Analysis of Sentiment by Binary Classification: Current Trends and Future Directions

[1]Deepti Sachin Deshmukh
[2]Dinkar Pandey, [3]Gaurav Sharma, [4]Hassan Rizvi
[1]Assistant Professor, Department of Computer Science and Engineering,
[2,3,4] Students
[2]pandey.dinkar2001@gmail.com

Mahatma Gandhi Mission College of Engineering & Technology Noida Sec-62

College code- 095 affiliated to Dr. A. P. J. Abdul Kalam Technical University, Lucknow

Abstract
Immersion of our society into the depth of digital era makes necessary to understanding the demands and the requirement of the public and the user .Human being critiques may be useful to organizations, governments, and individuals for accumulating statistics and making selections based totally on opinion and face several demanding situations. Due to sheer amount of data produced by the user it is difficult to achieve the goal with speed, accuracy and precision. To address this deep learning and machine learning are widely used for such cases to cater it more efficiently. The aim of our project is to build a binary review classification model for the client's requirement. The model will categorize reviews as either positive or negative, helping the client's in-house customer support team to address or resolve issues raised in negative feedback. By doing so, the client is able to improve customer satisfaction and encourage its domain.

Keywords: *Sentiment analysis, deciphering sentiments, content mining, multimodal records, accumulating statistics*

Introduction

Sentiment analysis, also known as opinion mining, is a computational approach to understanding and interpreting people's sentiments, attitudes, and emotions expressed in textual data. It emerges as a critical aspect in understanding how public opinion and sentiment on diverse topics for the duration of numerous domain names[1]. It has gained substantial prominence in various fields, including marketing, social sciences, and natural language processing. Sentiment evaluation identifies and extracts subjective records from the textual content the use of herbal language processing and textual content mining [2]. By employing a range of techniques, sentiment analysis offers a systematic and insightful way to analyze textual data, enabling researchers to extract valuable insights, trends, and patterns that can inform decision-making processes and further research in numerous domains.

In the study of sentiment analysis, it is crucial to elucidate the fundamental components and methodologies integral to this field. In this regard, a comprehensive understanding of the various techniques and algorithms employed in sentiment analysis is most important. These techniques are of following types:

1. Lexicon based methods
2. Hybrid approaches
3. Machine learning-based

Moreover, it is imperative to discuss the pre-processing steps that play a vital role in enhancing the accuracy of sentiment analysis. These steps typically include tasks such as tokenization, stemming, and lemmatization, which help normalize the text data and remove any noise or irrelevant information that could potentially skew the sentiment analysis results. Additionally, techniques such as part-of-

speech tagging and named entity recognition can be employed to gain a deeper understanding of the context and improve the overall sentiment classification accuracy.

Furthermore, we should emphasize the challenges and limitations associated with sentiment analysis which are;

1. The use of inherent complexity of human language-which often involves sarcasm, irony, and other forms of figurative speech that can be challenging for sentiment analysis models to accurately interpret.
2. The presence of noisy data, the lack of labelled datasets for specific domains.
3. Multilingual Analysis: Handling sentiment analysis for multilingual data introduces complexities related to language-specific nuances, idiomatic expressions, and cultural differences.
4. Imbalanced Datasets: In sentiment analysis, datasets may often be skewed towards a particular sentiment, leading to imbalanced datasets this imbalance can affect the model's ability to accurately predict sentiments to addressing this challenge we requires techniques such as data augmentation, resampling, or utilizing advanced algorithms capable of handling imbalanced datasets.

Application of Sentiment Analysis can be seen in various areas such as:

1. Brand Monitoring and Reputation Management: It helps to monitor and assess customer opinions, feedback, and reviews about their products and services.
2. Market Research and Competitive Analysis: It helps to gain a comprehensive understanding of market trends, consumer preferences, and competitor performance.
3. Customer Service and Support: It enables companies to analyze the customer inquiries, complaints, and feedback received through various communication channels such as emails, chat transcripts, and social media interactions.
4. Social Media Analysis and Campaign Evaluation: It plays a vital role in assessing the effectiveness of social media marketing campaigns and understanding public opinion about brands, products, or events. It helps marketers measure the success of their campaigns, track audience engagement, and adjust their strategies based on the sentiment expressed by the target audience.
5. Healthcare and Patient Feedback Analysis: Sentiment analysis is used in healthcare to analyze patient feedback, reviews of healthcare services, and sentiments expressed in medical surveys.

