SOCIAL BEHAVIOUR ANALYSIS DURING FARMER'S PROTEST

[1]Sarvesh Mungurwadi, [2]Prof.Neha Patil, [3]Manish Pradhan, [4]Savan Jadav, [5]Prathamesh Sanas

[1]Student, [2]Professor, [3]Student, [4]Student, [5]Student

[1]Department of Computer Engineering, AISSMS Institute of Information Technology, Pune, India
[2]Department of Computer Engineering, AISSMS Institute of Information Technology, Pune, India
[3]Department of Computer Engineering, AISSMS Institute of Information Technology, Pune, India
[4]Department of Computer Engineering, AISSMS Institute of Information Technology, Pune, India
[5]Department of Computer Engineering, AISSMS Institute of Information Technology, Pune, India

Abstract: Social media has developed into a forum where users may express their views and object to the decisions made by corporations and the government. Every internet user has the ability to express themselves on these sites. India's government recently approved three acts pertaining to farmers, which are being opposed by the country's citizens, especially its farmers. The farming community and other related communities are alarmed by the way these crimes are being carried out. At least 70% people in India have utilized social media platforms to express their displeasure with these acts and to voice their opposition to them. In more than one language, various communities have voiced their feelings over these acts. It is currently challenging for scholars to study the attitudes on these platforms due to the blending of many languages and grammatical systems. This study uses a statistical technique to analyze mood, emotion, and emoji in a collection of English-language tweets on farmers' protests. The performance and accuracy of the Models are also a major focus of this work.

KEYWORDS: Sentiment, Emoji , Emotion Analysis, statistical technique.

I. INTRODUCTION

In the northern areas of the nation, farmers in India were opposing three farm legislation that had been passed by Parliament in September 2020. The three laws in dispute are The Farmers' Produce Trade and Commerce (Promotion and Facilitation) Act, The Farmers' (Empowerment and Protection) Agreement of Price Assurance and Farm Services Act, and The Essential Commodities (Amendment) Act. The government claims that these Acts will "attract private investment" and "transform Indian agriculture." In a system known as contract farming, farmers produce crops in return for a salary that has been mutually agreed upon through contracts with corporate investors. The Farmers' (Empowerment and Protection) Agreement on Price Assurance and Farm Services Act, 2020 created contract farming. Farmers who are protesting worry that affluent investors would force them to sign unfair contracts created by prestigious corporate law firms that include liability clauses that, in most circumstances, are out of the price range of small farmers. Indian farmers are protesting against the three farm acts that were adopted by the Indian Parliament in September 2020. For more than a few months, there has been a standoff between the farmers and the federal government. Since January 2021, the ranch laws have been temporarily suspended by the courts. The demonstrations, also known as the Farm Bills, have drawn criticism from numerous farmer associations for being "against farmer laws," and opposition government legislators have noted that it would leave farmers "helpless facing corporations." In order to assure that businesses won't be able to pay the costs, the farmers have also requested the establishment of a Minimum Support Price (MSP) law. The government, however, argues that the legislation will make it easy for farmers to sell their produce honestly to significant buyers and asserts that disagreements are the result of deception. A few associations, the Indian government claims, have appeared to openly endorse protests. Laws for the area's advancement were provided a year later. Even though some farmer associations have been protesting the farm

regulations. Demonstrators in the Delhi region had submitted a number of petitions to the Supreme Court of India by the middle of December requesting that bars be removed. In order to assure that businesses won't be able to pay the costs, the farmers have also requested the establishment of a Minimum Support Price (MSP) law. The government, however, argues that the legislation will make it easy for farmers to sell their produce honestly to significant buyers and asserts that disagreements are the result of deception. A few associations, the Indian government claims, have appeared to openly endorse protests. Laws for the area's advancement were provided a year later. Even though some farmer associations have been protesting the farm regulations. Demonstrators in the Delhi region had submitted a number of petitions to the Supreme Court of India by the middle of December requesting that bars be removed. This Sentiment categorization has been approached from a variety of angles. The two primary techniques for sentimental analysis are NAVE BAYES. We'll use NLP data pre-processing techniques to understand the data. The study also discusses model-building strategies like Naive Bayes. The suggested approach generates more accurate emotive analyses of farmer protest data by combining Natural Language Processing (NLP) and Naive Bayes techniques.

