M.Sc. Mathematics

Vidipt Vashist

2022

Opinion mining and classification of New National Education Policy (2020-2022) using Twitter data

A DISSERATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE AWARD OF THE DEGREE

OF

MASTER OF SCIENCE

IN

MATHEMATICS

Submitted by:

Vidipt Vashist

2K20/MSCMAT/33

Under the supervision of

Dr. Goonjan Jain

DEPARTMENT OF APPLIED MATHEMATICS

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

MAY, 2022

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, Vidipt Vashist, 2K20/MSCMAT/33 student of M.Sc. Mathematics, hereby declare that the project Dissertation titled "Opinion mining and classification of New National Education Policy (2020-2022) using Twitter data" which is submitted by me to the Department of Applied Mathematics Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Science, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi **VIDIPT VASHIST**

Date: 5 May 2022

DEPARTMENT OF APPLIED MATHEMATICS

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled " Opinion mining and classification of New National Education Policy (2020-2022) using Twitter data" which is submitted by Vidipt Vashist, Roll No. 2K20/MSCMAT/33, Department of Applied Mathematics, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology/Bachelor of Technology, is a record of the project work carried out by the students under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

goonjan.

Place: Delhi

Date: 5 May 2022

DR. GOONJAN JAIN

SUPERVISOR

ASSISTANT PROFESSOR

DEPARTMENT OF APPLIED MATHEMATICS

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

ABSTRACT

With the growth of technology in recent years, there has also been a rapid increase in the usage of social media sites to exchange information and beliefs. Opinion mining, also known as sentiment analysis, is used to ascertain public opinion. It is a technique for natural language processing. Sentiment analysis can be characterised as a technique that utilises Natural Language Processing (NLP) to automate the mining of attitudes, opinions, perspectives, and emotions from text, audio, tweets, and database sources. We gathered data from the microblogging website Twitter regarding the New Education Policy (NEP 2020-2022) in order to have a better understanding of the public mood on a national level. Convenience of social media especially Twitter is that it empowers the swift collection of information about the opinions of the public and individual users on current continuing topics. We employed models to classify and assess the emotion elicited by a compilation of around 22,000 tweets about the topic of New Education Policy (NEP). We carried out our investigation using CountVectorizer and TF-IDF and discovered that CountVectorizer outperformed TF-IDF. Additionally, we used Naive Bayes, Decision Trees, Random Forests, Logistic Regression, Gradient Boosting, and Support Vector Machines, and discovered that Logistic Regression provided the highest assorting accuracy.

ACKNOWLEDGEMENT

The satisfaction of successful completion of any task would be incomplete without mentioning the people who made it possible and whose constant guidance and encouragement crown all of my efforts with success. I express my deep sense of gratitude to my parent Kamakshi and Gaurav, my brother Vedant and my guide Dr. Goonjan Jain, Assistant Professor, Department of Applied Mathematics, Delhi Technological University for her inspiration, constant assistance, valuable suggestions, sympathetic advice, fruitful conversation and unparalleled encouragement made throughout the course of study, without which this piece of work would not have taken its present shape.

VIDIPT VASHIST

CONTENTS

List of Figures

List of Tables

List of abbreviations

(CV)	CountVectorizer
(NBC)	Naïve Bayes Classifier
(NEP 20)	National Education Policy 2020
(NLP)	Natural Language Processing
(RFC)	Random Forest Classifier
(RQ)	Research Question
(SA)	Sentiment analysis
(SVM)	Support vector machine
(TFDIF)	Term Frequency – Inverse Document Frequency

CHAPTER 1: INTRODUCTION

Opinion mining, also known as sentiment analysis, is used to ascertain public opinion. It is a technique for natural language processing. Sentiment analysis can be characterised as a technique that utilises Natural Language Processing (NLP) to automate the mining of attitudes, opinions, perspectives, and emotions from text, audio, tweets, and database sources [1]. From a defining standpoint, it is a technique for extracting subjectivity and polarity from text; on the other hand, semantic orientation quantifies the text's polarity and strength [2]. Due to the enormous volume and frequency of user-generated information on social media sites, we rely on Machine Learning models to ascertain the opinion. In the modern era, information spreads extremely quickly digitally amongst users and has the potential to alter the overall feel/perception of an event. As a result, it is critical to comprehend/determine the popular opinion. It is applicable to a broad range of issues affecting practitioners and academics in human-computer interaction, as well as those in subjects such as sociology, marketing, advertising, psychology, economics, and political science [3]. When people live in a society, they form judgments about the world around them. They form opinions about people, things, places, and events. These are regarded to be sentimental attitudes. Sentiment analysis is the study of automated methods for extracting sentiments from written text. The proliferation of social media platforms has resulted in an explosion of freely accessible, user-generated text on the World Wide Web. These data and information may be used to provide real-time insights into people's attitudes. Social media includes blogs, online forums, comment areas on news websites, and social networking sites such as Facebook and Twitter. These social media platforms have the potential to capture the opinions or word of mouth of millions of individuals. The capacity to communicate and access these real-time opinions from people all across the world has revolutionised computational linguistics and social network analysis. Social networking is becoming a more critical source of information for businesses. On the other side, people are more willing and eager than ever before to share details about their lives, knowledge, experiences, and ideas with the rest of the world via social media. They take an active role in events by voicing their ideas and making observations on what is going on in society. This method of sharing their knowledge and feelings with society and social media compels businesses to gather additional information about their firms, goods, and the reputation they enjoy among the public, allowing them to make better informed business decisions.

Sentiment Analysis identifies the phrases in a text that bears some sentiment. The author may speak about some objective facts or subjective opinions. It is necessary to distinguish between the two. SA finds the subject towards whom the sentiment is directed. A text may contain many entities but it is necessary to find the entity towards which the sentiment is directed. It identifies the polarity and degree of the sentiment. Sentiments are classified as objective (facts), positive (denotes a state of happiness, bliss or satisfaction on part of the writer) or negative (denotes a state of sorrow, dejection or disappointment on part of the writer). The sentiments can further be given a score based on their degree of positivity, negativity or objectivity. The Web is a huge repository of structured and unstructured data. The analysis of this data to extract latent public opinion and sentiment is a challenging task.

1.1 Opinion Mining/Sentiment Analysis

Sentiment Analysis is a Natural Language Processing and Information Extraction work those entails examining a large number of documents to ascertain the writer's emotions represented in positive or negative comments, queries, and requests. In general, sentiment analysis seeks to ascertain a speaker's or writer's attitude toward a certain subject or the overall tone of a text. [4]The exponential growth in Internet usage and public opinion exchange in recent years is the driving force behind Sentiment Analysis nowadays. The Web contains a massive amount of structured and unstructured data. The examination of this data to ascertain hidden public sentiment and opinion is a difficult task. Sentiment analysis can be document-based, with the full document's sentiment summarized as positive, negative, or objective. It can be sentence-based, in which specific sentences in the text expressing sentiments are classified. [5]SA can be phrase-based, in which the polarity of the phrases in a sentence is determined. Sentiment Analysis is a technique that finds the sentences in a text that contain sentiment. The author may discuss objective facts or his or her subjective opinions. It is vital to differentiate the two. SA identifies the object of the sentiment. While a sentence may contain numerous entities, it is vital to determine which entity the feeling is intended toward. It establishes the sentiment's polarity and degree. Sentiments can be characterized as objective (denotes facts), positive (denotes the writer's sense of happiness, bliss, or satisfaction), or negative (denotes a state of sorrow, dejection or disappointment on part of the writer). Additionally, sentiments might be scored according to their degree of positivity, negativity, or objectivity.[6]

1.2 Classification Levels

Sentiment Analysis (SA) is considered a three-layered approach. The first of the layers is document based. The second being sentence based. The third is the aspect level, also considered as word or phrase based.[7]

1.1.1 Document level

The first level is considered the document level. At this level, an entire document is considered as a whole for SA. In this treatment for conducting SA, the one making an opinion is considered a single source or an individual entity

1.1.2 Sentence level

Research in SA techniques often place major thrust and emphasis at the sentence level. on classifying and analysing each sentence in a document or piece of text as either objective or subjective ones. research scholars have postulated a model that takes into account sentiments of all terms in each sentence to formulate an overall sentiment for a sentence that is under consideration. SA at sentiment level is not without its own shortcomings. There may be objective sentences that may actually have sentiments not detected. An example of such a sentence could be: “I bought a table from a reputed online store only to find that its legs are not stable enough”.

1.1.3 Aspect level

aims at addressing the shortcomings of document and sentence levels of SA. Fine-grained control can be exercise with the help of SA. The target focus of aspect level SA is to examine opinion in critically and exclusively. Aspect level SA assumes that an opinion can only be one among positive, neutral, negative or an outright objective sentiment expressed.

