Tourist Place Review System Using Machine Learning Techniques

[1]Shruti Rajendra Joshi[2]Swati Patil
[1]TE CSE, G.H.R.I.E.M., Jalgaon
2Assistant Professor,CSE Department, G.H.R.I.E.M., Jalgaon

Abstract: *There is an increasing preference for social media. Every day, millions of people evaluate and rate places on travel websites. You may do a sentiment analysis on these evaluations to better understand the popularity of your place. Tourists may readily determine where to go based on the analytical results. This study uses a machine learning technique to implement sentiment analysis. Data was acquired from a variety of travel review websites. Here, we compare two extraction algorithms: count vectorization and TFIDF vectorization. The algorithms include Naive Bayes (NB), Support Vector Machine (SVM), and Random Forest (RF). The efficacy of techniques has been evaluated based on a variety of criteria, including accuracy, recall, precision, and f1-score. We discovered that the TFIDF Vectorization extraction of features approach enhanced the classification algorithm's accuracy when compared to Count Vectorization for the supplied review dataset. TFIDF Vectorization +RF achieved the best accuracy of 86% in sentiment categorization of tourist site reviews using a study dataset.*

Keywords: *Naive Bayes (NB), Support Vector Machine (SVM), Random Forest (RF).*

I. INTRODUCTION

Nowadays, social media is rapidly expanding. Every day, millions of customers provide evaluations and ratings for tourist attractions on travel websites. This might be done to examine emotional evaluations. If you conduct appropriate research, you will notice patterns in the popularity of tourism locations. Sentiment analysis data might assist travelers choose places and arrange their travels. This study employed two extraction algorithms: calculus vectorization and TFIDF vectorization. In addition, three different methods were applied for the suggested classification: Naive Bayes (NB), Support Vector Machine (SVM), and Random Forest (RF). A performance comparison was done between picture extraction and algorithms are graded based on factors including execution speed, accuracy, recovery, and accuracy. In addition, the f1 score is evaluated. The paper's content is structured as follows. Part II covers the literature review on sentiment analysis. Part III discusses the fundamentals of machine learning. Chapter IV describes sentiment analysis, visualization, and performance evaluation for travel evaluation categorization. Section V describes an experimental implementation of machine learning techniques for calculating popularity distributions in tourist locations. Section VI gives the results of the implemented tests. Section VII provides a comparative examination of sentiment analysis utilizing machine learning methods from the research project. This article finishes with Section VIII. Section IX outlines the scope of future research efforts.

II.LITERATURE STUDY

1.Cognitive Research: A Comparison of Different Methods Author: M.D. Devika. Sunitha Amal Ganesh Sentiment analysis (SA) is a cognitive process that collects emotions and feelings from

people. Native language processing (NLP) is one of the areas of focus. The growth of Internet applications has resulted in numerous self-reviews of Internet-related content. These evaluations take numerous forms, including social media, blogs, wikis, and forum websites. Fans and consumers alike will find the information in these evaluations useful for understanding and planning the procedure. The rise of search engines like Yahoo and Google exposed people to a slew of critiques about certain areas that are beyond human comprehension. Sentiment analysis is an effective method for extracting and synthesizing information from reviews. In recent years, numerous strategies for doing this task have been identified. In this essay, we will evaluate and analyze several approaches utilized in random analysis collective sentiments of the reviews. Several methods have come to the limelight in recent years for accomplishing this task. In this paper we compare the various techniques used for Sentiment Analysis by analyzing various methodologies.

2. Comparative study of data from Twitter with classifiers under supervision Authors: Rohit Joshi and Rajkumar Tekchandani. Online microblogging on social networks has been used to express thoughts on certain entities in extremely brief messages. There are several prominent microblogs such as Twitter and Facebook, with Twitter receiving the most attention in terms of product research, movie reviews, stock exchange, and so on. We gathered data from Twitter, namely movie reviews, for sentiment prediction using machine-learning techniques. We used supervised machine-learning methods such as support vector machines (SVM), maximum entropy, and Naive Bayes to classify data with unigram, bigram, and hybrid (unigram + bigram) characteristics. The results demonstrate that SVM outperformed other classifiers with an impressive accuracy of 84% for movie reviews.

3. Research on issues Opinion Author: Doaa Mohey El-Din Mohamed Hussein. Websites, social networks, blogs, online platforms, reviews, opinions, suggestions, ratings, and comments on the Internet are the result of the author's surfing. This author did not develop the content for books, individuals, hotels, goods, research, or initiatives. This approach is extremely advantageous to businesses, governments, and individuals. While this information is valuable, most of what this author produces necessitates the use of text mining and sentiment analysis tools. However, assessing and interpreting emotions presents several obstacles. These difficulties impede assessing the underlying meaning of feelings and identifying the proper emotional state. Sentiment analysis is the process of identifying and extracting theme information from text by using text analysis and natural language processing tools. This research investigates methodological and technique-related problems in cardiac analysis.

