Technoarete Transactions on Advances in Computer Applications (TTACA)
Vol-2, Issue-1, March 2023
e-ISSN: 2583-3472

Sentiment Analysis Web App Using NLP

S.Venkata Lakshmi [1*], N.Geetha [2]

[1] Assistant Professor, CSE Department, Annamacharya Institute of Technology and Sciences, Tirupati, India
[2] Assistant Professor, CSE Department, KMM Institute of Technology and Sciences, Tirupati, India
*Corresponding Author Email: sibbalalakshmiaits@gmail.com

Abstract

Sentiment analysis is a crucial tool for understanding people's emotions, opinions, and attitudes towards products, services, and events, particularly in the business world. This paper presents a comprehensive study of sentiment analysis web applications, exploring the techniques and tools used for sentiment analysis. The study compares various web applications based on their accuracy, efficiency, and user-friendliness. The challenges and limitations of sentiment analysis, such as language and cultural barriers, are also discussed, and potential solutions are presented. The paper also describes a machine learning and natural language processing-based web application that has been developed for sentiment analysis. The application has been evaluated using a real-time dataset and has shown promising results. Finally, the paper presents the findings, highlighting the importance of sentiment analysis in the digital age, and suggests future research directions.

Keywords

Accuracy, Efficiency, Emotion, Natural Language Processing, Sentiment Analysis, Web Application.

INTRODUCTION

Sentiment analysis, also known as opinion mining, is a rapidly growing field that involves analyzing text data to determine the sentiment of a given text. This information can be used to understand people's emotions, opinions, and attitudes towards a variety of topics, including products, services, and events. With the explosion of social media and online reviews, sentiment analysis has become an essential tool for businesses, governments, and individuals to understand public opinion and make informed decisions. Web applications have become a popular way to perform sentiment analysis, as they offer a convenient and user-friendly way to interact with text data.

These applications often use machine learning and natural language processing techniques to analyze text and provide insights on sentiment. However, there are many challenges to developing accurate and effective sentiment analysis web applications, including language and cultural barriers, data bias, and the need for real-time analysis. Sentiment analysis is an exciting field of study that involves analyzing text data to determine the sentiment, attitude, and emotions behind the text.

The rise of social media, online reviews, and other forms of digital communication have made sentiment analysis increasingly important for individuals, businesses, and governments. The insights gained from sentiment analysis can be used to improve products and services, monitor brand reputation, and predict customer behavior. This paper aims to explore the various techniques and tools used for sentiment analysis in web applications. We will compare the accuracy, efficiency, and user- friendliness of several web applications and discuss the challenges of sentiment analysis. To address these challenges, we will present a machine learning and natural language processing-based web application for sentiment analysis.

Sentiment analysis can be defined as a classification task performed on text data with a significant pre-processing to classify it into two (binary classification) or many (multiple class classification) classes. The Sentiment analysis is a computational operation of analysing the sentiment in the text(information). It is a powerful machine learning application of the classification of text data into different classes. It is used in analysing the customer reviews, finding the sentiment, or the polarity of the information.

Motivation for Sentiment analysis:

In today's world, every person is connected with Social media and there is a lot of data created every day in the form of tweets, posts, messages, announcements, and news. All this data is a raw text which can be extracted or collected in various ways for analysing the content.

General workflow of Sentiment analysis:

Every Analytic task starts with the task of collecting data. Nowadays, many social media platforms provide an open and easy way to access data for research and analytics. Real-time Twitter tweets can be extracted for analysis with Twitter developer account and tweepy library in python. F

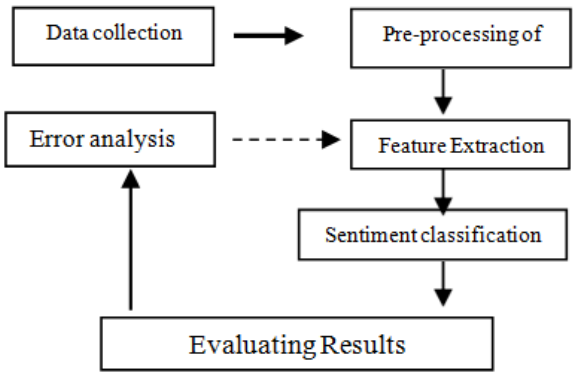

Figure 1. The Flow of Sentiment analysis

e-ISSN: 2583-3472

RELATED WORK

The field of sentiment analysis has garnered significant attention in recent years, as the explosion of social media and other digital communication channels have made it an essential tool for businesses, governments, and individuals. Researchers have explored various approaches to sentiment analysis, including machine learning and natural language processing techniques.