Citing Sources

In the course of this research, a multitude of sources and references have been extensively consulted and incorporated. Acknowledging the contributions of various scholars and researchers allows for the diversified and well-rounded elaboration on this topic and domain. Here is an inventory of the sources that have been referenced within this paper:

Table 1 . Literature Review of approaches in Sentiment Analysis

S.No.	Publication Title	Authors	Result
1	Sentimental Analysis Using Natural Language Processing	Vishal Tiwari, Sohan Singh, Dr. Raju Ranjan	The author analysis using natural language processing (NLP) that is more accurate and efficient than existing method. It combines a lexicon-based approach with a machine learning approach to achieve better

			result in less time.
2	Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naïve Bayes	RA Leksono, KR Sungkono, R Sarno, CS Wahyuni	The aim of this study is to analyze customer reviews from the best 10 restaurants from reviews. The paper compared the results from two algorithms and concluded the Naive Bayes algorithm has better accuracy than Text Blob sentiment analysis.
3	A survey on sentiment analysis methods, applications, and challenges	Mayur Wankhade,Annava--rapu Chandra Sekhara Rao & Chaitanya Kulkarni	This article gives the overview of the method for completing this task as well as the applications of sentiment analysis. It evaluates, compares, and investigates the approaches used to gain a comprehensive understanding of their advantages and disadvantages. Finally, the challenges of sentiment analysis are examined in order to define future directions.
4	Sentiment analysis: A survey on design framework, applications and future scopes	Monali Bordoloi and Saroj Kumar Biswas	It aims to study systematically and in-depth of different techniques, algorithms, and other factors associated with sentiment analysis. It does a critical assessment of different modules of a sentiment analysis framework with their shortcomings associated with them. It also project application of sentiment analysis based on the contents of data and provided in a particular directions

5	Sentiment Analysis in Social Media and Its Application: Systematic Literature Review	Zulfadzli Drus, Haliyana Khalid	This paper review sentiment analysis of social media in which it explored the various methods, social media platform used and their application. As Social media contains a large amount of raw data which is uploaded by the people in the form of text, videos, photos and audio. And this data can be used to get valuable information by using sentiment analysis approach. .

Methodology

A. **Naive Bayes Classifier**

Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.

It is mainly used in text classification that includes a high-dimensional training dataset. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which help in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object. Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles

Working of Naïve Bayes' Classifier follows the below steps:

1. Convert the given dataset into frequency tables.
2. Generate Likelihood table by finding the probabilities of given features.
3. Now, use Bayes theorem to calculate the posterior probability.

B. Binary Classification Model

Binary classification model is a supervised learning algorithm that categorizes new observation into one of two classes.

In the binary classifier model if it successfully predicts the review as positive, then it is called as True Positive (TP) and if the model successfully predicts review as negative then it is called as True Negative (TN). The binary classifier may interpret some of reviews as well. If a negative review is classified as positive by a negative test result then, this error is called as False Negative (FN). Similarly, if a positive review is classified as negative by a positive test result, this error is called False Positive (FP).

In machine learning various methods used for binary classification which are as follows

1. Support vector machine
2. Naïve Bayes
3. Nearest neighbour
4. Decision tree
5. Logistic regression
6. Neural network

But in our project we are using two of them as of now which are:
Naïve Bayes and logistic regression

```
# Naive Bayes
from sklearn.naive_bayes import GaussianNB
models['Naive Bayes'] = GaussianNB()

# Logistic Regression
from sklearn.linear_model import LogisticRegression
models['Logistic Regression'] = LogisticRegression()
```

Following are the are the steps used in functional model representation (figure 1)

Step 1: Data Extraction:
Data is extracted and stored locally, posing a challenge in diverse environments.

Step 2: Imputation:
Imputation techniques are applied to handle missing values, often utilizing mean, median, or mode calculations, and considering neighbouring values for linear data. However, imputation can introduce variability and data leakage during training.

Step 3: Data pre-processing:
Data pre-processing techniques address common issues like incompleteness, noise, and inconsistency. While removal of rows or columns with null values is common, careful consideration is required to prevent bias and information loss, ensuring the validity and reliability of the results.

Step 4: Tokenization
It involves breaking down unstructured data and natural language text into discrete elements. These tokens can be used as vectors for machine learning, triggering actions or aiding complex decision-making.

Step 5: Feature extraction
It transforms text data into suitable formats for sentiment analysis. Methods include Bag of Words, TF-IDF, word embedding, and sentiment lexicons.

Step 6: Model Selection
It entails choosing the best model for sentiment analysis, considering domain, accuracy, computational resources, and dataset size.

Step 7: Training Data
It split 80-20 for model training and testing, with machine learning algorithms applied to the generated tokens.

Step 8: Testing Data
It evaluates the trained model's performance using the remaining 20% of the dataset. If necessary, the model is redefined or tuned for improved accuracy.

Step 9: Sentiment Analysis
It applies the tested model to recognize and interpret user-provided data, predicting sentiment and offering insights on emotions, experiences, biases, or attitudes.

Step 10: Visual Representation
It presents results, accuracy, and precision to the user, often through symbols, graphs, and detailed reports, enhancing understanding and facilitating further analysis.

