TOURIST PLACE REVIEWS SENTIMENT CLASSIFICATION USING MACHINE LEARNING TECHNIQUES

Sai Bingi[1], Nadeem Mohammed[2], Naveen Mannem[3], Nilesh Rao Venepally[4], J.VijaySree[5]

[1,2,3,4] UG students, Dept of CSE, ANURAG Engineering College, Ananthagiri, Suryapet, TS, India.

[5] Assistant Professor, Dept of CSE, ANURAG Engineering College, Ananthagiri, Suryapet, TS, India.

ABSTRACT:

Nowadays, social media is becoming more popular. On tourism websites, millions of users review and rate tourist attractions every day. These may be subjected to sentiment analysis. Reviews that can assist you in determining the popularity of a tourist attraction. Tourists may simply make decisions based on the results of sentiment analysis. To be visited as part of a tour. In this study, we look at how people feel has been implemented with the help of a machine learning algorithm. The data set was compiled from a variety of tourism reviews websites. We have conducted a comparative analysis of CountVectorization, for example, is a feature extraction approach. TFIDF Vectorization. In addition to classification algorithms, Support Vector Machine (SVM), Naive Bayes (NB), and Forest of Chance (RF). Algorithm performance has been improved compared utilising multiple metrics such as recall, accuracy, and precision precision and f1-score are two terms that come to mind while discussing precision and f1-score We discovered via experimentation that TFIDFVectorization has a feature extraction approach called TFIDFVectorization better classification algorithm accuracy when compared to For a particular review dataset, use CountVectorization. In terms of feeling Reviews of tourist attractions are categorised.

Key words: RF, SVM, NB, TFIDF, High efficiency.

I. INTRODUCTION

Nowadays, social media is quickly expanding. Hundreds of millions On a daily basis, people publish reviews and rank tourism attractions. Websites dedicated to tourism For the purpose of analysing the sentiment of these reviews, It is possible to do research. Reviews will be properly analysed if they are properly analysed possible to identify a tourist attraction's popularity trend Summarized Tourists will be able to make better decisions

based on the findings of sentiment analysis Tour location and itinerary. Two feature extraction methods are presented in this study. CountVectorization and CountVectorization TFIDFVectorization is a vectorization algorithm developed by TFIDF. There are three classifications as well. Support Vector Machine (SVM), Naive Bayes (NB) For sentiment analysis, SVM and Random Forest (RF) were utilized classification. Performance has been compared conducted for the purpose of combining fea- ture extraction and classification algorithms based on variables like as Execution time, accuracy, recall, precision, and f1-score are all factors to consider. Nowadays, social media is quickly expanding. On a daily basis, millions of users publish-reviews & evaluate tourist attractions on tourism websites. This review may be analysed using sentiment analysis. A proper examination of evaluations will reveal a pattern in the famous places of tourist locations. The results of the sentiment research will be summarised to assist travellers in deciding on a holiday destination and itinerary. The CountVectorization and SVM Vectorization techniques were employed in this research study to extract features. For sentiment categorization,three classification algorithms were used: naive bayes (NB), Support-Vectors-Machines (SVM), & Random Forest (RF). Performance for a feature combination is calculated using variables such as output time, accuracy, recalls, precision, and scores.

2. LITERATURE SURVEY

Paper 1: Sentiment Analysis: A Compar-ative Study On Different Approaches: Sentiment-analysis (SA) is a method of eliciting a user's feelings and vibes using logic.It is one of most gains fields of (NLP). The expansion of Internet-based apps has resulted in a flood of personalised evaluations for a variety of internet resources. Sentiment Analysis is a sophisticated technique that allows users to extract necessary information as well as mixed reviews' collective sentiments .

Paper 2: A Novel Frame-work for aspect-based opinion categorization in tourist destinations: Before visiting a city or nation, tourists want to know the advantages and drawbacks of the tourist attractions. They usually utilise social media sites to read what previous guests have to say. The analysis in this work is based on three types of opinion mining techniques: trend-based opinion minings, aspect-based opinion mining, and sentence-based opinion mining. has collected useful information from tweets, reviews, and travel blogs.