II. LITERATURE SURVEY

Jetir et al. [1] have the progression of web innovation and its development, there is an enormous volume of information present in the web for web clients and a ton of information is created as well. In the proposed system we are mainly focuses mainly on sentiment analysis of farmer protest and using various algorithms like Support Vector Machine, Random Forest and KNN. An impasse between the focal government and the farmers has been seen for the beyond couple of months. The Indian government requested the police and law implementation of different states to assault the nonconformists utilizing water cannons, cudgel, and poisonous gas to keep the farmer associations from going into Haryana first and afterward Delhi. Later nonconformists arrived at Red Fort and introduced farmer association banners and strict banners on the pole on the bulwark of the Red Fort. In proposed framework to review and analyse the relationship of web-based media-based emergency correspondence with crisis informatics and its scientific classification and the connected emergency correspondence hypothetical models to infer the difficulties and constraints. Furthermore, taxonomy of the research dimensions of social media data management and analysis for disaster management was also proposed, which was then applied to a survey of existing literature and to discuss the core advantages and disadvantages of the various methodologies. The objective of this proposed framework is to execute a program where a web-based media post or comment will be input data to perform sentiment analysis and to classify its sentiment on nationwide protest. Investigating the subtleties of web-based media destinations will furnish one with a viewpoint on culture and the climate. We have explored ways to understand the sentimentality of people by building a sentiment analysis model and identifying the direction the protest is leading towards.

Sentimental analysis of social media et al.[2] social media platform would have the ability to perform sentiment analysis on all the activities of the users in a group and create a report according to their reactions and their posts on the platform. Existing social media platforms do not give us the privilege to track the activities of the users and analyse the user's behaviour for future predictions such as what to post when to post and whom to target. Sentiment analysis on a user's activities would generate a report for the admin of a group that the user belongs to; such as the Head of a department or a college counsellor. Reports can be generated by the analyses done on the data of the users on the platform. According to the posts and the actions of the users, the data that is generated is analyzed by the sentiment model. The existence of a system that facilitates this using the social media behaviour of the point of interest of the institute can be a game-changer when it comes to decision making. The users can also share the post privately with a group of users on the platform. The other aspect of the project is the sentiment analysis part where all the activities of a user in a group would be analyzed. The mechanism of the model is such that it analyses the likes, comments, and posts of the users.

Sentimental analysis and classification of Indian farmers protest et al. [3] They showed the three acts are: The Farmers' Produce Trade and Commerce Act, The Farmers' Agreement of Price Assurance and Farm Services Act, and The Essential Commodities Act. The farmers have rejected the government proposal of suspending the laws for 18 months and the government has insisted that the protests are a result of misinformation. The protests are an integral part of a democratic society and they can be fundamental in shaping the future of the society. Some groups also believed that the farm bills are in the favour of the farmers and the protests are being held because of a lack of information. The main objective of this research is to understand the sentiments of the public on farmers' protest shared on the microblogging website Twitter. In addition, we also aim to analyse the polarity and factuality of Twitter data regarding the demonstrations by extracting twitter data. The conclusion and the future scope of the research is covered towards the end. In Section 8.3. 1 shows the progressive steps we have taken to analyse and predict the sentiment of a particular

Twitter user. 25 Positive The true measure of the JUSTICE of a system is the amount of protection it guarantees to the weakest.