III.EXISTING SYSTEM

By leaving reviews of different goods and services that might benefit other prospective clients, the consumer turns into an active user. Nonetheless, a client finds it extremely time-consuming and difficult to read every single review that is available online due to the hundreds, thousands, or even more that are available for products and services. For the purpose of giving the user relevant information, gaps are required for suitable technologies that automatically categorize these evaluations as positive or unfavorable.

Disadvantages of Existing System

The following are some drawbacks of the existing system:

• Trust your instincts, but there hasn't been much research done on reviews from tourists.

• Xing Fang and Justin Zhan suggested a novel feature vector generation technique for sentiment pole categorization of product reviews, however the findings are not very accurate.

IV.PROPOSED SYSTEM

The suggested approach evaluated and examined a number of sentiment analysis methods. The evolved document level, sentence level, and form level are the various emotional levels. This article employs lexical, rule-based, and machine learning techniques for sentiment analysis. SVM (Support Vector Machine), NB (Naive Bayes), and feature-based sentiment analysis are only a few of the methods covered in detail under machine learning methodologies. The benefits and drawbacks of each sentiment analysis technique were thoroughly discussed and contrasted. The machine learning is determined to perform best based on a variety of standards, including efficiency, effectiveness, and accuracy.

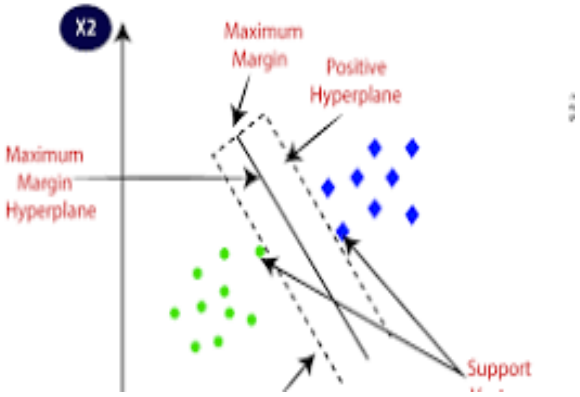
Advantages of Proposed System:

• A great deal of effort has been shown in reviews of various products, movies, restaurants, blogs, etc.

• In the travel sector, sentiment analysis Numerous sentiment analysis methods, including Naive Bayes and Support Vector machines, have been examined by researchers.

1.Support Vector Machine (SVM):

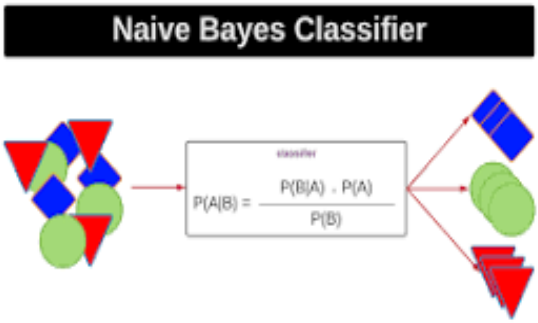
SVM is a powerful supervised learning approach that is used in classification and regression applications. In sentiment analysis, SVM determines the hyperplane that best differentiates between positive and negative sentiment classes. SVM aims to improve the margin between classes while minimizing classification errors. It is effective in high-dimensional domains and works well with sparse data, making it suitable for text classification applications like sentiment analysis. To increase separation, SVM may transform data into higher-dimensional spaces using a variety of kernel functions. In the context of the project, SVM may be trained on labeled tourist location reviews to determine if sentiments are positive, negative, or neutral.



2. NAIVE BAYES:

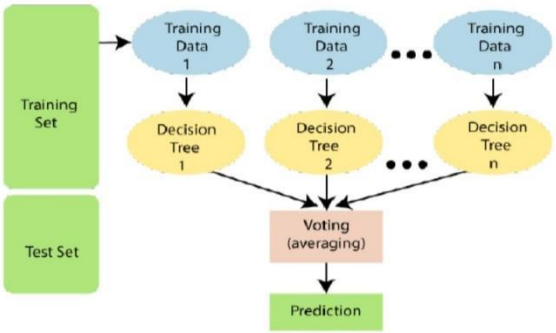
Naive Bayes is a probabilistic classifier that applies Bayes' theorem with the "naive" assumption that features are independent. Despite its simplicity, Naive Bayes consistently beats other text categorization algorithms like sentiment analysis. It assesses the chance that a certain review belongs to each sentiment class based on the existence of words or characteristics in the review. Naive Bayes classifiers are efficient

and scalable, making them suitable for large datasets. In sentiment analysis, Naive Bayes may be trained on labeled reviews to predict the likelihood of each sentiment class based on the observed words and attributes.



3. RANDOM FOREST:

Random Forest is an ensemble learning strategy that generates many decision trees during training and then combines their predictions to get the final classification. Each decision tree in the forest is trained on a random subset of the data and features, which reduces overfitting while enhancing generalization. Random Forest is noisy and outlier robust, and it can effectively handle high-dimensional data. It can detect complex relationships between features and attitudes, making it excellent for sentiment analysis applications. Random Forest may be trained using labeled tourist destination reviews to predict the sentiment (positive, negative, or neutral) of new reviews based on their qualities.



