

Twitter Network Sentimental Analysis on Vaccination

Deepak Nakka, Pavan Kalyan Nakka, Bhargav Pulipati

B. Tech Student, Department of CSE, Vellore Institute of Technology, Vellore, India

ABSTRACT

Globally, the COVID-19 pandemic has had an impact on daily life. Since the start of the pandemic, numerous research teams at significant pharmaceutical corporations and academic institutions throughout the world have been creating vaccines. Gender has an effect on vaccine responses, acceptability, and results. Additionally, the global advertising of the COVID-19 vaccine sparks a lot of conversations on social media outlets regarding the protection and effectiveness of vaccines, among other things. Twitter is viewed as one of the most popular social media sites that has been extensively used to communicate the public's thoughts on issues with the COVID-19 pandemic vaccination. However, there hasn't been enough research done to examine the analysis of the general public's view of the COVID-19 vaccine from a feminist perspective. The COVID-19 pandemic has been widely covered in social media, conventional print media, and electronic media since it first surfaced in December 2019. These sites provide data from reliable and unreliable medical sources. Additionally, the news from these mediums disseminates quickly. Spreading false information can cause anxiety, unintended exposure to medical treatments, digital marketing scams, and even lethal consequences. Therefore, it is imperative to develop a model for identifying bogus news in the news pool. The dataset employed in this work, which combines news about COVID-19 from various social media and news sources, is used for categorization.

Keywords: LSTM, CNN, Decision Tree, Random forest Classifier, Naïve bayes, SVM Logistic Regression, Gradient Boosting classifier, Sentiment Analysis and fake news detection.

Article Info

Publication Issue :

Volume 8, Issue 6

November-December-2022

Page Number : 493-504

Article History

Accepted: 05 Dec 2022

Published: 15 Dec 2022

I. INTRODUCTION

The globe has been severely afflicted by the new coronavirus disease (COVID-19), which has harmed billions of people. The COVID-19 virus has infected more than 260 thousand people, and it is responsible for more than 5 million fatalities, according to a new WHO report. In comparison to previous epidemics

like the Spanish Flu of 1918 and the Black Death of the middle ages, COVID-19 has significantly impacted public life and is regarded as the most devastating pandemic of this century. All nations now share the same objective of eradicating COVID-19 and resuming daily life. Closing borders, limiting work in public places (such as gyms, restaurants, and shopping malls), learning from home, limiting travel,

wearing masks, maintaining social distance, and maintaining personal hygiene are just a few of the official measures that many countries have adopted to slow the spread of COVID-19. These actions have successfully stopped the infection from spreading. However, the most recent COVID-19 Omicron variety raised new concerns for all nations. The most effective long-term method of managing the COVID-19 pandemic situation is vaccination since the existence of COVID-19 may be persistent in the future. Since the beginning of the pandemic epidemic, nations and pharmaceutical firms throughout the world have been developing vaccines and conducting clinical trials.

There are millions of posts regarding COVID-19 on social media. Even if some of the material shared on social media regarding COVID-19 is accurate and educational, the most of it is likely hearsay. Several fabricated images and films regarding the origin. Social media is utilised to convey information about COVID-19 fatalities, vaccinations, and viral transmission. The percentage of news, videos, and images shared on social media sites that are fraudulent is thought to be between 30 and 35 percent. This false information spreads more quickly than the virus itself, causing widespread fear. The fake news that circulated on social media between January and April 2020 may be divided into the following categories, according to research conducted by the International Fact Checking Network (IFCN): material on the virus's symptoms, causes, and treatments; official papers; its propagation; misrepresentations of videos and images; political commentary; and theories that blame certain groups, nations, or localities for the virus' spread. In certain nations, the economic crisis has been caused by the proliferation of bogus news on social media. For instance, false information about the possibility of COVID-19 infection in animals and birds and the risk of the virus spreading to humans caused some individuals in some countries to cease eating non-vegetarian food.