1.3 Applications

Thus, sentiment analysis is used in the consumer market for product reviews, marketing for understanding consumer attitudes and trends, social media for gauging public opinion on current hot issues, and film for determining whether a newly released film is a hit. Categorizes applications broadly into the following groups.

a. Submissions to review-related websites, such as movie and product reviews.

b. Applications as a Sub-Component Technology Detection of aggressive, heated language in e-mails, spam detection, and detection of context-sensitive information, among others.

c. Business and Government Intelligence Applications Recognize consumer sentiments and trends

d. Domain-Specific Applications Obtaining public opinion regarding political leaders or their perceptions of existing rules and regulations, etc.

1.4 Advantages

Sentiment analysis can assist you in determining how people perceive your brand or product on a large scale. This is frequently impossible to accomplish manually due to the sheer volume of data [8][9].

1.1.4 Non biased

Sentiment is a highly subjective concept. As humans, we communicate meaning through tone, context, and language. Our interpretation of its meaning is contingent upon our own experiences and unconscious biases. Consistent criteria are used in sentiment analysis systems to offer more accurate findings. For instance, a machine learning model can be trained to distinguish the existence of two distinct aspects, each with a distinct sentiment. It would consider the overall sentiment to be neutral, but would also keep track of specifics.

1.1.5 Fast Processing

Sentiment analysis enables organizations to make sense of massive amounts of unstructured data. When working with text, even 50 samples can feel like a large amount of data. Particularly when dealing with consumer opinions expressed in product reviews or on social media. Additionally, this type of study informs businesses of the number of clients who feel a certain way about their product. The amount of people and the overall polarity of feeling toward, say, "online documentation," might help a business choose its goals. For instance, they should concentrate on improving documentation in order to reduce client churn and remain competitive.

1.1.6 Real time insights

Machine Learning is used to automate sentiment analysis. This enables organizations to obtain real-time insights. This can be quite beneficial for spotting critical concerns that require immediate attention. For instance, a negative story spreading on social media can be swiftly identified and addressed. If one client complains about an account issue, it is possible that others will encounter the same difficulty. By immediately alerting the appropriate people to resolve the issue, businesses

1.5 Issues

Sentiment Analysis techniques seek to extract positive and negative sentiment-containing words from a text and to classify it as positive, negative, or otherwise objective if no sentiment-containing words are found. In this sense, it is comparable to a text categorization problem. While text classification has a large number of classes relating to various themes, sentiment analysis has only three basic categories. Thus, it appears as though Sentiment Analysis is simpler than text classification, however this is not the case. The following are some of the primary obstacles:

- Irony and Sarcasm

  People use irony and sarcasm to convey unfavourable events. Without prior understanding of what customers anticipate from airlines, it can be difficult for machines to comprehend this attitude. For example, terms such as "considerate" would be characterised as having a positive sentiment. However, it is self-evident to a human that the overall emotion is negative.

- Context

  Context is critical for comprehending sentiment. The polarity of opinion words varies according on circumstance. Machines must acquire contextual knowledge in order to appropriately classify a text. For instance, when asked "what did you appreciate about our product," the following responses may emerge: "Versatility" and "Features." The first response would be considered affirmative. The second response is similarly affirmative, although it is confusing on its own. If we substitute "what did you dislike," the polarity is fully inverted. Oftentimes, it is the rating that provides context, not the inquiry. The approach is to pre- or post-process the data in order to acquire the required context. This can be a time-consuming and complicated operation.

- Comparisons

Another potential stumbling barrier to accurate sentiment assessment is comparison. Consider the following online review examples: In the first instance, it is self-evident that sentiment is positive. The second is difficult, as it is based on comparisons. Without understanding the product against which these are being made, it's difficult to determine whether these are positive, negative, or neutral. It is dependent on the "alternatives" in the second sentence. If the individual believes the previous items they've used to be extremely poor, this sentence may be less favourable than it appears on the surface.

- Emojis

Emojis can be time-consuming to process, especially when data sources such as social media sites are used. Emojis are classified into two categories: Western emojis and Eastern emojis. Western emojis are composed of a few simple characters, such as :). Eastern emojis combine additional characters vertically, such as ¯_(ツ)_/¯ , which in Japan indicates "smiley sideways shrug.".

- Negation

Handling negation is a challenging task. Negation can be expressed in subtle ways even without the explicit use of any negative word. A method often followed in handling negation explicitly in sentences like “I do not like the movie”, is to reverse the polarity of all the words appearing after the negation operator (like not). But this does not work for “I do not like the acting but I like the direction”. So we need to consider the scope of negation as well, which extends only till but here. So the thing that can be done is to change polarity of all words appearing after a negation word till another negation word appears.

CHAPTER 2: NATIONAL EDUCATION POLICY

The new National Education Policy 2020, which the Union Cabinet approved, aims to highlight several changes to the country's education system. The primary purpose was to:[10][11][12].

1. The Cabinet has approved the Ministry of Human Resource Development's rebranding as the Ministry of Education.

2. The NEP 2020 aims to transform India into a "global knowledge force."

3. This is India's third major overhaul of its educational system since independence.

4. The policy is based on the Draft National Education Policy 2019, which the Committee for Draft National Education Policy – chaired by Dr. K. Kasturirangan, former chairman of the Indian Space Research Organisation – submitted to the Ministry of Human Resource Development on December 15, 2018.

5. Policy seeks to restructure school curricula and pedagogy in a new '5+3+3+4' design, so that school education can be made relevant to the needs and interests of learners at different developmental stages – a 'Foundational Stage' (five years), a 'Preparatory Stage' (three years), a 'Middle Stage' (three years) and the 'High Stage' (four years, covering grades nine, 10, 11 and 12).

6. Seeks to standardise the school curriculum for Indian Sign Language across the country.

7. M.Phil. programmes shall be discontinued.

8. National Research Foundation shall be established to facilitate “merit-based but equitable” peer-reviewed research funding.

9. Education institutions shall be held to similar standards of audit and disclosure as a 'not-for-profit' entity, says this policy. If the institution generates a surplus, it shall be reinvested in the educational sector.

10. Medium of expression until at least grade five – but preferably till grade eight or beyond – shall be the student’s mother tongue, or the local or regional language. The ‘three-language formula’ will continue to be implemented in schools, where two of the three languages shall be native to India.

11. Government of India shall constitute a 'Gender-Inclusion Fund' to provide equitable and quality education to all girls and transgender students. States shall use this fund to implement the central government's policies for assisting female and transgender students, such as provisions for toilets and sanitation, conditional cash transfers and bicycles. The fund will enable states to support 'community-based' interventions.

The issue with NEP:[13]

1. **Disparity between knowledge and jobs:** There is a persistent mismatch between the knowledge and skills staged and the jobs feasible. This has been one of the primary goals of the Indian education system since independence. NEP 2020 is ineffective in addressing this, as it is mute on education in emerging technological domains like as artificial intelligence, cyberspace, and nanotechnology, among others.
2. **The demand of Colossal Resources.** A sincere aim of 6% of GDP in public spending has been established. Mobilizing monetary resources will be a significant problem, given the low tax-to-GDP ratio and competing claims on the national healthcare fund, national defence fund, and other critical sectors.

2.1 Research Question (RQ)

To accomplish the goal of comprehending the public sentiments at the national level, this research generated the following research question (RQ):

- **RQ1:** Which categorization approach performs the best in terms of anticipating and interpreting opinion/sentiment in a Twitter post?
- **RQ2:** What information suggests the factors that contribute to public mood about the New Education Policy?

To address and resolve this issue, our research will seek out and identify the most effective classification and prediction methods for analysis. With this method, data indicating the causes influencing sentiment toward New Education Policy can be classified as positive or negative, as well as the time period during which the sentiment occurred.

2.2 Related Work

We reviewed several previous studies in the linked subject of knowledge and gathered ideas for how sentiment analysis can be performed across multiple domains and languages:

In [14] they used a quantitative data analysis approach to explain the critical area of focus of the policy document. They also used computer-assisted qualitative data analysis software to address the issue. Additionally, this paper identifies the three central aspects of the policy, which are course, language, and student for the higher education sector. In [15] they discovered that citizens and officials rely on Twitter to express their sentiments from their contexts; additionally, there is optimism in the minds of people regarding the new education policy as compared to previous policies; on the other hand, people express anger and negative thoughts, but in proportion to the positive and joy, and the government officials' Twitter handles play a significant role in spreading the positive and joy. The study [16] examines a careful examination of the collected tweets reveals that the community requires more practical-oriented instruction with more instances to tackle. Additionally, a significant amount of data is generated in the educational sector. Users contribute their learning experiences, such as the challenges they had, the quality of the content, and the instruction, which allows for the analysis of their valuable attitudes toward education. According to [17], social networking websites are constantly a breeding ground for misreporting, fraud, and fake news. Political Twitter accounts have grown in popularity over time, and they now play a critical role in shaping the general public's perceptions of problems. They employed a deep learning technique called long short-term memory. In [18] The result reveals that including emoji in addition to text has an effect on sentiment analysis. In [19], the study explores public opinion on China's government's higher education expansion programme, which was implemented primarily to mitigate graduate job challenges caused by COVID-19. [20] discussed the difficulties and an efficient technique for extracting opinions from Twitter tweets. Spam and widely disparate wording make opinion retrieval on Twitter difficult. Twitter is a data hub where users generate massive volumes of data. According to reports, Twitter users generate 12 GB of data per day. The general public makes extensive use of it to express their thoughts on a number of public issues and to lodge grievances with corporations and government bodies. As a social networking site, Twitter provides data that can be utilised for a variety of purposes, including study of certain topics, people, and so on. With the advancement of artificial intelligence, we now have state-of-the-art machine learning and natural language processing algorithms at our disposal for analysing the emotions expressed by users on social media platforms, where it has become a critical tool for understanding human behaviour, studying public relations, and assisting in the resolution of various issues.