Sentimental analysis of covid-19 lockdown et al. [4] In the present situation, the whole world is facing the crisis of the COVID-19 pandemic. To control the spread of coronavirus on March 25, the Indian government took a significant decision to lock down the whole nation for 21 days. Since March 25, 2020, the Indian government decided to slow down the spread of the virus. In all the lockdown phases, an immense contribution was from Indian Council for Medical Research -An apex body in India for the formulation, coordination, and promotion of research, which played an essential role in governing and controlling the spread of COVID-19. SENTIMENT ANALYSIS ON TWITTER DATA Sentiment analysis models primarily focus on the score of polarity (negative, positive, and neutral). The authors used NodeXL to extract and store the tweets of the related keywords. In our paper, the main objective is to perform sentiment analysis on "Nationwide Lockdown effect in India during COVID-19" while mainly focusing on the tweets of following topics: Effect of Lockdown 1. The following keywords were considered for the sentiment analysis in this research work and passed to extract the tweets as per date and number of tweets mentioned in Table 1. According to this, the subjectivity score is within the range of [0.0, 1. Sentiment Classification In this section, a calculation of polarity and subjectivity has been done to classify the tweets based on sentiment.

Twitter sentimental analysis during covid-19 et al. [5] The results of the study concludes that while majority of the people throughout the world are taking a positive and hopeful approach, there are instances of fear, sadness and disgust exhibited worldwide. China has witnessed benefits of the one of the largest lockdowns at the start of this pandemic, where it locked down 20 provinces and regions. This research study has been to identify the sentiments of the citizens of 12 different countries regarding COVID19 and identify what emotions people have been sharing from different parts of the world. Once the scoring of the tweets was done on the basis of sentiments and emotions, corpus was created in order to develop the word cloud for each country. In the first phase, the sentiments of the tweets from all the 12 countries will be discussed. In this process, it was observed that the USA, France and China had the highest number of tweets with anger. The country with most number of tweets associated with Trust and Surprise was Belgium while the anticipation quotient was highest in the tweets from Germany. In Belgium, words like Political, Government and Pandemic were most used words with the emotions of disgust and fear respectively. Netherland also had the words friends and share trending which were associated with emotions of joy.

Sentimental analysis using support vector machine et al. [6] Nurulhuda Zainuddin proposed that Sentiment analysis in reviews is the process of exploring product reviews on the internet to determine the overall opinion. The results were measured using Precision, Recall, Accuracy, F Measure and AUC for evaluating the effectiveness of the proposed method. The remainder of this paper is organized as follows; Section 2 describes related works performed by other researchers in this field; Section 3 describes the proposed method to perform the experiment; and Section 4 describes the primary results and discussion obtained from the experiments. The researches discover many issues of sentiment analysis in regards to domains of the datasets, corpus types and size and also multilingual context. In order to address these issues, a suitable method of feature selection is required to extract the useful features before the classification is done. The performance of the machine learning process depends heavily on its features so it is crucial to choose the extract features. The objectives are to summarize and transform input data into a set of representation features that work appropriately to the classifier. On the other hand, one of our main goals is to applied several n-gram models which is unigrams, bigrams and trigrams to compare the influence of using different n-gram schemes. 1) Term-weighting Scheme: The calculation of the term weighting scheme plays a crucial role in extracting the most classical features as an input to the classifier. The more classical the features, the better performance of the classifier will be.

