

Sentiment Connect: Opinion Mining on Education System In India

Ashutosh J¹, Chaitanya K¹, Himanshu N¹, Ritesh P¹, Tanmay T¹, Chandrayani Rokde²

¹Student, Department of Computer Science and Engineering, Government College of Engineering Nagpur (GCOEN), Rashtrasant Tukadoji Maharaj Nagpur University (RTMNU), Nagpur, Maharashtra, India.

²Professor, Department of Computer Science and Engineering, Government College of Engineering Nagpur (GCOEN), Rashtrasant Tukadoji Maharaj Nagpur University (RTMNU), Nagpur, Maharashtra, India.

Corresponding Author: riteshsparkhi@gmail.com

Abstract: - Social networking services have become a popular way of communication among internet users in recent years. They provide instruments for people to express themselves, discuss their feelings, and react to current events. These sites have a large volume of data, which has led to an increase in sentiment analysis and opinion mining studies. The project's goal is to do sentiment analysis utilizing Twitter, a popular microblogging and social communication network. To capture public sentiment on education policy, real-time data can be automatically streamed using the Twitter API. By taking into consideration the public's ideas and opinions, the initiative will aid in the resolution of issues that arise during the implementation of educational programs and reforms.

Key Words: — *Sentiment Analysis, Twitter opinion mining, Logistic Regression.*

I. INTRODUCTION

People may now contact with one another, seek opinions, and freely express their feelings and reactions to various events thanks to the rise of social media. With the massive Individuals and organizations are increasingly using social media on the internet as a result of the expansion of social media. Content in various mediums can be used to analyze public sentiment and make decisions. The many Social media posts have aided in the reshaping of businesses. Traditional communication tools, such as traditional telephones, are increasingly being abandoned by Internet users. To post on items and services they utilize, they can use blogs or mailing lists to microblogging services. Give vital sources of people's thoughts by expressing their political and religious perspectives as well as sentiment. People may now contact with one another, seek opinions, and freely express their feelings and reactions to various events thanks to the rise of social media.

With the massive Individuals and organizations are increasingly using social media on the internet as a result of the expansion of social media. Content in various mediums can be used to analyze public sentiment and make decisions. The many Social media posts have aided in the reshaping of businesses. Traditional communication tools, such as traditional telephones, are increasingly being abandoned by Internet users. To post on items and services they utilize, they can use blogs or mailing lists to microblogging services. Give vital sources of people's thoughts by expressing their political and religious perspectives as well as sentiment.

Twitter is the most widely used tool for exchanging ideas. These feelings are conveyed as soon as they are felt, providing an opportunity to mine text data in real time and analyze attitudes. The large amount of unstructured and informal data from the public on Twitter makes it difficult for governments and organizations to respond swiftly to input. In many developing countries, the government reforms education mostly by coercion. Furthermore, political elites obstruct education planning and reform by focusing education reforms on achieving political goals while neglecting to include other stakeholders such as educators, teachers, parents, and students. It is advised that public opinion be included in the development of policies and reforms that are in line with the public's opinions and needs in order to overcome challenges during the

Manuscript revised May 01, 2022; accepted May 02, 2022.

Date of publication May 04, 2022.

This paper available online at www.ijprse.com

ISSN (Online): 2582-7898; SJIF: 5.59

implementation of changes. As a result, we've established a foundation for the start of our project.

II. LITERATURE REVIEW

The majority of previous research has focused on the outdated educational system. Because the new education policy is being implemented for the first time, no research has been conducted on it. Attempts were made in the past to assign ratings to reviews or to categories entire periodicals as having overall positive or negative polarity. These systems were mostly built on supervised algorithms that depended on human labelled samples, such as movie or product reviews. The reviewer's overall positive or negative assessment was stated directly. Thoughts and sentiments, on the other hand, exist at a higher level than papers and aren't tied to a certain valence or target. Opposing or complementary opinions on the same or several themes may be found throughout a publication [1].

Later works used segment-level opinion analysis to distinguish sentimental from non-sentimental sections of a document—for example, by using graph-based techniques to segment sections of a paper based on their subjectivity or by classifying based on some fixed syntactic phrases that are likely to express opinions. Finally, granularity in text analysis has recently been reduced to the sentence level, for example, by detecting subjective sentences based on the presence of opinion-bearing lexical elements (single words or n-grams) or by using association rule mining for a feature-based analysis of product reviews.

Lexicon-Based Approach: To determine polarity, a lexicon-based system uses a sentiment dictionary including opinion phrases and matches them with the data. They then assign sentiment scores to the dictionary's opinion words, expressing how Positive, Negative, and Objective they are. Lexicon-based solutions, like the Opinion Finder lexicon, rely on a sentiment lexicon, which is a set of pre-compiled sentiment terms, phrases, and even idioms built for traditional communication genres [3].

