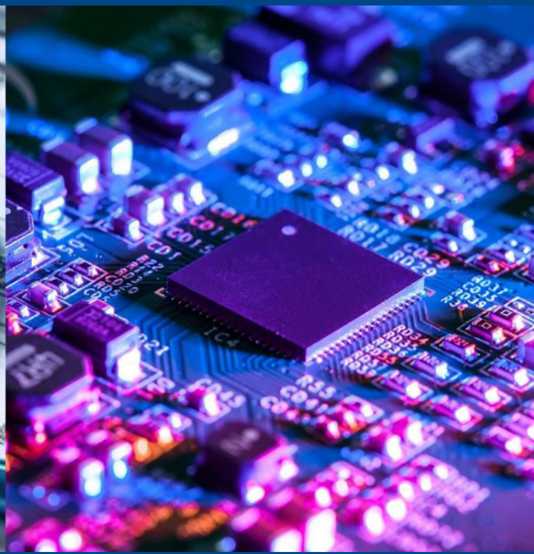
IJMRSET

ISSN: 2582-7219

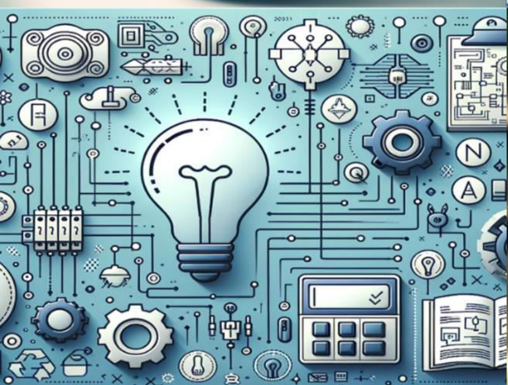




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 4, April 2025

www.ijmrset.com ijmrset@gmail.com +91 6381 907 438 +91 63819 07438

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Sentiment Analysis Based on Travelers' Reviews Using the SVM Model with Enhanced Conjunction Rule-Based Approach

Siva Prasad, Gaurav Tripathi, N. Sohan Mani, L.Sharath Chandra, B Sowmya Sree

Department of AI & ML, Malla Reddy University, Hyderabad, India

Dr. Sujit Das

Assistant Professor, Department of AI&ML, Malla Reddy University, Hyderabad, India

ABSTRACT: It is a tough task for tourism management to identify their user reviews and come up with solutions to the advancement of their tourism organizations. There are many social media reviews, and Tourism organizations face the challenge of evaluating numerous social media reviews to find solutions for advancing their organizations. It can be tough to physically assess all those reviews, social media has become a huge trend these days. People are constantly sharing their experiences and opinions on tourist places. By analyzing the sentiment of reviews, it may obtain useful information about the popularity of different tourist destinations. It's a great way to understand what people love about certain places. Sentiment classification is indeed a valuable tool in classifying reviews into different categories, aiding in decision-making. However, it's important to note that reviews often contain noisy content like typos and emoticons, which can affect the accuracy of the algorithms. Taking these aspects into consideration is crucial for achieving more accurate results.

KEYWORDS: Sentiment Analysis, Tourism, Support Vector Machine (SVM), Machine Learning, Conjunction Rule Base Approach

I. INTRODUCTION

Nowadays, social media is experiencing rapid growth, with millions of users posting reviews and rating visiting places available on different tourism websites. With the help of Sentiment Analysis, these reviews can be immensely helpful. By properly analyzing the reviews, we can identify trends in the popularity of tourist places, as mentioned in [1,2,3]. The summarized results from sentiment analysis can assist tourists in making decisions about their tour destination and planning. This project's research goal is to create a sentiment analysis system specifically designed for travelers' reviews, using a combination of Support Vector Machine (SVM) modeling and an enhanced conjunction rule-based approach.

This tailored system aims to provide valuable insights for travelers based on their reviews. In this study, a feature extraction algorithm was used i.e.., TFIDF Vectorization Algorithm. Three classifiers Support Vector Machine (SVM), SVM with N-grams, and SVM with Enhanced Conjunction Rule-Based classification were also utilized to classify sentiment. Based on variables like precision, F1-score, recall, execution time, and accuracy comparisons of feature extraction and classification algorithm combinations were carried out. In this paper Section II of the paper discusses the sentiment analysis literature surveys that the authors have undertaken.