III. METHODOLOGY

Here, we outline the rigorous procedure we followed to find out what people's opinions were on the Indian farmer protests. Gathering data from Twitter is the first step in the process, which is then followed by a few critical procedures like cleaning and preparing the data to make it computer intelligible. The following stage is to ascertain and classify a user's sentiment based on two factors. Additionally, the way people feel about certain situations can be studied using visualization approaches.
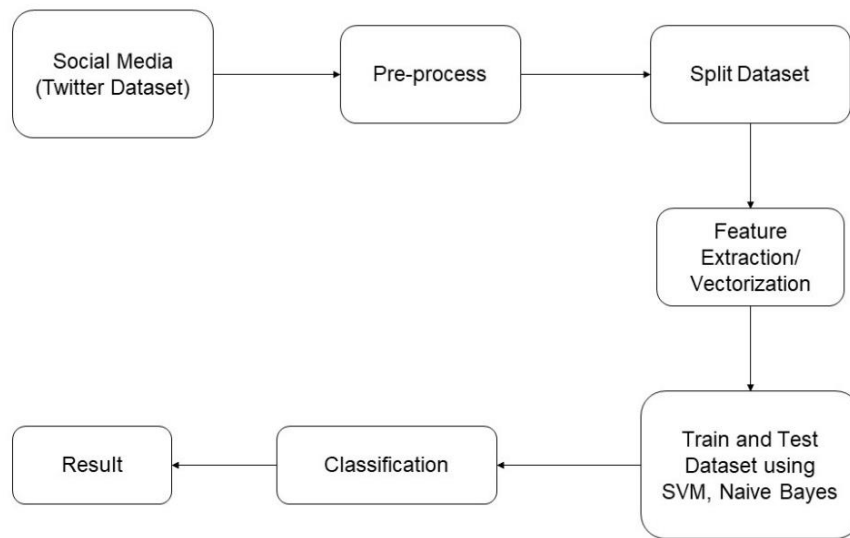
Fig.1.Sentiment/Emotion Analysis step by step approach

1)Data Collection: - Over the course of a year, 25,00,000 tweets in total were collected. The raw data was collected using the open-source Python application Tweepy, which was utilized to acquire direct access to the Twitter API.

Fig.2.Sample dataset of farmers protest tweet

2)Data Pre-Processing: - One of the most important components of data analysis is making sure that our data can be comprehended by computers. Text, photos, and videos cannot be understood by machines; they can only understand 1s and 0s. To be able to supply an input that just consists of 1s and 0s requires a number of stages. Pre-processing the data, also known as data cleaning, is necessary since it involves transforming raw data into a machine-understandable format. Our enormous text dataset of tweets needs to be cleaned up because it has certain discrepancies that need to be removed in order to avoid having inconsistent data.

3)Sentiment Analysis: - Sentiment analysis is the process of mining contextual text to find and extract individualised data from sources. It helps enterprises in observing online discussions and determining the social mood of their brand, product, or service. However, the only indicators commonly utilised in social media stream analysis are count-based metrics and basic sentiment analysis. Similar to simply scraping the surface and leaving out those extremely essential discoveries that are just waiting to be discovered, this.

4)Sentiment Classification: - It is an automated procedure that locates opinions in text and sorts them into three categories based on the emotions that the readers of the text are expressing: positive, negative, or neutral. To evaluate subjective data and better understand how customers feel about your products, services, or company, sentiment categorization employs NLP.

IV. MODELLING AND ANALYSIS

In this section, two well-known supervised machine learning algorithms—Naive Bayes and Support Vector Machine—are used to categorize and predict sentiment and emotion in tweets. Before machine learning models are trained on text data, the data must be cleaned up because text data cannot be processed by computers in its raw form. In order for the computer to understand the text, it must literally be broken down into a numerical structure. Due to this, we examine the results of the Bag of Words, Term Frequency, and Inverted Document Frequency Approach NLP approaches. NLP techniques BoW and TF-IDF aid in the conversion of tweets into numerical vectors.

1) Bag of Words:- The Bag of Words model can be used to extract features from a text while modelling, like in our example of machine learning techniques for tweet sentiment classification. It is a combination of words that are used to describe a sentence in a text with a word count. It consists of two parts: a list of well-known phrases and a metric to measure their frequency. BoW disregards the order in which they appear as well.

2) TF-IDF:- Based on how frequently a term appears in the document, a term's frequency score is calculated. Inverse Document Frequency is a metric that assesses a word's rarity based on a document.

$$TF(t, d) = N(t, d)/T \qquad (1)$$

Here, N(t,d) is the frequency of the term t in document d, T denotes the total number of terms in the document, and TF(t,d) denotes that frequency. As a result, a separate TF(t,d) value will be assigned to each document and word.

$$IDF(t) = logN/(N(t)) \qquad (2)$$

Equation (2) demonstrates how to calculate IDF(t), or the inverse document frequency of the term t. N is the total number of documents, and N(t) is the total number of documents that contain the term t.

$$TF - IDF = TF * IDF \qquad (3)$$

Equation (3) gives the calculation of TF-IDF.

