G-Tech : Jurnal Teknologi Terapan
Vol. 8, No. 3, July 2024, pp. 1656-1667
E-ISSN: 2623-064X | P-ISSN: 2580-8737

Optimizing Tourist Satisfaction at Mount Bromo: A Sentiment Analysis Approach

Alvin Lukita[1], Melissa Indah Fianty[2]✉

[1, 2] Department of Information System, Faculty of Engineering & Informatics, Universitas Multimedia Nusantara, Indonesia

Article Information

Article History

Received : Mei 30,2024
Revised : Juni 11, 2024
Accepted : Juni 12,2024

ABSTRACK

In the era of ever-growing globalization, information technology is the main driver of transformation in various sectors of life. Information technology plays an important role in the progress of the tourism sector by offering easy access to information. The Mount Bromo tourist attraction is one of the popular destinations in Indonesia which attracts many visitors every year. The large number of comments from visitors to the Mount Bromo tourist attraction on the Trip Advisor site means that there is no filtering of information or reviews so that the Mount Bromo National Park management does not know whether there are many positive or negative reviews. The method used in the research is the Naive Bayes method it produces an accuracy value of 0.85%. This accuracy value can be used to classify visitor review data at the Mount Bromo tourist attraction. The conclusion is that there are many positive reviews.

***Keywords*:**

Classification; Mount Bromo; Naive Bayes; Sentiment Analysis.

Kata Kunci:

Analisis sentimen; Gunung Bromo; Klasifikasi; Naive Bayes.

ABSTRAK

Teknologi infromasi berperan penting dalam kemajuan sektor pariwisata dengan menawarkan kemudahan akses informasi. Objek wisata Gunung Bromo adalah salah satu destinasi populer di Indonesia. Banyaknya komentar dari pengunjung Objek Wisata Gunung Bromo pada situs Trip Advisor menyebabkan tidak ada penyaringan informasi atau ulasan. Ulasan tersebut dapat dianalisa dengan menggunakan analisis sentimen. Metode yang digunakan dalam penelitan adalah Metode Naive bayes, data ulasan pengunjung objek wisata Gunung Bromo, menghasilkan nilai akurasi sebesar 0.85%. Hasil ulasan masyarakat mengenai objek wisata Gunung Bromo adalah terdapat 1001 ulasan, sebanyak 847 ulasan diklasifikasikan sebagai sentimen positif, dan 154 ulasan diklasifikasikan sebagai sentiment negatif. Kesimpulan menunjukan bahwa ulasan objek wisata Gunung Bromo banyak terdapat ulasan yang bersifat positif. Ulasan positif dan negaif tersebut dapat digunakan oleh pengelola objek wisata dalam memahami area mana yang memerlukan perbaikan atau pengembangan lebih lanjut.

***Corresponding Author* :**
Melissa Indah Fianty
Information System Department, Faculty of Engineering & Informatics, Universitas Multimedia Nusantara Provinsi Banten, Indonesia
Jl. Boulevard, Gading Serpong, Kel. Curug Sangereng, Kec. Kelapa Dua, Kab. Tangerang, Provinsi Banten, Indonesia
Email: melissa.indah@umn.ac.id

This is an open access article under the CC BY license

INTRODUCTION

The tourism industry is one of the many industries that has utilized information technology (Rachmawati et al., 2020). Information technology facilitates easy access to information, convenient services for tourists, and a number of other features that may be beneficial to tourists, all of which contribute to the growth of the tourism industry. Modern technology and the internet also have a significant positive impact on the tourism industry (Rahayu, 2019). For example, currently there are many online sites that provide booking services for hotels, restaurants, accommodation, train tickets, planes and other modes of transportation. Apart from that, currently many tourists are looking for photos and reviews from previous tourists(Ashoer et al., 2021).

Tourists can quickly research tourist destinations using various platforms, including TripAdvisor (Izza et al., 2022). The world's largest travel website, TripAdvisor can help domestic and international travelers in planning trips. Apart from that, TripAdvisor also offers suggestions regarding resorts, hotels, inns, holidays, travel packages, and travel guides (Assyafah et al., 2021). Prospective visitors can find out references to these tourist attractions based on the experiences of other tourists who have visited there. Location that attracts tourists because of its natural and artificial resources is called a tourist attraction.

Mount Bromo is one of the most famous tourist destinations in Indonesia, attracting thousands of tourists every year. Its popularity makes it an interesting subject to research in the context of sentiment analysis, due to the large number of reviews, comments and social media posts related to the experience of visiting Mount Bromo (Imron, 2019). As a popular tourist destination, Mount Bromo has many visitors with different backgrounds and experiences. Therefore, sentiment analysis towards Mount Bromo can provide rich insight into the various perceptions, views and experiences of visitors towards this place (Oktaviani et al., 2022).

