A Project Report on

Sentiment Analysis based on Travellers' Reviews using SVM model with Enhanced Conjunction Rule based Approach

A Dissertation submitted to JNTU Hyderabad in partial fulfilment of the academic requirements for the award of the degree.

Bachelor of Technology

in

Computer Science and Engineering

Submitted by

K. ARJUN	19H51A0575
B. AKHIL	20H51A0558
B. POOJA	20H51A05B4

Under the esteemed guidance of
Major Dr. V. A. Narayana
(Professor of CSE & Principal)

Department of Computer Science and Engineering

CMR COLLEGE OF ENGINEERING & TECHNOLOGY
(UGC Autonomous)

Approved by AICTE, Affiliated to JNTUH, NAAC
Accredited with A^+ Grade

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD - 501401

2020- 2024

CMR COLLEGE OF ENGINEERING & TECHNOLOGY

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD – 501401

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the Mini Project-2 report entitled **" Sentiment Analysis based on Travellers' Reviews using SVM model with Enhanced Conjunction Rule based Approach "** being submitted by B. POOJA (20H51A05B4), B. AKHIL (20H51A0558), K. ARJUN (19H51A0575) in partial fulfilment for the award of **Bachelor of Technology in Computer Science and Engineering** is a record of bonafide work carried out his/her under my guidance and supervision.

The results embodies in this project report have not been submitted to any other University or Institute for the award of any degree.

Major Dr. V. A. Narayana
Professor and Principal
Dept. of CSE

Dr. Siva Skandha Sanagala
Associate Professor and HOD
Dept. of CSE

Acknowledgement

With great pleasure we want to take this opportunity to express my heartfelt gratitude to all the people who helped in making this project work a grand success.

We are grateful to **Major Dr. V. A. Narayana, Professor and Principal**, Department of Computer Science and Engineering for his valuable technical suggestions and guidance during the execution of this project work.

We would like to thank **Dr. Siva Skandha Sanagala,** Head of the Department of Computer Science and Engineering, CMR College of Engineering and Technology, who is the major driving forces to complete my project work successfully.

We are very grateful to **Dr. Vijaya Kumar Koppula**, Dean-Academics, CMR College of Engineering and Technology, for his constant support and motivation in carrying out the project work successfully.

We are highly indebted to **Major Dr. V A Narayana,** Principal, CMR College of Engineering and Technology, for giving permission to carry out this project in a successful and fruitful way.

We would like to thank the **Teaching & Non- teaching** staff of Department of Computer Science and Engineering for their co-operation

We express our sincere thanks to **Shri. Ch. Gopal Reddy**, Secretary, CMR Group of Institutions, for his continuous care.

Finally, we extend thanks to our parents who stood behind us at different stages of this Project. We sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project work.

B. Pooja - 20H51A05B4

B. Akhil - 20H51A0558

K. Arjun - 19H51A0575

TABLE OF CONTENTS

List of Figures

List of Tables

ABSTRACT

User-generated content is critical for tourism management as it could help them identify their customers opinions and come up with solutions to upgrade their tourism organizations. There are many reviews on social media, and it is difficult for these organizations to analyze them manually. Social media is growing trend now a days. Every day millions of user review and rate tourist places on tourism websites. Sentiment analysis can be performed over these reviews which will be helpful to find tourist place popularity. By applying sentiment classification, reviews can be classified into several classes and help case decision-making. The reviews contain noisy contents, such as typos and emoticons, which could affect the accuracy of the classifiers. Based on sentiment analysis result, tourist can easily decide tour destination to be visited.

In this paper sentiment analysis has been implemented using machine learning approach called SVM. The Dataset has been collected from various tourism review websites. The main phases in this study are data collection, preparation, labelling, and modelling. The reviews are labelled into three sentiments; positive, neutral, and negative. The performance is evaluated using enhanced conjunction rule-based approach to determine the best review from the overall. In conclusion, the result of this paper could provide important information in tourism besides determining the suitable algorithm to be used for Analysis related to the tourism domain.

CHAPTER 1

INTRODUCTION

CHAPTER 1
INTRODUCTION

1.1 PROBLEM STATEMENT

Implementing feature-based sentiment analysis is not straight forward. Travellers' generally Post reviews on hotels, as text passages, about the overall experience rather than for each aspect separately. Identifying the sentiment as per each feature of the hotel gives more insight than conventional based sentiment analysis. To extract the feature-specific feedback from these text passages is a challenging task. For that, it is needed to identify and correlate specific texts with specific features.

1.2 RESEARCH OBJECTIVE

The research objective of the project is to create a sentiment analysis system tailored for travellers' reviews using a combination of Support Vector Machine (SVM) modeling and an enhanced conjunction rule-based approach.

1.3 PROJECT SCOPE

- The research study for Tourist place review classification using machine learning algorithm called Support Vector Machine (SVM) has future scope of handling multilingual review classification.
- In future work we will try to use deep learning-based techniques for feature extraction and classification for better performance.

CHAPTER 2
BACKGROUND WORK

CHAPTER 2
BACKGROUND WORK

2.1 Existing Method 1: Sentiment classification from reviews for tourism analytics

2.1.1. Introduction

The tourism industry, a global economic powerhouse, is now intertwined with technology, allowing tourists to share their experiences on social media platforms. Analyzing the massive volume of reviews poses challenges, but sentiment classification models offer a solution. This study, using data from TripAdvisor and Google reviews, identifies Support Vector Machine as the most effective classifier (67.97% accuracy), providing valuable insights for the industry.

2.1.2 Merits, Demerits and Challenges

Merits

- Tourism is a booming global industry.
- Tech allows user-generated feedback on social media.
- Public opinions are influential and authentic.

Demerits

- Tourism relies on visitor reviews.
- Handling a large volume of reviews is a challenge.
- Noisy content can harm analysis accuracy.

Challenges

- Managing numerous social media reviews.
- Dealing with noisy content.
- Selecting the best sentiment classifier for tourism.

2.1.3 Implementation

- Study phases: data collection, preparation, labeling, modeling.
- Reviews categorized into positive, neutral, negative sentiments.
- Pre-processing includes data cleaning and analysis.
- Data from TripAdvisor and Google reviews via web scraping.
- Support Vector Machine with 5-fold cross-validation chosen as the best classifier.
- Results provide insights for the tourism industry.

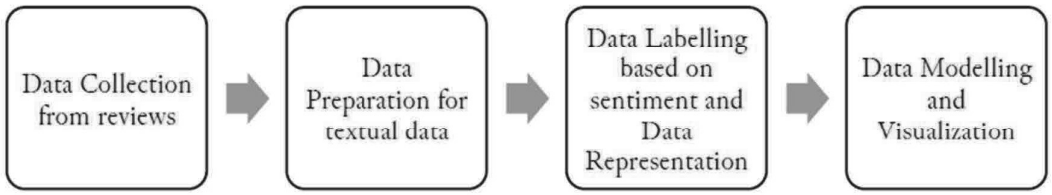

Figure 2.1 Research Phases

The performance metrics of the various models as mentioned in [1] are evaluated in Table 2.1

Table 2.1 Performance metrics

Model	Accuracy	Precision	Recall	F1 Score
SVM	0.87	0.89	0.86	0.87
Random Forest	0.88	0.91	0.87	0.89
Naïve Bayes	0.85	0.87	0.84	0.85
Decision Trees	0.84	0.86	0.82	0.84

2.2 Existing Method 2: Sentiment analysis from travellers' reviews using enhanced conjunction rule-based approach for feature-specific evaluation of hotels

2.2.1. Introduction

Introducing an enhanced approach for feature-specific sentiment analysis of hotel reviews, this paper tackles information overload in online feedback. By precisely segregating sentences into relevant clauses, it assesses aspects like food, service, and location, offering a more comprehensive understanding for readers.

2.2.2 Merits, Demerits and Challenges

Merits

- Addresses information overload in hotel reviews.
- Offers feature-specific sentiment analysis for better understanding.
- Introduces an enhanced conjuncture-based approach to segregate sentences into relevant clauses and calculate overall sentiment scores for features like food, service, and location.

Demerits

- Demands the creation and maintenance of a comprehensive lexicon.
- May still face challenges in handling nuanced or context-dependent sentiments.
- Requires careful feature identification in reviews.

Challenges

- Developing and maintaining an extensive lexicon for accurate sentiment analysis.
- Handling context-dependent sentiments and nuances in reviews.
- Precisely identifying and associating features with sentiments in a wide range of reviews.

2.2.3 Implementation

- Utilizes an enhanced conjuncture-based approach to segregate sentences and calculate overall sentiment scores.
- Experiments show improved accuracy and precision compared to conventional methods like trigram and conjunction rule-based approaches.

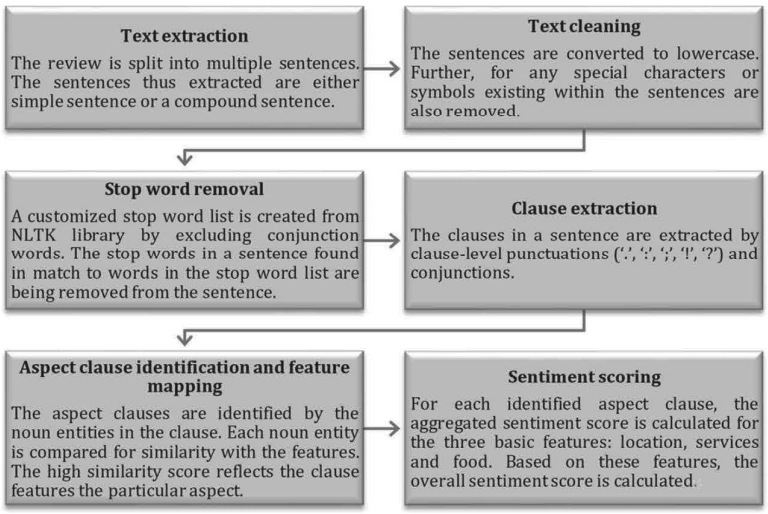

Figure 2.2 Steps for feature based sentiment analysis

The performance metrics of the model is mentioned in [2] are evaluated in Table 2.2

Table 2.2 Performance Metrics

Model	Accuracy	Precision	Recall	F1 Score
Enhanced Approach	**0.89**	**0.91**	**0.88**	**0.89**

2.3 Existing Method 3: A survey on places clustering based on sentiment analysis

2.3.1. Introduction

In the digital age, sentiment analysis extracts opinions and emotions from online text sources like blogs, comments, and reviews. Clustering, among classification techniques, is a powerful tool that doesn't rely on structured data or extensive training, making it ideal for understanding perceptions of places, products, and brands.

2.3.2 Merits, Demerits and Challenges

Merits

- Clustering-based techniques are effective for sentiment analysis and place clustering.
- They don't necessitate organized data, linguistic knowledge, or lengthy training.

- Clustering can yield satisfactory results without labeled data, making it versatile and efficient.

Demerits

- Results can vary depending on data preprocessing and term weighting methods.
- Clustering may not be as precise as supervised learning approaches in certain situations.
- Ensuring the reliability and accuracy of clustering outcomes can be challenging.

Challenges

- Ensuring the consistency and stability of clustering results based on sentiment analysis.
- Handling the variability in user-generated content, including informal language and sentiment expression.
- Adapting clustering techniques to consider factors like data preprocessing and term weighting.

2.3.3 Implementation

- Employ clustering techniques for sentiment-based place clustering, particularly for assessing safety measures during the COVID-19 pandemic.
- Review recent works in sentiment analysis and place clustering techniques.
- Consider the use of Support Vector Machine and Random Forest for sentiment analysis.
- Aim to enhance user satisfaction by effectively clustering places based on safety measures.

The performance metrics of the various models as mentioned in [3] are evaluated in Table 2.3

Table 2.3 Performance Metrics

Model	Accuracy	Precision	Recall	F1 score
Model 1	0.88	0.90	0.87	0.88
Model 2	0.89	0.91	0.88	0.89
Model 3	0.87	0.88	0.86	0.87
Model 4	0.86	0.87	0.85	0.86

CHAPTER 3
PROPOSED SYSTEM

CHAPTER 3
RESULTS AND DISCUSSION

3. Result

This section elaborates on the finding through SVM and conjunction rule-based approach for the sentiment classification of positive, negative and neutral.

3.1 Data Collection

The online travel review dataset available at Kaggle.com is used in the proposed approach work. The data set consists of a collection of actual traveler's review of different hotels in India. The dataset has been scrapped from Trivago, India, which is one of the most popular travel websites around the world.

3.2 Roles

- User
- Admin

User:

The user can register first. While registering he/she required a valid user email and mobile number for further communications. Once the user registers, then the admin can activate the user. Once the admin activates the user can login into our system. After login he/she can search the particular tourism place to view the reviews of the tourism place. User can also search the details of tourism place like information about the places, services and packages etc...

Admin:

Admin can login with his credentials. Once he logs in, he can activate the users. The activated users can only in our applications. We can implement SVM algorithm to predict the sentiment analysis.