In one study [1], the authors proposed a machine learning-based framework for sentiment analysis of Twitter data. The framework used supervised learning techniques to classify the sentiment of tweets and achieved an accuracy of 85% on a large Twitter dataset.

In another study [2], the authors employed deep learning techniques to perform sentiment analysis on product reviews. Their neural network-based approach utilized word embeddings and achieved an accuracy of 90%. Corpus is the collection of data. Text corpus is used most commonly in Annotations. An example for this annotation is parts-of-speech taggers (POS taggers). Taggers are commonly used in many applications and blogs for the allocation of the data. The most commonly used websites for the taggers include Instagram, Facebook, and many social websites.

There has also been research into sentiment analysis for languages other than English, such as the study by, which focused on sentiment analysis of Arabic news articles. The authors [3] used a hybrid approach combining lexicon-based and machine learning techniques and achieved an accuracy of 80%. While these studies offer valuable insights, they have limitations in terms of language and cultural context, as well as dataset size. Our paper presents a machine learning and natural language processing-based web application for sentiment analysis that has been evaluated using a real-time dataset. We aim to compare our approach with other sentiment analysis web applications and propose potential solutions to the challenges of sentiment analysis.

The Machine learning strategies [4] have speed up the once gruelling data crunching tasks. It works by training an algorithm with a training data set developed from examples and experiences. This means the algorithm learns automatically from the data that is provided and this is called the training phase of the algorithm followed by a testing phase where we feed some test data to the algorithm to get the expected outcome. The approach can be further categorised as Supervised and unsupervised.

ANALYSIS

A sentiment analysis web application is a software tool that allows users to input text data and receive an analysis of the sentiment expressed within that text. These web applications use algorithms and machine learning techniques to classify text into different categories of sentiment, such as positive, negative, or neutral. Users can then use the results of the sentiment analysis to make data-driven decisions or gain insights into their data.

Steps in Sentiment analysis tasks:

i. Data collection
ii. Pre-processing of data
iii. Feature extraction and Selection
iv. Sentiment classification
v. Evaluating Results and Error analysis

One potential challenge with a sentiment analysis web application is the need for constant updates and maintenance. As language and context evolve over time, the algorithms and training data used by the application may become outdated, resulting in a decrease in accuracy. Therefore, it is important for the developers of the application to keep up with the latest advancements in natural language processing and machine learning to ensure that the application remains up-to-date and effective [4].

Natural Language Processing

Sentiment analysis web applications use a variety of NLP techniques to accurately classify text into different categories of sentiment. One such technique is part-of-speech (POS) tagging, which involves labelling each word in a text with its corresponding part of speech, such as noun, verb, adjective, or adverb. This information is used by the sentiment analysis algorithm to identify the most important words and phrases in the text and to understand their relationship to each other.

Sentence level analysis aims at identifying the polarity of a sentence thus is widely applied to microblogging posts. This kind of analysis segregates objective sentences having concrete information from the subjective ones having subjective views and opinion. The whole "emotional" document is under scrutiny in Document-level analysis and the overall polarity.