The Mount Bromo tourist attraction is a famous tourist attraction for domestic and foreign tourists, but in the last five years the number of visitors to Mount Bromo has decreased and is inconsistent, and has only slightly increased. Based on information from the Bromo Tengger Semeru National Park Center (BBTNBTS), as many as 690,831 tourists visited Bromo Tengger Semeru National Park in 2019 (Adilarif, 2020). Compared to 2018, when there were 825,206 people visiting, the number of visits has dropped significantly. The TNBTS visitor data graph for the next four years also shows a varying trend that is not always stable and only has a slight increase (Hilyah et al., 2021). This problem occurs because the development of tourist attractions has not been managed properly, which has an impact on different Bromo entrance ticket prices (Singgalen, 2022). This problem has occurred since ticket prices were increased by the management fraudulent practices often occur when purchasing tickets. If you want to visit the Mount Bromo area, there are also differences in certain entry prices that apply. The management of tourist destination products and services can be assessed using the sentiment analysis method for tourism image. Machine Learning methods are being developed to estimate public opinion based on tourists' evaluations in digital media. Furthermore, a classification approach algorithm can be used to evaluate social media user review data about tourist attractions to extract review information with positive, neutral and negative meaning (Benoît, 2023).

Improvements in services for visitors continue to be made. Comments given by visitors to tourist destinations can be used as input for improving services and developing tourist attractions. Since TripAdvisor is a widely used platform for travel-related reviews and recommendations, reviews on the site may offer useful information. Information regarding tourist sentiment towards Mount Bromo was carried out using the sentiment analysis method (Kuncoro, 2022). Positive and negative sentiment will be checked using sentiment analysis. Additionally, it will utilize supervised learning to assist with sentiment analysis procedures called Supervised Learning. Naive Bayes is a classifier method that is often used in developing classification models (Astawa & Sudibia, 2021). Additionally, exploring review data can improve sentiment analysis results, including the average rating given by visitors and the terms included in each sentiment. It is hoped that these findings can provide the knowledge needed by tourist attraction managers to develop Mount Bromo tourist destinations (Hidayat et al., 2021).

Based on this background, this research aims to conduct sentiment analysis using the Naive Bayes method. This researched report can provide useful information for the Malang City Culture and Tourism Office, Mount Bromo tourist attraction managers, and other parties. Therefore, this report will analyze the sentiment of reviews of the Mount Bromo tourist attraction. The comments column on TripAdvisor, written by visitors to the Mount Bromo tourist attraction, is the data source for this review.

RESEARCH METHODS

CRISP-DM is the research methodology used in this study. As a technique, CRISP-DM provides an overview of the data mining life cycle when viewed as a process model. It also includes the phases of the project, the work performed in each phase, and an explanation of the relationships between these activities (Damayanti et al., 2020). The stages of CRISP-DM are shown in Figure 1.