On the basis of the aforementioned works, we propose assessing the attitudes expressed in new education policy tweets and classifying them according to their polarity scores. The best strategies for classifying tweets with a balanced degree of accuracy are selected from a pool of techniques using a few characteristics.

2.3 Methodology

We have provided a full explanation of the methodology used to ascertain public opinion on the new education policy shown in Figure & Figure . The approach to extract sentiment from tweets is as follows:

1. Data collecting through the use of the TWINT library.
2. Remove any stop words from the tweets.
3. Tokenize and feed each word in the dataset to the algorithm.
4. Using a lexicon-based sentiment analyzer, compare each word to the dictionary's positive and negative attitudes. Then increment either the positive or negative count.
5. Using the positive and negative counts, we can calculate the sentiment proportion and so get the polarity score of a tweet.
6. Divides the data set into two sections (namely - train and test dataset).
7. Finally, the machine learning model is fitted using a training dataset and used to predict using the test dataset in order to evaluate the model's performance.

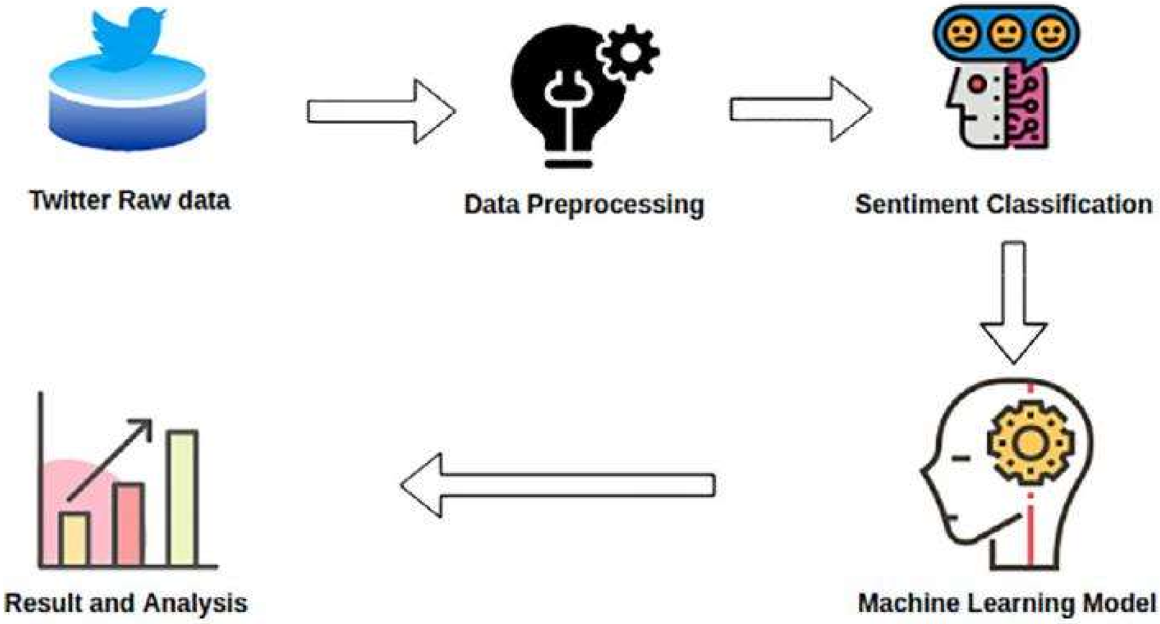


Figure 1. Sentiment Analysis Step by Step Approach.

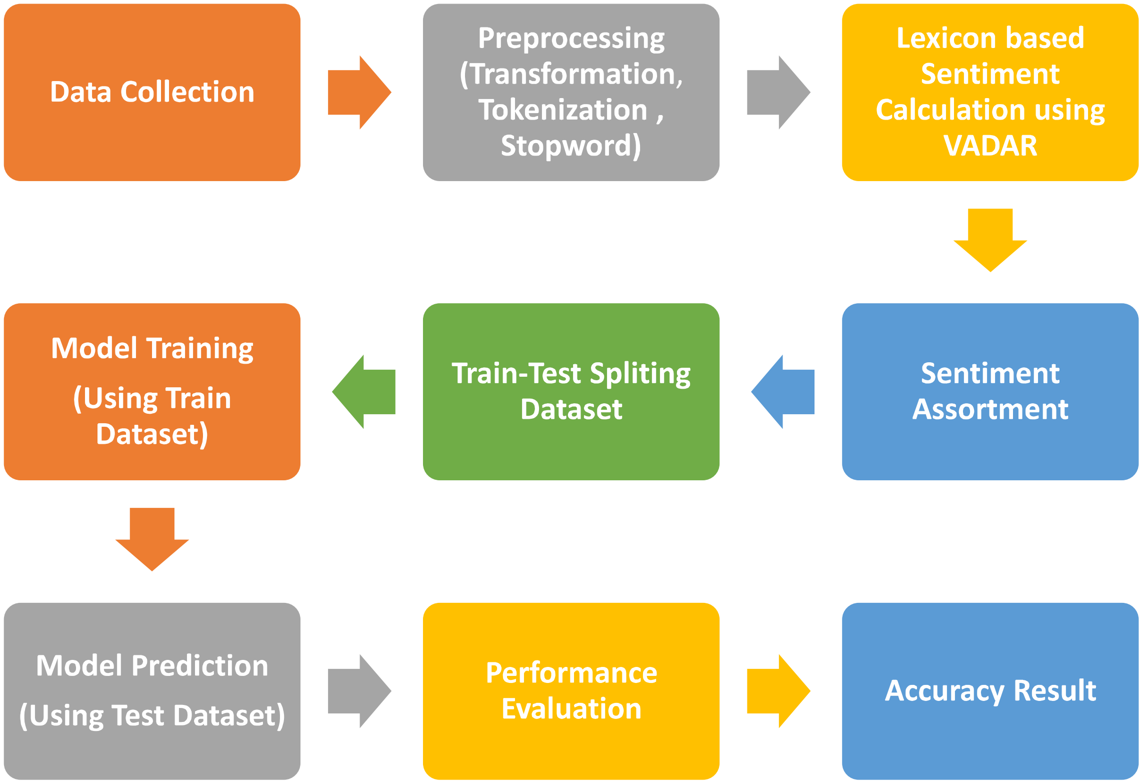

Data Collection
Preprocessing (Transformation, Tokenization , Stopword)
Lexicon based Sentiment Calculation using VADAR
Sentiment Assortment
Train-Test Spliting Dataset
Model Training (Using Train Dataset)
Model Prediction (Using Test Dataset)
Performance Evaluation
Accuracy Result

Figure 2, Flowchart of Methodology

2.4 Dataset

Overall, **22990 tweets** were collected from Twitter between **2019-12-15 17:32:43+00:00** and **2022-02-10 16:05:26+00:00** using an open-source Python package named TWINT - Twitter Intelligent Tool, which was available on GitHub and can be used to obtain data from Twitter without limitation. (GitHub - twintproject/twint: An advanced Twitter scraping & OSINT tool written in Python that doesn't use Twitter's API, allowing you to scrape a user's followers, following, Tweets, and more while evading most API limitations., n.d.). We chose the keyword **'education policy'** as a search term, which encompassed any tweets that contained the terms 'education', 'policy', and 'education.[22].

2.5 Data pre-processing

We collected a total of 22990 tweets, from which we removed duplicates using an excel property, leaving only 22488 tweets. We have shown a sample of the dataset

Then we further clean data by implementing various steps namely:

1. Transformation, in which we make data (i.e., tweets) more consistent by converting it to lowercase, removing URLs, and special symbols.
2. Stopword, in which we omit words that have no bearing on the meaning of a phrase, such as that, the. Etc.
3. Tokenization, in which we turn tweets to words by separating them with delimited spaces and the underscore character ' '.

This contributes to insight and data cleansing, which are then used for sentiment analysis. As a result, I've added a new column called 'clean tweets' to store newly processed tweets.

Table 1. Sample of the dataset.

Datetime	Tweet id	Username	Tweets
2021-11-10,13:54:07	1460000000000000000	dpradhanbjp	india national education policy 2020 has opened new pathways for india and the us to work together and facilitate mutually beneficial partnerships in research two way mobility of students and teachers institute to institute collaborations etc.
2020-09-20,09:13:51	1310000000000000000	rkramkumar91	when english has been made compulsory in the new education policy then why cant it be used as a link language across india why should the students be forced to learn a new language for empty nationalism can anyone answer this stop_hindi_imposition.
2021-09-01,19:03:21	1430000000000000000	MyRayagada	wah this new education policy is a revolutionary step in education reform in india and mp enacted this law in entire state will see the outcome very soon of such good education policy.
2020-09-16,05:21:38	1310000000000000000	abvp4dehra	new education policy is made to enter the roots of southern and eastern india becaz english and local language is what restrict the bashan philosophy of hypocrisy and hypnotism rahulgandhi restrict them

2.6 Sentiment Classification

VANDER classifies tweets into three categories: favorable, negative, and neutral. Each tweet is assigned a polarity value between -1 and 1, with -1 representing the tweet with the most negative sentiment and +1 representing the tweet with the most positive sentiment; additionally, we can adjust the classification's sensitivity by selecting an appropriate interval value classification, i.e., if the polarity score is greater than 0.05, the tweet is labelled positive; if the polarity score is less than -0.05, the tweet is labelled negative; and any score between -0.05 (Table (maximum)) After distributing the sentiments, we calculated the percentage of each type of attitude and discovered that a sizable proportion of individuals (about 63.07 percent) have a favorable opinion of the new education policy, showing they agree with the government's policy decision.

Table 2. Sample tweets to show polarity and sentiment (maximum and minimum)

Tweets	Polarity	Sentiment
new education policy every revolution in education has long lasting impact with modern technology and enriched traditions great great efforts.	+0.9571	Positive
dr kalam said the need of education is to make good human beings with skill and expertise enlightened human beings can be created by teachers change in the education policy is a big way to provide the nation better students professionals amp better human being youthcanlead	+0.9169	Positive
narendramodi please make a very sensible education policy for educating to people leaders in politics and holding responsible positions before any educating policy for the nation people being a very long term dream protect and would unitedly they allow the country to be survived by than	+0.9081	Positive
sophiawanuna ktnnewske 12k worth of text books for a grade 1 student that they cant share with others from parents who cant afford a meal cbc will be the most unequalizing education policy	+0.0031	Neutral
vivekagnihotri to make bharat once again the land of knowledge wisdom philosophy and dharma vedic education policy should be adopted the colonial education system is designed to corrupt and pollute young minds	0.0258	Neutral
mahatmagandhi always said hate the sin amp not the sinner i hate caa, new farm laws and new education policy rise in unemployment inflation etc let truth amp nonviolence be our weapons to fight against sins	-0.9638	Negative
new education policy is not discussed in public neither teachers are clear about it we are engaged by govt in none important things its not good	-0.7536	Negative
narendramodi in every where and everyday general suffer from this cast and reservation system of the government we are finished from this reservation scheme of india thats why our education policy weak our research field weak i am very disappoint from this reservation scheme hate	-0.9403	Negative

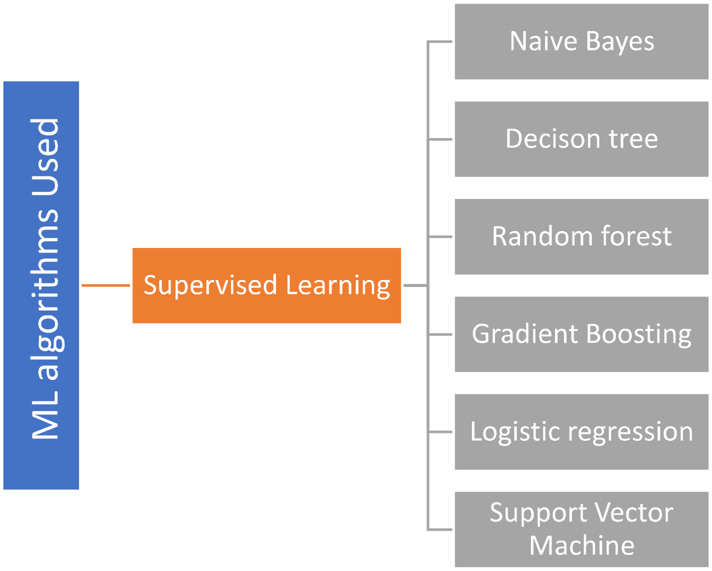


Figure 3. ML algorithms used for classification.

2.7 Model building

This section discusses how to classify and forecast tweets' sentiment using six machine learning algorithms: Naive Bayes, Decision Tree, Random Forest, Logistic regression, Gradient boosting, and Support Vector Machine. (Figure). Then, we compare the performance of two natural language processing strategies, the Count Vectorizer and the Term Frequency and Inverse Document Frequency approaches, to discover which performs better.

2.7.1 Numerical/Vector Representation

As we need to transform word strings into numerical/vectorial representation of text strings for running into machine learning models. We converted it into 2 ways (namely- TFIDF and Count Vectorizer) and compare them which performed better when different Machine Learning models are fitted on them.

2.7.1.1 TF-IDF

It means Term Frequency – Inverse Document Frequency. This is a statistic that is established on the frequency of a word in the corpus but it also provides a numerical representation of how important a word is for statistical analysis. As the term implies, TF-IDF determine values for each word in a document through an inverse proportion of the frequency of the word in a particular document to the percentage of documents the word is present in. Words with high TF-IDF numbers imply a firm relationship with the document they appear in, suggesting that if that word were to appear in a query,[23]This approach not only leads us to better understandings of the commonly practiced heuristic measures but also enables us to propose and verify different heuristics, in connection with other research fields such as probabilistic language modelling. [24] But its major constraint is It fails to provide linguistic information about the words such as the real meaning of the words, similarity with other words, the algorithm cannot identify the words even with a slight change in its tense.[25]

2.7.1.2 CountVectorizer

It is a way to convert a given set of strings into a frequency representation.[26]. Texts can be converted into count frequency using the CountVectorizer function of the sklearn library but its major constraint are: Its Inability in identifying more important and less important words for analysis, It will just consider words that are abundant in a corpus as the most statistically significant word and also doesn't identify the relationships between words such as the linguistic similarity between words.

2.7.2 Naive Bayes

Naïve Bayes Classifier (NBC) is a text mining method that can be used to solve opinion mining problems by classifying opinions into positive and negative. NBC can function properly as a method of text classifiers [4]. The Bayes' Theorem allows one to quantify the likelihood of a hypothesis given prior knowledge Bayes' Theorem is stated as The Naive Bayes algorithm that is used for text mining is Multinomial Naive Bayes. Multinomial Naïve Bayes is also a supervised learning machine in the process of classifying text by using the probability value of a class in a document. [27] they found the accuracy of naive Bayes is not directly correlated with the degree of feature dependencies measured as the class-conditional mutual information between the features. Instead, a better predictor of accuracy is the loss of information that features content about the class when assuming the naive Bayes model.

2.7.3 Decision Tree

Decision Tree Classifier (DTC) basic idea involved in any multistage approach is to break up a complex decision into a union of several simpler decisions, hoping the final solution obtained this way would resemble the intended desired solution. This method classifies a population into branch-like segments that construct an inverted tree with a root node, internal nodes, and leaf nodes. [28] [5] leaf nodes are the result of those decisions and tell us whether the sentiment is positive, negative, or neutral and has no additional branches. [29] mentions potential limitations as that Errors may accumulate from level to level in a large tree and when the number of classes is large, can cause the number of terminals to be much larger than the number of actual classes and thus increase the search time and memory space requirements.

2.7.4 Random Forest

Random Forest Classifier (RFC) approaches by combining several randomized decision trees and aggregates their predictions by averaging.[30] . The Random Forest algorithm has two stages: random forest construction and prediction using the random forest classifier generated in the first stage [31]. Also, it is easily adapted to various ad-hoc learning tasks and returns measures of variable importance.[5] *Select "K"* features at random from a total of "m" features such that k<<m. b. Calculate the node "d" using the best split point of the "K" functions. c. Using the optimal break, divide the network into daughter nodes d. Repeat measures 1 to 3 until the l number of nodes is reached. e. Build a forest by repeating steps 1 to 5 for "n" number of times to create an "n" number

2.7.5 Logistic Regression

The central mathematical concept that underlies logistic regression is the logit—the natural logarithm of an odds ratio[8] A logistic regression will model the chance of an outcome based on individual characteristics. works very similar to linear regression, but with a binomial response variable. It is a powerful tool, especially in epidemiologic studies, allowing multiple explanatory variables to be analyzed simultaneously, meanwhile reducing the effect of confounding factors[32]

2.7.6 Gradient Boosting

Are a family of powerful machine-learning techniques that have shown considerable success in a wide range of practical applications. It is a boosting-like algorithm for regression They are highly customizable to the particular needs of the application, like being learned concerning different loss functions.[33] the learning procedure consecutively fits new models to provide a more accurate estimate of the response variable. [34]The principle idea behind this algorithm is to construct the new base-learners to be maximally correlated with the negative gradient of the loss function, associated with the whole ensemble.

2.7.7 Support vector machine

is a mathematical entity, an algorithm for maximizing a particular mathematical function with respect to a given collection of data.[35] SVM boasts a strong theoretical underpinning, coupled with remarkable empirical results across a growing spectrum of applications. It is used to classify the texts as positives or negatives. SVM works well for text classification due to its advantages such as its potential to handle large features.[36]

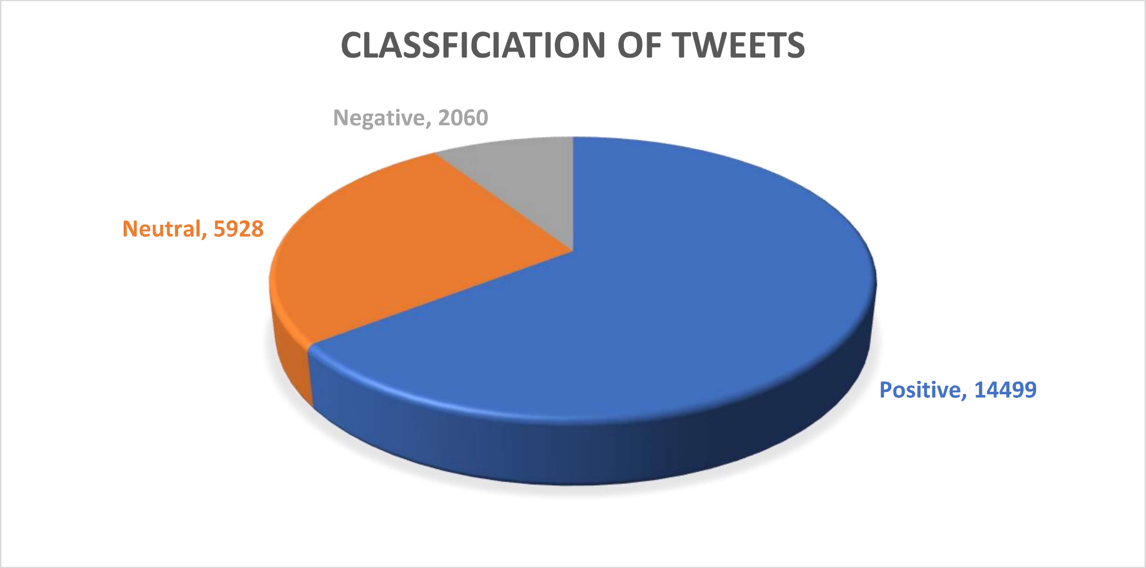


Figure 4. Pie Plot of Sentiment Count

2.8 Model prediction

We did this by fitting our data set into machine learning algorithms using a python-based toolkit called scikit-learn. Additionally, this section discusses forecasting, visualisation, and reasoning with regard to the gathered data, as well as evaluating the performance of machine learning algorithms on the following parameters:

2.8.1 Accuracy

measures how often the classifier makes the correct prediction.

$$=\left[\frac{\textbf{number of correct predictions}}{\textbf{total of predictions}}\right]$$

2.8.2 Precision

tells us what proportion of tweets we classified as positive or negative or neutral were positive or negative or neutral.

$$=\left[\frac{\textbf{True Positives}}{\textbf{True Positives + False Positives}}\right]$$

2.8.3 Recall

tells us what proportion of tweets that were positive or negative or neutral was classified by us as positive or negative or neutral.

$$=\left[\frac{\textbf{True Positives}}{\textbf{True Positives + False Negatives}}\right]$$

2.8.4 F1 score

It is a common way to measure classifier performance in sentiment analysis as it computes the harmonic mean between precision and recall.

$$= \mathbf{2} * \left[\frac{(\textbf{precision}*\textbf{recall})}{(\textbf{precision}+\textbf{recall})}\right]$$

CHAPTER 3: RESULT AND ANALYSIS

3.1 Results

Table , summarises the precision, recall, and f1-score for each type of machine learning method, including Naive Bayes, Decision Tree, Random Forest, Logistic Regression, Gradient Boosting, and SVC utilising TFIDF.

Table , summarises the precision, recall, and f1-score for each type of machine learning method, including Naive Bayes, Decision Tree, Random Forest, Logistic Regression, Gradient Boosting, and SVC with CountVectorizer.

Table , compares the accuracies of the TF-IDF with the Count Vectorizer.

Table 3. Precision, Recall and F1 score in each class for all ML algorithms using TFIDF.

ML algorithm	Negative			Neutral			Positive		
	Precision	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
Naive Bayes	0.10	1.00	0.18	0.32	0.88	0.47	0.99	0.71	0.83
Decision tree	0.42	0.67	0.51	0.80	0.87	0.73	0.85	0.87	0.86
Random Forest	0.36	0.96	0.52	0.80	0.80	0.80	0.94	0.86	0.90
Logistic Regression	0.40	0.93	0.56	0.81	0.85	0.83	0.96	0.87	0.92
Gradient Boosting	0.37	0.91	0.52	0.64	0.79	0.71	0.94	0.81	0.87
SVC	0.38	0.96	0.55	0.80	0.87	0.83	0.97	0.86	0.91

Table 4. Precision, Recall and F1 score in each class for all ML algorithms using CountVectorizer.

ML algorithm	Negative			Neutral			Positive		
	Precision	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
Naive Bayes	0.44	0.74	0.55	0.66	0.79	0.72	0.93	0.83	0.88
Decision tree	0.56	0.71	0.62	0.88	0.84	0.86	0.93	0.92	0.93

Random Forest	0.39	0.94	0.56	0.88	0.82	0.85	0.95	0.89	0.92
Logistic Regression	0.57	0.84	0.68	0.91	0.83	0.87	0.94	0.93	0.94
Gradient Boosting	0.40	0.89	0.55	0.33	0.89	0.48	0.98	0.73	0.84
SVC	0.34	0.96	0.51	0.87	0.79	0.83	0.94	0.89	0.91

Table 5. Accuracy Chart of ML algorithms achieved through sentiment analysis.

	Accuracy		
	TFIDF	**CountVectorizer**	**Average**
Naive Bayes	73.31	81.50	77.40
Decision tree	79.82	88.38	84.10
Random Forest	84.65	87.36	86.01
Logistic Regression	86.99	89.77	88.38
Gradient Boosting	80.60	75.32	77.96
SVC	86.73	86.17	86.45

3.2 Analysis

Figure and Figure shows the positive and negative word cloud collected from the dataset. A word cloud is a visual representation of frequently used words that illustrates both positive and negative emotion from the dataset. [37] The entire word cloud depicts the most frequently used words. The larger font size words occur more frequently than the smaller font size words. Word clouds are increasingly being employed as a straightforward method for determining the central theme of written material. It aids in comprehending the underlying sentiments around the new education policy. Increase foreign investment, digital India, improved health, Ayushman scheme, low inflation, sector improvement, and start-up improvement are the most often utilized words in our positive word cloud. And the most often used words in our negative word cloud are: battle poverty, poor, ruin language, and bad development.

Figure depicts the polarity scale for each tweet in the dataset, i.e., Polarity vs. Tweet Serial Number. While the polarity is concentrated in the [0,0.80] range, this indicates an overall favorable attitude toward the new education program. However, the sentiment expressed by individuals ranges from negative to favorable. [38]. Around 5000 tweets, we find a quick surge in good sentiment, implying that government official accounts aided in the dissemination of new policies, but after a while, the influence reduces, revealing an increasing amount of unfavorable sentiment from the public. As a result, the media/politicians play a critical role in propagating either good or negative opinion. And in the current time period, roughly 20000 tweets, we have a sentiment split in half and a majority of tweets classified as neutral.

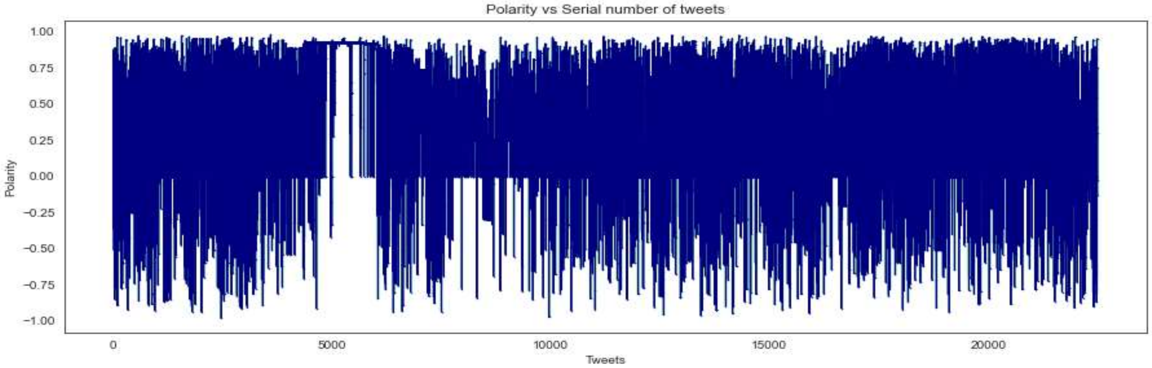


Figure 5. Polarity scale of each tweet in the dataset

Figure represents the confusion matrix for logistic regression using TFIDF. The true label indicates the tweet's real sentiment, while the predicted label indicates the tweet's predicted sentiment. [39] The confusion matrix is a special type of table that contains data on actual and expected classifications. Additionally, it enables display of the performance of a classification algorithm or approach. According to the confusion matrix, 3.88 percent + 21.13 percent + 61.99 percent = 87 percent of tweets were predicted with their correct label, whereas 13% were forecasted with the incorrect label.

Similarly, Figure illustrates the logistic regression confusion matrix using CountVectorizer. We conclude that 5.48 percent + 23.55 percent + 60.74 percent = 89.77 percent of tweets were predicted with their correct label, while 10.23 percent were forecasted with the incorrect label.

Figure and Figure illustrates a comparison of the accuracy of various machine learning algorithms used for sentiment analysis. Logistic regression has the maximum accuracy of 89.77 percent, gradient boosting has the lowest accuracy of 75.32 percent, and support vector machine has a high accuracy when both TFIDF and CV are used.

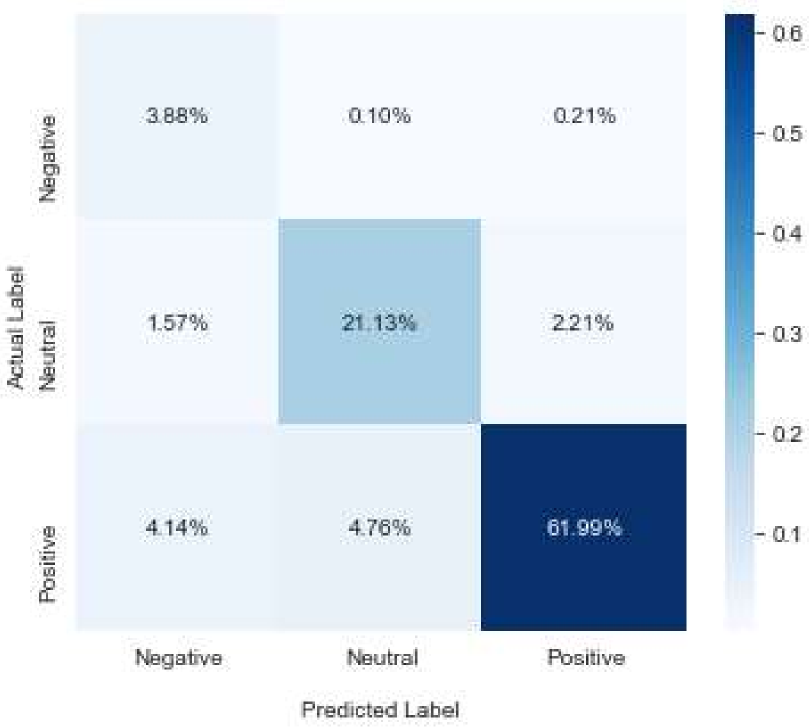


Figure 6. Confusion Matrix of Logistic Regression using TF-ID

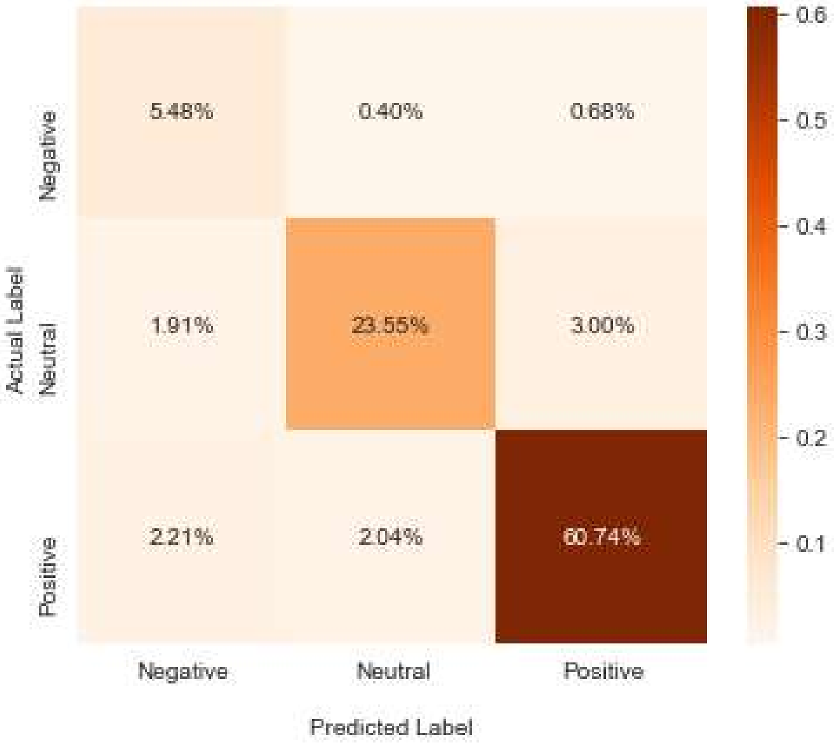


Figure 7. Confusion Matrix of Logistic regression using CountVectorizer

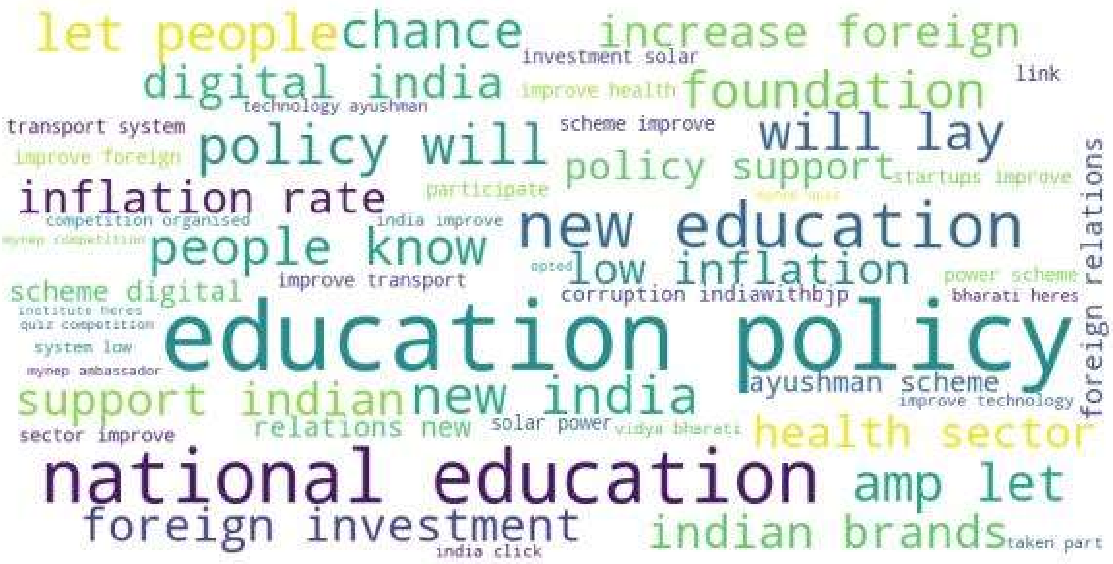


Figure 8. Positive word cloud.

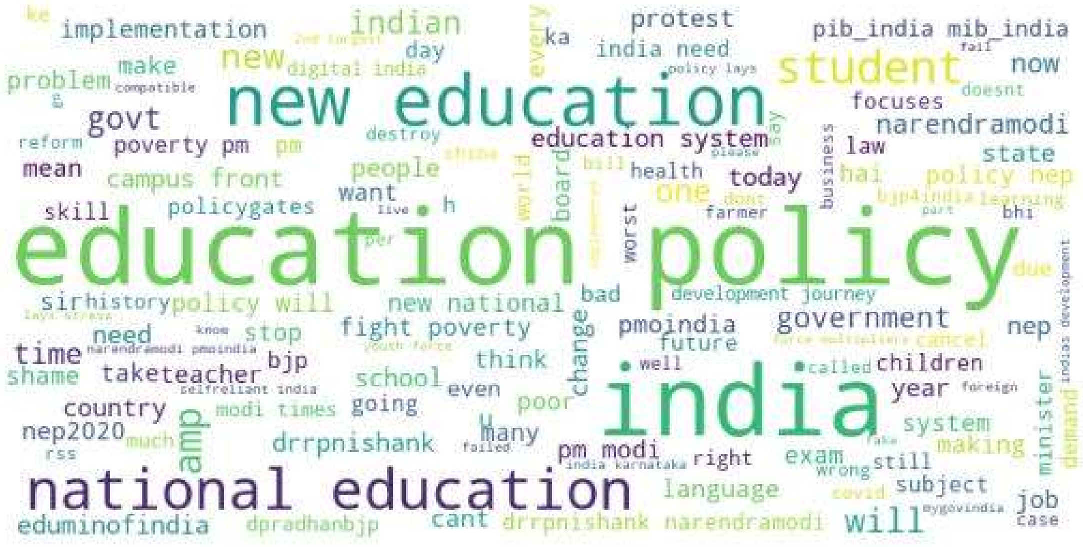
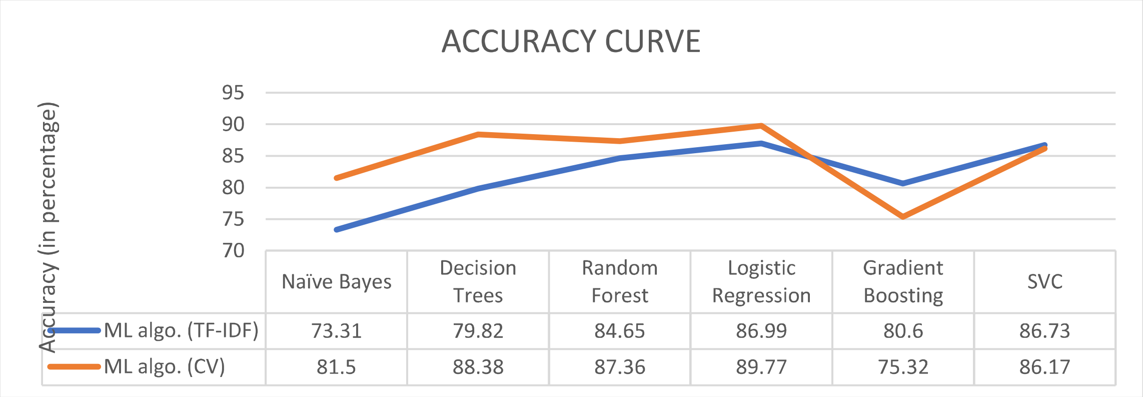


Figure 9. Negative word cloud.



	Naïve Bayes	Decision Trees	Random Forest	Logistic Regression	Gradient Boosting	SVC
ML algo. (TF-IDF)	73.31	79.82	84.65	86.99	80.6	86.73
ML algo. (CV)	81.5	88.38	87.36	89.77	75.32	86.17

Figure 10. Accuracy Curve of ML algorithms (TF-IDF vs CV).

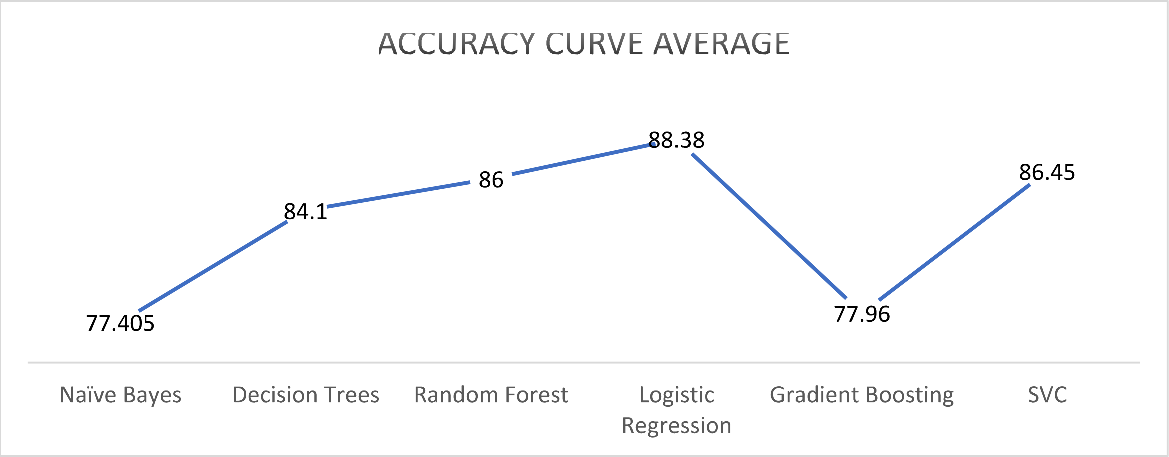


Figure 11. Accuracy Curve of ML algorithms Average

3.3 Discussion

This research focused on extracting tweets about India's new education policy in order to examine the user's views. A new education strategy that began in July 2020 has garnered widespread attention and support from members of all segments of society, both directly and indirectly via Twitter exchanges. As data continues to be created in abundance across many social media sites, most notably Twitter. As a result, processing such enormous amounts of data using standard methods is extremely time consuming, and we thus require new technical resources and new methodologies to process it efficiently. As a developing country, India's primary focus is on strengthening the education sector and implementing policies that benefit the population, and a new education policy has been discussed and debated throughout the country. Confusion over new policies has resulted in an imbalanced public sentiment; thus, we see that the official account of government and politicians is critical in effectively communicating policy to the general public and thus contributing to a healthy relationship between the government machinery of the policymaker and public opinion.

3.3.1 Contribution to literature

Our work focused on grasping different sentiments by analyzing tweets and classifying them according to their polarity score using multiple machine-learning algorithms. We implemented various machine learning algorithms using two distinct numerical/vector representations, namely Count Vectorizer and TF-IDF, and demonstrated that Count Vectorizer outperforms TF-IDF in terms of accuracy. And logistic regression outperforms other machine learning algorithms in terms of assortment accuracy, with an average accuracy of 88.38 percent.

3.3.2 Implication to practice

With the always-increasing of data on social media platforms in the current age, especially on Twitter, there is constantly a need to analyze tweets systematically and assort them. This research will be beneficial to governments, who are the compelling forces after the amendment of the Education policy of India, as it will give a more advanced order-level understanding to them regarding the situation at the ground level opinion about the new policy and more rigorous analysis of policy pros and cons. Thus, policymakers can make subsequently new decisions that are better for both government and the public.

3.4 Conclusion and future scope

This study contributes to the contemporary literature on education policy by investigating the opinion of the public on new education policy and provides timely suggestions to the Government of India to promote its policy practices based on public feedback. It also offers an informative case study for other countries with new education policy reforms in the forthcoming. As social media platforms have given the general masses a chance to be heard and give their immediate and timely reaction to reach millions.[40] There is across the board generation and consumption of content on social media,[41] Thus convenience of social media especially Twitter is that it empowers the swift collection of information about the opinions of the public and individual users on current continuing topics.

Furthermore, 6 machine learning models were used for assorting and prediction. We saw that Logistic Regression performed the best with an accuracy of 88.38 percent. The keywords used ('education policy 'as a search query wherein all the tweets containing words 'education', 'policy', and 'education policy' combine) to collect tweets admissible are one of the work's limitations. If the keywords were not used in the tweet, some admissible tweets were expectedly skipped and were not collected by us thus we could have extracted a comparatively large number of tweets given that crores of people expressed their opinions about the new policy. A higher number of tweets might have been resourceful in determining a rather large number of opinions, but we lacked the computational resources to collect and process such a large number of tweets. This study might be valuable in the field of public policy with the development of social media (e.g., Twitter), the interest in sentiment analysis (SA) has increased, allowing the summary of opinions from large amounts of opinionated tweets and thus support decision-making in several domains, including Government policy decision making.[42]

Figure shows the polarity scale of each tweet in the dataset around 5000 tweets we see a sudden increase in positive sentiment thus implying government official accounts helped provide a better understanding of the new policy but after some while, the impact decreases show the more and more negative sentiment of people. Hence media/ politician plays a very important role in propaganda whether positive or negative sentiment. And in current time around 20000 tweets, we see the sentiment is half divided and mostly neutral classification.

There are approximately 22500 tweets in the dataset and the graph shows the range of sentiments shown by these correspondents to Figure , the average positive sentiment is expressed in as many as 14599 tweets, thus ranking highest among all sentiments. Then neutral sentiment with 5928 tweets and then negative sentiment with 2060 tweets.

In future work, we will also explore other text processing possibilities in the local language thus getting more grass-root opinions of the masses and covering more vertical channels of information exchange platforms (i.e., Twitter, Facebook, YouTube-comment, reputed newspapers, and other information exchange platforms).

References

[1] V.A. Kharde, S.S. Sonawane, Sentiment Analysis of Twitter Data: A Survey of Techniques, International Journal of Computer Applications. 139 (2016) 975–8887. http://ai.stanford. (accessed February 10, 2022).

[2] A. Szabolcsi, Positive polarity - negative polarity, Natural Language and Linguistic Theory. 22 (2004). https://philpapers.org/rec/SZAPP (accessed October 22, 2021).

[3] C.J. Hutto, E. Gilbert, VADER: A parsimonious rule-based model for sentiment analysis of social media text, Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014. (2014) 216–225.

[4] H. Saif, Y. He, H. Alani, Semantic sentiment analysis of twitter, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 7649 LNCS (2012) 508–524. https://doi.org/10.1007/978-3-642-35176-1_32.