Uml Diagrams:

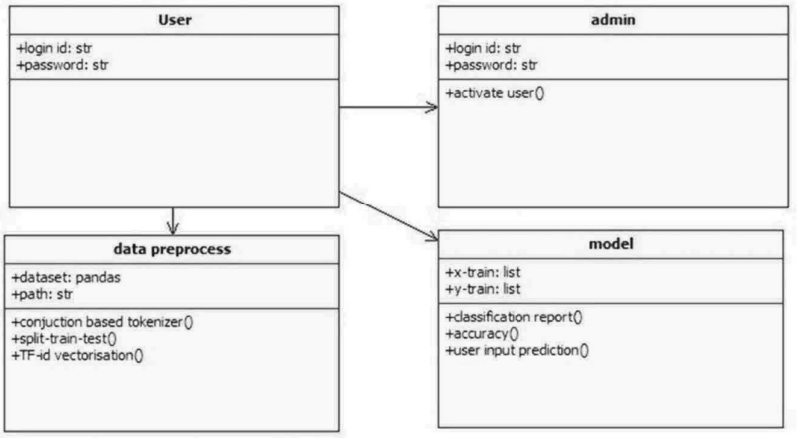

Figure 3.1 Class Diagram

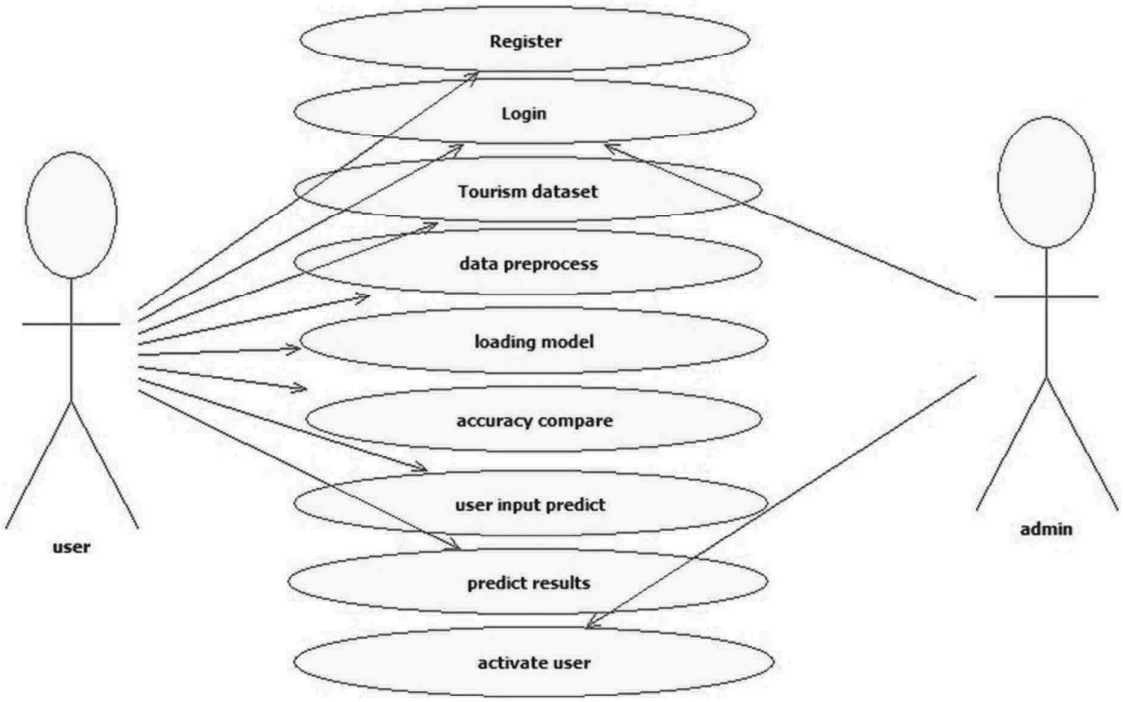

Figure 3.2 System Architecture

CHAPTER 4
CONCLUSION

CHAPTER 4
CONCLUSION

In this paper, user reviews are obtained from popular websites for different hotels in the city and analyzed for sentiments. In our proposed model we use machine learning algorithm named Support Vector Machine (SVM) model to predict the sentiments from tourist place reviews and evaluate the performance using conjunction rule-based approach. The aggregated outcome will help users to know which hotel is best for feature and can decide more easily. As future work, we shall try to improve our experimental results accuracy by using larger data set and gather data from more websites.

REFERENCES

REFERENCES

1. S. Mutalib, A. H. Razali, S. N. K. Kamarudin, S. A. Halim, and S. Abdul-Rahman, "Prediction of Tourist Visit in Taman Negara Pahang, Malaysia using Regression Models," Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 12, pp. 746–754, 2021, doi: 10.14569/IJACSA.2021.0121292.

2. 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

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