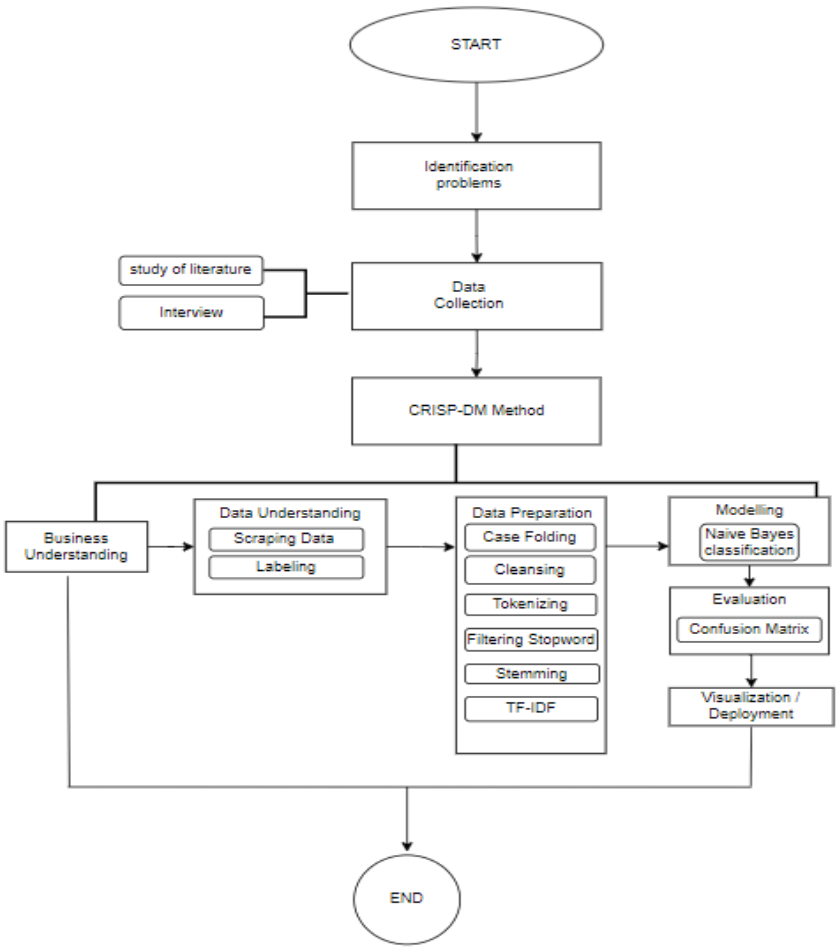

Figure 1. Research Flowchart

Identification Problems

Interviewing the parties involved to identify the problem is the first step in this procedure. This phase is useful in identifying existing problems. This problem can be used as a topic of discussion for research. From the results of interviews conducted, problems occurred because the development of tourist attractions had not been managed properly.

Data Collection

At the data collection stage there were 2 methods used, namely literature study and interviews. This literature study method is used to present different ideas related to the problems being researched or faced in order to provide context in discussing research findings, while the interview method is carried out with parties related to the management of the Mount Bromo tourist attraction by discussing the problems currently occurring at the tourist attraction Mount Bromo and the impact of this problem. The research method used by researchers to complete research tasks is by collecting data from books, journals and other data sources from literature (Kurniawan & Yasir, 2022).

Business Understanding

Afterwards, this information is converted into a description of the data mining challenge and an initial project plan aimed at achieving that goal. Determining the context of the discussion focuses on Mount Bromo tourist destination management issues which can be reviewed based on tourists' perceptions through text data in the form of visitor reviews (Steven & Wella, 2020).

Data Understanding

This research utilizes information from TripAdvisor user evaluations of the Mount Bromo tourist destination. TripAdvisor site scraping techniques were used to collect data. User, date, rating, and review are the attributes collected. As many as a thousand reviews of Mount Bromo tourist destinations were collected (Taecharungroj & Mathayomchan, 2019).

Data Preparation

The Data Preparation stage involves processing the text to eliminate unused features. This stage is very important in the CRISP-DM framework because good and relevant data quality determines the success of the analysis process. The Data Preparation stages used consist of case folding, cleansing, tokenizing, stop word filtering, stemming (Wilim & Oetama, 2021).

Modelling

At this stage, the review data that has been previously processed will be used for modeling. The review data will now be divided into two stages namely separate data and cross validation. This will be separated into training and testing data. Supervised learning is the machine learning algorithm used in this research. Naive Bayes is a popular machine learning technique for classification (Ananda & Dirgahayu, 2021).

Evaluation

Confusion matrix is a method for assessing classification methods. After effective data collection and pre-processing, the final dataset is displayed with its attributes organized according to sentiment classes. The visualization results show which terms are frequently used in each type of sentiment (Firdaus et al., 2021).

Deployment

The final stage in the CRISP-DM model is Deployment. Planning for Deployment begins during Business Understanding and must incorporate not only how to generate model values, but also how to convert decision scores, and how to incorporate decisions in operational systems (Assyafah et al., 2021).

RESULTS AND DISCUSSION

Identification Problems

The results of problem identification were carried out through online interviews with the management of Mount Bromo National Park, to find out what problems were occurring. This problem can be used as a topic of discussion for research. From the results of interviews conducted,

problems occurred because the development of tourist attractions had not been managed properly (Heliyanti Susana, 2022).

Business Understanding

Business understanding aims to define data mining challenges and create an initial project plan that will help the project achieve its goals. The background to the discussion is based on examining the difficulties of managing the Mount Bromo tourist destination, which can be studied based on tourists' opinions expressed in textual data in the form of visitor evaluations. To determine the application of sentiment analysis using the Naive Bayes classification method to determine the sentiment of visitors to the Mount Bromo tourist attraction, the analysis carried out in this research was carried out using Anaconda Navigator software with Jupyter Notebook tools and the Python programming language (Meray et al., 2016).

Data Understanding

Using Octoparse for web scraping, the data used in this research is taken from the TripAdvisor comments section, Ratings, dates, and reviews will be collected. A total of one thousand reviews were extracted from the internet using web scraping (Panyahuti & Yadi, 2022).

Data Analyst exploration

Through summary statistics and graphical (visual) representations, an exploratory data analyst examines the important process of initial exploratory testing on data to detect patterns, discover anomalies, test hypotheses, and verify assumptions. What data sets can be further exposed through formal data modeling or hypothesis testing activities can be determined using EDA (Ananda & Dirgahayu, 2021).

```
#Shape of the dataset, And Breakdown of the classes
print(f"Input data has {len(df)} rows and {len(df.columns)} columns")
print(f"rating 1.0 = {len(df[df['Rating']==1.0])} rows")
print(f"rating 2.0 = {len(df[df['Rating']==2.0])} rows")
print(f"rating 3.0 = {len(df[df['Rating']==3.0])} rows")
print(f"rating 4.0 = {len(df[df['Rating']==4.0])} rows")
print(f"rating 5.0 = {len(df[df['Rating']==5.0])} rows")
```

```
Input data has 1001 rows and 9 columns
rating 1.0 = 31 rows
rating 2.0 = 48 rows
rating 3.0 = 75 rows
rating 4.0 = 270 rows
rating 5.0 = 577 rows
```

