pp. 191–204, 2022, doi: <https://doi.org/10.51519/journalisi.v4i2.243>.
- [37] E. Widodo and Y. A. Singgalen, “Strategic Planning of Regional Tourism Information System of Morotai Island Regency



- using Ward and Peppard Framework,” *J. Inf. Syst. Informatics*, vol. 4, no. 3, pp. 556–573, 2022, doi: 10.51519/journalisi.v4i3.282.
- [38] A. R. Kadafi, “Perbandingan Algoritma Klasifikasi Untuk Penjurusan Siswa SMA,” *J. ELTIKOM*, vol. 2, no. 2, pp. 67–77, 2018, doi: 10.31961/eltikom.v2i2.86.
- [39] B. S. Prakoso, D. Rosiyadi, H. S. Utama, and D. Aridarma, “Klasifikasi Berita Menggunakan Algoritma Naive Bayes Classifier Dengan Seleksi Fitur Dan Boosting,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 2, pp. 227–232, 2019, doi: 10.29207/resti.v3i2.1042.
- [40] R. Fatmasari, V. M. Ayu, B. Pratama, and W. Gata, “Analisis Sentimen Dalam Pengkategorian Komentar Youtube Terhadap Layanan Akademik dan Non-Akademik Universitas Terbuka Untuk Prediksi Kepuasan,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 2, pp. 395–404, 2022, doi: 10.47065/bits.v4i2.1738.
- [41] W. Hadi and H. Widyarningsih, “Implementasi Penerapan Sapta Pesona Wisata Terhadap Kunjungan Wisatawan Di Desa Sambirejo Kecamatan Prambanan Kabupaten Sleman Daerah Istimewa Yogyakarta Wisnu,” *Khasanah Ilmu J. Pariwisata Dan Budaya*, vol. 11, no. 2, pp. 127–136, 2020, doi: 10.31294/khi.v11i2.8862.