3) Naïve bayes: - The supervised learning algorithm Naive Bayes is based on the Bayes' Theorem, which implies predictor independence. In layman's terms, a Naive Bayes classifier falsely assumes that the presence of one function in a class has no bearing on the presence of any other features. It is therefore possible to comprehend what the Bayes theorem says. It is frequent in machine learning to need to select the best hypothesis(h) given the dataset (d). One of the simplest ways to choose a hypothesis is to use what we already know about the situation. The Bayes' Theorem makes it possible to determine the likelihood of a hypothesis given prior knowledge. What is meant by the Bayes theorem is:

$$P(h/d) = (P(d/h) * P(h))/(P(d)) \qquad (4)$$

The value for P(h/d), which represents the probability of hypothesis h given the data d, is provided by equation (4). The posterior likelihood refers to this. P(d/h) represents the likelihood that data d would exist if hypothesis h were correct. The likelihood that an assertion is correct is expressed as P(h). The probability previous to h is what we refer to as. The probability of the facts is P(d).

4) Support Vector Machine: - The SVM training algorithm, a non-probabilistic conditional linear classifier, takes a set of training examples, each labelled as belonging to one of two categories, and builds a model that assigns new examples to one of the two categories. The purpose of utilising SVMs to assist us categorise our space is to find the best line in two dimensions or the best hyperplane in more than two dimensions. For both types of data, the hyperplane is identified using the maximum margin, or the greatest distance between data points (line).

5) Decision tree: - A tree-structured classifier is made up of core nodes, which reflect dataset properties, branches, which represent decision rules, and leaf nodes, which represent the outcome. A decision tree has two nodes: the leaf node and the judgement node. Decision nodes are used to make decisions and contain several branches, in contrast to leaf nodes, which are the results of decisions and tell us if the emotion is good, negative, or neutral. The main node or the starting point for further information is initially thought of as our dataset of tweets.

6) Random forest: - Random Forest is yet another supervised machine learning technique. A random forest is only a collection of decision trees, to put it simply. The Random Forest algorithm consists of two steps: building the random forest and making predictions using the random forest classifier created in the first step. A. Select "K" features at random from a total of "m" features such that "K"="M". b. Using the ideal division point of the "K" functions, identify the node "d." c. Use the best break to divide the network into daughter nodes. d. Steps 1 through 3 should be repeated until the lth node is reached. e. Repeat steps 1 through 5 as necessary to create a "n" number of trees.
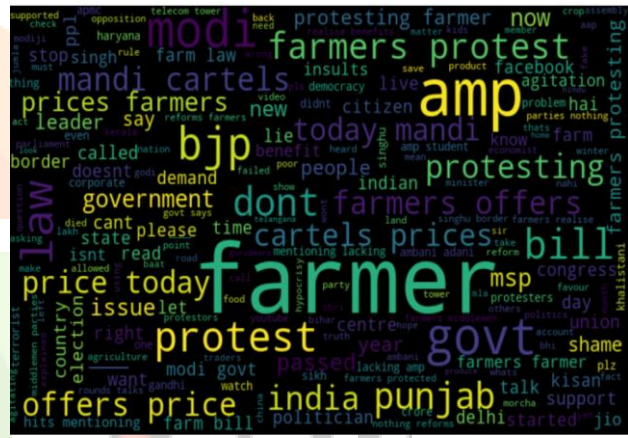
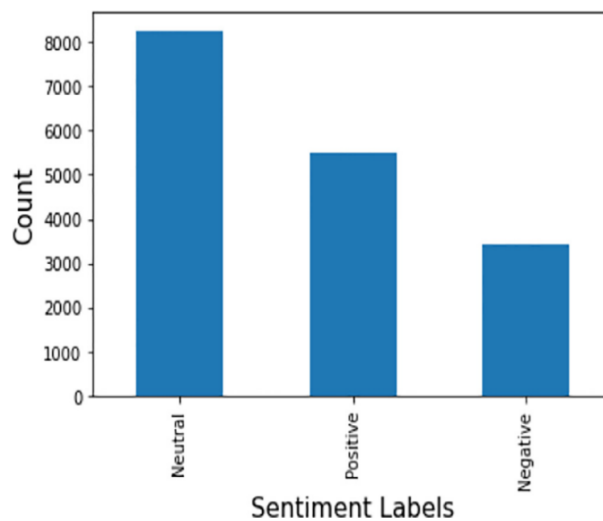|	Algorithms	Accuracy
1)	Naïve Bayes	71.33
2)	Support Vector Machine	83.05
3)	Random Forest	92.62
4)	Decision Tree	76.68