Figure 2. Visitor Ratings Coding Process

Figure 2 above explains that the code provided displays various statistics about the DataFrame named df. Initially, it prints the total number of rows and columns in the DataFrame. After that, it will sequentially print the number of rows for each specific rating value in the "Rating" column, ranging from 1.0 to 5.0. For example, it counts and displays how many rows have a rating of 1.0, then does the same for ratings of 2.0, 3.0, 4.0, and 5.0.

```
#Missing Values in the dataset
print(f"Number of null in label: { df['Rating'].isnull().sum() }")
print(f"Number of null in text: { df['Review / Ulasan '].isnull().sum()}")
sns.countplot(x='Rating', data=df);
```

```
Number of null in label: 0
Number of null in text: 0
```

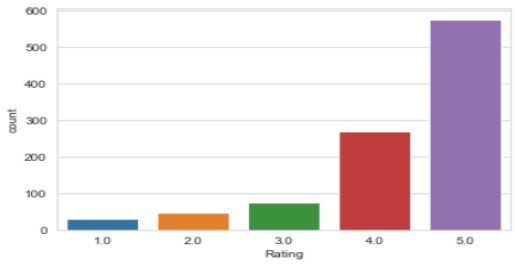

Figure 3. Visitor Ratings

Figure 3 above shows that the visitor review data for the Mount Bromo tourist attraction consists of 1001 rows and 11 columns. From the analysis results, it can be seen that rating 1.0 has 31 lines, rating 2.0 has 48 lines, rating 3.0 has 75 lines, rating 4.0 has 270 lines, and rating 5.0 has the largest number of lines, namely 577 lines. (Prestasi & Lalu, 2020) In general, the highest ratings such as 4.0 and 5.0 have a significant number of rows meaning that the majority of visitors give positive ratings.

Data Labelling

Data labelling is the process of giving labels or tags to data to provide information about certain characteristics or classes. These labels are used to train machine learning models so they can recognize patterns or correlations between data features and the class or target they want to predict. These labels are used to train machine learning models so they can recognize patterns or correlations between data features and the class or target they want to predict (Santoso et al., 2020). The data is then given a label according to the predetermined category or class. For example, in sentiment classification on product reviews, data can be labelled positive, negative.

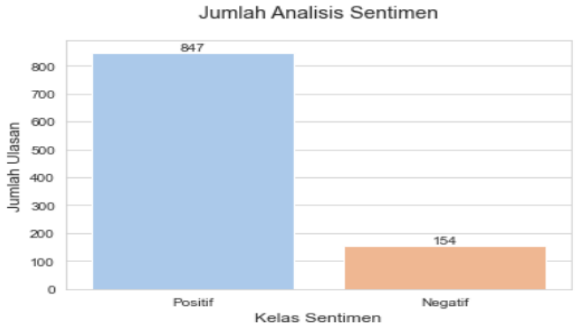

Figure 4. Results of Data Analysis

In figure 4 above, it is explained that 154 texts are classified as having negative sentiment, and 847 texts are classified as having positive sentiment. From these results it can be concluded that the majority of reviews provide positive results. Bar charts are created using sns.barplot() with the x-axis as the sentiment_count index (category) and the y-axis as the sentiment_count value (the number of sentiment analyzes for each category).

Data Preparation

Data pre-processing is a step that eliminates various issues that might hinder data processing. This is due to the fact that many data sets have uncertain formats. Before doing data mining, the first thing to do is pre-process the data. Pre-processing of this data is done to handle it before proceeding to the next stage. Below are several stages of Data Preparation. Data pre-processing involves a number of steps, including CaseFolding, tokenization, filtering, and stemming. Case Folding is a text preparation step called "transforming case" trying to convert all document characters to lowercase (small text).

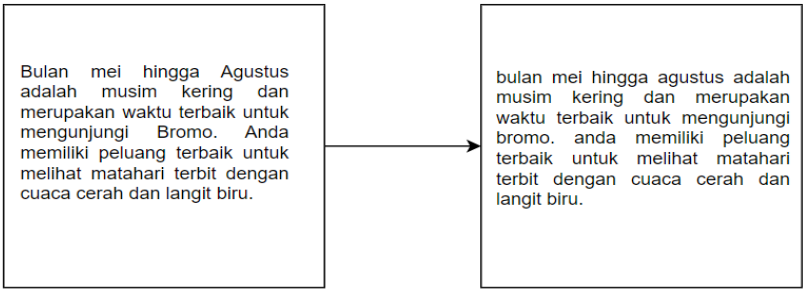

Figure 5. Result of Case Folding

Figure 5 above explains that by converting all the letters to lowercase, letter folding helps in normalizing the text. This ensures that two words that are actually the same but written in different font styles will be treated as the same word. For example, "Data" and "data" will be considered the same after the case folding process. Cleaning is deleting non-capital characters, done before cleaning the text. The goal of the cleanup text preparation step is to remove tabs, newlines, backslashes, mentions, links, hashtags, and URLs from the text. The main purpose of data cleaning is to ensure that the data used in analysis or further processing is accurate, complete, consistent and relevant. This process is the initial stage in text analysis and is often carried out before feature extraction or modeling.