Paper 3: A Comparative analysis of Twitter data using supervise-classifiers: Online social media microblogging is used to share opinions on specific topics in very short messages.There are some famous microblogs such as Twitter, Facebook, etc.where Twitter gets the most attention in areas like product research, movie reviews, stock market, etc. The applied supervise ML algorithms are (SVM), maximum-entropy & naïve-bayes for classification of data using the unigram, bigram and hybrid.

EXISTING SYSTEM

Customers can become active users by providing reviews of various products/services that other potential customers may find useful, but there are thousands or even more product/service reviews on the Internet, and reading all of these reviews is a very important task. And it's hard for clients. However, there has been almost little study done on tourism assessments to define their feelings. To classify product evaluations based on their sentiment polarity, The findings of a unique feature vector generating approach were unreliable. The study's purpose is to determine how comfortable people are with technology. This involves the process of instructing the user on how to utilize the system effectively. Rather then being fearful of the system, the user should accept it as a need. The amount of acceptability by consumers is totally determined by the tactics used to teach them about it. As a result, the produced system was also developed within the plan, which was accomplished because the majority of the technologies used were freely viewable. Tourism industry is also based nature calamities like floods, storms, tsunamis, volcanic eruption etc so weather forecast will be one of the important factor in tourist places as tourists safety is prior first in the tourism industry.

PROPOSED SYSTEM

Nowadays, social media is becoming more popular. On tourism websites, millions of users review and rate tourist attractions every day. These may be subjected to sentiment analysis.

Reviews that can assist you in determining the popularity of a tourist attraction. Tourists may simply make decisions based on the results of sentiment analysis. To be visited as part of a tour. In this study, we look at how people feel has been implemented with the help of a machine learning algorithm. The data set was compiled from a variety of tourism reviews websites. We have conducted a comparative analysis of CountVectorization, for example, is a feature extraction approach. TFIDF Vectorization. In addition to classification algorithms, Support Vector Machine (SVM), Naive Bayes (NB), and Forest of Chance (RF). Algorithm performance has been improved compared utilising multiple metrics such as recall, accuracy, and precision precision and f1-score are two terms that come to mind while discussing precision and f1-score We discovered via experimentation that TFIDFVectorization has a feature extraction approach called TFIDFVectorization better classification algorithm accuracy when compared to For a particular review dataset, use CountVectorization. In terms of feeling Reviews of tourist attractions are categorised.

Operation:

Data collected from various tourism websites was raw so need of data preprocessing to eliminate irrelevant and unuseful words from reviews. In data preprocessing stop words, punctuation mark, short word has been removed. Also tokenization, lemmatization, stemming has been performed. Data preprocessing is crucial step which help to feature reduction and better performance of

machine learning algorithms. After data cleaning, Feature extraction algorithm has been implemented from scratch. CountVectorization and TFIDFVectorization has used for feature-extract.
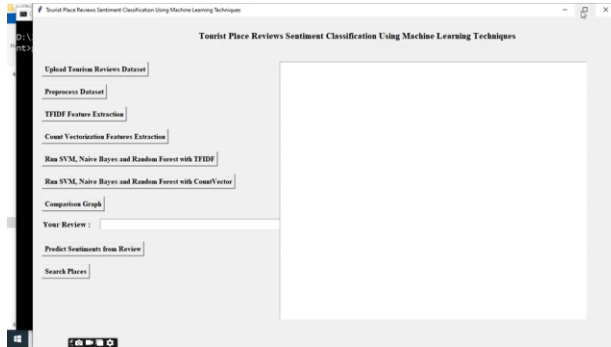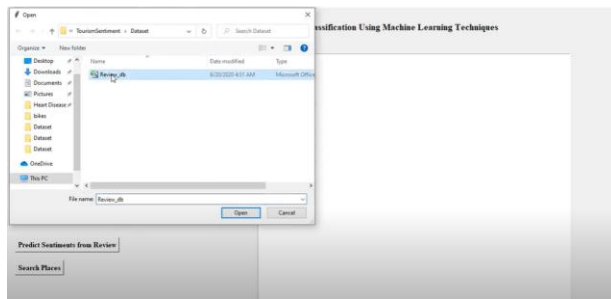

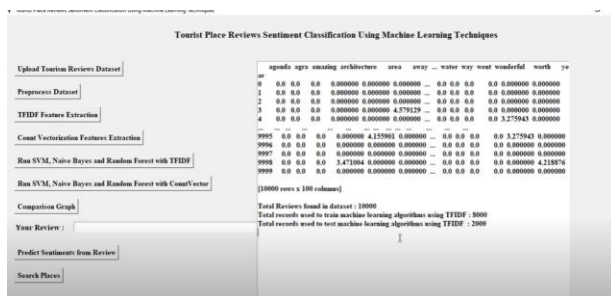
Fig.1. Home page.