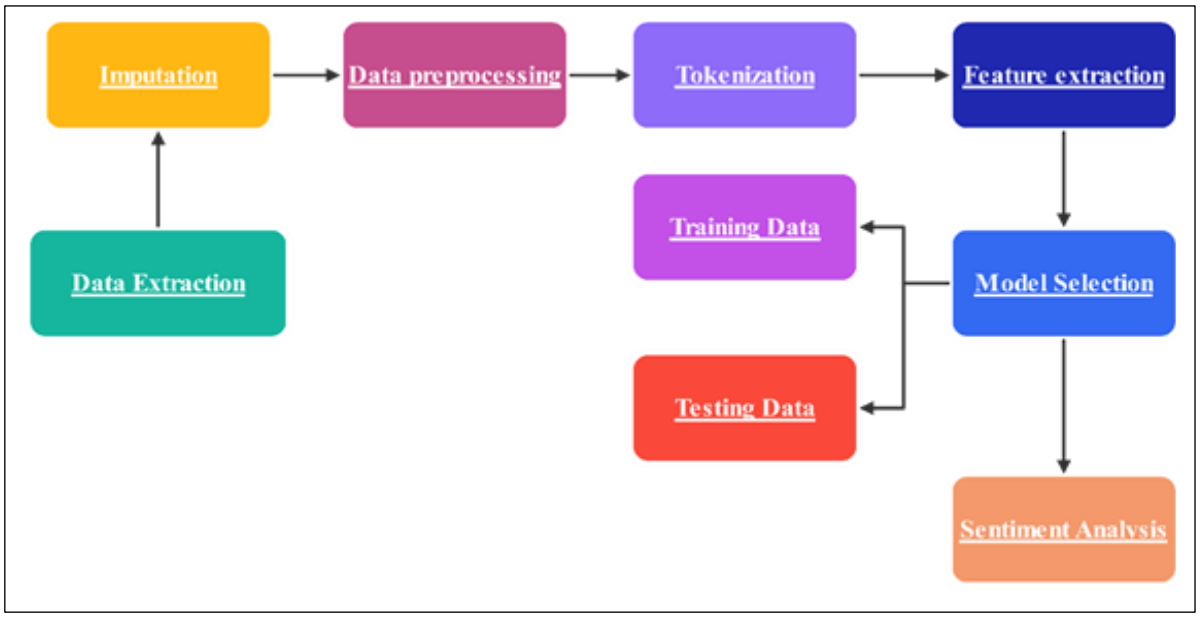

Figure 1 . Functional model representation

Conclusions

1. Sentiment analysis holds significant importance across various domains.
2. Deep learning currently offers the most effective solutions in this field, catering to diverse domains.
3. Further development of models for other domains is imperative, especially for sarcasm detection, multimodal data, and resource-poor languages.

Future Scope

1. Introduction of a versatile model capable of real-time analysis for various data formats like images, video, and audio.
2. Development of cross-lingual processing capabilities to target a global audience and diverse platforms.
3. Creation of more precise and adaptable models tailored to specific domains.
4. Implementation of a model to identify and filter out fake and misleading information.
5. Enhancement of the model's ability to comprehend sarcasm and associated lexicons for improved sentiment prediction.

Conflict of Interest

This work has no associated conflicts of interest.

References

[1] Sentiment analysis using product review data by Xing Fang & Justin Zhan

[2] A survey on sentiment analysis methods, applications, and challenges by Mayur Wankhade, Annavarapu Chandra Sekhara Rao & Chaitanya Kulkarni

[3] Sentimental Analysis Using Natural Language Processing by Vishal Tiwari, Sohan Singh, Dr. Raju Ranjan

[4] Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naïve Bayes by RA Leksono, KR Sungkono, R Sarno, CS Wahyuni

[5] A survey on sentiment analysis methods, applications, and challenges by Mayur Wankhade,Annava-rapu Chandra Sekhara Rao & Chaitanya Kulkarni

[6] Sentiment analysis: A survey on design framework, applications and future scopes by Monali Bordoloi and Saroj Kumar Biswas

[7] Sentiment Analysis in Social Media and Its Application: Systematic Literature Review by Zulfadzli Drus, Haliyana Khalid

[8] Jiayi Chen, Tianyu Zhao, Yifan Peng, and Yunbo Cao, "SentiBERT: A Pretrained Language Model for Joint Sentiment and Emotion Analysis," in Proceedings of the 2023 Conference of the Association for Computational Linguistics, 2023.

[9] Zhehao Xiao, Xiaoliang Dai, Yifan Peng, and Yunbo Cao, "Graph-Based Multi-Task Learning for Aspect-Based Sentiment Analysis," in Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 2023.

[10] Wenpeng Yin, Haoli Dong, Xiangyang Li, and Daxiang Dong, "Aspect-Based Sentiment Analysis with Multi-Head Attention and Graph Convolutional Networks," in Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 2023.

[11] Hu, M., & Liu, B. (2010). Aspect extraction and sentiment analysis of Chinese online customer reviews. 23rd International Conference on Computational Linguistics (COLING 2010)