Because they rely on knowledge bases that are too restricted to process text at the sentence level efficiently, traditional techniques are still incapable of inferring the cognitive and emotive content associated with natural language. Furthermore, text analysis at this level of granularity may be insufficient, because a single sentence may contain opposing perspectives on different parts of the same product or service.

The Lexicon method, which is based on dictionaries, is unable to handle domain and context-specific orientations [5].

III. PROPOSED SYSTEM

The three categories of sentiment analysis methodology include statistical methods, lexicon-based approaches (knowledge-based methods), and hybrid approaches. Scalable and computationally efficient techniques rely on lexicons. The lexicon-based approach, on the other hand, fails to recognize sentiments when language norms are taken into account. Statistical procedures, on the other hand, use machine learning. Deep learning methods have become increasingly popular among researchers, despite the fact that both approaches require labelled training data for polarity recognition.

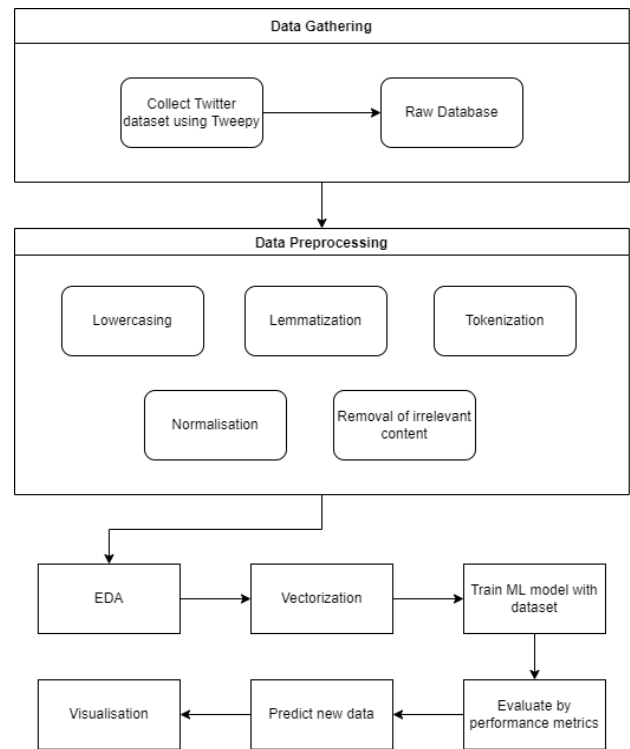


Fig.1. Data flow diagram

In this proposed system, we adopt a hybrid technique to implement sentiment analysis. The hybrid technique of sentiment analysis uses both statistical and knowledge-based methods to detect polarity. Machine learning (statistical techniques) has a high level of accuracy, whereas the lexicon-based approach has a high level of consistency.

3.1 Data Gathering:

We used the Twitter API to get data from Twitter in this module. We got access to the Twitter API by signing up for a developer account on Twitter's official website. We utilized Python's tweepy package to interface with the Twitter API. We used hashtags like #NEP2020, #NewEducationPolicy, #NewBharatNewNEP, #educationpolicy, and #NEP to collect tweets about the New Education Policy 2020 after authenticating the Twitter API. The tweets are saved in a CSV file and utilized as data frames in Python with the pandas module.

3.2 Data Pre-Processing:

Preprocessing improves the accuracy of classification. This module removes links and special characters from tweets using regular expressions. We also eliminated the stopwords and lowercased all of the words, as well as applying the following processes:

Tokenization is the process of breaking down a stream of text into smaller pieces called tokens, which are typically words or phrases. Emoticons, Twitter @usernames, URLs, and hashtags are all considered and treated as distinct tokens during tokenization. Tokenization makes it simple to remove extraneous symbols and punctuation, leaving only the words that contribute to the text's emotive polarity score.

Normalization: All upper-case words are converted to lower case characters in this step (for example, EDUCATION SYSTEM should be reformed). Within a tweet, abbreviations are noted and replaced with the true meaning they convey (for example, OMG > Oh My God), while terms with redundancy or repeated letters (such as I've received my exams results!! Im reeeeaallly happyyyy) are replaced by a single character.

Stemming refers to the process of creating morphological variants of a root/base word. Stemming programs are referred to as stemming algorithms or stemmers. The words "chocolates," "chocolatey," and "choco" are all reduced to the root "chocolate," whereas "retrieval," "retrieved," and "retrieves" are all reduced to the stem "retrieve." Stemming is an integral part of the pipelining process in natural language processing.