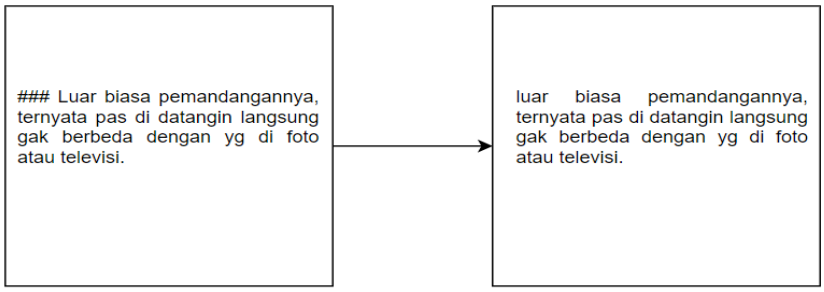

Figure 6. Result of Cleaning

Figure 6 above has the main objective in data cleaning is to ensure that the data used in analysis or further processing is accurate, complete, consistent and relevant. Apart from that, Cleaning the text also provides additional features such as the length of the document being reviewed, as well as the percentage of punctuation marks that appear in the document. Tokenization, in this section the text is tokenized—that is, divided into token segments or attributes. The text is divided into several parts according to the space used for tokenization in this study.

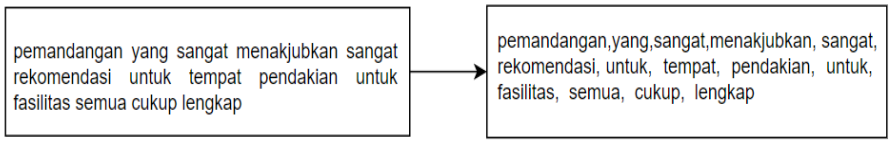

Figure 7. Result of Tokenization

Figure 7 above explains that through this procedure, documents that previously contained sentences now consist of word fragments. We can differentiate between word separators and non-word separators by tokenizing. Tokenization often involves removing unnecessary spaces, symbols, and punctuation. Filtering Stop word means words are changed to their most basic form at this point. This needs to be done because words that have affixes derived from the same root word are treated as separate words.

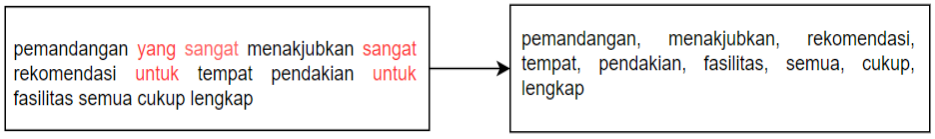

Figure 8. Result of Filtering Stop word

Figure 8 above explains that words that have affixes originating from the same root are treated as separate words. Lemmatization is a procedure that reduces meaningful words to a single word, and Removing Stop words is a method that filters a data frame. The TF-IDF approach creates a term weight for each word in the document by multiplying the TF and IDF values. First,

the data must be word weighted with TF-IDF and then converted into numerical form. TF-IDF assigns weight to words in a document based on their importance in the context of the document. Words that appear frequently in the document will have a higher weight than words that appear rarely.

Table 1. TF-IDF Weighting results

No	Term	Weight
1	Gunung	0.046098
2	Indah	0.042785
3	Pandang	0.033611
4	Gunung Bromo	0.029541
5	Yang	0.029279
6	Sunrise	0.026956
7	Jalan	0.026146
8	Banget	0.025939

Table 1 is the result of weighting the TF-IDF words in the form of a table which displays the words that appear most frequently, the words that appear most frequently are mountain, namely 0.046098. First, the data must be word weighted with TF-IDF and then converted into numerical form. TF-IDF assigns weight to words in a document based on their importance in the context of the document. Words that appear frequently in the document will have a higher weight than words that appear rarely.

