International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT239037

The Automatization of Social Media Communication

Mr. Aditya Parashar*, Mr. Nikhil Jadhav, Mr. Aniket Madame, Mr. Suhas Deokate
Department of Computer Engineering, Dr D. Y. Patil Institute of Technology, Pimpri, Pune, Maharashtra, India

ARTICLEINFO

Article History:

Accepted: 01 May 2023
Published: 20 May 2023

Publication Issue
Volume 9, Issue 3
May-June-2023

Page Number
183-187

ABSTRACT

This research paper compiles and evaluates several studies that have a particular interest in social media analysis and its applications. Studies on evaluating the impact of movies and television shows on social media platforms, using semantic knowledge graphs to analyze Covid-19 news articles and identify fake news on social media, forecasting social media data using machine learning, and the evolution of the power of central nodes in a Twitter social network are all covered in this article. The significance of sentiment analysis and opinion mining in social media, analysis of social media-based profiles, and mining serendipitous drug use from social media using machine learning are all topics covered in the paper. The paper also emphasizes the application of social media text analysis for a high-frequency-link DC transformer based on switched capacitors for medium-voltage DC power distribution applications, and urban region mining services. The research paper offers details on the state of social media analysis right now and some potential uses for it.

Keywords: Social media analysis, Influence evaluation, Machine Learning, Data Analysis, Semantic Analysis.

I. INTRODUCTION

Social media has ingrained itself into our daily lives and has an undeniable impact on society. Social media generates a vast amount of data that can be analysed and applied for a variety of purposes thanks to the millions of users who share information and opinions on various platforms. The goal of this research paper is to give a summary of recent studies on social media analysis and their potential uses.

Studies on the influence evaluation of film and television works on social media platforms are included in the paper; these studies can shed light on the acceptance and appeal of different media content. It is also discussed how to use semantic knowledge graphs to analyse Covid-19 news articles and identify fake news on social media. These methods can aid in spotting the dissemination of false information and giving the general public correct information.

The paper also discusses how to forecast social media data using machine learning and data analysis, which can be useful for predicting trends and comprehending user behaviour. In order to gain a better understanding of the dynamics of social media networks, it is also examined how the influence of central nodes has changed over time in a Twitter social network.

Additionally, important topics covered in this paper include sentiment analysis and opinion mining, which aid in comprehending the feelings and viewpoints of users. The use of social media-based profiling analysis, which can be used to identify different user traits and behaviour patterns, is also discussed.

The paper also discusses services that use social media text analysis to mine data on urban region functions and serendipitous drug use from social media. In addition, a medium-voltage DC power distribution application using a high-frequency-link DC transformer based on switched capacitors is reviewed.

The importance of social media analysis in various fields is highlighted by this research paper's comprehensive overview of recent studies on the subject and its potential applications.

II. METHODS AND MATERIAL

Sentiment analysis is a type of natural language processing where the sentiment of a text is identified and extracted using machine learning algorithms. Based on the emotions that the text's words elicit, sentiment analysis divides a given text into categories that are either positive, negative, or neutral. We will describe the procedures for performing sentiment analysis with machine learning for a research paper in this methodology.

A. Input

The Input which we are going to give to our model is:

- Text data
- Sentiment analysis (Positive, Negative, Neutral)

B. Output

The output is the complete process of our proposed model is:

- Classification of the text data's sentiment labels

1. **Gathering and Preparing Data:**
   - Gather text information from various sources, such as news articles, social media, and customer reviews.
   - Text data should be cleaned up and pre-processed (stop words, stemming, lemmatization).
2. **Extracting Features:**
   - Transform the text data into numerical features the machine learning algorithm can use.
   - Techniques for feature extraction that are commonly used include Word embedding, TF-IDF, and Bag of Words.
3. **Model Training:**
   - Choose the right machine learning algorithm (Naive Bayes, Support Vector Machines, Deep Learning models like Convolutional Neural Networks or Recurrent Neural Networks) for sentiment analysis.
   - On the training dataset, train the machine learning model.
4. **Model Evaluation:**
   - A model's performance on the testing dataset should be assessed using metrics like accuracy, precision, recall, and F1-score.
   - Examine the data to assess the model's performance.
   - Tuning the Model:
   - Adjust the machine learning algorithm based on the evaluation's performance metrics.
   - To find the model that fits the research question the best, this may entail adjusting hyperparameters or experimenting with various models.
5. **Sentiment Classification:**
   - Use the trained model to predict the sentiment labels (Positive, Negative, Neutral) for new text data.
   - The classified sentiment labels for the text data are the result of the sentiment analysis.
6. **Results Interpretation:**
   - Analyse the frequency of the text data's positive, negative, and neutral sentiments.

