A Project Report on

SENTIMENT ANALYSIS BASED ON TRAVELERS' REVIEWS USING THE SVM MODEL WITH ENHANCED CONJUNCTION RULE-BASED APPROACH

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

Bachelor of Technology

In

Computer Science and Engineering

Submitted by

B. POOJA
(20H51A05B4)
B. AKHIL
(20H51A0558)
Under the esteemed guidance
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 Major Project report entitled **"Sentiment Analysis based on Travelers' Reviews using the SVM model with Enhanced Conjunction Rule-Based Approach"** being submitted by Belde Pooja (20H51A05B4) and Bachupally Akhil Goud (20H51A0558) in partial fulfillment for the award of **Bachelor of Technology in COMPUTER SCIENCE AND ENGINEERING** is a record of bonafide work carried out under my guidance and supervision.

The results embodied 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

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

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

We are grateful to our guide **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 force to complete our project work successfully.

We are very grateful to **Dr. Ghanta Devadasu**, 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& Correspondent, CMR Group of Institutions, and **Shri. Ch. Abhinav Reddy**, CEO, CMR Group of Institutions for their continuous care and support.

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 or indirectly in completion of this project work.

B. POOJA – 20H51A05B4

B. AKHIL – 20H51A0558

TABLE OF CONTENTS

List of Figures

List of Tables

ABSTRACT

It is a tough task for tourism management to identify their user reviews and come up with solutions for 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. Decision-making and multiple class classification of reviews are possible with the application of sentiment classification. Sentiment analysis plays a crucial role in helping tourists make informed decisions about their travel destinations. In this particular paper, Using an Enhanced Conjunction Rule-Based Approach and Support Vector Machine (SVM) machine learning technique, the authors conducted sentiment analysis. They collected the dataset from different tourism review websites to train and evaluate their model. The findings of this paper are significant as they not only provide valuable insights in the field of tourism but also help in identifying the most appropriate algorithm for tourism-related analysis. The findings of this paper are significant as they not only provide valuable insights in the field of tourism but also help in identifying the most appropriate algorithm for tourism-related analysis. This information can greatly contribute to the development and improvement of tourism strategies and decision-making processes.

CHAPTER 1

INTRODUCTION

CHAPTER 1

INTRODUCTION

1.1 PROBLEM STATEMENT

Sentiment analysis based on traveler evaluations is difficult to implement since textual data contains a wide range of frequently subtle expressions of thoughts. Although Support Vector Machine (SVM) models provide a strong foundation for sentiment categorization, current methods might not be able to fully capture the nuanced emotions found in traveler reviews. Moreover, conjunction rules are difficult to include in typical SVM models and are crucial for deciphering sentiment nuances in travel-related information. Consequently, to increase sentiment analysis accuracy in the context of traveler evaluations, an improved conjunction rule-based method is required, one that can supplement SVM models. The following major issues are the focus of this problem statement:

Complexity of Travel Reviews: Sentiment analysis is difficult since traveler reviews frequently contain a wide range of linguistic idioms, slang, and contextual nuances. The sentiment nuances indicated in these reviews may be too delicate for current SVM models to fully grasp.

Conjunction rules should be used: Conjunction rules are essential for detecting subtleties in sentiment in textual data, such as negations and amplifications. Nevertheless, the incapacity of conventional support vector machine (SVM) models to efficiently incorporate conjunction rules curtails their precision in sentiment analysis of traveler reviews.

Need for Improved Sentiment Analysis: Due to the tourism industry's reliance on online evaluations and client input, there is a rising need for sentiment analysis techniques that are more precise and perceptive. This need can be met with an improved strategy that combines the advantages of SVM models with sophisticated conjunction rule-based approaches, offering insightful information to travel industry companies.

1.2 RESEARCH OBJECTIVE

This study's main goal is to improve sentiment analysis techniques by combining an enhanced conjunction rule-based strategy with the Support Vector Machine (SVM) model. These techniques are then customized to the unique context of traveler evaluations. Acknowledging the difficulties presented by the varied and complex expressions included in textual data connected to travel, this study aims to provide a strong framework that can precisely classify sentiment polarity while capturing minute linguistic details. This study intends to enhance the granularity and accuracy of sentiment analysis in traveler evaluations by integrating sophisticated conjunction rules for sentiment interpretation with the strengths of SVM models in pattern recognition.

The creation of a sentiment analysis model tailored for traveler evaluations is one of the specific goals. This model will involve feature selection, preprocessing, and SVM classifier training. To handle linguistic constructions like negations and amplifications, which are crucial for sentiment interpretation, emphasis is placed on integrating conjunction rules. Additionally, the study intends to assess the suggested approach's effectiveness through extensive testing with benchmark datasets from the travel industry. To determine the model's robustness and generalization abilities, this evaluation will determine how well it can predict sentiment polarity across a variety of review types, lengths, and linguistic styles.

The ultimate goal of this research is to boost sentiment analysis techniques in the travel sector, which will have a useful impact on companies looking to use consumer input for better customer service and decision-making. The methodology that has been established intends to give businesses practical information to improve customer happiness and increase competitiveness in the dynamic travel market by offering a more nuanced understanding of the emotion conveyed in traveler evaluations.

1.3 PROJECT SCOPE

This project will use an upgraded conjunction rule-based approach to the Support Vector Machine (SVM) model to build, implement, and evaluate a sentiment analysis framework specifically designed for traveler evaluations. The following crucial areas will be the project's main focus:

1. **Data Gathering and Preprocessing**: The project will entail gathering a varied dataset of reviews left by travelers from many websites, including Yelp, Booking.com, and TripAdvisor. A variety of travel-related topics will be covered in these reviews, such as lodging, dining options, tourist destinations, and transit options. To clean and ready the textual material for sentiment analysis, the gathered data will go through preprocessing procedures such text normalization, noise removal, and tokenization.
2. **Model Development:** The main goal of the project is to create a sentiment analysis model that is strengthened by an improved conjunction rule-based methodology and is based on the SVM algorithm. To guarantee correct sentiment polarity (positive or negative) classification in traveler evaluations, this will involve feature selection, model training, and optimization. Conjunction rule integration will serve as a key component in capturing sentiment subtleties found in the textual data.
3. **Implementation of Conjunction Rules:** The project will investigate and put into practice conjunction rules that are specifically adapted to the linguistic subtleties included in traveler evaluations. These guidelines will cover things like intensifiers, negations, amplifications, and other language constructions that affect how sentiment is interpreted. Creating algorithms to successfully

identify and apply conjunction rules during sentiment analysis will be part of the implementation.

4. **Performance Assessment**: To determine the effectiveness and success of the created sentiment analysis framework, a thorough examination will be conducted. This will entail running tests to gauge classification accuracy, precision, recall, and F1 score using benchmark datasets and approved assessment criteria. To determine whether the suggested method is superior to current sentiment analysis techniques, a comparison analysis will also be part of the review process.
5. **Deployment and Application:** The sentiment analysis system will be put to use in real-world travel industry scenarios following its successful development and assessment. This entails examining sentiment patterns, producing useful insights, and assisting companies in making decisions by utilizing the sentiment analysis data obtained from visitor reviews. To aid in the acceptance and application of the created framework, documentation, and user guides will also be supplied.

CHAPTER 2

BACKGROUND WORK

CHAPTER 2
BACKGROUND WORK

We carried out a thorough examination of previous research on techniques and literature as part of our preparation phase for investigating sentiment analysis based on traveler evaluations utilizing the SVM model with an improved conjunction rule-based approach. By looking through scholarly publications, books, and databases, we were able to compile a plethora of information about sentiment analysis, natural language processing, and tourism analytics. This thorough assessment offered insights into the many theories, concepts, and methods used in sentiment analysis in a variety of contexts, from social media sentiment tracking to product reviews.

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

2.1.1. Introduction

The tourism industry stands as a cornerstone of the global economy, continuously evolving alongside technological advancements. With the proliferation of social media platforms and online review websites, tourists now have unprecedented avenues to share their travel experiences in real time. These platforms serve as virtual forums where travelers can voice their opinions, offer recommendations, and provide feedback on various aspects of their journeys, ranging from accommodations and attractions to dining experiences and transportation services.

However, with this vast influx of user-generated content comes the challenge of effectively harnessing and analyzing the wealth of information contained within these reviews. The sheer volume and diversity of opinions expressed in travel reviews make manual analysis impractical and time-consuming. As a result, the tourism industry increasingly relies on automated sentiment classification models to extract valuable insights from these textual data streams.

In this study, we aim to address this challenge by leveraging sentiment classification techniques to analyze reviews sourced from prominent platforms such as TripAdvisor and Google. By employing machine learning algorithms, specifically the Support Vector Machine (SVM), we seek to classify sentiments expressed in these reviews with high accuracy. Our research endeavors to provide actionable insights for stakeholders in the tourism industry, enabling them to better understand customer sentiments, identify areas for improvement, and tailor their offerings to meet the evolving needs and preferences of travelers. Through this endeavor, we aim to contribute to the burgeoning field of tourism analytics, where data-driven insights pave the way for enhanced customer experiences and sustainable business growth.

2.1.2 Merits, Demerits, and Challenges

Merits

- **Tourism is a booming global industry:** The tourism sector experiences continuous growth and contributes significantly to the global economy, making it an attractive domain for sentiment analysis.
- **Tech allows user-generated feedback on social media:** With the advent of technology and social media platforms, users can easily share their experiences and opinions about tourism destinations, providing a rich source of user-generated feedback.
- **Public opinions are influential and authentic**: User-generated feedback on social media platforms is often perceived as more authentic and trustworthy by potential travelers, influencing their decisions regarding destination choices and experiences.

Demerits

- **Tourism relies on visitor reviews:** The tourism industry heavily relies on visitor reviews and recommendations to attract new visitors and retain existing ones, making sentiment analysis crucial for understanding customer sentiments and preferences.
- **Handling a large volume of reviews is a challenge:** The sheer volume of user-generated reviews on social media platforms poses a challenge for tourism analytics, requiring efficient methods for data collection, processing, and analysis.
- **Noisy content can harm analysis accuracy:** User-generated reviews may contain noisy or irrelevant content, such as spam, irrelevant comments, or biased opinions, which can negatively impact sentiment analysis accuracy and reliability.

Challenges

- **Managing numerous social media reviews**: Aggregating and managing the vast amount of user-generated reviews from various social media platforms can be challenging, requiring robust data collection and management strategies.
- **Dealing with noisy content:** Filtering out noisy or irrelevant content from user-generated reviews is essential to ensure accurate sentiment analysis results, necessitating the development of effective noise reduction techniques.
- **Selecting the best sentiment classifier for tourism:** Choosing the most suitable sentiment classifier for tourism analytics requires careful consideration of various factors, such as classification

accuracy, computational efficiency, and scalability, among others.

2.1.3 Implementation

- Study phases: data collection, preparation, labeling, and modeling: The implementation of sentiment classification for tourism analytics involves several phases, including data collection from social media platforms, data preparation, labeling reviews with sentiment labels (positive, neutral, negative), and building sentiment classification models.
- Reviews categorized into positive, neutral, and negative sentiments: User-generated reviews are categorized into three sentiment categories: positive, neutral, and negative, based on the sentiments expressed within the reviews.
- Pre-processing includes data cleaning and analysis: Before sentiment analysis, the collected data undergoes pre-processing, which involves tasks such as data cleaning to remove noise and irrelevant information and data analysis to gain insights into the characteristics of the dataset.
- Data from TripAdvisor and Google reviews via web scraping: User-generated reviews are collected from popular platforms such as TripAdvisor and Google reviews using web scraping techniques, enabling access to a diverse range of reviews from different destinations.
- Support Vector Machine with 5-fold cross-validation chosen as the best classifier: Among various sentiment classification algorithms, the Support Vector Machine (SVM) algorithm is chosen as the best classifier for tourism analytics, and 5-fold cross-validation is used to evaluate its performance.
- Results provide insights for the tourism industry: The results obtained from sentiment analysis provide valuable insights and actionable recommendations for stakeholders in the tourism industry, helping them make informed decisions to enhance customer satisfaction and improve overall tourism experiences.

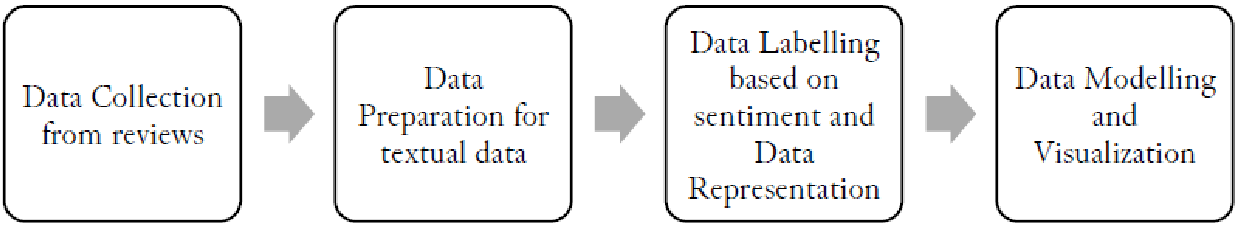

Figure 2.1 Research Phases

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 travelers' 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. In the world of hospitality, the growth of internet review sites has resulted in a flood of feedback from tourists, bringing both benefits and challenges to hoteliers. In the middle of such a large number of information, precisely deciphering the subtle sentiments expressed toward certain hotel features becomes critical for businesses looking to effectively understand and satisfy consumer preferences. In response to this necessity, this article presents an improved approach for feature-specific sentiment analysis of hotel reviews, to mitigate the impacts of information overload and provide useful insights to stakeholders in the hospitality industry.

2.2.2 Merits, Demerits, and Challenges

Merits

- **Addresses information overload in hotel reviews:** The enhanced conjunction rule-based approach helps tackle the challenge of information overload by precisely segregating sentences into relevant clauses, allowing for more focused analysis.
- **Offers feature-specific sentiment analysis for better understanding:** By providing feature-specific sentiment analysis for aspects like food, service, and location, the approach offers deeper insights into travelers' experiences, enhancing understanding and decision-making.
- **Introduces an enhanced conjuncture-based approach:** The enhanced conjuncture-based approach enables the calculation of overall sentiment scores for specific features, offering a more nuanced understanding of sentiments expressed in hotel reviews.

Demerits

- **Demands the creation and maintenance of a comprehensive lexicon:** Developing and maintaining an extensive lexicon for sentiment analysis requires significant effort and resources, which can be a drawback of this approach.

- **May still face challenges in handling nuanced or context-dependent sentiments:** Despite enhancements, the approach may still struggle with capturing nuanced or context-dependent sentiments accurately, potentially affecting the overall analysis.
- **Requires careful feature identification in reviews:** Accurately identifying and associating features with sentiments in a wide range of hotel reviews requires careful consideration and may pose challenges, especially in reviews with ambiguous or complex language.

Challenges

- **Developing and maintaining an extensive lexicon for accurate sentiment analysis:** Creating and updating a comprehensive lexicon of sentiment-bearing words and phrases is crucial for accurate sentiment analysis but presents challenges in terms of resource allocation and ongoing maintenance.
- **Handling context-dependent sentiments and nuances in reviews:** Dealing with context-dependent sentiments and nuanced expressions in hotel reviews requires advanced natural language processing techniques and may pose challenges in accurately capturing and interpreting subtle nuances.
- **Precisely identifying and associating features with sentiments:** Precisely identifying and associating specific features such as food quality, service level, and location with sentiments expressed in hotel reviews can be challenging due to variations in language and expression styles across different reviews.

2.2.3 Implementation

- **Utilizes an enhanced conjuncture-based approach:** The implementation of the approach involves leveraging an enhanced conjuncture-based method to segregate sentences and calculate overall sentiment scores for specific features in hotel reviews.
- **Experiments show improved accuracy and precision:** Experimental results demonstrate improved accuracy and precision compared to conventional methods such as trigram and conjunction rule-based approaches, highlighting the effectiveness of the enhanced approach in sentiment analysis for hotel reviews.

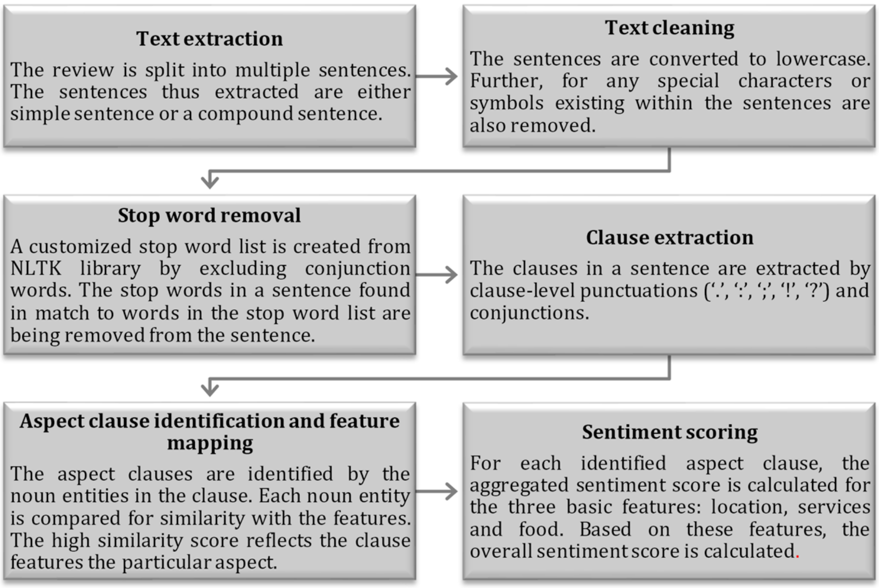

Figure 2.2 Steps for feature-based sentiment analysis

The performance metrics of the model 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. In the digital age, sentiment analysis has developed as a critical technique for collecting opinions and feelings from a variety of online textual sources such as blogs, comments, and reviews. Among the various methodologies, clustering stands out as a powerful strategy for assessing views of places, products, and brands because it does not require organized data or substantial training. Instead, clustering algorithms put comparable instances of textual information together based on shared qualities and underlying patterns, providing a versatile and data-driven strategy for sentiment analysis. This survey delves into the application of clustering techniques in sentiment analysis specifically for understanding perceptions of places, aiming to provide a comprehensive overview of existing methodologies, their strengths and limitations, and their implications for decision-making across various domains, such as tourism and hospitality. Through this investigation, the survey hopes to contribute to a better understanding of sentiment analysis in the digital era, as well as its role in informing strategic efforts and improving consumer experiences.

2.3.2 Merits, Demerits, and Challenges

Merits

- **Clustering-based techniques are effective:** Clustering techniques prove to be effective for sentiment analysis and place clustering, providing valuable insights into the sentiment distribution across different locations.
- **Do not necessitate organized data or linguistic knowledge:** Unlike some other approaches, clustering methods do not require organized data or extensive linguistic knowledge, making them accessible and applicable in various scenarios.
- **Versatile and efficient:** Clustering can yield satisfactory results even without labeled data, making it a versatile and efficient approach for sentiment analysis and place clustering tasks.

Demerits

- **Results dependent on data preprocessing and term weighting methods:** The effectiveness of clustering results can vary depending on the techniques used for data preprocessing and term weighting, potentially affecting the accuracy and reliability of the outcomes.
- **May not be as precise as supervised learning approaches:** While clustering techniques are versatile, they may not always be as precise as supervised learning approaches, particularly in scenarios where labeled data is available and highly accurate predictions are required.
- **Challenges in ensuring reliability and accuracy:** Ensuring the reliability and accuracy of clustering outcomes can be challenging, especially when dealing with noisy or ambiguous data and when evaluating the quality of clustering results.

Challenges

- **Ensuring consistency and stability of clustering results:** Maintaining consistency and stability in clustering results based on sentiment analysis poses challenges, particularly when dealing with dynamic datasets and evolving sentiment patterns.
- **Handling variability in user-generated content:** User-generated content often exhibits variability in terms of language style, sentiment expression, and content structure, making it challenging to adapt clustering techniques effectively.
- **Adapting techniques to consider preprocessing and term weighting:** Adapting clustering techniques to consider factors like data preprocessing methods and term weighting schemes requires careful consideration and may pose challenges in implementation.

2.3.3 Implementation

- **Employ clustering techniques for sentiment-based place clustering:** Implement clustering techniques specifically tailored for sentiment-based place clustering, with a focus on applications such as assessing safety measures during the COVID-19 pandemic.
- **Review recent works in sentiment analysis and place clustering:** Stay updated on recent advancements in sentiment analysis and place clustering techniques to incorporate the latest developments into the implementation process.
- **Consider the use of Support Vector Machine and Random Forest:** Explore the use of advanced machine learning algorithms such as Support Vector Machine and Random Forest for sentiment analysis tasks, complementing clustering techniques to enhance accuracy and efficiency.
- **Aim to enhance user satisfaction:** The ultimate goal of the implementation is to enhance user satisfaction by effectively clustering places based on safety measures, providing valuable insights for travelers and decision-makers.

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
PROPOSED SYSTEM

3.1. Objective of Proposed Model

The proposed system for sentiment analysis on tourist place reviews endeavors to address the nuanced nature of sentiment expression in such reviews by integrating both machine learning and linguistic rules. Through the utilization of a Support Vector Machine (SVM), the system seeks to harness the power of supervised learning to effectively classify sentiments conveyed in reviews as positive or negative. Furthermore, the inclusion of an enhanced conjunction rule-based approach aims to capture the subtle nuances in sentiment by considering the impact of conjunctions on the overall tone of the text. This approach acknowledges that conjunctions play a crucial role in connecting ideas and can significantly influence the sentiment conveyed. By synergistically combining SVM with linguistic rules, the system aims to achieve a more nuanced understanding of sentiment in tourist place reviews, thereby enhancing the accuracy of sentiment classification. Through rigorous model development and evaluation of diverse datasets of reviews from various tourist destinations, the system endeavors to provide a robust sentiment analysis solution tailored specifically for the tourism domain. This solution holds the potential to offer valuable insights to stakeholders such as tourism boards, travel agencies, and individual travelers, facilitating more informed decision-making processes and ultimately enhancing the overall tourism experience.

3.2. Algorithms Used for Proposed Model

In the proposed system for sentiment analysis on tourist place reviews, three key algorithms play crucial roles: Support Vector Machine (SVM) for classification, n-gram for feature extraction, and an enhanced conjunction rule-based approach for refining sentiment analysis. Here's how each algorithm contributes to the system:

1. **Support Vector Machine (SVM)**:

In the proposed system, SVM serves as the primary classification algorithm. It is utilized to categorize reviews into positive, negative, or neutral sentiments based on the features extracted from the text data. SVM works by finding the optimal hyperplane that best separates the data points belonging to different sentiment classes in a high-dimensional feature space. In the context of tourist place reviews, SVM learns to distinguish between positive reviews expressing satisfaction, negative reviews indicating dissatisfaction, and neutral reviews with no strong sentiment. Through training on labeled data, the SVM model learns to generalize patterns and make accurate predictions on unseen reviews, thereby enabling sentiment classification in the proposed system.

2. **N-gram Approach**:

N-grams are sequences of N consecutive words or tokens extracted from the text data. In the proposed system, the n-gram approach is employed for feature extraction to capture the contextual information and linguistic patterns present in the tourist place reviews. By considering sequences of words rather than individual words, n-grams provide a more comprehensive representation of the text data, enabling the model to capture dependencies and nuances in language usage. For example, bigrams (2 grams) such as "beautiful scenery" or "friendly staff" convey specific aspects of the tourist experience that contribute to overall sentiment. By incorporating n-gram features, the proposed system enhances the representation of textual data, thereby improving the performance of sentiment analysis.

3. **Enhanced Conjunction Rule-Based Approach**:

The enhanced conjunction rule-based approach supplements the SVM classifier by incorporating linguistic rules related to conjunctions to refine sentiment analysis. Conjunctions such as "but," "and," or "however" play a crucial role in connecting ideas and can significantly influence the sentiment expressed in a sentence. In the proposed system, conjunctions are identified within the reviews, and specific rules are applied to adjust the sentiment labels based on the context provided by the conjunctions. For instance, if a conjunction like "but" follows a positive sentiment phrase, it may indicate a contrast or qualification, potentially altering the overall sentiment expressed in the sentence. By considering conjunctions and their impact on sentiment, the proposed system improves the granularity and accuracy of sentiment analysis, thereby providing more nuanced insights into the opinions and experiences conveyed in tourist place reviews.

3.3. Designing

The architecture of sentiment analysis on reviews contains various roles that help easy of understand the flow process of our application they are

- **Upload Dataset**
- **Preprocess dataset**
- **Train SVM**
- **SVM with n-grams**
- **SVM with Conjunction**
- **Predict Review**
- **Display Results**
- **Plot Accuracy**

1. **Upload Dataset:**

- Allow users to upload a dataset containing tourist place reviews. This dataset should include text data along with sentiment labels.

2. **Preprocess Dataset:**

- Preprocess the uploaded dataset by removing noise such as HTML tags, punctuation, and stop words.
- Tokenize the text into individual words.
- Optionally, apply techniques like stemming or lemmatization to normalize the text data.

3. **Train SVM:**

- Train a Support Vector Machine (SVM) classifier using the pre-processed dataset.
- Use the labeled data to learn the decision boundary between different sentiment classes (positive, negative).

4. **SVM with N-grams:**

- Enhance the SVM classifier by incorporating n-grams, which capture sequences of N consecutive words.
- Re-train the SVM classifier using feature vectors that include both unigrams (single words) and n-grams.

5. **SVM with Conjunction:**

- Further refine the SVM classifier by incorporating rules to handle conjunctions within the reviews and their impact on sentiment expression.
- Integrate these rules into the sentiment analysis process to improve sentiment classification accuracy.

6. **Predict Review:**

- Allow users to input new reviews or select reviews from the dataset for sentiment analysis.
- Utilize the trained SVM classifier to predict the sentiment of the selected reviews.

7. **Display Results:**

- Display the predicted sentiment labels (positive, negative) for the selected reviews.
- Provide additional information such as review text, predicted sentiment probability scores and any relevant conjunction rules applied.

8. **Plot Accuracy:**

- Calculate the accuracy of the trained SVM classifiers using evaluation metrics such as accuracy, precision, recall, and F1-score.
- Plot the accuracy of each SVM classifier (SVM, SVM with n-grams, SVM with conjunction) to

compare their performance.

3.3.1 Diagrams

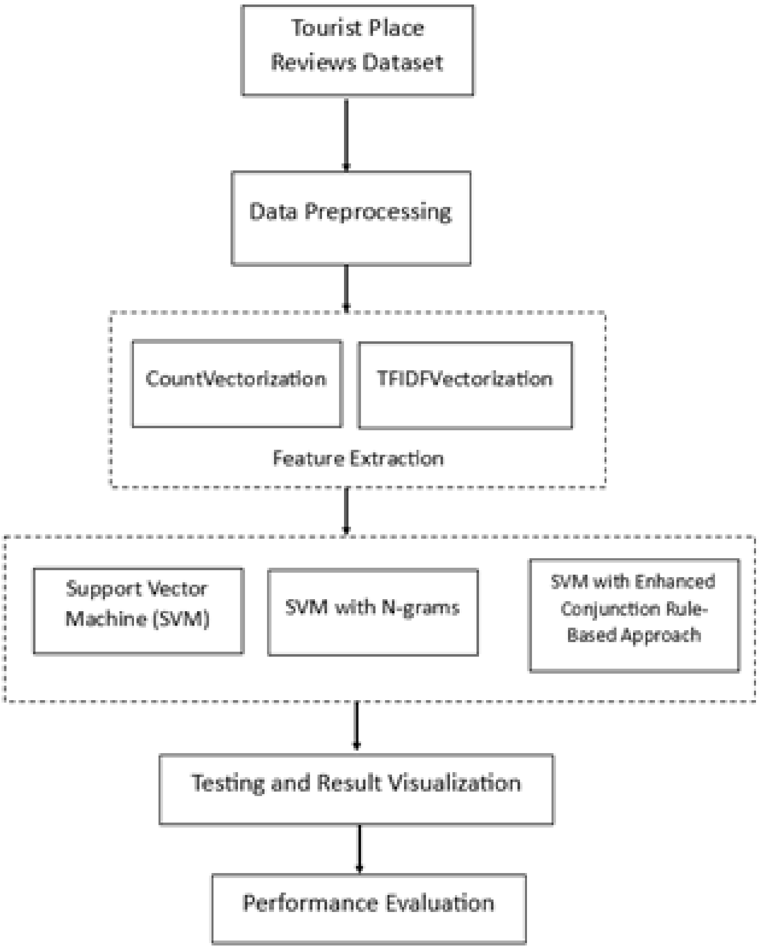

Figure 3.1 Data Flow Diagram

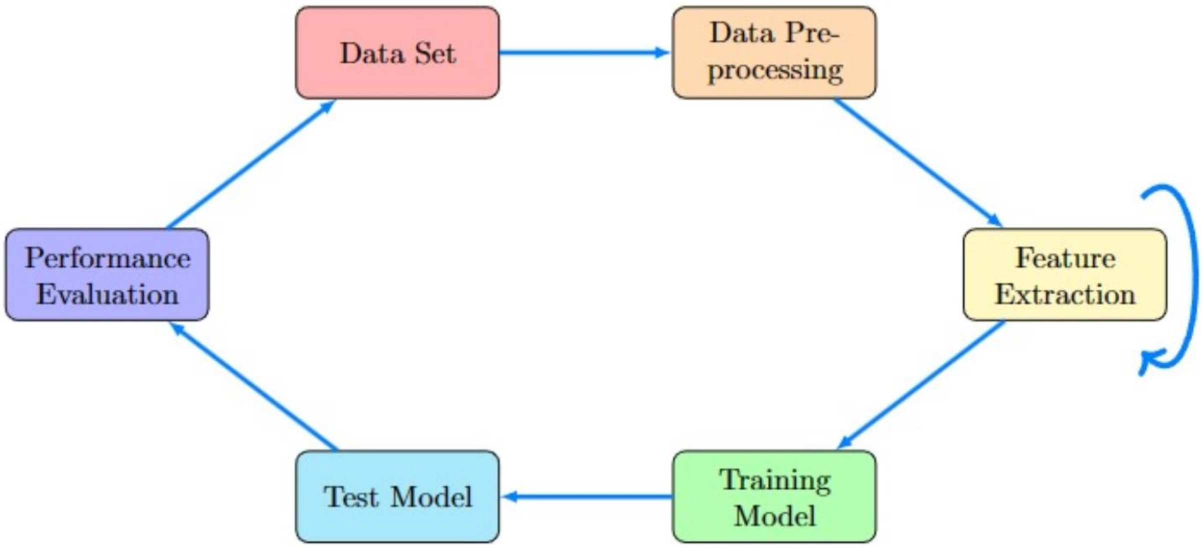

Fig 3.2 UML Diagram

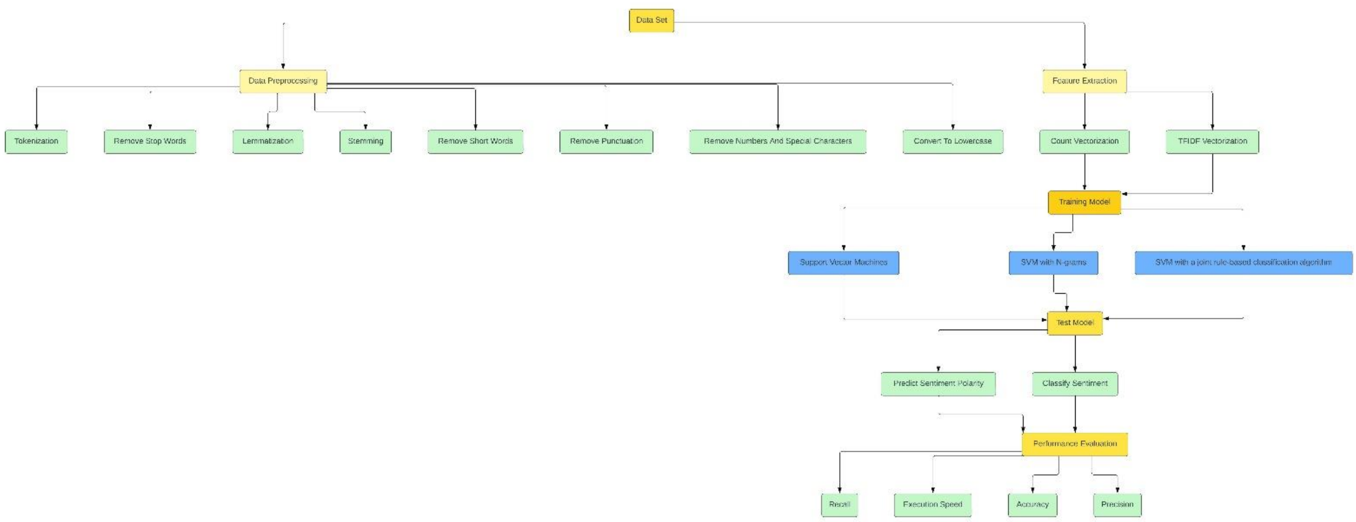

Fig 3.3 System Architecture

3.5 Stepwise Implementation

1. Import necessary libraries: pandas, scikit-learn (CountVectorizer, train_test_split, SVC), imbalanced-learn (SMOTE), and time.

2. Load dataset:

Assuming a dataset named df with columns 'ProcessedReview' and 'ReviewType' indicating sentiment (positive/negative).

3. Split the dataset into training and testing sets:

Use the train_test_split function to split the dataset into X_train, X_test, y_train, and y_test.

4. Vectorization:

- Perform TF-IDF vectorization on text data:

- Initialize TfidfVectorizer and fit_transform on X_train and transform on X_test to convert text into numerical features.

5. Train SVM models:

- Initialize SVM models:

- svm_model: SVM model without n-grams using TF-IDF vectorization.

- svm_model_ngrams: SVM model with n-grams (1, 2) using TF-IDF vectorization.

- Fit SVM models on the training data.

6. Feature Engineering:

- Read positive and negative words from files.

- Create modified sets of positive words with appended negation words for handling sentiment negation.

- Create a CountVectorizer using Bag of Words (BoW) with modified positive and negative words.

7. Model Evaluation:

- Provide options for users to choose between positive and negative reviews.

- Provide options to select a model

- SVM model (without n-grams)

- SVM model with n-grams

- SVM model with conjunctions using Bag of Words

- Evaluate the chosen model's performance on the selected review type:

- Iterate through reviews

- Apply the chosen model to each review.

- Make predictions.

- Calculate performance metrics

- True positives, true negatives, false positives, false negatives

- Accuracy

- Execution time

- Store metrics for each model run.

8. Results Reporting:

- Store performance metrics in a Pandas DataFrame.

- Print the DataFrame to display the results.

3.4.1 N-gram based Approach

The trigram approach favored grouping words, which conveys the opinion of the aspect. The scope of negation could also be taken into account. However, a major drawback found is impracticality in differentiating which opinion words contributed to revealing the sentiment behind the aspect. This approach generates trigrams containing opinion words that do not belong to that aspect. Table 3.1 shows a sample of trigrams generated from some sample user reviews. It can be seen that the generated trigrams provide contradictory information. For instance, the trigrams (very, good, food) and (food, awfully, bad) for the same review are confusing. This may lead to incorrect sentiment score calculation for the feature.

User reviews	Generated trigrams
"The rooms are very good, food awfully bad"	(rooms, very, good), (very, good, food), (food, awfully, bad).
"The amenities are very appealing, but the service is not okay"	(amenities, very, appealing) (very, appealing, service), (service, not, okay).
"Nice, spacious rooms and friendly staff"	(nice, spacious, rooms), (spacious, rooms, friendly), (rooms, friendly, staff).

Fig 3.4 Trigram segregation Approach

N-gram Approach Algorithm:

1. Start: Begin the algorithm.
2. Load Dataset: Load the dataset containing 'ProcessedReview' and 'ReviewType' columns.

3. Data Splitting: Split the dataset into training and testing sets using train_test_split function from scikit-learn, with a test size of 20% and a random state of 42.
4. Vectorization with n-grams:
    - Initialize a TfidfVectorizer with n-grams range of (1, 2).
    - Fit and transform the training data.
    - Transform the test data.
5. SMOTE Resampling:
    - Use SMOTE to oversample the minority class in the training data to address class imbalance.
    - Resample the training features and labels.
6. Train and Evaluate SVM:
    - Train and evaluate an SVM classifier using the resampled training data and the original test data.
    - Within the training_and_evaluate_svm function:
        - Start a timer to measure execution time.
        - Initialize an SVM classifier with a linear kernel.
        - Fit the SVM classifier on the resampled training data.
        - End the timer after training.
        - Make predictions on the test data.
        - Calculate accuracy using the accuracy_score function.
        - Store the classifier name, accuracy, execution time, and the trained model in results dictionary.
        - Return the results.
7. Display SVM Results:
    - Print the classifier name, accuracy, and execution time.
8. Input a Review for Prediction:
    - Prompt the user to input a review for prediction.
9. Vectorize Input Review:
    - Vectorize the input review using the same TfidfVectorizer.
10. Predict Sentiment:
    - Use the trained SVM model to predict the sentiment of the input review.
11. Display Prediction:
    - Print the predicted sentiment (positive/negative).
12. End: End the algorithm.

3.4.2 Conjunction Rule-Based Approach

Unlike trigram, the conjunction rule-based approach segregates clauses from sentences based on conjunction. The sentiment score is calculated on the whole clause altogether, but it is found that most of the time, the returned score is neutral. It occurs that in an attempt to find the sentiment of the entire clause, most words in the clause that do not have any contribution towards the sentiment are also being considered. The other problem seen in conjunction with rule-based sentiment analysis is that the predefined feature set often fails to account for the wide range of a word that suggest the same meaning as it is not feasible to consider and confine all the words relating to a feature. And, thus a lack of suitable mapping of the clause for a particular aspect. Table 2 illustrates the problems conjunction rule-based approach.

User reviews	Clauses
"The food was very tasty and the buffet spread was enormous. The room service was quick and every food items were excellent."	[food very tasty], [buffet spread enormous], [room service quick], [every food items were excellent]
"Absolutely fantastic superb hospitalities shown by all the staffs. Will be staying there only whenever comes to Darjeeling."	[absolutely fantastic superb hospitalities will shown by all the staffs], [Will be staying there only whenever comes to Darjeeling]

Fig 3.5 Conjunction Rule-Based Approach

Conjunction Rule-Based Approach Algorithm:

1. Start: Begin the algorithm.
2. Load Dataset: Load the dataset containing the 'ProcessedReview' and 'ReviewType' columns.
3. Data Splitting: Split the dataset into training and testing sets using the train_test_split function from sci-kit-learn, with a test size of 20% and a random state of 42.
4. Vectorization with n-grams:
    - Initialize a TfidfVectorizer with n-grams range of (1, 2).
    - Fit and transform the training data.
    - Transform the test data.
5. SMOTE Resampling:
    - Use SMOTE to oversample the minority class in the training data to address class imbalance.

- Resample the training features and labels.

6. Conjunction Rule-based Sentiment Analysis:
    - Define a conjunction rule function to predict sentiment based on specific conditions such as the presence of positive or negative keywords.
    - Implement the conjunction rule function to predict sentiment for each review in the test data.
7. Display Conjunction Rule Results:

- Print or display the results of sentiment prediction using the conjunction rule, including the number of positive and negative predictions.

8. Input a Review for Prediction:

- Prompt the user to input a review for prediction.

9. Predict Sentiment using the Conjunction Rule:

- Apply the conjunction rule function to predict the sentiment of the input review.

10. Display Prediction:

- Print the predicted sentiment (positive/negative) based on the conjunction rule.

11. End: End the algorithm.

CHAPTER 4

RESULTS AND DISCUSSION

CHAPTER 4

RESULTS AND DISCUSSION

4.1. Performance metrics

This section focuses on analysing the results obtained from sentiment classification using various methods such as SVM, SVM with n-gram, and SVM with a conjunction rule-based approach. The dataset used for analysis exhibited class imbalance, with a significantly higher number of positive reviews compared to negative ones. The class distribution before applying the SMOTE (Synthetic Minority Over-sampling Technique) approach was imbalanced, with the majority class (positive reviews) having 760,000 instances and the minority class (negative reviews) having only 178,645 instances.

To address this class imbalance and ensure a more robust analysis, researchers applied the SMOTE technique to balance the dataset. After applying SMOTE, the class distribution became balanced, with both positive and negative reviews having an equal number of instances. The class distribution after SMOTE showed that both the positive and negative classes now had 760,000 instances each.

The balancing of the dataset using SMOTE ensures that the classification models are trained on a more representative and balanced dataset, which can lead to more accurate and reliable results. This balanced dataset allows for a fair comparison of the performance of different sentiment classification methods and provides more meaningful insights into the effectiveness of each approach.

Class distribution before SMOTE: Counter ({1: 760000, 0: 178645})

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

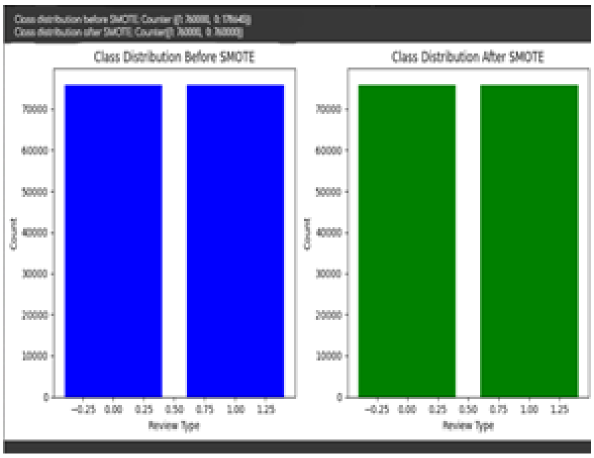

Fig. 4.1 Data Imbalance Graph

Based on the outputs obtained from the execution of Tourist Place Reviews using SVM, SVM with N-grams, and SVM with Conjunction Rule-Based algorithms, as depicted in Figure 4.2, several insights can be derived regarding the performance of each approach.

Figure 4.2 likely displays the outputs generated by each algorithm for the sentiment analysis of tourist place reviews on the 8 lakhs review dataset. These outputs may include the predicted sentiment labels (positive or negative) assigned to each review by the respective algorithms. Upon analyzing the outputs presented in Figure 4.2, several observations can be made regarding the performance of each algorithm:

- SVM: The SVM algorithm likely provides a baseline performance in terms of sentiment classification. The outputs generated by SVM can serve as a reference point for comparing the effectiveness of the other approaches.
- SVM with N-grams: This approach likely incorporates n-grams, capturing sequences of words as features for sentiment analysis. The outputs may demonstrate improved performance compared to the baseline SVM approach, as n-grams can capture more nuanced linguistic patterns.
- SVM with Conjunction Rule-Based: This approach likely integrates conjunction rules alongside SVM to enhance sentiment classification. The outputs may exhibit further performance improvements, particularly in capturing complex linguistic structures and context-dependent sentiments.

	Model	True Positive	True Negative	False Positive	False Negative	Accuracy	Total Positive Reviews	Total Negative Reviews	Total Reviews	Execution Time (s)
0	SVM	656825	0	0	103175	0.864243	760000	0	760000	1739.295545
1	SVM with n-grams	662342	0	0	97658	0.871503	760000	0	760000	2731.811074
2	SVM with conjunctions (Bag of Words)	710399	0	0	49601	0.934736	760000	0	760000	3.181110
3	SVM	0	124016	54629	0	0.694204	0	178645	178645	436.475842
4	SVM with n-grams	0	111138	67507	0	0.622116	0	178645	178645	614.953740
5	SVM with conjunctions (Bag of Words)	0	127357	51288	0	0.712911	0	178645	178645	0.942182

Fig 4.2 Comparison of Algorithms

By examining the outputs displayed in Figure 4.2, researchers can gain insights into the effectiveness of each algorithm in accurately predicting sentiment from tourist place reviews. These outputs can guide further analysis and refinement of the sentiment analysis models to improve their performance and reliability. Additionally, visualizing the outputs allows for easier comparison and interpretation of the results, facilitating informed decision-making in sentiment analysis tasks. From the above output generated the accuracy, precision, recall, and F1-Score are calculated and are as follows.

A. Accuracy Score

Accuracy serves as a straightforward and intuitive metric for evaluating performance, representing the proportion of correctly predicted observations out of all observations. In essence, it quantifies the model's ability to correctly classify instances into their respective classes. Visualizing accuracy through graphs for SVM, SVM with N-grams, and SVM with Conjunction Rule-Based is as shown in Fig 4.3, facilitates a more straightforward interpretation of the model's performance. Utilizing visual representations aids in enhancing data analysis accessibility and comprehension. The formula for accuracy is as follows:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

Visualizing accuracy provides valuable insights into the effectiveness of the classification models in accurately predicting sentiments from the reviews. By comparing accuracy scores across different models, researchers can assess which approach performs better in terms of overall prediction accuracy. This analysis aids in identifying the most suitable method for sentiment classification in the given context and dataset.

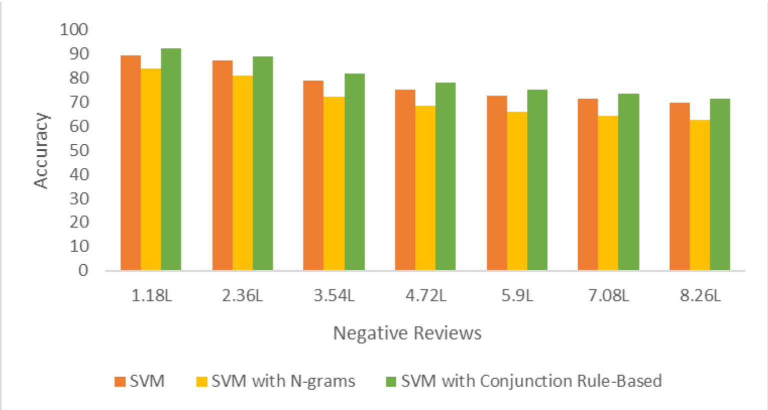

Fig. 4.3 Accuracy comparison of algorithms

B. Precision

Precision, as depicted in the comparison graph Fig. 4.4 for SVM, SVM with N-grams, and SVM with Conjunction Rule-Based algorithms , serves as a valuable visual indicator of performance. Also known as predictive value, precision is a crucial performance metric that emphasizes the proportion of accurately predicted positive instances among all instances predicted as positive. Visualizing precision aids in comprehending the model's ability to correctly identify positive sentiments while minimizing false positives.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives + False Positives}}$$

Where:

- True Positives (TP) are the number of correctly predicted positive instances.

- False Positives (FP) are the number of negative instances incorrectly predicted as positive.

The precision comparison graph provides insights into the accuracy performance of the model by illustrating how well it identifies positive instances without misclassifying negative instances as positive. By examining precision across different models, researchers gain a deeper understanding of each model's effectiveness in accurately predicting positive sentiments from the reviews. This analysis helps in selecting the most suitable approach for sentiment classification based on the precision performance metric and the specific requirements of the task at hand.

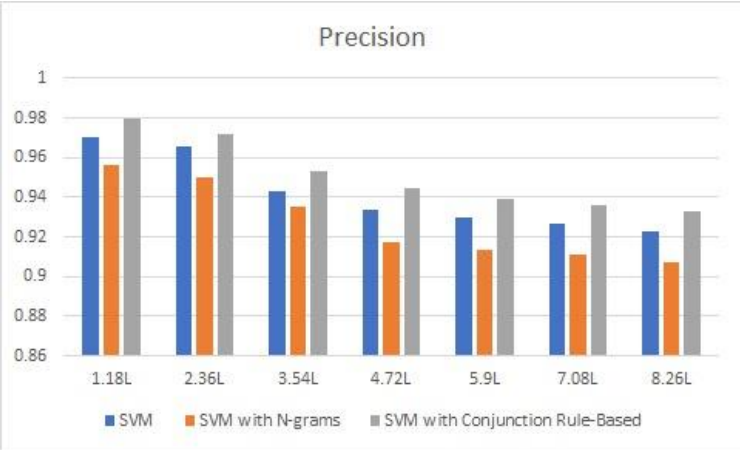

Fig. 4.4 Precision comparison of algorithms

C. Recall

Recall, also known as sensitivity, is a significant performance metric that measures the model's ability to correctly identify all relevant instances of a positive class. It focuses on the proportion of all actual positive observations that were correctly predicted as positive by the model. The recall comparison graph, as depicted in Fig. 4.5 for SVM, SVM with N-grams, and SVM with Conjunction Rule-Based algorithms , provides valuable insights into the model's memory capabilities by showcasing how well it captures positive instances while minimizing false negatives.

The formula for recall is as follows:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

Where:

- True Positives (TP) are the number of correctly predicted positive instances.
- False Negatives (FN) are the number of positive instances incorrectly predicted as negative.

Recall measures the model's ability to recall or retrieve all relevant positive instances, regardless of whether some negative instances are incorrectly classified as positive. It is particularly important in scenarios where missing positive instances is costly, such as in medical diagnosis or anomaly detection. By analyzing recall across different models, researchers can assess each model's effectiveness in capturing positive instances and make informed decisions about model selection and optimization.

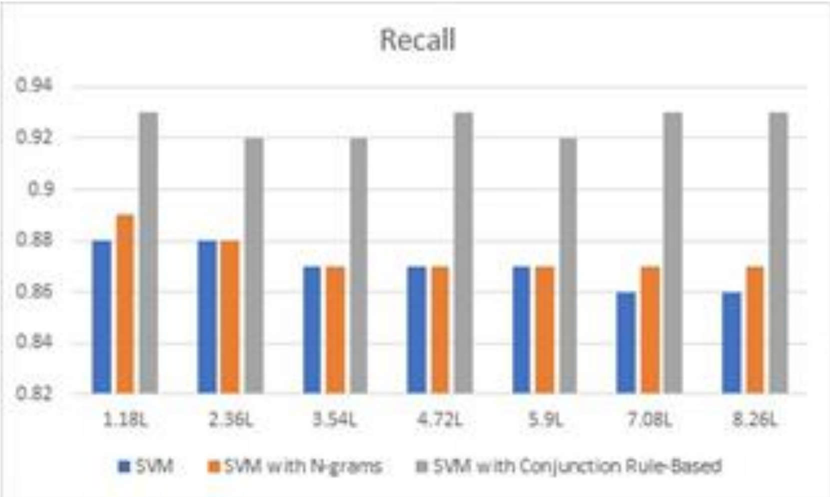

Fig. 4.5 Recall comparison of algorithms

D. F1-Score

The F1 score is a metric that combines both precision and recall into a single value, offering a balanced measure of performance. It considers the trade-off between precision and recall, making it particularly useful for tasks where both false positives and false negatives are equally important. The F1-score comparison graph, presented in Fig. 4.6 for the following algorithms SVM, SVM with N-grams, and SVM with Conjunction Rule-Based, offers insights into the model's overall performance in a balanced manner.

The formula for calculating the F1 score is as follows:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

The F1-score ranges from 0 to 1, with higher values indicating better performance. It is particularly useful when there is an uneven class distribution or when false positives and false negatives have different costs. By visualizing the F1-score through graphs, researchers can gain a comprehensive understanding of the model's overall performance and make informed decisions regarding model selection and optimization.

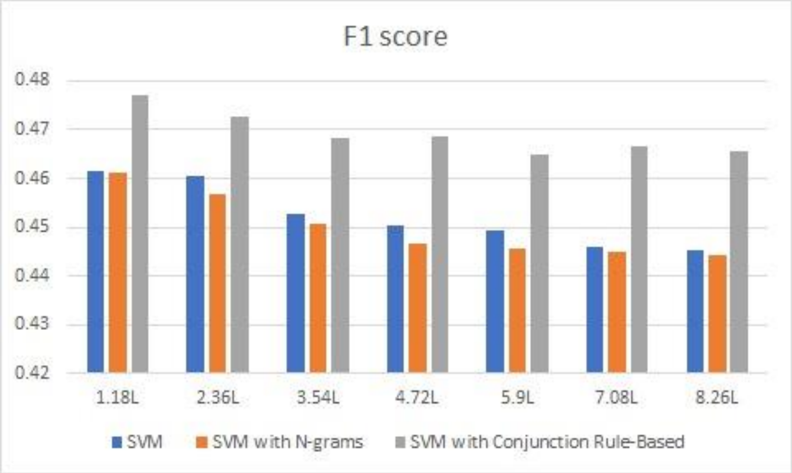

Fig. 4.6 F1-Score comparison of algorithms

CHAPTER 5

CONCLUSION

5.1 CONCLUSION

In this paper, we focus on analyzing user reviews sourced from popular websites covering various tourism destinations. Our primary objective is to discern the sentiments expressed within these reviews. To achieve this, we propose a machine learning model, specifically the Support Vector Machine (SVM) model, to predict the sentiments embedded within tourist place reviews. Additionally, we augment the SVM model with an Enhanced Conjunction rule-based approach to further refine sentiment analysis accuracy. The performances of these classifiers are compared to determine the most suitable classifier to be used in sentiment analysis for tourism reviews on social media. The SVM with Enhanced Conjunction Rule Based model is chosen as the most suitable classifier to be used with the tourism dataset. It outperforms other classifiers with a 71.12% accuracy, while SVM with N-grams and SVM has 62.21% and 69.12% accuracy rates.

Our proposed model serves a dual purpose. Firstly, it aids in providing valuable insights to users seeking information about different tourism destinations. By analyzing sentiments expressed in reviews, our model helps users gauge the suitability of various places for traveling. Secondly, it facilitates the identification of users based on their sentiments and preferences. By understanding the sentiments associated with different destinations, our model enables personalized recommendations tailored to individual user preferences.

Looking ahead, our future work will focus on enhancing the accuracy of our experimental results. We plan to achieve this by leveraging a larger and more diverse dataset. Additionally, we aim to expand our data collection efforts by gathering reviews from an even broader array of websites. By incorporating more comprehensive and varied data, we anticipate further refining our sentiment analysis model and providing even more accurate insights for users.

CHAPTER 6

REFERENCES

6.1 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 (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 places. In 2015 Tenth International Conference on Digital Information Management (ICDIM) (pp. 1-9). IEEE.

[10] Wankhede, R., & Thakare, A. N. (2017, April). Design approach for accuracy in movies reviews using sentiment analysis. In 2017 international conference of electronics, communication and aerospace technology (ICECA) (Vol. 1, pp. 6-11). IEEE.

[11] Wawre, S. V., & Deshmukh, S. N. (2016). Sentiment classification using machine learning techniques. International Journal of Science and RAllahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). A brief survey of text mining: Classification, clustering and extraction techniques. arXiv preprint arXiv:1707.02919. esearch (IJSR), 5(4), 819-821.

[12] Kaur, H., Mangat, V., & Krail, N. (2017). Dictionary-based sentiment analysis of Hinglish text and comparison with machine learning algorithms. International Journal of Metadata, Semantics and Ontologies, 12(2-3), 90-102.

[13] Nally, R. M., & Fleishman, E. (2004). A successful predictive model of species richness based on indicator species. Conservation biology, 18(3), 646-654.

[14] Garau-Vadell, J. B., Díaz-Armas, R., & Gutierrez-Taño, D. (2014). Residents' perceptions of tourism impacts on island destinations: A comparative analysis. International Journal of Tourism Research, 16(6), 578-585.

[15] Nilashi, M., Yadegaridehkordi, E., Ibrahim, O., Samad, S., Ahani, A., & Sanzogni, L. (2019). Analysis of travellers' online reviews in social networking sites using fuzzy logic approach. International Journal of Fuzzy Systems, 21, 1367-1378.

[16] Iqbal, S., Hassan, S. U., Aljohani, N. R., Alelyani, S., Nawaz, R., & Bornmann, L. (2021). A decade of in-text citation analysis based on natural language processing and machine learning techniques: An overview of empirical studies. Scientometrics, 126(8), 6551-6599.

[17] Sarkar, K., & Bhowmick, M. (2017, December). Sentiment polarity detection in bengali tweets using multinomial Naïve Bayes and support vector machines. In 2017 IEEE Calcutta Conference (CALCON) (pp. 31-36). IEEE.

[18] Laksono, R. A., Sungkono, K. R., Sarno, R., & Wahyuni, C. S. (2019, July). Sentiment analysis of restaurant customer reviews on tripadvisor using naïve bayes. In 2019 12th international conference on information & communication technology and system (ICTS) (pp. 49-54). IEEE.

[19] Mirzaalian, F., & Halpenny, E. (2021). Exploring destination loyalty: Application of social media analytics in a nature-based tourism setting. Journal of Destination Marketing & Management, 20, 100598.

[20] Warsito, B., & Prahutama, A. (2020). Sentiment analysis on Tokopedia product online reviews using random forest method. In E3S Web of Conferences (Vol. 202, p. 16006). EDP Sciences.

[21] https://www.tripadvisor.in/

[22] https://www.kaggle.com

CHAPTER 7
APPENDIX

7.1 Code:

```
import pandas as pd
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, classification_report
from imblearn.over_sampling import SMOTE
import time

# Load your dataset (assuming df contains 'ProcessedReview' and 'ReviewType'
columns)
# Example: df = pd.read_csv('your_dataset.csv')

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(df['ProcessedReview'],
df['ReviewType'], test_size=0.2, random_state=42)

# Vectorization with TfidfVectorizer
tfidf_vectorizer = TfidfVectorizer()
X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)
X_test_tfidf = tfidf_vectorizer.transform(X_test)

# SVM model
svm_model = SVC()
svm_model.fit(X_train_tfidf, y_train)

# SVM with n-grams (1,2)
tfidf_vectorizer_ngrams = TfidfVectorizer(ngram_range=(1, 2))
X_train_ngrams = tfidf_vectorizer_ngrams.fit_transform(X_train)
X_test_ngrams = tfidf_vectorizer_ngrams.transform(X_test)

svm_model_ngrams = SVC()
svm_model_ngrams.fit(X_train_ngrams, y_train)

# Read positive words from file
positive_words_file = "positive-words.txt"
with open(positive_words_file, 'r', encoding='utf-8') as file:
    positive_words = set(file.read().splitlines())
```

```
# Append negation words to positive words
negation_words_files = ["not", "no", "never"]  # List of negation word files
negation_words = set()
for negation_word in negation_words_files:
    try:
        with open(f"{negation_word}-positive-words.txt", 'r', encoding='utf-8')
as file:
            negation_words.update(file.read().splitlines())
    except FileNotFoundError:
        print(f"Warning: {negation_word}-positive-words.txt not found.")

modified_positive_words = set()
for word in positive_words:
    modified_positive_words.add(word)
    for negation_word in negation_words:
        modified_positive_words.add(negation_word + " " + word)

# Read negative words from file
negative_words_file = "negative-words.txt"
with open(negative_words_file, 'r', encoding='utf-8') as file:
    negative_words = set(file.read().splitlines())

# Vectorizer with modified positive words and negative words
vectorizer =
CountVectorizer(vocabulary=list(modified_positive_words.union(negative_words)))
X_train_bow = vectorizer.fit_transform(X_train)
X_test_bow = vectorizer.transform(X_test)

# SVM model with conjunctions using Bag of Words
svm_model_conjunctions_bow = SVC()
svm_model_conjunctions_bow.fit(X_train_bow, y_train)

# Define conjunction rule function with negation handling
def conjunction_rule(review):
    negation_detected = False
    for negation_word in negation_words:
        if negation_word in review.lower().split():
            negation_detected = True
            break

    if negation_detected:
```

```
        negative_detected = False
        for word in review.lower().split():
            if word in negative_words:
                negative_detected = True
                break
        return int(negative_detected)
    else:
        positive_detected = False
        for word in review.lower().split():
            if word in positive_words:
                positive_detected = True
                break
        return int(positive_detected)

# Results DataFrame
rows = []
while True:
    print("\nChoose an option:")
    print("1. Positive")
    print("2. Negative")
    print("3. Exit")

    choice = input("Enter your choice (1/2/3): ")

    if choice == '1' or choice == '2':
        if choice == '1':
            reviews_file = '95k-positivereviews.txt'
            review_label = 'positive'
        else:
            reviews_file = '21k-negativereviews.txt'
            review_label = 'negative'

        # Read reviews from the respective file
        with open (reviews_file, 'r', encoding='utf-8') as file:
            reviews = file.readlines()

        num_reviews = len(reviews)

        print('\nSelect a Model:')
        print('1. SVM')
        print('2. SVM with n-grams')
```

```
        print('3. SVM with conjunctions (Bag of Words)')

        model_choice = input('Select a model from this list: ')
        model_name = ''

        true_positive_review = 0
        false_negative_review = 0
        true_negative_review = 0
        false_positive_review = 0

        start_time = time.time()

        for i in range(num_reviews):
            review_input = reviews[i].strip()

            if model_choice == '1':
                model_name = 'SVM'
                review_vec = tfidf_vectorizer.transform([review_input])
                prediction = svm_model.predict(review_vec)
            elif model_choice == '2':
                model_name = 'SVM with n-grams'
                review_vec_ngrams =
tfidf_vectorizer_ngrams.transform([review_input])
                prediction = svm_model_ngrams.predict(review_vec_ngrams)
            elif model_choice == '3':
                model_name = 'SVM with conjunctions (Bag of Words)'
                prediction = conjunction_rule(review_input)

            if choice == '1':
                if prediction == 1:
                    true_positive_review += 1
                else:
                    false_negative_review += 1
            elif choice == '2':
                if prediction == 0:
                    true_negative_review += 1
                else:
                    false_positive_review += 1

        # Calculate accuracy
        total_reviews = num_reviews
```

```
        accuracy = (true_positive_review + true_negative_review) / total_reviews

        # Calculate execution time
        execution_time = time.time() - start_time

        # Append a dictionary to the list for each model after processing all
reviews
        rows.append({'Model': model_name,
                     'True Positive': true_positive_review,
                     'True Negative': true_negative_review,
                     'False Positive': false_positive_review,
                     'False Negative': false_negative_review,
                     'Accuracy': accuracy,
                     'Total Positive Reviews': true_positive_review +
false_negative_review,
                     'Total Negative Reviews': true_negative_review +
false_positive_review,
                     'Total Reviews': total_reviews,
                     'Execution Time (s)': execution_time})

    elif choice == '3':
        print("Exiting the program.")
        break

    else:
        print("Invalid choice. Please enter 1, 2, or 3.")

# Create the final results DataFrame
results_df = pd.DataFrame(rows)

print("\nFinal Results DataFrame:")
print(results_df)
```

8 GITHUB LINK

https://github.com/pooja2918/major-project.git

9 DOI

https://doi.org/10.22214/ijraset.2024.59108

10 PUBLISHED PAPER

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

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

Major Dr. V.A. Narayana[1], Belde Pooja[2], Bachupally Akhil Goud[3]
[1]Professor, [2, 3]UG Student, Department of CSE, CMR College of Engineering and Technology, 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. Decision-making and multiple class classification of reviews are possible with the application of sentiment classification. Sentiment analysis plays a crucial role in helping tourists make informed decisions about their travel destinations. In this particular paper, Using an Enhanced Conjunction Rule-Based Approach and Support Vector Machine (SVM) machine learning technique, the authors conducted sentiment analysis. They collected the dataset from different tourism review websites to train and evaluate their model. The findings of this paper are significant as they not only provide valuable insights in the field of tourism but also help in identifying the most appropriate algorithm for tourism-related analysis. The findings of this paper are significant as they not only provide valuable insights in the field of tourism but also help in identifying the most appropriate algorithm for tourism-related analysis. This information can greatly contribute to the development and improvement of tourism strategies and decision-making processes.

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. In Section III, the methodology for sentiment analysis has also been outlined, along with features like performance evaluation, visualization, and classification of reviews of tourist attractions. In Section IV, they present the results of their experiment. It's great that they have provided a comprehensive overview of their approach and findings in different sections of their paper.

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.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

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 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.

Fang-Zhan [5] even found that implicit neutral emotions were marked as positive. Finding the polarity of a particular target is difficult. For example, even with the same word "rich", "rich service" has a positive meaning, but "rich food" has a negative meaning. In summary, assigning polarity to a specific target may also require some common sense of aspect-oriented polarity calculations. Timor Kadir et al. [6] describe the growing challenge of processing large amounts of unstructured text. This requires effective text-mining techniques and algorithms to uncover meaningful patterns. Text mining is essential for extracting valuable insights from text data, and is especially important in the biomedical and healthcare fields. Rohit Joshi et al. [7] investigate the use of his Twitter data for sentiment prediction by supervised machine learning algorithms. This research focuses on sentiment analysis of film reviews and uses classifiers such as SVM and Naive Bayes for classification. SVM outperforms other classifiers with an impressive 84% accuracy in predicting sentiment in movie reviews. M.D. Devika et al., [8] Describe sentiment analysis as a process of interpreting user emotions, which falls under the domain of Natural Language Processing (NLP). The rise of internet-based applications has led to an increase in personalized reviews to assist travelers and customers in their decision-making process. Sentiment analysis can prove to be an invaluable tool for extracting and summarizing useful insights from overwhelming online reviews. In a study [9], Twitter data about two major international clothing Using Naive Bayes and the Lexicon Dictionary, brands were contrasted and examined. The purpose of this was to find out how the people felt about the two brands. A different document [10] focused on classifying travelers' ratings into four categories: flight comfort, staff service, food and entertainment, and price.

Bayesian and SVM techniques were used to determine passenger fulfillment in these categories. Additionally, opinions on the Indonesian vaccine were collected on Twitter [11], and support vector machines and random forests were used to predict people's sentiments on vaccines. It's interesting to note that Naive Bayes, SVM, and RF are commonly used as machine learning classifiers in sentiment analysis research. There are various opinion mining strategies, as shown in [12]. Trend management, aspect management, set management. The authors proposed not only tourist destinations but also aspect-based opinion mining. Information on related topics was retrieved from visitor reviews. The ratings were then divided into two categories of positive and negative emotions on various topics. For aspect extraction and opinion trends, they use Tagger and WordNet. They extracted tweets based on that. Positive, negative, and neutral classifications are used. Machine learning helps improve system performance. A learning approach was developed. Text mining is also used in biomedicine. As mentioned in [13], the authors have conducted a study on sentiment analysis and the movie examines the dataset. He noted that prior studies have concentrated on maximum entropy, Naive Bayes, and SVM classification techniques. In this paper, the author has categorized reviews using sentiment classification. The most accurate classifier was the RF classifier with a 90% accuracy rate. The authors of this study [14] used multiple features such as Unigram for the movie review dataset Top 2633, Bigram, Unigrams + Bigrams, POS, adjectives for numerous classification methods, and Unigrams and Unigrams + Position maximum entropy. The authors used sentiment analysis. Naive Bayes and he used SVM at all. He made an accurate comparison. Research has shown that Naive Bayes has the lowest level of accuracy, while SVM provides the highest level of accuracy. Similar to [12], there are various opinion mining strategies. B. Trend management, aspect management, set management. Various Researchers from [15] to [20] have worked on tourist place review analysis using various technologies like machine learning, and opinion mining.

III. METHODOLOGY

The act of computationally examining a text to determine people's thoughts, judgments, points of view, feelings, and sentiment polarity (positive, negative, or neutral) toward topics, situations, events, and themes is known as sentiment analysis [6][7]. 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.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

Machine learning techniques including SVM, SVM with N-Gram, and SVM with conjunction rule-based are being used to predict feelings from the tourist reviews dataset to close gaps and enhance the performance and accuracy of the predictions. After that, the reviews' performance is assessed.

Sentiment analysis was done in this article using the procedures listed below. Figure 1 displays the architecture of the system. The following procedures are used in this paper to do sentiment analysis.

A. Data Set

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 Preprocessing

When dealing with social media data, it's crucial to perform data preprocessing to clean up the raw data. This involves several such as lemmatization, stemming, tokenization, and the removal of stop words.

These steps help to refine the data and make it more suitable for sentiment analysis.

1) *Tokenization:* It converts the sequence of reviews into smaller parts which are considered tokens.
2) *Remove Stop Words:* Words that have little to no meaning are known as "stop words." These are the words like 'the', 'a, 'an', 'in', 'of', and 'and'. This study used a custom stop-word list containing words that are unrelated and occur very frequently in the corpus. This reduced the size of the feature vector and improved the performance of the system.
3) *Lemmatization:* It reduces the word to its root form to identify similarities.
   Example: A pile of tokens is converted to a pile.
4) *Stemming:* Discovering the root word of a token is called stemming.
   Example: Token treks will be converted to treks. was performed along with the above steps to improve the performance of the machine learning algorithm. Remove short words, remove punctuation, remove numbers and special characters, and convert to lowercase.

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.

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.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

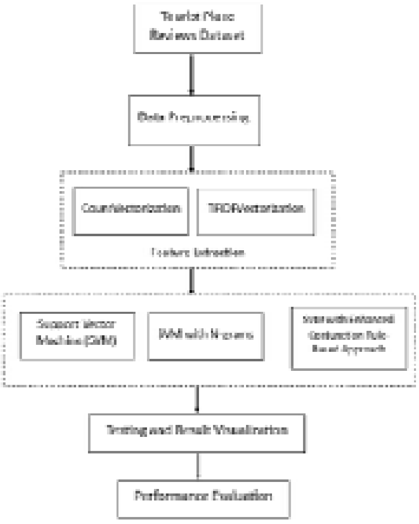

Fig. 1 Data Flow Diagram

In this research, they used to review data from different tourism websites, that was gathered in CSV file format. The information contained review text and related ratings. They classified the reviews as favourable or negative based on these ratings, which indicated the sentiment of the reviews. If the rating was greater than 3, it was considered positive and if it was less than or equal to 3, it was considered negative. It's a clever way to categorize the sentiments expressed in the reviews. The users can log in to the website and can find the tourism data, they have access to search about various places which they want to do analysis. They will get the prediction result based on the data they have searched.

In this research study, the team collected a whopping 800,000 reviews from different tourism websites. Each review was accompanied by corresponding ratings. By analyzing these reviews, they were able to calculate the sentiment expressed in them. It's incredible how they gathered such a large dataset to gain insights into the sentiments of the reviewers.

1) Reviews with ratings higher than 3 indicate a favorable opinion and are assigned a 1
2) If the rating of the review is less than or equal to 3 then the sentiment is negative and marked as 0.

In this way, a labeled dataset for review sentiment analysis was obtained.

Hardware and Software Requirements for Application Development are

a) *Processor Tools:* Intel i5 processor with 32 GB RAM a minimum of 1 TB space on the Hard Disk is needed.
b) *Software Requirements:* The application is developed using Python, with Windows 10 64-bit OS support.
c) *Technologies and Languages:* Python is the primary language used for development.
d) *Dataset Collection:* The data is gathered from Kaggle.

IV. IMPLEMENTATION

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.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

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. 2 Tourist Place Review Sentiment Analysis Example

V. RESULTS AND DISCUSSION

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})

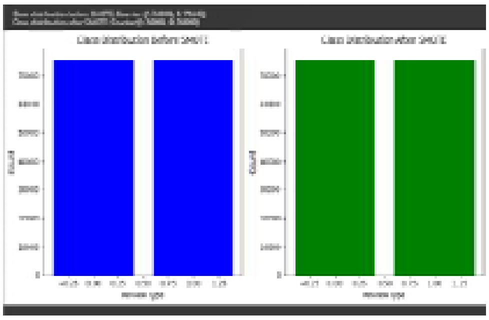

Fig. 3 Data Imbalance Graph

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.

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

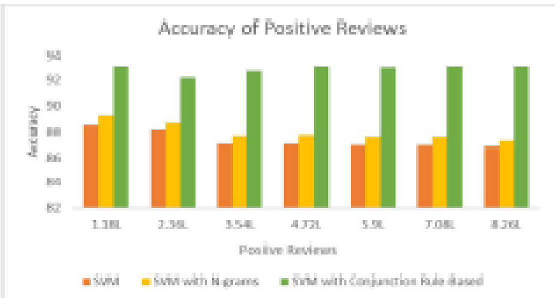

Fig.4 Accuracy of Positive Reviews

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

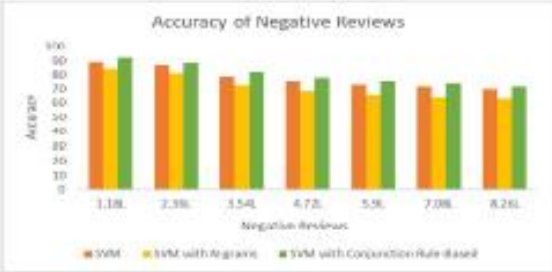

Fig. 5 Accuracy of Negative Reviews

B. Precision

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.

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

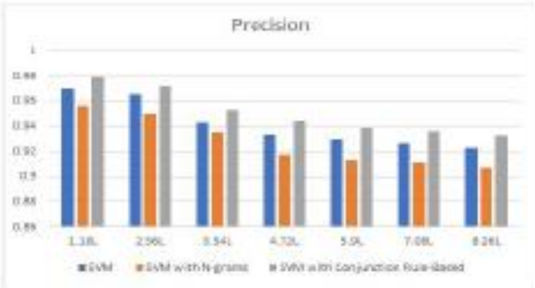

Fig. 6 Precision comparison of algorithms

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.

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

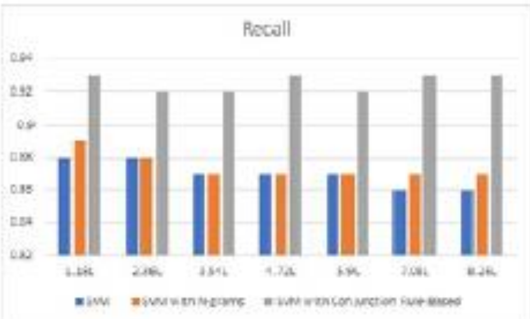

Fig. 7 Recall comparison of algorithms

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

D. F1-Score

The F1-score is a metric that considers both precision and recall, providing a balanced measure of performance. It takes into account the trade-off between precision and recall, making it a valuable evaluation metric. The F1-score comparison graph is provided in Fig. 9. Visualizing the F1 score can help us understand the overall performance of the model in a balanced way.

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

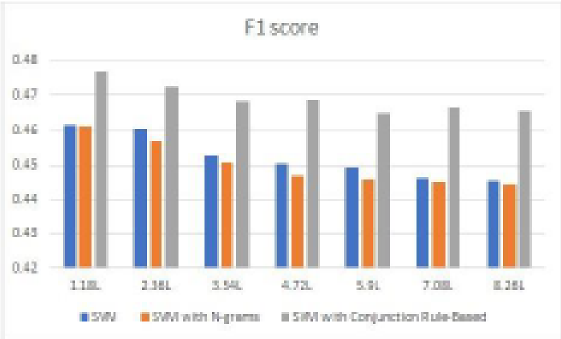

Fig. 8 F1-Score comparison of algorithms

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 approach. The 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.

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 (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 places. In 2015 Tenth International Conference on Digital Information Management (ICDIM) (pp. 1-9). IEEE.

[10] Wankhede, R., & Thakare, A. N. (2017, April). Design approach for accuracy in movies reviews using sentiment analysis. In 2017 international conference of electronics, communication and aerospace technology (ICECA) (Vol. 1, pp. 6-11). IEEE.

[11] Wawre, S. V., & Deshmukh, S. N. (2016). Sentiment classification using machine learning techniques. International Journal of Science and RAllahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). A brief survey of text mining: Classification, clustering and extraction techniques. arXiv preprint arXiv:1707.02919. esearch (IJSR), 5(4), 819-821.

[12] Kaur, H., Mangat, V., & Krail, N. (2017). Dictionary-based sentiment analysis of Hinglish text and comparison with machine learning algorithms. International Journal of Metadata, Semantics and Ontologies, 12(2-3), 90-102.

[13] Nally, R. M., & Fleishman, E. (2004). A successful predictive model of species richness based on indicator species. Conservation biology, 18(3), 646-654.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

[14] Garau-Vadell, J. B., Diaz-Armas, R., & Gutierrez-Taño, D. (2014). Residents' perceptions of tourism impacts on island destinations: A comparative analysis. International Journal of Tourism Research, 16(6), 578-585.
[15] Nilashi, M., Yadegaridehkordi, E., Ibrahim, O., Samad, S., Ahani, A., & Sanzogni, L. (2019). Analysis of travellers' online reviews in social networking sites using fuzzy logic approach. International Journal of Fuzzy Systems, 21, 1367-1378.
[16] Iqbal, S., Hassan, S. U., Aljohani, N. R., Alelyani, S., Nawaz, R., & Bornmann, L. (2021). A decade of in-text citation analysis based on natural language processing and machine learning techniques: An overview of empirical studies. Scientometrics, 126(8), 6551-6599.
[17] Sarkar, K., & Bhowmick, M. (2017, December). Sentiment polarity detection in bengali tweets using multinomial Naïve Bayes and support vector machines. In 2017 IEEE Calcutta Conference (CALCON) (pp. 31-36). IEEE.
[18] Laksono, R. A., Sungkono, K. R., Sarno, R., & Wahyuni, C. S. (2019, July). Sentiment analysis of restaurant customer reviews on tripadvisor using naïve bayes. In 2019 12th international conference on information & communication technology and system (ICTS) (pp. 49-54). IEEE.
[19] Mirzaalian, F., & Halpenny, E. (2021). Exploring destination loyalty: Application of social media analytics in a nature-based tourism setting. Journal of Destination Marketing & Management, 20, 100598.
[20] Warsito, B., & Prahutama, A. (2020). Sentiment analysis on Tokopedia product online reviews using random forest method. In E3S Web of Conferences (Vol. 202, p. 16006). EDP Sciences.
[21] https://www.tripadvisor.in/
[22] https://www.kaggle.com


doi
crossref
10.22214/IJRASET
INDEX COPERNICUS
45.98
ISRA JIF
IMPACT FACTOR:
7.129
TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429
ISI
INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY
Call : 08813907089 (24*7 Support on Whatsapp)

CERTIFICATES:

ISSN No. : 2321-9653

iJRASET

International Journal for Research in Applied Science & Engineering Technology

IJRASET is indexed with Crossref for DOI-DOI : 10.22214

Website : www.ijraset.com, E-mail : ijraset@gmail.com

ISRA Journal Impact Factor: 7.429

45.98 INDEX COPERNICUS

THOMSON REUTERS

TOGETHER WE REACH THE GOAL SJIF 7.429

Certificate

It is here by certified that the paper ID : IJRASET59108, entitled

by

Major Dr. V. A. Narayana

after review is found suitable and has been published in
Volume 12, Issue III, March 2024
in
International Journal for Research in Applied Science & Engineering Technology
(International Peer Reviewed and Refereed Journal)
Good luck for your future endeavors

Editor in Chief, iJRASET

ISSN No. : 2321-9653

iJRASET

International Journal for Research in Applied Science & Engineering Technology

IJRASET is indexed with Crossref for DOI-DOI : 10.22214

Website : www.ijraset.com, E-mail : ijraset@gmail.com

ISRA Journal Impact Factor: 7.429

45.98 INDEX COPERNICUS

THOMSON REUTERS

TOGETHER WE REACH THE GOAL SJIF 7.429

Certificate

It is here by certified that the paper ID : IJRASET59108, entitled

by

Belde Pooja

after review is found suitable and has been published in
Volume 12, Issue III, March 2024
in
International Journal for Research in Applied Science & Engineering Technology
(International Peer Reviewed and Refereed Journal)
Good luck for your future endeavors

Editor in Chief, iJRASET


iJRASET
International Journal for Research in Applied Science & Engineering Technology
IJRASET is indexed with Crossref for DOI-DOI : 10.22214
Website : www.ijraset.com, E-mail : ijraset@gmail.com
ISSN No. : 2321-9653
ISRA Journal Impact Factor: 7.429
45.98 INDEX COPERNICUS
THOMSON REUTERS
TOGETHER WE REACH THE GOAL
SJIF 7.429
Certificate
It is here by certified that the paper ID : IJRASET59108, entitled
by
Bachupally Akhil Goud
after review is found suitable and has been published in
Volume 12, Issue III, March 2024
in
International Journal for Research in Applied Science & Engineering Technology
(International Peer Reviewed and Refereed Journal)
Good luck for your future endeavors
Editor in Chief, iJRASET