Table 1: Accuracy of Models

V. RESULTS
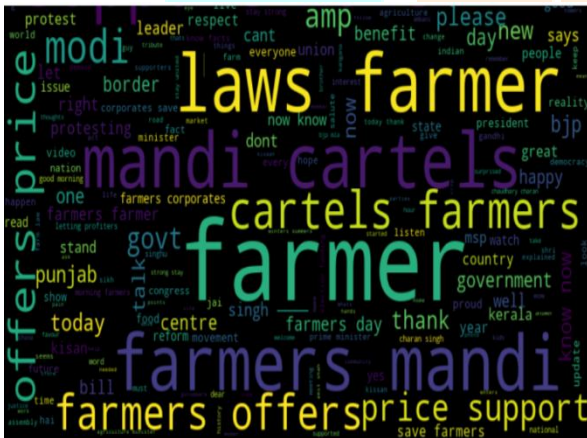
Fig.3 Word cloud of positive Sentiment

Fig.4. Word cloud of negative Sentiment

Fig.5.Bar plot of sentiment analysis

VI. REFERENCES

[1] Ashwin Sanjay Neogi, Kirti Anilkumar Garga, Ram Krishn Mishraa, Yogesh K Dwivedi, "Sentiment analysis and classification of Indian farmers' protest using twitter data".

[2] Anshuman Sahooi, Balaji Kharat and Devansh Gupta, Sameer Inamdar, "A Survey Paper on Sentimental Analysis on Nation-wide Protest using Social Network Analysis".

[3] Jishnusri Ojaswy Akella, LN Yashaswy Akella, "Sentiment Analysis Using Naïve Bayes Algorithm With Case Study".

[4] Dr. Akash D Dubey, "Twitter Sentiment Analysis during COVID19 Outbreak". DOI: 10.2139.

[5] Nida Afroz, Vijeta Sharma, Moinak Boral , Manjari Gupta, "Sentiment Analysis of COVID-19 Nationwide Lockdown effect in India ". DOI: 10.1109.

[6] Nida Afroz, Vijeta Sharma, Moinak Boral , Manjari Gupta, "Sentiment Analysis of COVID-19 Nationwide Lockdown effect in India

[7] Evgeny Gamletovitch Kim,Roman Klinger, "A Survey on Sentiment and Emotion Analysis for Computational Literary Studies".

[8] Munir Ahmad, Shabib Aftab, Muhammad Salman Bashir, Noureen Hameed, "Sentiment Analysis using SVM: A Systematic Literature Review".(IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 9, No. 2, 2018.

[9] Nurulhuda Zainuddin,Ali Selamat, "Sentiment analysis using Support Vector Machine".

[10] Tita Tosida, Erniyati, Krisna, "Sentiment Analysis Using the Support Vector Machine For Community Compliance Representation in The Covid-19 Pandemic Period". Proceedings of the 11th Annual International Conference on Industrial Engineering and Operations Management Singpore, March 7-11, 2021