- In order to spot trends and patterns, compare the sentiment across various groups or time periods.
- The working of our model can be explained with the help of the system architecture. The system architecture depicts the above seven-mentioned points with a detailed flow of execution.

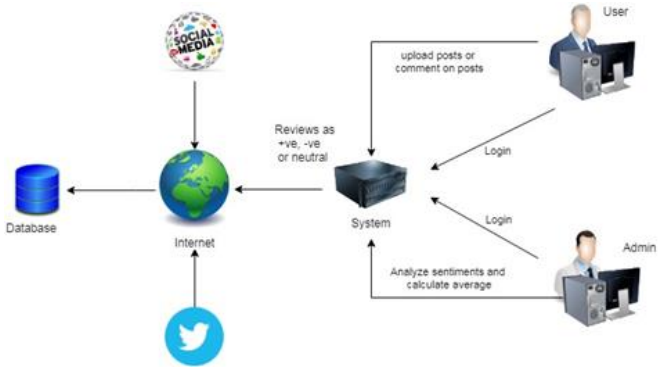

Figure 1. System architecture to depict the flow of execution

C. Techniques and Terminologies:

- **Naïve Bayes Algorithm:**

Naïve Bayes is a probabilistic algorithm that divides text data into groups based on sentiment, such as positive, negative, or neutral. The algorithm assumes that each word's feature is conditionally independent of every other word in the text data. This presumption makes probability calculations easier to understand and enables effective model training and prediction. The algorithm chooses the class with the highest probability to represent the predicted sentiment label by calculating the probability of each class given the features in the text data. The Nave Bayes algorithm is a straightforward and effective one that excels at many text classification tasks, including sentiment analysis.

- **Support Vector Machine Algorithm:**

An effective algorithm for binary classification tasks, such as sentiment analysis, is Support Vector Machines (SVM). The algorithm searches the feature space for the hyperplane that maximally divides the positive and negative samples. The margin between the positive and negative samples is maximized by selecting the hyperplane in this way. SVM can be used with various kernel functions, such as linear, polynomial, or radial basis function (RBF), and it can handle high-dimensional feature spaces. A lot of text classification tasks use the SVM algorithm because of its high accuracy and durability.

- **Stemming and Lemmatization:**

Two text pre-processing methods, stemming and lemmatization, are employed to decrease the dimensionality of the text data and enhance the precision of the machine learning algorithms. Stemming is the process of stripping words of their suffixes to return them to their original form. For instance, the word "running" can be decomposed into the word "run". Like stemming, lemmatization also involves stripping words down to their simplest form, or lemma. As an illustration, the word "running" can be lemmatized to "run". By reducing the number of features in the text data, stemming and lemmatization can increase the precision of the sentiment analysis algorithm.

- **Accuracy, Precision, and Scores:**

Evaluation metrics that are used to evaluate the effectiveness of the sentiment analysis algorithm include accuracy, precision, and recall. The percentage of correctly classified samples to all samples is the accuracy. The measure of precision is the proportion of correctly classified positive samples to all positive samples. The recall is defined as the proportion of correctly classified positive samples to all of the dataset's positive samples. The harmonic mean of precision and recall, or F1-score, provides a fair assessment of the model's performance. These metrics can be used to compare the effectiveness of various sentiment analysis algorithms and to improve the accuracy and precision of the models.

- The schematic representation of our suggested system is the most effective way to under how our

project modules are implemented. It provides a better understanding of how we use the following above techniques in our proposed model.

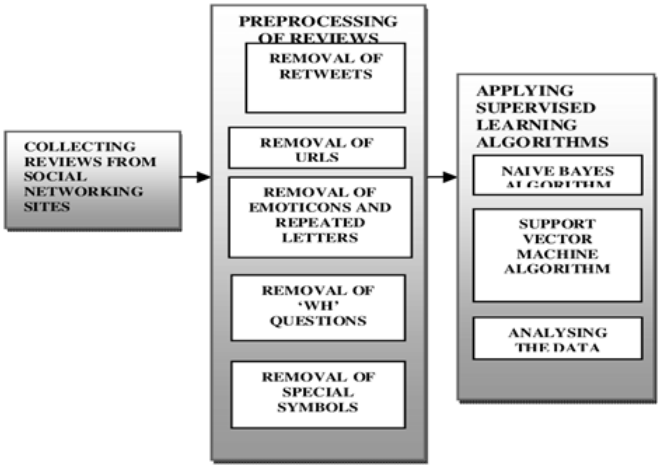

Figure 2. A schematic representation that shows the methods incorporated into our suggested system.