Modeling

K-Fold Cross Validation is the initial modeling step that has been completed. The process of separating training data from testing data is called K-Fold Cross Validation. The goal is to identify the optimal data set, which may include the most accurate, precise, error-free data, and so on. The information will be divided into training (to train the model) and test (to test the model) sets. This data will then be iterated continuously until it meets the predetermined K value.

```
from sklearn.model_selection import cross_val_score
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.neighbors import KNeighborsClassifier

models = [
          MultinomialNB(),
         ]
names = ["Naive Bayes"]
for model, name in zip(models, names):
    print(name)
    for score in ["accuracy", "precision", "recall", "f1"]:
        print(f" {score} - {cross_val_score(model, X_train_vect, y_train, scoring=score, cv=10).mean()} ")
    print()
```

```
Naive Bayes
 accuracy - 0.8442857142857143
 precision - 0.8454865424430642
 recall - 0.9983050847457626
 f1 - 0.9155504621347644
```

Figure 9. K-Fold Cross Validation Process

Figure 9 above explains that Accuracy, precision, recall, F1 is a loop that iterates over each performance evaluation metric that you want to calculate. Additionally, cross_val_score to perform cross-validation with 10 folds, calculating the evaluation metric score (accuracy, precision, recall, or F1 value) for each fold.

Evaluation

This procedure looks at the output of the model developed using the Confusion Matrix to provide information on how to compare the model output with the actual classification results. The Confusion Matrix is used in the assessment stage of the model assessment report to measure how well the Naive Bayes classifier performs. calculations to find Positive Precision, Negative Precision, Positive Precision, and Accuracy. Confusion Matrix provides a detailed picture of how well the model performs classification on the test data set. Confusion matrices are usually used to evaluate models on data that has been grouped (labeled), where the actual values of the classes are.

	precision	recall	f1-score	support
0	1.00	0.02	0.04	46
1	0.85	1.00	0.92	255
accuracy			0.85	301
macro avg	0.93	0.51	0.48	301
weighted avg	0.87	0.85	0.78	301

Figure 10. Confusion Matrix Results

Figure 10 above explains that the positive and negative sentiment class information that was predicted correctly was 255 positive data, 0 positive data, 1 negative data and 45 negative data. There were also 255 positive data which were correctly predicted to fall into the positive sentiment class and 0 positive data which were predicted to fall into the negative sentiment class. These results are displayed in plot form in figure 11. below

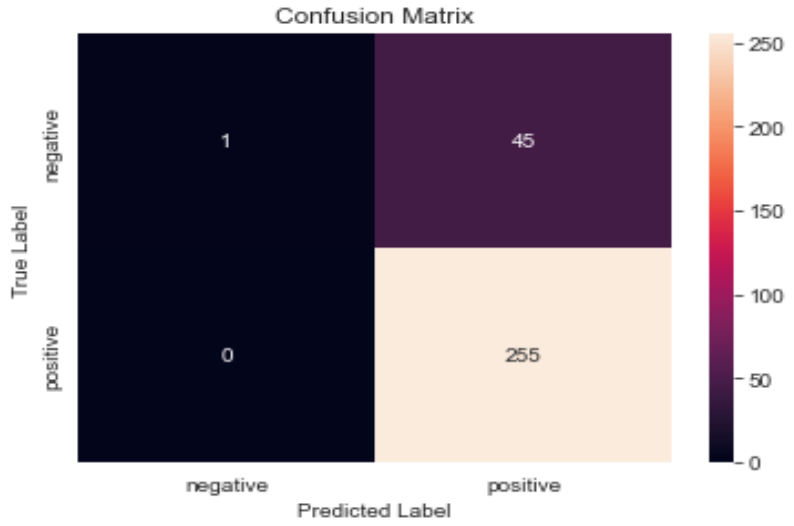

Figure 11. Confusion Matrix Plot

Figure 11 above explains that the confusion matrix provides information about how well the model is able to predict the actual classes. Confusion_matrix(y_test, naive_bayes_pred) calculates the confusion matrix using the actual class labels (y_test) and the predicted results from the Naive Bayes model (naive_bayes_pred). X_train_vect is a feature matrix of training data that has been converted into a vector representation (for example, using TF-IDF or CountVectorizer). By using a Naive Bayes Multinomial classification model, we can train it with training data and then use it to make predictions on test data. These predictions can be used to evaluate model performance. The confusion matrix plot is a visualization of the confusion matrix which makes it easier to understand how the classification model works.

Deployment

Deployment process that involves the final review of the application or system to be deployed. At this stage, we have to prepare the source code that will be deployed and pay attention to libraries, configuration files, and other resources needed by the application so that it can work properly.

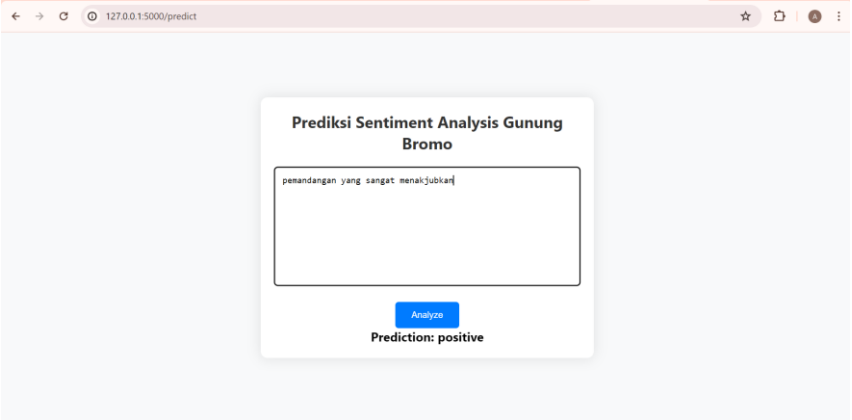

Figure 12. Website Deployment Positive

Figure 12 above explains that on this website, the user enters text into the input box, and the system will predict the sentiment of the text. In the example shown, the text entered is "a truly amazing view," and the prediction result is "positive." The benefit of this website is that companies or individuals can use sentiment analysis to understand public feelings or opinions towards certain products, services or phenomena, which can help in decision making. One of the main challenges is ensuring that the sentiment analysis model has a high level of accuracy and this can be achieved through training the model using large and diverse datasets and using sophisticated machine learning techniques.

CONCLUSION

Information technology plays a crucial role in the development of the tourism industry. Easily accessible information and services through online platforms like TripAdvisor allow tourists to plan their trips better, receive accommodation recommendations, and read reviews from other travelers. In the context of sentiment analysis, these platforms can provide deep insights into the perceptions and experiences of visitors to tourist destinations like Mount Bromo. Data from these online reviews, when analyzed using machine learning methods such as Naive Bayes, can help destination managers understand and improve the quality of services and tourist experiences.

The CRISP-DM methodology and data mining techniques were used to evaluate the sentiment of tourist reviews of Mount Bromo. The analysis results showed that the majority of reviews tend to be positive. However, there are also a number of negative reviews indicating issues in the management and development of the tourist destination. By using the Naive Bayes model for sentiment classification, this research successfully identified areas needing improvement. These findings can be used by the Malang City Culture and Tourism Office and Mount Bromo destination managers to develop better strategies for enhancing services and tourist attractions in the future.

ACKNOWLEDGEMENTS

This research was successfully completed with the assistance and support from Multimedia Nusantara University. We extend our gratitude for the guidance provided during the preparation of this article.

REFERENCES

Adilarif, M. B. (2020). Pengaruh Daya Tarik Wisata, Lokasi, dan Harga Terhadap Keputusan Berkunjung Wisatawan Pada Taman Nasional Bromo Tengger Semeru. *Jurnal Ilmiah Mahasiswa FEB*, *8*(2), 1–17.

Ananda, I., & Dirgahayu, T. (2021). Pemanfaatan Teknologi Informasi Pada Desa Wisata di Indonesia: A Systematic Literature Revies. *Jurnal Teknik Informatika Dan Sistem Informasi*, *8*(4), 2291–2302.

Ashoer, M., Revida, E., Dewi, K. I., Simarmata, M. M., Nasrullah, Mistriani, N., Samosir, R. S., Purba, S., Islahuddin, Meganingratna, Permadi, L. A., Purba, B., Murdana, I. M., & Simarmata, H. M. P. (2021). Ekonomi Pariwisata. In *Suparyanto dan Rosad (2015* (Vol. 5, Issue 3).

Assyafah, H. B., Trivena, D., & Kom, Y. S. (2021). *Analisis Dataset menggunakan Sentiment Analysis (Studi Kasus Pada Tripadvisor)*. *3*(November), 320–331.

Astawa, I. P. P., & Sudibia, I. K. (2021). Sikap dan Kepedulian Masyarakat Terhadap Objek Wisata dan Pembangunan Berkelanjutan di Bali. *Widya Manajemen*, *3*(1), 15–26. https://doi.org/10.32795/widyamanajemen.v3i1.1139

Benoît, G. (2023). Data mining. *Annual Review of Information Science and Technology*, *36*, 265–310. https://doi.org/10.1002/aris.1440360107

Damayanti, S. E., Informatika, P. T., Tinggi, S., Bandung, T., & Swasta, P. T. (2020). *Analisis Dan Implementasi Framework CRISP-DM (Cross Industry Standard Process For Data Mining)*.

Firdaus, A., Firdaus, W. I., Studi, P., Informatika, T., Digital, M., & Sriwijaya, P. N. (2021). *Text Mining*. *13*(1), 66–78.

Heliyanti Susana. (2022). Penerapan Model Klasifikasi Metode Naive Bayes Terhadap Penggunaan Akses Internet. *Jurnal Riset Sistem Informasi Dan Teknologi Informasi (JURSISTEKNI)*, *4*(1), 1–8. https://doi.org/10.52005/jursistekni.v4i1.96

Hidayat, T. F. T., Garno, G., & Ridha, A. A. (2021). Analisis Sentimen Opini Pemindahan Ibu Kota Pada Twitter Dengan Metode Support Vector Machine. *Jurnal Ilmu Komputer*, *14*(1), 49. https://doi.org/10.24843/jik.2021.v14.i01.p06

Hilyah, A., Fajar, M. H. M., Ikmaluhakim, D. R., Hawan, S. I., Purwanto, M. S., & Bahri, A. S. (2021). Studi Geologi dan Geofisika Batuan Gunung Bromo dan Sekitarnya. *Sewagati*, *5*(2), 156. https://doi.org/10.12962/j26139960.v5i2.8248

Imron, A. (2019). Analisis Sentimen Terhadap Tempat Wisata di Kabupaten Rembang Menggunakan Metode Naive Bayes Classifier. *Teknik Informatika*, 10–13. https://dspace.uii.ac.id/handle/123456789/14268

Izza, A. N., Ratnawati, D. E., & Putra, W. H. N. (2022). Analisis Sentimen Objek Wisata Di Provinsi Sulawesi Sentiment Analysis of Tourist Attractions in the Province of South Sulawesi Based on Visitor Reviews. *Jurnal Sistem Informasi, Teknologi Informasi, Dan Edukasi Sistem Informasi (JUST-SI)*, *3*(2), 97–105.

Kuncoro, C. P. (2022). Analysis Of UMN Student Graduation Timeliness Using Supervised Learning Method. *IJNMT (International Journal of New Media Technology)*, *8*(2), 89–95. https://doi.org/10.31937/ijnmt.v8i2.2366

Kurniawan, D., & Yasir, M. (2022). Optimization Sentimen Analysis using CRISP-DM and Naive Bayes Methods Implemented on Social Media. *Cyberspace: Jurnal Pendidikan Teknologi Informasi*, *6*(2), 74. https://doi.org/10.22373/cj.v6i2.12793

Meray, J. G., Tilaar, S., & Takumansang, E. D. (2016). Partisipasi Masyarakat Terhadap Pengembangan Pariwisata Pantai Mahembang Kecamatan Kakas. *Open Journal Systems*, *3*(3), 47–55. https://ejournal.unsrat.ac.id/v3/index.php/spasial/article/view/13429

Oktaviani, Y., Putu, I. G., Wedashwara, W., & Zubaidi, A. (2022). *Klasifikasi Teks Ulasan Pada Web TripAdvisor Tentang Wisata Alam Pulau Lombok Menggunakan Metode Naive Bayes CLASSIFIER Tripadvisor Web Review Text Classification Of Lombok Island Nature Tourism Using Naive Bayes Classifier Method*. *4*(2), 253–262.

Panyahuti, P., & Yadi, Y. (2022). Pengembangan Aplikasi E-Assessment Skill Programming berbasis Web. *Edumatic: Jurnal Pendidikan Informatika*, *6*(1), 78–87.

https://doi.org/10.29408/edumatic.v6i1.5393

Prestasi, D. A. N., & Lalu, M. (2020). *Data Mining Untuk Memprediksi Prestasi Siswa Data Mining To Predict Students Achievement Based On Socio-Economic , Motivation , Discipline And*. *4*, 222–231.

Rachmawati, E., Faizah, N. K., & Muntasib, E. H. (2020). Potensi bahaya di kawasan wisata Gunung Bromo, Resort Tengger Laut Pasir, Taman Nasional Bromo Tengger Semeru, Jawa Timur. *Media Konservasi*, *12*(3), 140–145.

Rahayu, S. D. (2019). *Potensi Wisata Danau Kualomudo di Kelurahan Balai Makam Kecamatan Mandau Kabupaten Bengkalis*. https://repository.uir.ac.id/5727/

Santoso, J. T., Kom, S., & Kom, M. (2020). *ANALISIS Big Data P Y Yayasan Prima Agus Teknik Yayasan Prima Agus Teknik*.

Singgalen, Y. A. (2022). Analisis Sentimen Wisatawan Melalui Data Ulasan Candi Borobudur di Tripadvisor Menggunakan Algoritma Naïve Bayes Classifier. *Building of Informatics, Technology and Science (BITS)*, *4*(3). https://doi.org/10.47065/bits.v4i3.2486

Steven, C., & Wella, W. (2020). The Right Sentiment Analysis Method of Indonesian Tourism in Social Media Twitter. *IJNMT (International Journal of New Media Technology)*, *7*(2), 102–110. https://doi.org/10.31937/ijnmt.v7i2.1732

Taecharungroj, V., & Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. *Tourism Management*, *75*(April), 550–568. https://doi.org/10.1016/j.tourman.2019.06.020

Wilim, N. N., & Oetama, R. S. (2021). Sentiment Analysis About Indonesian Lawyers Club Television Program Using K-Nearest Neighbor, Naïve Bayes Classifier, And Decision Tree. *IJNMT (International Journal of New Media Technology)*, *8*(1), 50–56. https://doi.org/10.31937/ijnmt.v8i1.1965