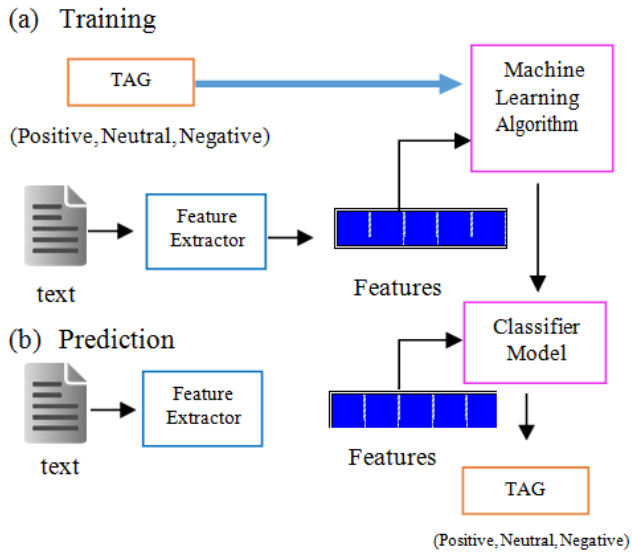

Figure 2. Natural Language Processing Work Flow

Lexicon based Dictionary Approach

In Lexicon-based Dictionary approach, the algorithm analyzes the text by counting the number of positive and negative words in the text. The algorithm then calculates the overall sentiment score of the text by subtracting the number

e-ISSN: 2583-3472

of negative words from the number of positive words. The resulting score provides an estimate of the sentiment expressed in the text.

Figure 3. Lexicon based Dictionary Approach

Machine Learning Approach

The machine learning approach to sentiment analysis involves several steps, including a labelled dataset divided into training and testing sets. Text data is pre- processed by removing stop words and feature extraction is done using techniques such as bag-of- words or TF-IDF. The algorithm is trained on the labelled training data using a classification algorithm such as logistic regression, support vector machines, or neural networks, and learns to associate certain features with positive, negative, or neutral sentiment to classify new text data.

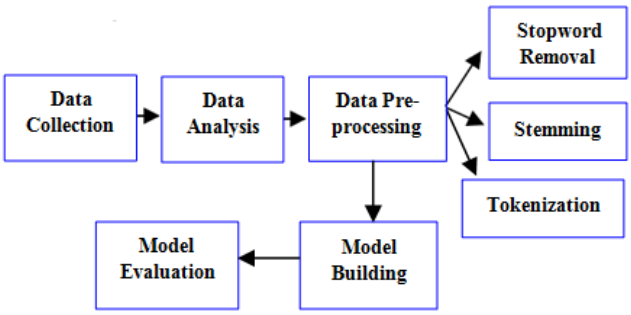

Figure 4. Machine Learning Approach

Hybrid Approach:

The hybrid approach begins by pre-processing the text data using techniques such as removing stop words and feature extraction. Next, the text data is fed into a rule-based system that uses predefined rules to classify the text as positive, negative, or neutral. These rules may be based on linguistic rules, such as the presence of certain keywords, or on patterns in the text data.

RESULTS

Sentiment analysis research typically evaluates the accuracy of machine learning algorithms in correctly classifying text as positive, negative, or neutral. Results are often presented using graphs that compare the performance of different algorithms [5].

Performance on IMDB Dataset

To demonstrate the performance of classifiers, with Social media text data, IMDB reviews of the standard machine learning dataset is considered. It consists of 50,000 reviews, divided equally for training and testing.

Tabular overview of the comparison of classifiers performance are provided in table (1). The performance of the above models shows sequence- based models show good accuracy. Naive Bayes and SVM model being low complex, good processing, and parameter tuning helps in an increase in their performance

Implementing MNB and SVM with Bag of words

Bag of Words (BOW) are discussed in the pre-processing section. It is Being a prominent method for identifying features, the results also show good accuracy in the feature extraction.

Table 1. Performance of Classifiers

Classifier	Evaluation metrics			
	Accuracy	Precision	Recall	FI score (avg)
Naïve Bayes	0.75	0.75,0.75	0.76,0.75	0.75,0.75
SVM (BOW)	0.58	0.7,0.3	0.18,0.99	0.3,0.77
Random forest	0.76	0.79,0.76	0.92,0.57	0.65,0.85
DNN	0.86	0.85,0.87	0.88,0.84	0.86,0.86
RNN	083	0.81,0.82	0.82,0.81	0.81,0.81
LSTM	0.85	0.87,0.84	0.83,0.87	0.85,0.85
CNN	0.89	0.89,0.89	0.89,0.89	0.89,0.89
CNN-LSTM	0.96	0.93,0.97	0.84,0.99	0.88,0.98

Based on the results shown in figure (5), RNN has poor performance compared to DNN in four epochs, but there are a steady increase and less overfitting nature compared to remaining sequence models. CNN, DNN, LSTM model's performance is good. Finally, the CNN-LSTM combined model being a more complex cause in the increase of parameters, but pooling can decrease the number of parameters and have shown good accuracy levels.