Lemmatization is the process of merging many inflected versions of a word into a single item that may be analysed. Lemmatization is similar to stemming but adds context to the words. As a result, it combines similar-sounding words into a single term.

3.3 Sentimental Analysis:

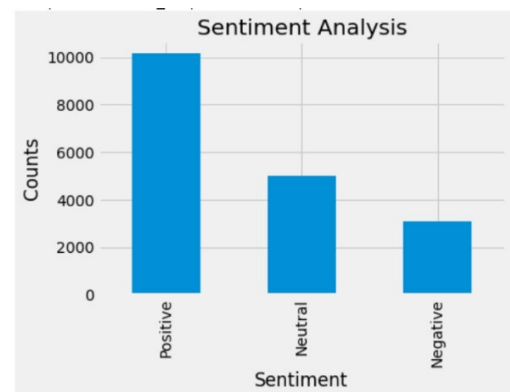
Our sentiment classifier, which uses a training set, was built using supervised learning. A training set and a testing set of data are included for categorization. The training set is used to build the classification model, which classifies the input feature vectors into their respective class labels. The testing set is used to validate the model by predicting the class labels of unseen feature vectors. Sentiment detection is a technique for extracting subjective perspectives from a dataset, with subjective statements being kept for further analysis and objective sentences being dismissed. A logistic regression model was used to categorize good, negative, and neutral tweets.

3.4 Output Visualization:

In this module, we used the matplotlib package to visualize the results graphically and the wordcloud library to visualize the frequency of terms.

IV. RESULTS AND DISCUSSION

The mood analysis of the tweets revealed that the majority of the tweets about India's new education policy are positive. Many individuals are still waiting to see how the new strategy would affect the country. People are taken aback by the sensational move, and they are placing their faith in the government's bold move. Some people, on the other hand, were critical of the project, citing their uncertainty about what it means for the future of Indian education. People are also skeptical because the intended government plans are good, yet problems develop during execution. Some individuals also believe that this was a step that should have been taken a long time ago.



positive = 71.4 % , Negative = 8.6 % , Neutral = 20.0

Fig.2. Final output graph of the analysis done

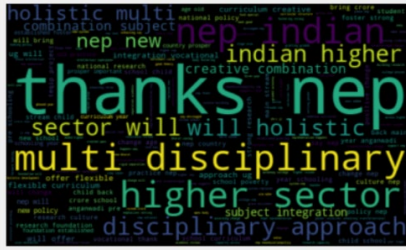


Fig.3. Positive word cloud extracted from the positive tweets.

If we look at the figure 3, frequency with which certain words appear in the term-document matrix, we can see that they are quite related to "New Education Policy." The most often used words in the title are disciplinatory, NEP, higher sector, skills, and progressive. These comments reflect the public's opinion, which is that the new education strategy is directly linked to improving, progressing, and transforming India in terms of skills, education, and knowledge. Any government must prioritize increasing the country's production, which can only be accomplished by providing residents with high-quality education.

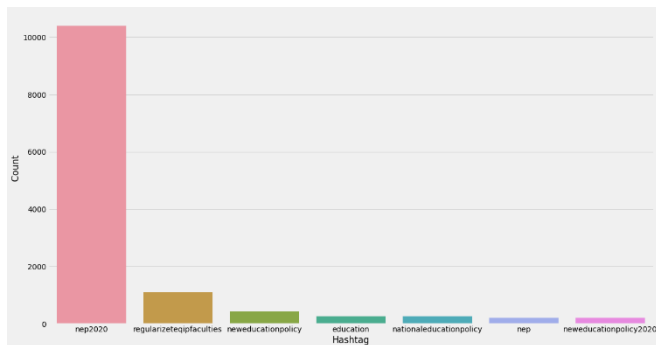


Fig.4. Graphical distribution of positive hashtags.

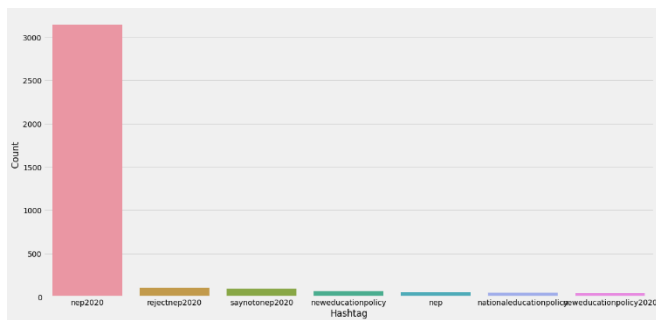


Fig.5. Graphical distribution of negative hashtags.