[5] A.S. Neogi, K.A. Garg, R.K. Mishra, Y.K. Dwivedi, Sentiment analysis and classification of Indian farmers' protest using twitter data, International Journal of Information Management Data Insights. 1 (2021) 100019. https://doi.org/10.1016/j.jjimei.2021.100019.

[6] D. Halpern, S. Valenzuela, J.E. Katz, We Face, I Tweet: How Different Social Media Influence Political Participation through Collective and Internal Efficacy, (2017). https://doi.org/10.1111/jcc4.12198.

[7] R. Kumar, D. Sarddar, I. Sarkar, R. Bose, S. Roy, A Literature Survey On Sentiment Analysis Techniques Involving Social Media And Online Platforms, INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH. 9 (2020) 5. www.ijstr.org (accessed May 1, 2022).

[8] J. Peng, An Introduction to Logistic Regression Analysis and Reporting Effect size View project Power Analysis View project, Article in The Journal of Educational Research. (2002). https://doi.org/10.1080/00220670209598786.

[9] Sentiment Analysis Report – Delacon, (n.d.). https://support.delaconcorp.com/hc/en-us/articles/360050332013-Sentiment-Analysis-Report (accessed May 1, 2022).

[10] National Education Policy 2020, (n.d.). https://ruralindiaonline.org/en/library/resource/national-education-policy-2020/?gclid=Cj0KCQjw37iTBhCWARIsACBt1IxsHge7Dy2xnP5CxIPgA2jspKhtnqsY48nUWEZ7Ems_ug86ud-ksOwaAncmEALw_wcB (accessed May 1, 2022).

[11] L. Devi, . Cheluvaraju, A Study on Awareness about the Impact of National Education Policy-2020 Among the Stakeholder of Commerce and Management Disciplinary, European Journal of Business and Management Research. 5 (2020). https://doi.org/10.24018/EJBMR.2020.5.6.640.

[12] P.S. Aithal, Student Centric Curriculum Design and Implementation – Challenges & Opportunities in Business Management & IT Education, IRA International Journal of Education and Multidisciplinary Studies (ISSN 2455-2526). 4 (2016) 423. https://doi.org/10.21013/JEMS.V4.N3.P9.

[13] P.S. Aithal, S. Aithal, Munich Personal RePEc Archive Analysis of the Indian National Education Policy 2020 towards Achieving its Objectives Analysis of the Indian National Education Policy 2020 towards Achieving its Objectives, (2020).

[14] R. Pratap, S. Kaurav, K.G. Suresh, NEW EDUCATION POLICY: QUALITATIVE (CONTENTS) ANALYSIS AND TWITTER MINING (SENTIMENT ANALYSIS), Community & Communication Amity School of Communication. 12 (2020) 2456–9011. https://doi.org/10.31620/JCCC.12.20/02.

[15] N. Malhotra, T. Goyal, Sentiment Analysis using Twitter Information Flow about the New Education Policy Introduced in India in 2020, International Journal of Management (IJM). 11 (2020) 2411–2416. https://doi.org/10.34218/IJM.11.12.2020.228.

[16] J. Sultana, M. Usha, R. Sri, P.M. Visvavidyalayam, H. Farquad, Sentiment Analysis based Recommender System for Reforming Indian Education using Multi-Classifiers Sentiment Analysis View project Big Data Analytics View project, (n.d.). https://www.researchgate.net/publication/342851828 (accessed February 10, 2022).

[17] P. Grover, A.K. Kar, S. Gupta, S. Modgil, Influence of political leaders on sustainable development goals – insights from twitter, Journal of Enterprise Information Management. 34 (2021) 1893–1916. https://doi.org/10.1108/JEIM-07-2020-0304/FULL/XML.

[18] R. Biswas, N. Vyas, M. Baskar, Sentiment Analysis on National Education Policy Change 2020, Turkish Journal of Computer and Mathematics Education (TURCOMAT). 12 (2021) 1480–1488. https://doi.org/10.17762/TURCOMAT.V12I11.6063.

[19] X. Yu, S. Wu, W. Chen, M. Huang, Sentiment Analysis of Public Opinions on the Higher Education Expansion Policy in China:, Https://Doi.Org/10.1177/21582440211040778. 11 (2021). https://doi.org/10.1177/21582440211040778.

[20] Z. Luo, M. Osborne, T. Wang, An effective approach to tweets opinion retrieval, World Wide Web. 18 (2015) 545–566. https://doi.org/10.1007/S11280-013-0268-7.

[21] GitHub - twintproject/twint: An advanced Twitter scraping & OSINT tool written in Python that doesn't use Twitter's API, allowing you to scrape a user's followers, following, Tweets and more while evading most API limitations., (n.d.). https://github.com/twintproject/twint (accessed September 16, 2021).

[22] S. Kumar, A.K. Kar, P.V. Ilavarasan, Applications of text mining in services management: A systematic literature review, International Journal of Information Management Data Insights. 1 (2021) 100008. https://doi.org/10.1016/j.jjimei.2021.100008.

[23] J. Ramos, Using TF-IDF to Determine Word Relevance in Document Queries, (n.d.).

[24] A. Aizawa, An information-theoretic perspective of tf-idf measures q, (n.d.). www.elsevier.com/locate/infoproman (accessed December 7, 2021).

[25] R. Ali, S. Qaiser, U. Utara, M. Sintok, M. Kedah, A. Ramsha, T. Analytics, Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents Text Mining, Article in International Journal of Computer Applications. 181 (2018) 975–8887. https://doi.org/10.5120/ijca2018917395.

[26] I. v Bogach, V.A. Kovenko, RECOMMENDATION SYSTEM BASED ON NLP TECHNIQUES, (n.d.). https://papers.nips.cc/paper/5021-distributed-representations-of-words-and- (accessed February 12, 2022).

[27] I. Rish, An empirical study of the naive Bayes classifier, (n.d.).

[28] Y.Y. Song, Y. Lu, Decision tree methods: applications for classification and prediction, Shanghai Archives of Psychiatry. 27 (2015) 130–135. https://doi.org/10.11919/j.issn.1002-0829.215044.

[29] S.R. Safavian, D. Landgrebe, A Survey of Decision Tree Classifier Methodology, (1990).

[30] L. Breiman, Random Forests, 45 (2001) 5–32.

[31] G. Biau, E. Scornet, A Random Forest Guided Tour, (n.d.). http://www.kaggle.com/c/dsg-hackathon (accessed February 12, 2022).

[32] S. Sperandei, Understanding logistic regression analysis, Biochemia Medica. 24 (2014) 12–20. https://doi.org/10.11613/BM.2014.003.

[33] A. Natekin, A. Knoll, Gradient boosting machines, a tutorial, Frontiers in Neurorobotics. 0 (2013) 21. https://doi.org/10.3389/FNBOT.2013.00021.

[34] C. Bentéjac, A. Csörg˝ O B Gonzalo Martínez-Muñoz, A Comparative Analysis of XGBoost, (2019).

[35] W.S. Noble, What is a support vector machine?, NATURE BIOTECHNOLOGY. 24 (2006). http://www.nature.com/naturebiotechnology (accessed February 12, 2022).

[36] N. Zainuddin, A. Selamat, Sentiment analysis using Support Vector Machine LEARNING ANALYTICS FRAMEWORK TO SUPPORT SELF-REGULATED LEARNING STRATEGIES IN MASSIVE OPEN ONLINE COURSES (MOOC) IN REDUCING DROPOUT View project Requirements Prioritization View project Sentiment Analysis Using Support Vector Machine, (2014). https://doi.org/10.1109/I4CT.2014.6914200.

[37] R. Atenstaedt, Word cloud analysis of the BJGP, British Journal of General Practice. 62 (2012) 148–148. https://doi.org/10.3399/BJGP12X630142.

[38] S. Urologin, Sentiment Analysis, Visualization and Classification of Summarized News Articles: A Novel Approach, IJACSA) International Journal of Advanced Computer Science and Applications. 9 (2018). www.ijacsa.thesai.org (accessed March 3, 2022).

[39] F. Sağlam, H. Sever, B. Genç, Developing Turkish sentiment lexicon for sentiment analysis using online news media Turkish Speech Recognition View project Kavram Tabanlı Bilgi Geri Getirim Yaklaşımı View project Developing Turkish Sentiment Lexicon for Sentiment Analysis Using Online News Media, (n.d.). https://doi.org/10.1109/AICCSA.2016.7945670.

[40] D.M. Romero, W. Galuba, S. Asur, B.A. Huberman, Influence and Passivity in Social Media, (n.d.).

[41] M. Ristova, Economic Development No.1-2, (2014) 181–191. http://www.statisticbrain.com/facebook-statistics/] (accessed March 12, 2022).

[42] N. Oliveira, P. Cortez, N. Areal, Stock market sentiment lexicon acquisition using microblogging data and statistical measures, (n.d.).