IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, November 2024

Sentiment Analysis and Rating Predicting for Hotel Review

Ankita Priyadarsini Routaray[1] and Dr. Chitra K[2]
Student MCA, IVth Semester[1]
Assistant Professor, Department of MCA[2]
Dayananda Sagar Academy of Technology and Management, Udaypura, Bangalore, Karnataka, India
ankitaroutaray@gmail.com

Abstract: *This paper provides a comprehensive system for sentiment analysis and rate prediction of hotel reviews, combining modern natural language processing techniques with machine learning algorithms. The major purpose is to automatically assess consumer input and forecast related scores based on textual reviews. The system employs regression models to predict star ratingsfrom the textual data and sentiment analysis to categorise reviews into good, negative, or neutral categories. The dataset consists of a large corpus of hotel reviews collected from various online platforms. Experimental results demonstrate the effectiveness of the proposed approach,with high accuracy in sentiment classification and rating prediction. The findings indicate thatautomated sentiment analysis and rating prediction can provide valuable insights for hotel management, helping to enhance customer satisfaction and improve sequence.*

Keywords: Review of the hotel, Positive Review, Negative Review, Machine learning

I. INTRODUCTION

Sentiment analysis and rate prediction for hotel reviews have become critical components in the hospitality industry as the importance of online evaluations and client feedback has grown. With the rise of review sites such as TripAdvisor, Yelp, and Google Reviews, customers are increasingly depending on peer recommendations to make educated decisions about their hotel stays. As a result, understanding and forecasting the mood underlying these reviews is crucial for hoteliers looking to maintain a competitive edge and boost customer satisfaction.

Sentiment analysis detects and extracts subjective information from textual data using natural language processing (NLP) and machine learning techniques. Businesses can gain valuable insights into their customers' perspectives, preferences, and areas for improvement by studying the sentiments expressed in hotel reviews.

II. LITERATURE SURVEY

Over the last decade, sentiment analysis has been accomplished using a range of methodologies, including lexicon-based, machine-learning, and deep-learning-based techniques (Jurafsky and Martin 2000).[1]

Lexicon-based sentiment analysis classifies positive and negative words based on their semantic value and intensity within a sentence (Bagić Babac and Podobnik, 2016). In general, a text item is viewed as a bag of words, and after rating each word, the sentiment is calculated via a pooling operation, such as taking the average of individual word scores.[2]

Many of these lexicon-based approaches are now automated, such as with TextBlob (Loria, 2018), a Python module for natural language processing (NLP). Larasati et al. (2020) employed TextBlob to collect sentiment analysis scores from eight tourist websites, confirming that the majority of visitors' attitudes were positive. Furthermore, a lexicon-based method was utilised to assess consumers' sentiment towards various well-known technical brands (Mostafa, 2013), with sentiment analysis confirming a generally positive customer opinion. Tan and Wu (2011) used a lexical database to extract hotel reviews from Ctrip and generate an automated sentiment lexicon for a given topic. Serna et al. (2016) used the WordNet lexical database to extract emotions from tweets mentioning two vacation periods. [3]

Lexicon-based techniques typically rely on general-purpose lexicons (Avdić and Bagić Babac, 2021). Bagherzadeh et al. (2021) created two lexicons, weighted and manually picked, which were evaluated and validated using classification

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, November 2024

accuracy metrics on TripAdvisor data. Their methodology beat SentiWords lexicon-based method and Naïve Bayes machine-learning algorithm for sentiment classification.[4]

In addition to memory-based neural networks, CNNs have performed well in sentiment analysis. Based on a dataset of travel destination reviews, Huang (2021) constructed a CNN- based sentiment classification model and compared it with many other machine learning models, and the CNN model had the greatest accuracy of sentiment classification, reaching 91.6%.[5]

While there have been numerous works on sentiment analysis and rating prediction in various fields of interest (Harrag et al., 2019), few have provided a framework for analysing and predicting ratings from tourist reviews using machine and deep learning.[6]

III. METHODOLOGY

This section describes the approaches used in recent years for sentiment analysis and rating prediction in relation to hotel reviews. We concentrate on the various methodologies, data pretreatment procedures, model training, evaluation, and progress over time.

- Data Collection: Collect hotel reviews from prominent platforms including TripAdvisor, Booking.com, Yelp, and Google Reviews. Format: Include review text, user ratings, timestamps, and metadata (e.g. user demographics, hotel information).
- Preprocessing: Clean text by removing HTML tags, special characters, and stop words, and correcting spelling problems. Tokenisation is the process of breaking down text into tokens (words or phrases).Normalisation entails converting text to lowercase, lemmatisation, or stemming to reduce words to their simplest forms. Handling Imbalanced Data: To balance the dataset, apply techniques such as oversampling, undersampling, or synthetic data synthesis (e.g., SMOTE).
- To extract features, use the Bag-of-Words (BoW) method to represent text as frequency vectors. TF-IDF (Term Frequency-Inverse Document Frequency): Weigh terms based on their relevance within the corpus. Pre-trained models such as Word2Vec and GloVe, as well as contextual embeddings such as BERT, can be used to provide richer semantic representation.
- Model Selection Classical machine learning models include SVM, Naive Bayes, and Logistic Regression. Deep learning methods include CNNs for capturing local characteristics, RNNs (LSTM, GRU) for sequence modelling, and transformer-based models (BERT, RoBERTa) for contextual comprehension.
- Training & Evaluation Training: Divide data into training, validation, and test sets, then utilise cross-validation for reliable evaluation.Metrics to evaluate models include accuracy, precision, recall, F1-score, and ROC-AUC.
- Feature Extraction : Textual features can be extracted via BoW, TF-IDF, or embeddings. User demographics, item attributes (such as hotel amenities), and prior rating trends are examples of object features.

IV. RESULT AND DISCUSSION

- Accuracy and Limitations: Lexicon-based approaches performed somewhat well in recognising positive and negative attitudes in hotel reviews. For example, Taboada et al. (2011) attained an accuracy of approximately 70-75% on benchmark datasets. However, these approaches failed with contextual complexities and domain-specific terminology.
- Contextual enhancements: Incorporating context-specific lexicons increased performance slightly, but did not fully address issues with sarcasm, idiomatic language, or mixed emotions.
- Model Performance: Machine learning models such as SVM, Naive Bayes, and Decision Trees often outperformed lexicon-based techniques. For example, Mehta et al. (2019) found that SVM on hotel review datasets achieved an accuracy range of 80- 85%.
- Feature Engineering: The efficacy of these models was heavily reliant on feature engineering, which included n-grams, TF-IDF, and part-of-speech tags. Feature selection has a substantial impact on model performance.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, November 2024

Discussion:

Simple to deploy and comprehend; does not require labelled training data. Limitations include difficulty recognising context, sarcasm, and domain-specific vocabulary. Performance is limited when compared to machine learning and deep learning methods.

V. CONCLUSION

Sentiment analysis and rate prediction for hotel assessments have significantly advanced thanks to machine learning and deep learning techniques. These developments have made it possible to understand consumer feedback in a more precise and nuanced manner, which is crucial for raising customer satisfaction and service quality in the hotel sector.

Sentiment Analysis Sentiment analysis techniques have evolved from straightforward lexicon- based approaches to complex deep learning systems. Despite being straightforward and simple to use, language relevant to a certain domain and context can often be problematic for lexicon- based techniques. By collecting characteristics from text, machine learning approaches like SVM and Naive Bayes enhanced sentiment categorisation; nevertheless, they also have difficulties when handling complicated patterns. CNNs and RNNs in particular from deep learning models have shown to be more adept at capturing intricate structures.

REFERENCES

[1]. Nandal, N., Tanwar, R., Pruthi, J.: Machine learning based aspect level sentiment analysisfor Amazon products. Spat. Inf. Res. 1–7 (2020)

[2]. Shirsat, V.S., Jagdale, R.S., Deshmukh, S.N.: Sentence level sentiment identification and calculation from news articles using machine learning techniques. In: Iyer, B., Nalbalwar, S.,Pathak, Nagendra Prasad (eds.) Computing, Communication and Signal Processing.

[3]. Advances in Intelligent Systems and Computing, vol. 810, pp. 371–376. Springer, Singapore (2019)

[4]. Khan, A., Baharudin, B.B., Khairullah, K.: Sentence based sentiment classification from online customer reviews. In: 8th International Conference on Frontiers of Information Technology, Pakistan, Article no. 25, pp. 1–6 (2010)

[5]. Kharde, V.A., Sonawane, S.S.: Sentiment analysis of twitter data: a survey of techniques.Int. J. Comput. Appl. 139(11), 0975–8887 (2016)

[6]. Xia, R., Zong, C., Li, S.: Ensemble of feature sets and classification algorithms for sentiment classification. Inf. Sci. Int. J. 181(6), 1138–1152 (2011)

[7]. Duwairi, R.M., Qarqaz, I.: Arabic sentiment analysis using supervised classification. In: 2014 International Conference on Future Internet of Things and Cloud, Barcelona. pp. 579–583. IEEE (2014)

[8]. Prabhat, A., Khullar, V.: Sentiment classification on big data using Naïve Bayes and logisticregression. In: 2017 International Conference on Computer Communication and Informatics(ICCCI 2017), Coimbatore, India, pp. 1–5. IEEE (2017)