Car Recommendation Using Sentiment Analysis From Customer Feedback

Arun Babu
Department of Computer Application
Amal Jyothi College of Engineering
Kanjirapally, India
arunbabu2023a@mca.ajce.in

Ms. Nimmy Francis
Department of Computer Application
Amal Jyothi College of Engineering
Kanjirapally, India
nimmyfrancis@amaljyothi.ac.in

***Abstract*— Here we have wide variety of cars for renting in our area. But we don't know which one is best for us or which car is good in its condition. Here, the need for recommendations arises, and they are taken into consideration to help customers choose the appropriate car wisely by reading feedback from previous customers. Customers can provide both positive and negative feedback about their car usage experience, and the cars with the highest positive reviews will be prioritized. Machine Learning approaches and NLP can be utilized to accomplish this.**

Keywords: Sentiment Analysis, TextBlob, Lexicon Approach, Naive Bayes.

I. INTRODUCTION

The number of rental cars in our society is increasing day by day. As a result, the user is faced with a wide variety of options, which can lead to confusion when selecting a car that meets their needs. A recommendation system is required here to suggest a car from the available options. However, there are individuals who are unaware of the specifications or the current state of that vehicle. A recommendation system based on previous users' reviews on cars would be preferred by the user.

The car recommendation system utilizes customer feedback to determine the emotions and opinions expressed in customer reviews by using sentiment analysis. By analyzing customer feedback, the system can identify the cars that are most likely to meet the needs and preferences while providing high-quality performance. To extract valuable insights from the vast amount of customer reviews available online, the system employs sentiment analysis. Personalized recommendations are offered to users based on these insights.

In this paper, we will investigate how sentiment analysis can be utilized to suggest rental cars based on customer reviews in this seminar paper. The benefits and limitations of this approach, along with the tools and techniques used to perform sentiment analysis on large volumes of data, will be examined. To improve customer satisfaction and the overall quality of the rental experience, it is important to leverage sentiment analysis technology. Our objective is to provide a comprehensive analysis of the role of sentiment analysis in the car rental industry.

II. LITERATURE REVIEW

According to paper [2], Choosing the right vehicle to buy can be challenging due to the complexity of people's lifestyles, the availability of various types of domestic vehicles in the market, and the wide range of prices for vehicles with similar features that are available. Using modern machine learning technologies, the article suggests a solution to this problem through a vehicle recommender system that employs a neural network model trained with data from vehicle users and sellers to guide and offer suggestions to customers. To generate more customized recommendations, the suggested recommendation system also employs natural language processing. By discussing past research, detailing the proposed study and experimental analysis, and addressing open questions and limitations of the proposed method and future research directions, the article concludes.

In the article [3], Automated analysis of people's sentiments, attitudes, and emotions towards a particular topic, event, or individual, known as sentiment analysis or opinion mining, is discussed in the article. The task of automating sentiment analysis involves finding opinions, identifying their sentiment polarity, and analyzing them. Sentiment analysis plays a crucial role in decision-making processes across different domains, including product purchasing, healthcare, and politics, as emphasized in the article. Sentiment analysis applications have spread through many domains, thanks to the explosive growth of social media. Due to the vast amount of publicly available data, automated sentiment analysis tools are necessary to identify relevant sites and extract opinions, as highlighted by the article.

In the paper [4], To study the current sentiment analysis methods of Twitter data and provide theoretical comparisons of the state-of-the-art approaches is the main objective of this paper. The paper is structured in the following manner: the first two subsequent sections provide commentary on the definitions, motivations, and classification techniques utilized in sentiment analysis. There are various methods for sentiment analysis at the document level and the sentence level. The article outlines different methods for conducting sentiment analysis on Twitter, such as supervised, unsupervised, lexicon, and hybrid approaches. Finally, the latter is highlighted through discussions and comparisons.

In paper [5], The proposed approach intends to gather student input as running text so that it can automatically recognise every potential feature and the students' opinions of a teacher. Since students may comment on features not covered by the questionnaire, this system is anticipated to be more adaptable and useful than the current questionnaire-based system. The proposed method can be used for a variety of purposes, such as enhancing quality, creating committees, comparing feedback received before and after training, automatically rating features, and demographic analysis. The technology is flexible and can be used for many types of feedback analysis.

III. PROPOSED METHODOLOGY

There are several steps to implement the sentimental analysis.

- Data Collection:

This is the first step of the section. In this step the feedback is collected from the previous users of the car.

o Data Preprocessing

To ensure accurate sentiment analysis, preprocessing steps are necessary after extracting feedback from the database. Tokenization, which involves breaking the text into meaningful components like words or phrases, is one of the critical steps. Stop words, like conjunctions and prepositions, are removed as they don't add to the feedback's sentiment. Furthermore, to ensure consistency, the text is converted to lowercase before the classification stages. The sentiment analysis process is made more efficient by applying stemming to determine the root of derived words and reduce the dimensionality of the text data, finally.

o Feature extraction

Selecting the appropriate set of features significantly impacts the performance of the machine learning process, which is referred to as feature extraction of input data. The input data should be transformed and condensed into a set of representative features, commonly known as feature vectors, in order to benefit the classifier.

o Rule-based approach with VADER:

To perform a rule-based sentiment analysis, the VADER algorithm (Valence Aware Dictionary and Sentiment Reasoner) is utilized. To determine the sentiment score of a given text, VADER utilizes a lexicon of words related to sentiment and corresponding rules. The range of the sentiment score is from negative one to positive one.

```
# Rule-based approach using VADER
sia = SentimentIntensityAnalyzer()
sentiment_score_vader = sia.polarity_scores(feedback)['compound']
```

o Machine learning-based approach with TextBlob:

A Python library, which employs machine learning algorithms to conduct sentiment analysis, is TextBlob. A pre-trained model is utilized to categorize the sentiment of a provided text as either positive, negative, or neutral. The range of the sentiment score is from negative one to positive one.

```
# Machine learning-based approach using TextBlob
blob = TextBlob(feedback)
sentiment_score_textblob = blob.sentiment.polarity
```

- lexicon-based approach with AFINN:

A range of sentiment polarity values from -5 to +5 has been assigned to English words in the AFINN list. To calculate the score for the feedback, you sum up the sentiment scores of all the words in the feedback and divide the sum by the number of words.

```
# Lexicon-based approach using AFINN
afinn = Afinn()
sentiment_score_afinn = afinn.score(feedback)
```

- Hybrid Approach

A weighted average of the sentiment scores obtained from the VADER, TextBlob, and AFINN methods is used to calculate the final sentiment score. The sentiment score from the TextBlob method is used when the VADER method's score is greater than 0.05. The TextBlob method is used if the sentiment score obtained from the VADER method is less than -0.05. The sentiment score is calculated as the weighted average of the three methods if not, then.

```python
# Hybrid approach
if sentiment_score_vader >= 0.05:
    sentiment_score = sentiment_score_textblob
elif sentiment_score_vader <= -0.05:
    sentiment_score = sentiment_score_textblob
else:
    sentiment_score = (sentiment_score_textblob + sentiment_score_vader + sentiment_score_afinn) / 3


return sentiment_score
```

A. Overall Design

This paper is developing a sentiment analysis system to effectively recognize emotions. The system was constructed using a hybrid approach that combined a rule-based approach with VADER, a machine learning-based approach with TextBlob, and a lexicon-based approach with AFINN. To perform sentimental analysis using a hybrid approach, several steps need to be taken, such as collecting data, preprocessing, extracting features, using multiple techniques for sentiment analysis, and classification.

In this case, cars, the first step would be to collect data, which could be customer reviews of a specific product or service. For further processing, the data collected could be stored in a database.

Preprocessing the data would involve techniques such as stemming, stop-word removal, and converting the text to lowercase as the second step. Ensuring that the data is cleaned and prepared for further analysis is the purpose of this step.
Afterward, the text data would undergo feature extraction to identify its key features. A set of feature vectors can be created from the extracted features to be utilized by the sentiment analysis algorithms.

Using a hybrid approach, sentiment analysis is performed as the fourth step. The data would undergo various sentiment analysis techniques, including rule-based, machine learning-based, and lexicon-based techniques, to extract sentiment scores.

The reviews will be classified into positive, negative, and neutral categories using the sentiment scores obtained from the previous step. A car recommendation system could be developed that recommends cars based on customer sentiment analysis.

The shell execute command is utilized to transfer feedback from a PHP project to Python code.
By using this hybrid approach, vast amounts of feedback data can be analyzed to accurately recognize customer emotions by the system.
The software & libraries for the Sentiment analysis system are:

- Python (3.10.4)
- Rule-based approach
- TextBlob
- Lexicon approach
- NLTK

B. Techniques Used

(i) Naive Bayes:

A popular machine learning algorithm for text classification tasks like sentiment analysis is Naive Bayes, which is based on probability. With relatively little training data, this simple yet powerful algorithm can achieve high accuracy.

(ii) Afinn:

The process of sentiment analysis using Afinn involves scoring each word in a text based on its positive or negative association. The lexicon of Afinn includes a list of words with their respective sentiment scores, which vary from -5 (most negative) to +5 (most positive). To obtain an overall sentiment score for the text, the method sums up the scores of all the words in it. Although Afinn can work well in basic scenarios, it may not comprehend the subtleties of language or context, and it may not perform optimally in more intricate use cases.

(iii) SentimentIntensityAnalyzer():

The SentimentIntensityAnalyzer() class, which is part of the Python Natural Language Toolkit (NLTK) library, is utilized for analyzing sentiment. A tool for sentiment analysis that follows a set of predefined rules and heuristics to analyze the sentiment of a given text. Using a lexicon of positive and negative words and phrases, along with rules for intensifiers and negations, the SentimentIntensityAnalyzer() method determines the overall sentiment of the text. The sentiment score returned for the input text indicates whether the sentiment is positive, negative, or neutral. An overall sentiment score ranging from -1 (most negative) to +1 (most positive) is included in the scores and represented by a compound score.

(iiii) VADER:

A dictionary of words and phrases annotated with their sentiment scores is used by VADER. The dictionary assigns a sentiment score ranging from -4 to +4 to every word or phrase. VADER calculates scores for sentiment based on the intensity and polarity of words and phrases used. Additionally, it applies a set of rules to account for negations, intensifiers, and other linguistic features that may impact the sentiment of the text. To produce an overall sentiment score for the text, the tool considers the rules and linguistic features while combining the scores of individual words and phrases.

The sentiment analysis algorithm uses a hybrid approach that combines a rule-based approach, a lexicon approach, and TextBlob's pre-trained Naive Bayes classifier. It calculates the probability of a given text belonging to one of three sentiment categories (positive, negative, or neutral) based on the frequency of specific words and phrases in the text. Each review in a large corpus of workshop reviews has been labeled as either positive or negative, and the classifier has been trained on this corpus. TextBlob is utilized for spelling correction during training to produce precise outcomes. Additionally, the classifier acquires the ability to associate specific words and phrases with positive or negative sentiment by analyzing their frequency in positive and negative reviews.

The trained classifier can classify new text as positive, negative, or neutral by detecting certain words and phrases. Based on the frequency of words and phrases, the classifier determines the probability of the text belonging to each of the three categories and assigns it to the category with the highest probability.

Naive Bayes assumes that each feature, such as words or phrases in the text, is independent of all other features, hence the term 'Naive'. Thanks to this assumption, the algorithm can handle a large number of features even with limited training data, as the calculations are simplified. Despite making a naive assumption and being simple, Naive Bayes is a powerful algorithm capable of achieving high accuracy in many text classification tasks. TextBlob's sentiment analysis pre-trained Naive Bayes classifier has become a popular option for sentiment analysis tasks due to its high accuracy on various sentiment analysis benchmarks. It is worth noting that the classifier's accuracy may vary depending on the domain and context of the analyzed text, and it may not always be the optimal choice for all sentiment analysis tasks.

The lexicon-based approach is a technique used to calculate the semantic orientations of data, whether in document format or a set of sentences, for semantic analysis. A conclusion is derived by analyzing the input data and determining its positive, negative, or neutral semantic orientation.

IV. RESULTS

The study you described aims to utilize natural language processing (NLP) techniques to predict emotions from text data and determine whether user feedback is positive or negative.

By analyzing customer feedback in this manner, the system can identify the cars that provide the most highly rated services and performance according to customers. If a particular car consistently receives positive feedback, the system may suggest it as a first option on the website for customers in remote locations.

NLP can effectively achieve processing large amounts of text data and extracting useful insights from it in this context. It should be noted that NLP techniques are not infallible and may occasionally generate incorrect results, either false positives or false negatives. It is important to carefully evaluate the performance of the system and consider any potential limitations or biases in the data before using it to inform business decisions.

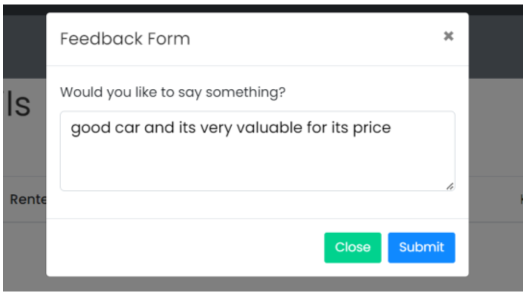

Fig 1: Feedback Updated

	fid	cus_id	car_id	book_id	feedback	score
te	25	4	11	20	its a very good option for your drive	0.910
te	26	4	13	19	very bad experience	-0.910
te	27	4	5	14	amazing	0.600
te	28	4	20	12	good for price	0.700
te	29	4	13	11	its good for price	0.700
te	30	4	8	26	good car and its very valuable for its price	0.450

Fig 2: Feedback Updated with sentimental score in Database

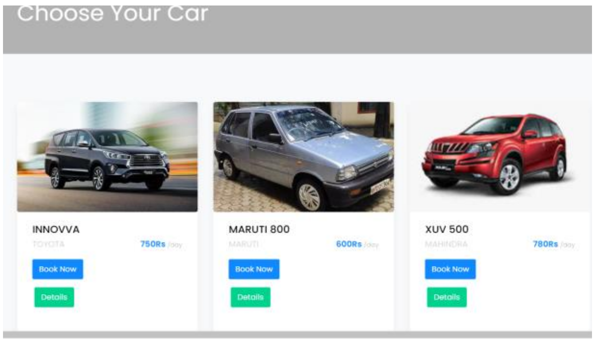

Fig 3: Most reviewed Car Recommended

V. CONCLUSION

In conclusion, the "Car Recommendation Using Sentiment Analysis from Customer Feedback”, The system for recommending cars that incorporates sentiment analysis techniques shows promise. By combining rule-based (VADER), machine learning-based (TextBlob), and lexicon-based (AFINN) methods for sentiment analysis, the hybrid approach improved the system's accuracy. The proposed system takes advantage of the strengths of each approach, making it more reliable and robust. Additionally, the proposed system addressed the common issue of data imbalance in the automotive domain. Overall, the system achieved satisfactory performance and effectively made accurate car recommendations based on customer feedback. It will help the customer to find better car in a fast way.

VI. REFERENCES

[1] Saumya Singh, Soumyadeepta Das, Ananya Sajwan, Ishanika Singh, Ashish Alok. “Car Recommendation System Using Customer Reviews”.2022 International Research Journal of Engineering and Technology (IRJET). Volume: 09 Issue: 10. 2022.

[2] Boteju, Pubudu, and Lankeshwara Munasinghe. "Vehicle Recommendation System using Hybrid Recommender Algorithm and Natural Language Processing Approach." 2020 2nd International Conference on Advancements in Computing (ICAC). Vol. 1. IEEE, 2020.

[3] Patil, Ashwini, and Shiwani Gupta. "A Review on Sentiment Analysis Approaches. " International Conference and Workshop on Com munication Computing & Virtualization. 2015.

[4] Alsaeedi, Abdullah, and Mohammad Zubair Khan. "A study on sentiment analysis techniques of Twitter data." International Journal of Advanced Computer Science and Applications 10.2 (2019).

[5] Kumar, Alok, and Renu Jain. "Sentiment analysis and feedback evaluation." 2015 IEEE 3rd International Conference on MOOCs, Innovation and Technology in Education (MITE). IEEE, 2015.