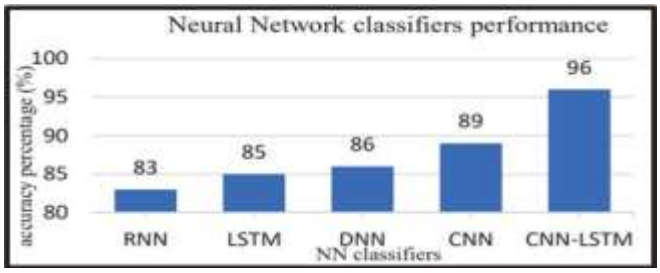

Figure 5. The comparison of Neural Network classifiers performance

e-ISSN: 2583-3472

A common graph used to compare the performance of different algorithms is the Receiver Operating Characteristic (ROC) curve, which plots the true positive rate against the false positive rate. The area under the curve (AUC) is a measure of overall performance, with a higher AUC indicating better performance. Another common graph used is the confusion matrix, which shows the number of true positives, true negatives, false positives, and false negatives for each algorithm.

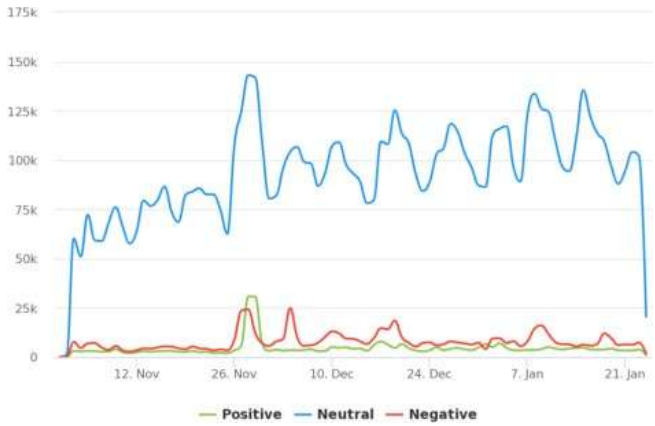

Figure 6. Sentiment Analysis Plot

In addition to these graphs, charts or tables are often presented that show the distribution of positive, negative, and neutral sentiment in the dataset, as well as the accuracy of the algorithms in correctly classifying each sentiment category. This information helps in understanding the overall performance of the algorithms and their ability to accurately classify sentiment.

Figure 7. Result of an Input Text

The accuracy of machine learning algorithms and natural language processing in classifying text as positive, negative, or neutral is often evaluated in sentiment analysis research. Performance comparison of different algorithms is usually presented through graphs and emojis with text will be displayed.

CONCLUSION

In conclusion, sentiment analysis is a powerful tool that can provide valuable insights into public opinion and sentiment towards a variety of topics. There are different approaches to sentiment analysis, including the lexicon-based dictionary approach, machine learning approach, and hybrid approach. The lexicon- based dictionary approach uses predefined dictionaries to classify text as positive, negative, or neutral based on the presence of specific words. The machine learning approach involves training a machine learning algorithm on labeled data to classify text, while the hybrid approach combines both approaches to improve accuracy. Each approach has its advantages and disadvantages, and the choice of approach depends on the specific application and the available data. With the right approach and tools, a sentiment analysis web application can provide valuable insights into public opinion and sentiment towards a variety of topics.

REFERENCES

[1] Asif,M,Atiab Ishtiaq,H.Ahmad,Hanan Aljuaid and Jalal Shah."Sentiment analysis of extremism in social media from textual information. "Telematics Informatics 48(2020): 101345.

[2] S.Amara and R.R Subramanian,"Collaborating personalized recommender system and content-based recommender system using TextCorpus," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020.

[3] Subramaniyaswamy V,Logesh R,VijayaKumar,V, & Indra Gandhi,V. (2015).

[4] International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-1, October 2019.

[5] 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) | 978-1-6654-1451-7/20/$31.00 ©2021 IEEE.
DOI: 10.1109/Confluence51648.2021.9377136.