Fig.2. Upload the data set.
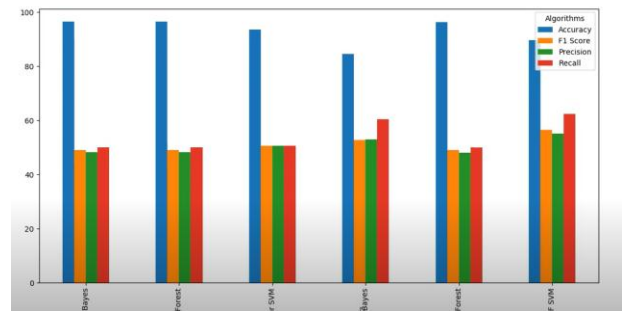

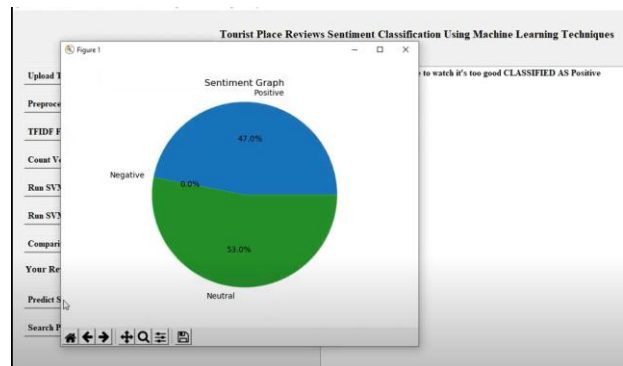**Fig.3. Prediction page.**

Fig.4. Output graphs.

Fig.5. Output results.

V. CONCLUSION

From research study, we can infer that TFIDFVectorization has outperformed over CountVectorization feature extraction algorithm by increasing accuracy of classification. But feature extraction using TFIDFVectorization requires more execution time than CountVectorization algorithm. In research, classification algorithms Support Vector Machine(SVM), Naïve Bayes(NB), Random Forest(RF) has been used. It has found that

TFIDFVectorization+RF outperformed over other algorithms used on bases of several evaluation parameters like accuracy, precision, recall and f1-score.

REFERENCES

[1] M.D.Devika, C.Sunitha, Amal Ganesh "Sentiment Analysis: A Comparative Study on Different Approaches" ScienceDirect Fourth Interna- tional 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, SeyedaminPouriyeh, Mehdi Assefi, Saied Safaei, Elizabeth D. Trippe, Juan B. Gutierrez, KrysKochut, "A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques", arXiv:1707.02919 [cs.CL], July 2017

[5] Robert Dzisevič , DmitrijŠešok "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. \

[7] Rasika Wankhede, Prof. A.N.Thakare "Design Approach for Accuracy in Movies Reviews Using Sentiment Analysis". International Conference on Electronics, Communication and Aerospace Technology ICECA 2017

[8] Bo Pang and Lillian Lee, ShivakumarVaithyanathan "Sentiment Classifi- cation using Machine Learning Techniques " Proceedings of the Conference on Empirical Methods in

Natural Language Processing (EMNLP), Philadelphia, July 2002, pp. 79-86. Association for Computational Linguistics.

[9] Muhammad Afzaal, Muhammad Usman "Novel Framework for Aspect-based Opinion Classification for Tourist Places" The Tenth International Conference on Digital Information Management (ICDIM 2015)

[10] Upma kumari, Dr. Arvind K Sharma, Dinesh Soni "Sentiment analysis of smart phone product reviews using SVM classification techniques" 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS).