CLASSIFICATION OF SENTIMENT IN TOURIST DESTINATION REVIEWS USING MACHINE LEARNING TECHNIQUES

T.Shashi Rekha[1]., G.Sai Vaishnavi[2].,M.Pujitha[3].,P.Sneha[4].,S.Kruthika Priya[5]
1Assistant Professor, Department Of CSE., Malla Reddy College Of Engineering For Women., Maisammaguda.,Medchal., Ts, India (shashirekha.tejavath@gmail.Com)
2, 3, 4, 5 B.TechCSE, (19RG1A05E7, 19RG1A05F2, 19RG1A05G2, 19RG1A05H2),
Malla Reddy College Of Engineering For Women., Maisammaguda., Medchal., Ts, India

ABSTRACT :

The use of social media is growing in popularity these days. Millions of people every day use travel review sites to rank and comment on local attractions. Sentiment analysis might be used to these ratings to better understand how well-liked certain vacation spots are. The findings of sentiment analysis may help tourists decide where to go on vacation. In this study, sentiment analysis was performed using a machine learning approach. The information was collected from a wide range of tourist evaluation sources. In this study, we examined the similarities and differences between two feature extraction methods: Count Vectorization and TFIDF Vectorization. Classification methods such as Random Forest (RF), Support Vector Machine (SVM), and Naive Bayes (NB). Algorithms were compared using a number of different metrics, including accuracy, recall, precision, and f1-score. TFIDF Vectorization feature extraction was shown to be more effective than Count Vectorization feature extraction for the provided review dataset. For a research dataset used in sentiment classification of travel site evaluations, TFIDF Vectorization RF achieved the highest accuracy of 86%.

KEY WORDS :

Tourism-related website sentiment analysis is a hot topic.

INTRODUCTION

The social media industry is growing rapidly at the moment. Millions of travellers share their opinions on travel review sites every day. Sentiment analysis may be used to this review. There is a pattern in the most visited places that may be seen by carefully analysing reviews. Travelers will be provided with a summary of the sentiment analysis results to aid them in selecting a destination and planning their itinerary. This study used two different feature extraction algorithms: Count Vectorization and TFIDFVectorization. Naive Bayes, Support Vector Machine, and Random Forest were the three classification algorithms used for sentiment analysis. Combinations of feature extraction and classification algorithms have been evaluated based on metrics such as execution time, accuracy, recall, precision, and f1- score. Big data is now available to researchers in the tourist industry. Big data analysis has improved the ability of academic institutions and private companies to study tourist behaviour and the tourism industry as a whole. According to Li et al., big data analysis may provide sufficient information on visitor behaviour without introducing sampling bias and can make up for the limitations of the survey data's sample size. According to Sivarajah et al., big data analysis has become the norm since it has the potential to provide novel insights. One of the most important types of data for big data is user-generated content (UGC), which includes anything from blog posts and social network updates to online travel reviews. Insightful commentary provided in real time by users is what we call "user-generated content" (UGC). In most cases, collecting and analysing this kind of feedback information is free or somewhat inexpensive. Financial benefit may also accrue in areas like targeted advertising, customer-company interactions, and brand communication as a result of this kind of feedback.

TripAdvisor and similar sites generate vast amounts of valuable user-generated content (UGC) in the form of text-based online travel reviews. Researchers and professionals may get insight into travellers' tastes and needs by analysing the language of internet reviews. The opinions voiced by users in comments are also influential on the choices of potential visitors. Large data's characteristics have complicated the process of information extraction. The question of how to transform raw data into actionable insights is of paramount importance in the age of big data. Previous research on online reviews has mostly ignored the language of online reviews in favour of the website's quantitative ratings. While star ratings can't reveal which aspects of a product customers adore or despise, the text of reviews often reveals this information. In addition, the abundance of reviews on travel review websites might make some customers feel overwhelmed. Researchers in other fields have also brought forth similar concerns. Ali et al. noted that a city's rating score is insufficient to provide accurate information; nonetheless, comments or tweets may help travellers and traffic managers understand all aspects of the city, where congestion is rapidly growing. Therefore, there is a need for a reliable method to aid users in extracting the essential content and feelings from the review language.

ANALYSIS OF READINGS

Copyright @ 2023 Authors

Anshul Jain et al. (2012) looked at the problem of emotion detection in user comments automatically. Analysing how customers feel about a product is a simple way to tell which ones are effective and which ones aren't. This may be uncovered via the use of sentiment analysis. Because of this, we need to develop a system that can automatically identify words that express an opinion and classify them according to their polarity. This saves time compared to manually analysing and assessing these opinions. It also stresses the need of using unstructured text in place of costly training data. Herein is revealed a rule-based procedure for collecting and analysing product reviews from online communities of shared opinion. Therefore, it should be clear whether a reviewer has a positive, negative, or neutral outlook on a given product. The system uses real-time data to determine rating strengths and weaknesses. The program also determines the efficiency of Lexical resources over the training data. A method for determining opinion goals using the WTM (Word-based Translation Model) was suggested by Kang Liu et al. (2012). This model is first created in a monolingual scenario to extract the associations between opinion targets and opinion terms. Combining the candidate importance with the candidate opinion importance from the derived relationships yields a global measure. Using WTM in this way may be more effective in conquering opinion relations, particularly in the long run. When dealing with informal writings in large online corpora, our approach may be more effective than other current syntax-based techniques at removing noise from parsing failures. By employing a graph-based strategy to generate opinion targets globally, we are able to significantly mitigate the error propagation issues that plague conventional bootstrap-based methods like double propagation. Trials on three real-world datasets of varying sizes and languages demonstrate the superiority of this approach.

Analysis of the Present System

One way for a buyer to become an engaged user is to provide feedback on items and services they've used. However, the client has the option of reading any of the hundreds, thousands, or even millions of reviews related to the product or service in question on the internet. Therefore, there is a need for methods that can automatically summarize these reviews into a positive or negative category, so that the user can get useful feedback.

NEW SYSTEM PLAN

The proposed methodology compares and contrasts a number of different approaches to sentiment analysis. Several tiers of emotion have been developed, including the document level, the phrase level, and the aspect level. In this study, we used machine learning, a rule-based method, and a lexical approach to sentiment analysis. Some examples of techniques used in machine learning include the Support Vector Machine (SVM), the Naive Bayes (NB), and feature-driven sentiment analysis. There has been a lot of research done on the advantages and disadvantages of various approaches to sentiment analysis. Machine learning has been proved to be the most effective method by a number of measures of performance, efficiency, and accuracy.

IMPLEMENTATION

Architecture:

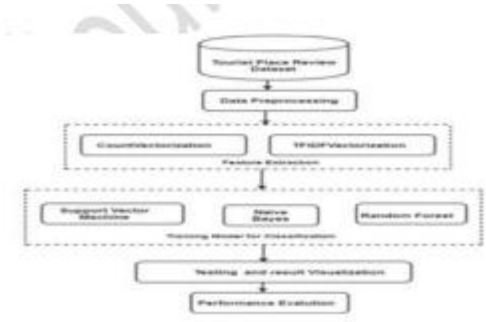


Fig-1. architectures of the system

MODULES:

- User.
- Tourist guide
- Admin
- Machine learning

DESCRIPTION OF MODULES:

The user might be the one to create the initial account. A valid User email and mobile phone number were required during registration for future communication. After registration, the administrator may choose to make the User active. After the administrator has given the user permission to access our system, the user can do so. After signing in, he may look up a certain tourist spot and see what other users have to say about it. The user may also be interested in learning more about vacation spots and vacation packages. Tour guides may be the first to sign up for a new account. A valid User email and mobile phone number were required during registration for future communication. The administrator may start using the instructions after registration is complete. The

Copyright @ 2023 Authors

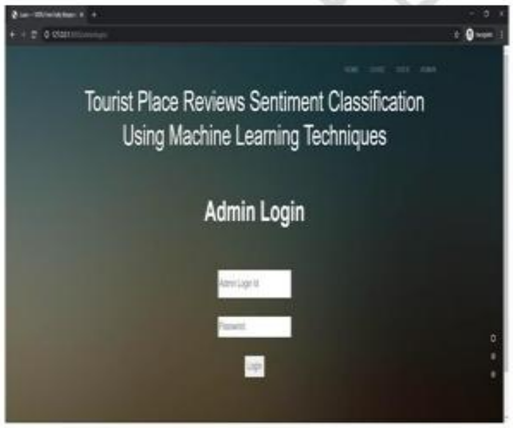
guide will be able to access our system after the administrator has given them permission to do so.

The tour leader has the option of adding the locations to this page. Admin: The administrator may sign in using his credentials. After signing in, he has the option of making the users and tour guide active. Our applications are restricted to authenticated users and administrators. We may utilize naïve bayes algorithms, support vector machines, and random forests to make predictions about emotional analysis. Learning machines: By analysing and learning from massive amounts of data, machines may improve their ability to forecast outcomes and choose optimal solutions. Examples of representation algorithms include deep learning, artificial neural networks, decision trees, and enhancement algorithms. One of the most important ways computers may learn to mimic human intellect is via machine learning. These days, you may find applications of machine learning all throughout the artificial intelligence field. Search engines, biometric identification systems, autonomous vehicles, robots on Mars, presidential elections in the United States, military decision support systems, and many more areas may all benefit from machine learning. Set of Data The research makes use of feedback gathered from a wide variety of travel sites. Reviews and star ratings were collected in a comma-separated values (CSV) file. Based on the score, we identified whether the feeling was positive, negative, or neutral. An evaluation more than three is considered positive. If it's fewer than 3, it's negative; if it's the same as or more than 3, it's neutral. Preparing the Data Raw data from social media platforms required cleaning before it could be used. Data preparation includes operations like as tokenization, stop word deletion, stemming, and lemmatization. Tokenization is the process of separating a sentence into its individual words. Each word is "tokenized" throughout this procedure. Change "stop" to "x" Stop words are common terms found in texts that should not be transmitted to text mining algorithms. Stop words include a, the, this, you, in, is, was, and so on. In the investigation, we used a custom stop word list consisting of frequently occurring but unrelated phrases from the corpus. Since this has occurred,

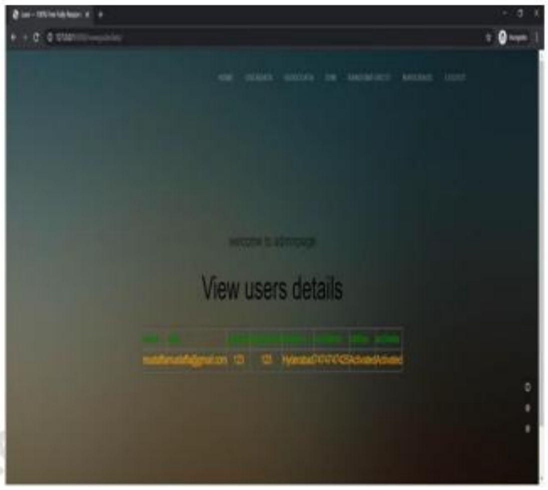
RESULT AND DISCUSSION

Home page:

Admin page:



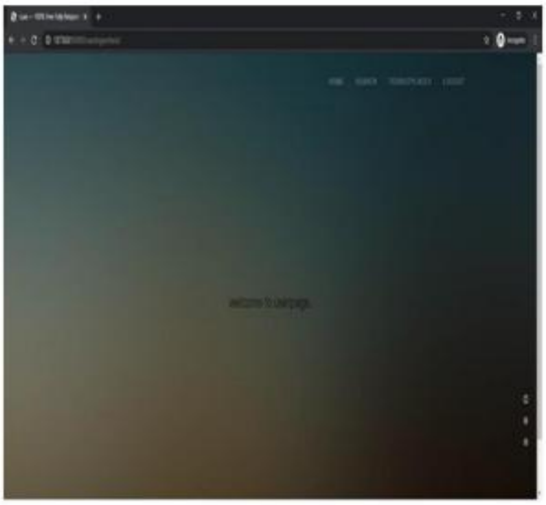
Guide-Details:



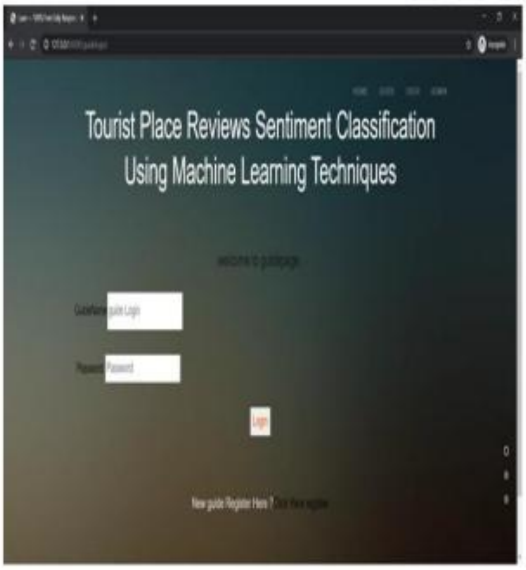
User login

Copyright @ 2023 Authors

User home:



Guide login:



CONCLUSION

Based on the results, TFIDF Vectorization is the superior feature extraction method to Count Vectorization when it comes to classification accuracy. In contrast to Count Vectorization, TFIDF Vectorization requires more time to complete. Classification strategies used in the research included Support Vector Machines (SVM), Naive Bayes (NB), and Random Forests (RF). TFIDF Vectorization was deemed superior to other methods based on a battery of evaluation criteria, including accuracy, precision, recall, and f1- score.

REFERENCES

[1] M.D.Devika, C.Sunitha, and Amal Ganesh, "Sentiment Analysis: A Comparative Study of Different Approaches." Fourth International Conference on Recent Trends in Computer Science Engineering (https://doi.org/10.1016/j.procs.2016.05.12)

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

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

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

[5] Dmitrij Sesok, Robert Dzisevic, "Text Classification Using Different Feature Extraction Approaches" "Text Classification Using Various Feature Extraction Methodologies" is the title of the 2019 Open Conference of Electrical, Electronic, and Information Sciences (eStream)

Copyright @ 2023 Authors

[6] "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), Coimbatore, TN, India, 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),

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

[8] "Sentiment Classification Using Machine Learning Techniques," Proceedings of the Conference on Empirical Methods in Natural Language Processing(EMNLP), Philadelphia, July 2002, pp. 79- 86. Bo Pang and Lillian Lee, Shivakumar Vaithyanathan, "Sentiment Classification Using Machine Learning Techniques," Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Philadelphia, July 2002, pp. 79-

[9] Upma Kumari, Arvind K Sharma, and Dinesh Soni, "Sentiment analysis of smart phone product evaluations using SVM classification approaches." International Conference on Energy, Communication, Data Analytics, and Soft Computing (Energy, Communication, Data Analytics, and Soft Computing) (ICECDS)

[10] "Sentiment analysis utilising product review data," by Xing Fang and Justin Zhan. DOI: 10.1186/s40537-015-0015-2 Springer a Journal of Big Data (2015) 2:5 DOI: 10.1186/s40537-015-0015-2

[11] Leo Breiman, Statistics Department, University of California Berkeley, CA 94720, "RANDOM FORESTS"

[12] Tree-based Machine Learning Algorithms: Decision Trees, Random Forests, and Boosting, C. Sheppard, CreateSpace is a self-publishing platform that was launched in 2017.

[13] "Detecting Sentiment Polarity in Bengali Tweets Using Character N-gram Features and Multinomial Naive Bayes," Kamal Sarkar, 2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT)

[14] Naive Bayes and Text Classification NB.pdf is available at https://web.stanford.edu/juraf sky/slp3/slides/7

Copyright @ 2023 Authors