II. LITERATURE REVIEW

Sentiment analysis of reviews of Tourist Place reviews has been studied and obtained results by many authors. Each author used a different approach when analyzing the reviews. In [1] the author Nur Aliah Khairina Mohd Haris, et.al., have evaluated the sentiment polarity of tourist destination reviews focusing on Taman Negara using an SVM and RF classifier. Contrasting these classifiers' performances shows that SVM is the best option for analyzing sentiment in tourism reviews on social media platforms. The results achieve an impressive accuracy of 67.97% for the 5-fold cross-validation model, which outperforms RF's accuracy of 63.55%. The project also provides a sentiment analysis

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804309

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

dashboard that visually displays visitor reviews of Taman Negara. In [2] the authors Aranyak Maity, et.al., Their study compares N-grams and Conjunction Rule-based Approaches, concluding that the conjunction rule-based approach yields superior results. By enhancing feature-specific sentiment analysis of hotel reviews, this study addresses the challenge of information overload in online feedback, providing a detailed assessment of aspects like food, service, and location. In another study [3], the authors provide an original approach that applies a dual prediction technique. In this method, the unigram and inverse unigram models calculate both positive and negative polarity weights to calculate average spending sentiment. An experimental study [4] shows that in automatic aspect-based sentiment analysis, many important aspects are missed and instead, various irrelevant aspects are extracted and taken into consideration, thereby deteriorating the performance, which can be addressed by providing a pre-built feature list and dictionaries based on those features.

III. PROBLEMSTATEMENT

With the growing reliance on online reviews and user-generated content, tourists and businesses alike struggle to extract meaningful insights from vast amounts of textual data. Traditional sentiment analysis methods face challenges in handling ambiguous trends, data integrity issues, and feature- specific evaluations, limiting their effectiveness in decision- making. While various techniques like aspect-based opinion mining, sentence-based sentiment analysis, and rule-based classification have been applied, their accuracy and interpretability remain key concerns.

IV. METHODOLOGY

Sentiment analysis involves analyzing user opinions from text data, referred to as opinion mining. There are two primary approaches utilized for sentiment analysis i.e., Supervised methodology for machine learning and unsupervised lexicon- based methodology.

A. Real Data Collection

In this study, the researchers collected review data from different tourism websites. The data was stored in a CSV format and included review text along with corresponding ratings. They were able to ascertain if the reviewers' sentiments were favorable or negative by examining the text. A rating score greater than 3 is considered a positive review, and a score less than or equal to 3 is considered a negative review.

B. Data Processing

Data preprocessing is a crucial step when dealing with social media data as it cleans and refines raw data, making it suitable for sentiment analysis. It involves steps like tokenization, which breaks down reviews into smaller parts called tokens, and the removal of stop words—common words like 'the,' 'a,' and 'and' that carry little meaning—using a custom stop-word list to reduce the feature vector size and enhance system performance. Lemmatization and stemming are employed to identify root forms of words; lemmatization focuses on reducing words to their base form (e.g., "a pile of tokens" becomes "a pile"), while stemming simplifies tokens to their roots (e.g., "treks" becomes "trek").
Additional preprocessing includes removing short words, punctuation, numbers, and special characters, and converting text to lowercase—all aimed at improving the performance of machine learning algorithms.

C. Feature Extraction

For feature extraction from evaluation data, count vectorization and TFIDF vectorization algorithms were used. Vectorization of counts is the same as the Bag of Word (BoW) approach. This shows how often the text appears in a particular document and how often it occurs. TFIDF vectorization is an extension of count vectorization, which also considers inverse document frequencies in parallel to term frequencies.

D. Training Model

The dataset was used for the training model. The study uses support vector machines, SVM with N-grams, and SVM with a joint rule-based classification algorithm on the training validation dataset.

E. Test Model

From the entire evaluation data set, data was used for testing. To predict sentiment polarity, a test was performed on new, unseen reviews. The trained model classifies the sentiment of reviews into two classes: positive and negative.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

F. Performance Evaluation

Performance evaluation is a crucial step. In this study, the researchers evaluated the performance utilizing metrics such as recall, execution speed, accuracy, and precision. It's important to assess how well the model performs and how long it takes to execute. These evaluations help us understand the effectiveness and efficiency of the machine-learning approach.

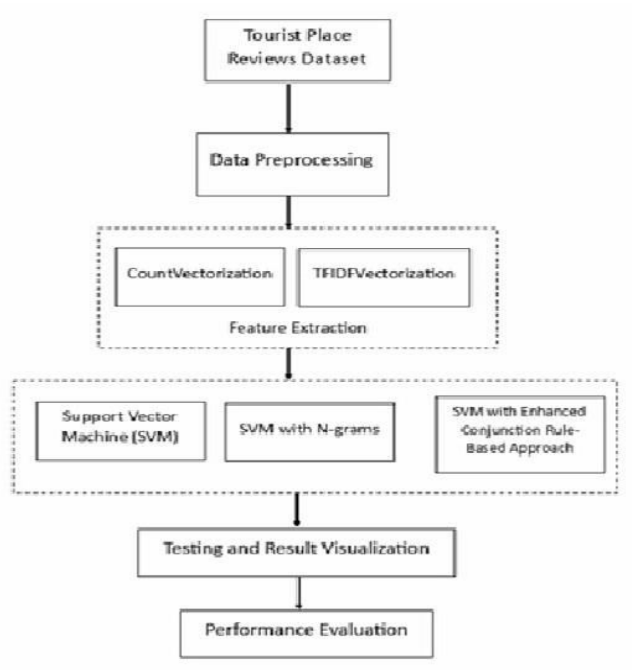


Fig 3.1 Data Flow Diagram

V. EXPERIMENTAL RESULTS

Pre-processing the data was necessary to eliminate superfluous and extraneous terms from the reviews because the information gathered from different travel websites was unprocessed. A data preprocessing stop removed words, punctuation marks, and short words. Tokenization, lemmatization, and stemming were also performed. The significance of data pretreatment for feature reduction and enhanced machine learning algorithm performance was acknowledged in this work. They started with data purification and then implemented a feature extraction algorithm from scratch. Count Vectorization and TFIDF Vectorization were used to extract features. The classification algorithm was implemented using the Python sklearn package, which offers various routines for different classifiers. They trained the classification algorithm on the labeled training dataset and evaluated its performance utilizing metrics such as F1 score, recall, and precision.

Review	Class label	Sentiment
Best time is go just after rainy season to enjoy fog & green carpet over the mountain. Good trekking area.	1	Positive
Security at this place is not good at all	0	Negative

Fig 5.1 Tourist Place Review Sentiment Analysis Example

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

This section will delve into the results obtained from analyzing positive and negative reviews using an SVM, SVM with n-gram, and SVM with a conjunction rule-based approach for sentiment classification. The dataset from which obtained results has inconsistency. Therefore, researchers balanced the dataset using the SMOTE approach. Following the dataset's use of the Smote.

This is the graph that balances out next. Class distribution before SMOTE: Counter ({1: 760000, 0: 178645})

Class distribution after SMOTE: Counter ({1: 760000, 0: 760000})

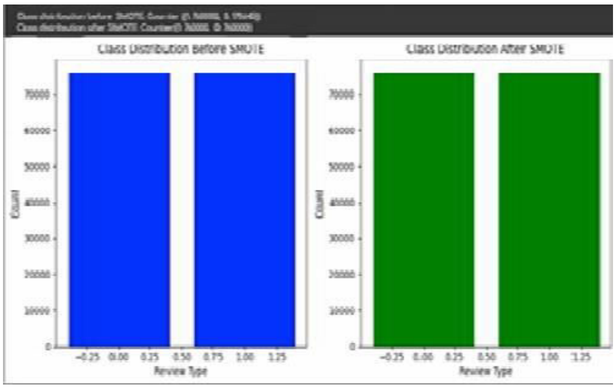


Fig 5.2 Data Imbalance Graph

A. Accuracy Score

To be sure, accuracy is a simple, common-sense way to assess performance. It shows the proportion of accurately anticipated observations to all observations. The graphs that compare accuracy are shown in Figures 5 and 6. Visualizing the accuracy can help us understand the performance of the model more easily. It is always helpful to have visual representations to make data analysis more accessible. The formula for accuracy is given below.

$$accuracy = \frac{true\ positives + true\ negatives}{true\ positives + true\ negatives + false\ negatives + false\ positives}$$

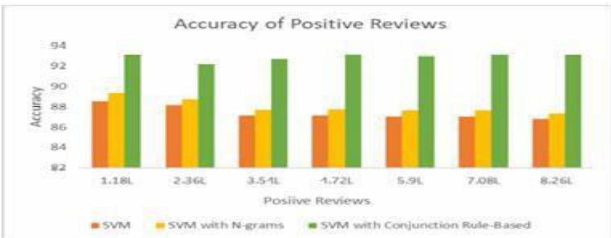


Fig 5.3 Accuracy of Positive Reviews

B. Precision Recall

The precision comparison graph, which is shown in Fig. 7, is a valuable visual representation. Predictive value, another name for precision, is an important performance measure. It focuses on the proportion of accurately forecasted positive observations to all scheduled positive observations. The accuracy performance of the model can be understood by looking at a precision comparison graph.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

$$precision = \frac{true\ positives}{true\ positives + false\ positives}$$

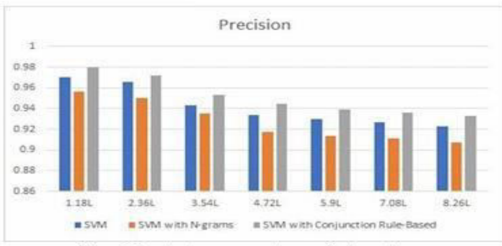


Fig. 6 Precision comparison of algorithms

Fig 5.4 Precsion comparaison of algorithms

C. Recall

Recall, sometimes called sensitivity, is an important performance measure. Its main focus is on the proportion of all actual positive observations to all correctly predicted positive observations. Fig. 8 is a recall comparison graph that ought to offer insightful information about the model's memory capabilities.

VI. CONCLUSION

In this paper, user reviews are obtained from popular websites for different tourism destinations and the sentiments of the reviews are analyzed. In our proposed model we use a machine learning algorithm named Support Vector Machine (SVM) model to predict the sentiments from tourist place reviews and evaluate the performance using an Enhanced Conjunction rule- based approachThe aggregated outcome will help the users to know which place is suitable for traveling and it helps to identify the users more easily. In future work, we shall try to improve our experimental results accuracy by using a larger data set and gathering data from more websites.

VII. FUTURE ENHANCEMENT

Integration of Additional Features: Incorporate advanced linguistic features like sarcasm detection and context-aware analysis to improve sentiment accuracy. This can help address ambiguous sentiments commonly found in traveler reviews.
Multilingual Support: Enhance the model to process and analyze traveler reviews in multiple languages, enabling a broader range of insights from global datasets.
Hybrid Models: Combine SVM with deep learning techniques such as LSTM or Transformers for better handling of complex data structures and deeper sentiment predictions.
Real-Time Sentiment Analysis: Develop systems for real-time sentiment analysis, allowing users to instantly analyze newly added reviews and make quicker decisions.
Visualization Tools: Introduce sophisticated data visualization methods like interactive sentiment graphs and heatmaps to help users easily understand trends in traveler sentiments.
Personalized Analysis: Tailor sentiment analysis outputs based on user preferences (e.g., focusing on specific tourism attributes such as accommodation, dining, or activities).

REFERENCES

1. Khairina Mohd Haris, N. A., Mutalib, S., Ab Malik, A. M., Abdul-Rahman, S., & Kamaliah Kamarudin, S. N. (2023). Sentiment classification from reviews for tourism analytics. International Journal of Advances in Intelligent Informatics, 9(1).
2. Maity, A., Ghosh, S., Karfa, S., Mukhopadhyay, M., Pal, S., & Pramanik, P. K. D. (2020). Sentiment analysis from travellers' reviews using enhanced conjunction rule based approach for feature-specific evaluation of hotels. Journal of Statistics and Management Systems, 23(6), 983-997.
3. Anto, M. P., Antony, M., Muhsina, K. M., Johny, N., James, V., & Wilson, A. (2016, March). Product rating using sentiment analysis. In 2016 International Conference on Electrical, Electronics, and Optimization Techniques

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

(ICEEOT) (pp. 3458-3462). IEEE
4. Afzaal, M., & Usman, M. (2015, October). A novel framework for aspect-based opinion classification for tourist places. In 2015 Tenth International Conference on Digital Information Management (ICDIM) (pp. 1-9). IEEE.
5. Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. Journal of Big Data, 2(1), 1- 14.
6. Shaik, K. B., Ganesan, P., Kalist, V., Sathish, B. S., & Jenitha, J. M. M. (2015). Comparative study of skin color detection and segmentation in HSV and YCbCr color space. Procedia Computer Science, 57, 41-48.
7. Rasool, A., Tao, R., Marjan, K., & Naveed, T. (2019, March). Twitter sentiment analysis: a case study for apparel brands. In Journal of Physics: Conference Series (Vol. 1176, p. 022015). IOP Publishing.
8. Satria, A. T., & Nugraheni, D. M. K. (2020). Implementation of Integrated Bayes Formula and Support Vector Machine for Analysing Airline's Passengers Review. In E3S Web of Conferences (Vol. 202, p. 15004). EDP Sciences
9. Afzaal, M., & Usman, M. (2015, October). A novel framework for aspect-based opinion classification for tourist place

INTERNATIONAL JOURNAL OF

MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com