Figure.4. visualizes the graphical distribution of positive hashtags. In the graph we can see that the hashtag - "#nep2020" was the most trending positive hashtag in the year 2020. Similarly, one can also observe various other positive hashtags

that were trending in 2020. Figure 5 visualizes the graphical distribution of negative hashtags. In the graph we can see that the hashtag - "#nep2020" was again the most trending negative hashtag in the year 2020.

Algorithm	Accuracy
Linear Regression	0.9293094072865687
Bernoulli Naive Bayes	0.8896139206090267
Support Vector Machine (SVM)	0.9252311038607939
Random Forest	0.9418162044589451

Fig.6. Accuracy scores of different models.

The accuracy scores of various models are shown in Figure 6, as we can see the accuracy achieved by random forest algorithm is the highest followed by Linear Regression which is followed by SVM and the least is given by Bernoulli Naïve Bayes.

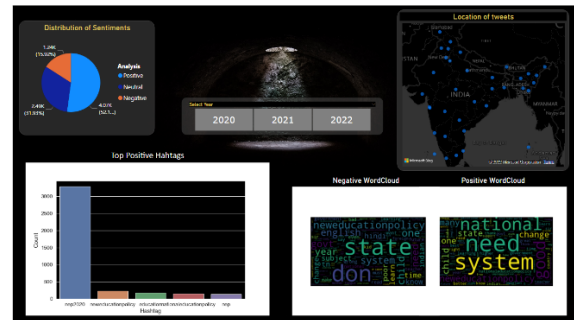


Fig.7. Dashboard

Fig.7. show the dashboard containing various graphical representations of the analysis made on the twitter data of NEP 2020. The dashboard is implemented using PowerBI tool from Microsoft.

V. CONCLUSION

The study concludes that the material available on Twitter concerning the new education policy was the most accurate, relevant, and accurately portrayed people's emotions and opinions, based on the results of the twitter analysis. People are more upbeat about the new education policy than they were about prior approaches. People, on the other hand, express anger and opposing ideas in lower amounts than happiness and delight; government officials have also utilized their Twitter accounts to spread positivism about the new education policy.

Positive sentiments outnumber negative sentiments, as one might expect given the government's decision after nearly 34 years. We can see how enthusiastic the public is about this

policy. Some of the most regularly used terms, such as "nep," "disciplinary," and others, helped to make this research more easy and conclusive. Citizens and government officials have been witnessed using Twitter to convey their emotions. Analysts may also consider Twitter to be the most preferred social media site over others for studying the current emotional state of the population in a country, such as politics, education, or current events.

REFERENCES

- [1]. Nikhil Malhotra and Twinkle Goyal, Sentiment Analysis using Twitter Information Flow about the New Education Policy Introduced in India in 2020, *International Journal of Management*, 11(12), 2020, pp. 2411-2416.
- [2]. Omar, Mwana & Njeru, Alexander & Paracha, Samiullah & Wannous, Muhammad & Yi, Sun. (2017). Mining tweets for education reforms. 416-419.
- [3]. Archana Rao P N & Kishore Baglodi, Role of Sentiment Analysis in Education Sector in The Era of Big Data: A Survey, *International Journal of Latest Trends in Engineering and Technology Special Issue SACAIM 2017*, pp. 022-024.
- [4]. C. Havasi, B. Schuller, H. Wang, B. Liu and E. Cambria, "Knowledge-Based Approaches to Concept-Level Sentiment Analysis" in *IEEE Intelligent Systems*, vol. 28, no. 02, pp. 12- 14, 2013.
- [5]. Kharde, Vishal & Sonawane, Sheetal. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. *International Journal of Computer Applications*. 139. 5-15.
- [6]. Gupta, Itisha and Joshi, Nisheeth. "Enhanced Twitter Sentiment Analysis Using Hybrid Approach and by Accounting Local Contextual Semantic" *Journal of Intelligent Systems*, vol. 29, no. 1, 2020, pp. 1611-1625.
- [7]. P. H. Rahmath and T. Ahmad, "Sentiment analysis Techniques – A Comparative Study.", *International Journal of Computational Engineering & Management*. vol. 17, (2014), pp. 2230-7893.
- [8]. A. Goel, J. Gautam and S. Kumar, "Real time sentiment analysis of tweets using Naive Bayes," 2016 2nd International Conference on Next Generation Computing Technologies (NGCT), 2016, pp. 257-261.
- [9]. Gayathiri.R et al, *International Journal of Computer Science and Mobile Computing*, Vol.5 Issue.3, March- 2016, pg. 512-516.