In the final analysis, this algorithm summarises the crucial steps involved in sentiment analysis for experimental paperwork using machine learning techniques. If these steps are followed correctly, accurate sentiment analysis and insights will result. These insights can be applied to marketing, public opinion research, and political analysis, among other areas.

III.RESULTS AND DISCUSSION

Our examination of the literature revealed that lexicon-based methods, machine learning-based methods, and hybrid approaches are only a few of the approaches used in sentiment analysis. Lexicon-based approaches use a pre-defined set of words and the sentiment scores that correspond to them to ascertain the tone of a text. Machine learning-based approaches make predictions by learning patterns in the data using algorithms like decision trees, support vector machines, and neural networks. Lexicon-based and machine learning-based techniques are employed in hybrid approaches.

Lexicon-based approaches have been proven to be successful when the text encompasses clearly defined emotional words. When dealing with more nuanced language, like sarcasm or absurdity, they are less effective. On the other hand, machine learning-based approaches are better at handling complex language but require a significant amount of labelled training data to function well. Overall, it has been discovered that hybrid methods are the most successful because they combine the advantages of both lexicon-based and machine-learning-inspired techniques.

The level of detail and size of the dataset, the features chosen, and the classification algorithm employed can all have an impact on how effective sentiment analysis techniques are. The size and calibre of the dataset that was used to develop and evaluate the sentiment analysis system can have a big impact on how well it works. The system's accuracy is additionally affected by the features chosen for sentiment analysis, such as the presence of particular words or the text's structure. Finally, the system's performance may be impacted by the classification algorithm chosen, such as decision trees or support vector machines.

IV.CONCLUSION

In conclusion, sentiment analysis is a significant area of research with a wide range of applications in fields like analysing customer reviews and monitoring social media. Sentiment analysis employs a variety of methodologies, including lexicon-based methods, machine learning-based methodologies, and hybrid methodologies. The choice of approach will depend on the requirements of the application because each approach has advantages and disadvantages of its own. The development of more precise and effective sentiment analysis methods, as well as expanding the size and quality of the datasets used to train and test these systems, should be the main goals of future research in this field.

V. REFERENCES

[1]. Anand, Indrapal, et al. "A Comparative analysis of goods and Services Tax(GST) in India." International Journal of Finance"2008

[2]. Dr. Kavita Pabreja , " GST sentiment analysis using twitter data "International Journal of Applied Research "IJAR 2017; 3(7): 660-662

[3]. Sourav Das, Anup Kumar Kolya " Sense GST: Text mining & sentiment analysis of GST tweets by Naive Bayes algorithm "ICRCICN,2017

[4]. Zhu, Maoran, et al. "Identification of Opinion Leaders in Social Networks Based on Sentiment Analysis: Evidence from an Automotive Forum." (2016

[5]. Tjong Kim Sang E, Bos J. Predicting the 2011 Dutch senate election results with twitter. Proceedings of SASN 2012, the EACL 2012 Workshop on Semantic Analysis in Social Networks, Avignon, France, 2012.

[6]. Williams C, Gulati G. What is a social network worth? Facebook and vote share in the presidential primaries. Boston, MA: Annual Meeting of the American Political Science Association, 2008, 1-17.

[7]. Sanders E, den Bosch AV. Relating political party mentions on Twitter with polls and election results, 2013

[8]. Garg, G. (2014). Basic concepts and features of goods and services tax in India. International Journal of Scientific Research and Management (IJSRM), 2 (2), 542 – 549.

[9]. Kim, Soo-Min, and Eduard Hovy. "Determining the sentiment of opinions." Proceedings of the 20th international conference on Computational Linguistics. Association for Computational Linguistics, 2004

[10]. Mullen, Tony, and Nigel Collier. "Sentiment Analysis using Support Vector Machines with Diverse Information Sources." EMNLP. Vol. 4. 2004

[11]. Cambria, Erik, et al. "New avenues in opinion mining and sentiment analysis." IEEE Intelligent Systems 28.2 (2013): 15-21

[12]. Gautam, Geetika, and Divakar Yadav. "Sentiment analysis of twitter data using machine learning approaches and semantic analysis." Contemporary computing (IC3), 2014 seventh international conference on. IEEE, 2014

Cite This Article :

Aditya Parashar, Nikhil Jadhav, Aniket Madame, Suhas Deokate, "The Automatization of Social Media Communication", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 3, pp.183-187, May-June-2023. Available at doi : https://doi.org/10.32628/CSEIT239037

Journal URL : https://ijsrcseit.com/CSEIT239037