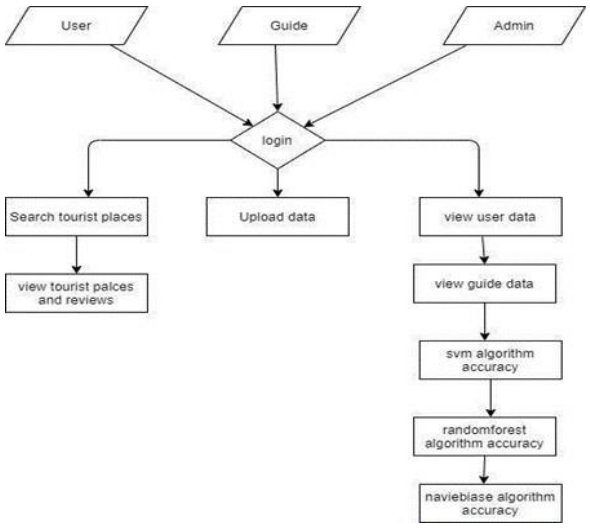
V.DATA FLOW DIAGRAM

- The DFD is referred to as a bubble graph. It is a basic graphical formalism that may be used to depict a system in terms of input data, various processing of this data, and output data created by thesystem.

- The data flow diagram (DFD) is an essential modeling tool. It is used to represent the system components. These components include the system process, the data utilized by the process, an external entity that interacts with the system, and the flow of information inside the system.

- DFD demonstrates how information flows across the system and undergoes modifications. It is a graphical tool for representing information flow and the changes that occur when data goes from input to output.

DFD is often referred to as bubble chart. A DFD may be used to depict a system at any degree of abstraction. DFD may be divided into levels that indicate increased information flow and functional.



VI. MODULES

- User
- Tourist guide
- Admin
- Machine learning

MODULES DESCRIPTION:

User:

The user might be the first to register. When registration, he needed a valid User email and mobile number for future conversations. Once the User has registered, the administrator can activate the User. Once the admin has activated the User, the User will be able to log into our system. After logging in, he may search for a certain tourist attraction and read opinions about it. The user may also search for information on tourist destinations and packages, among other things.

Tourist guide:

The guide can be the first to register. When registration, he needed a valid User email and mobile number for future conversations. Once the guide has registered, the administrator can activate it.Once the admin has activated the guide, the guide will be able to log into our system. Here, the guide can post tourism destinations.

Admin:

Admin may login using his credentials. Once logged in, he may activate the users and the tourist guide. Only activated users and guides can login to our applications. We can use naïve bayes, svm, and random forest to forecast emotional analysis.

Machine learning:

Machine learning is the process by which a computer gains the capacity to make predicted judgments and make the best decisions by evaluating and learning from a significant amount of existing data. Deep learning, artificial neural networks, decision trees, and improvement algorithms are some examples of representation algorithms. Machine learning is the primary mechanism by which computers develop

artificial intelligence. Machine learning is being used extensively in a variety of artificial intelligence applications. Machine learning may be utilized in a variety of applications, including internet search, biometric identification, self-driving cars, Mars robots, American presidential elections, military decision aides, and so on.

VII. CONCLUSION

According to the research findings, TFIDF Vectorization surpassed the Count Vectorization feature extraction approach in terms of classification accuracy. However, feature extraction using TFIDF Vectorization takes longer to execute than the Count Vectorization approach. In study, classification methods such as Support Vector Machine (SVM), Naive Bayes (NB), and Random Forest (RF) were utilized. TFIDF Vectorization +RF beat competing methods based on numerous assessment factors, including accuracy, precision, recall, and f1-score.

VIII. REFERENCES

[1] M.D.Devika, C.Sunitha, Amal Ganesh “Sentiment Analysis: A Comparative Study on Different Approaches” ScienceDirect Fourth International Conference on Recent Trends in Computer Science Engineering https://doi.org/10.1016/j.procs.2016.05.124.

[2] Rohit Joshi, Rajkumar Tekchandani” Comparative analysis of Twitter data using supervised classifiers” 2016 International Conference on Inventive Computation Technologies (ICICT)DOI:10.1109/ INVENTIVE.2016.7830089.

[3] Harpreet Kaur, Veenu Mangat, Nidhi ”A Survey of Sentiment Analysis techniques ” 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) **DOI: 10.1109/ISMAC.2017.8058315.**

[4] Mehdi Allahyari, Seyedamin Pouriyeh, Mehdi Assefi, Saied Safaei, Elizabeth D. Trippe, Juan B. Gutierrez, Krys Kochut, “A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques”, arXiv:1707.02919 [cs.CL], July 2017.

[5] Robert Dzisevicˇ , Dmitrij Sˇesˇok ”Text Classification using Different Feature Extraction Approaches Text Classification using Different Feature Extraction Approaches” 2019 Open Conference of Electrical, Electronic and Information Sciences (eStream).

[6] Seyyed Mohammad Hossein Dadgar, Mohammad Shirzad Araghi, Morteza Mastery Farahani ”A Novel Text Mining Approach Based on TF-IDF and Support Vector Machine for News Classification” 2nd IEEE International Conference on Engineering and Technology (ICETECH), 17th 18thMarch 2016, Coimbatore, TN, India.