II. RELATED WORKS

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy n company strength. Once these things r satisfied, ten next steps are to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

Opinion mining and sentiment analysis on a Twitter data stream: The new Coronavirus Disease 2019 (COVID-19) epidemic has had a profound influence on people's daily lives all across the world. Governments enact a number of tactics and policies, including lockdown and social isolation, to battle the illness during the pandemic era. People may experience various mental health problems including depression, anxiety, melancholy, etc. as a result of these laws and regulations as well as the virus itself. In this study, we leverage the vast amounts of text data that Twitter users have uploaded to analyse the sentiment dynamics of residents of the Australian state of New South Wales (NSW) throughout the epidemic. We examine the dynamics of sentiment at the fine-grained local government levels, in contrast to the prior work, which mostly focuses on static sentiment analysis at the national level. We discovered that people in NSW displayed an overall positive sentimental polarity, and that the COVID-19 pandemic reduced the overall positive sentimental polarity even during pandemic period. This finding was based on the analysis of approximately 94 million tweets that were posted by approximately 183 thousand users situated in various LGAs in NSW in 5 months. Despite the predominant positive attitude predominating most days during the research period,

the fine-grained examination of sentiment in LGAs revealed that several LGAs saw large sentiment swings from positive to negative. This study also examined the emotive dynamics generated by popular Twitter subjects, including governmental initiatives (such as Australia's JobKeeper programme, lockdowns, and social-distancing) and targeted social gatherings (e.g. the Ruby Princess Cruise).

How bad is omicron? What scientists know so far:

The deadliest severe acute respiratory crisis coronavirus 2 (SARS-CoV-2) variant Omicron (B.1.1.529) has caused widespread concern due to its infectious and vaccine escape alterations. Critical pathogenicity and antibody resistance of the SARS-CoV-2 variant are controlled by mutations in the spike (S) protein integrin region. Omicron is a highly divergent variant with several mutations, including 26–32 alterations in the spike protein. Some of these mutations are associated with innate immunity escape potential and higher transmissibility. Four lineages make up the Omicron variation: B.1.1.529, BA.1, BA.2, and BA.3. According to the information that is now available, Omicron's overall risk is still rather high. Omicron has a growth advantage over Delta, leading to rapid community development and higher incidence levels than have previously been noted in this pandemic. The very high levels of transmission have increased hospitalization, continue to put a tremendous burden on healthcare systems in most countries, and may cause significant morbidity, especially in vulnerable populations, despite a lower risk of severe illness and death following infection than previous SARSCoV-2 variants.

The Impact of COVID-19 on Women and Children Experiencing Domestic Abuse, and the Life-Saving Services that Support Them:

For women and girls who already have experienced violence during the current COVID-19 epidemic, this brief emphasizes growing trends and their consequences for the delivery of key services (health, police and justice, social services, and coordination of these services). It details the steps done in collaboration with

governments, civil society groups, and UN bodies at the regional, national, and local levels. 1 These consist of the EU-UN Spotlight Initiative's promising practices, pilot and self-starter nations taking part in the UN Joint Global Programme on Essential Services for Women and Girls Subject to Violence, and the deployment of the Essential Services Package, the program's principal directive. It makes use of the expertise and experience of several a wide spectrum of stakeholders are in favour of ending intimate partner violence other types of abuse against women and girls, taking each particular case into consideration the national setting in which the crisis is taking place. It suggests actions to take. many partners, including major line ministries, civil society groups, and foreign organizations that want to enhance the standard of service all women and girls have access to integrated services across sectors, including those who are most vulnerable to violence and most marginalized.

Coronavirus disease 2019 (COVID-19) in Italy. JAMA:

The 2019 coronavirus illness (COVID-19) is now receiving international attention. Five to twenty percent of all COVID-19 positive individuals needed admission to an intensive care unit (ICU), and critically sick patients who developed acute respiratory distress syndrome had a death risk of almost fifty percent [1–5]. The World Health Organization (WHO), which was extremely alarmed by the spread and severity, declared COVID-19 to be a pandemic in March 2020. Due to the COVID-19 outbreak, Wuhan had a significant lack of medical personnel in February. Ten of the 3019 Chinese healthcare professionals who contracted COVID-19 and died as a result were reported to have perished by the Chinese health authorities [6]. The danger of infection for front-line healthcare workers is significant. Patient overcrowding, staff exhaustion, and inadequate precautions are viewed as important causes for health worker infections. We discovered that in the event of heavily delayed action, the potential benefits of restrictive pandemic P&C

measures would almost entirely disappear, and while a partially counterproductive work wait-and-see control policy might still be able to help contain the extent of epidemic spread even though its efficacy might be significantly reduced in comparison to a scenario of early diagnosis along with stringent P&C measures. From the standpoint of optimum management, the governments and the health authority's laissez-faire approach to dealing with the uncertainty of a COVID19-type pandemic emergence during the early stages of the pandemic comes out to be an high risk strategy since severe harm would result.

COVID-9 patient health prediction using boosted random forest algorithm: Artificial intelligence (AI) approaches are increasingly being integrated into wireless infrastructure, real-time data collecting, and end-user device processing. Today, it is exceptional to utilise AI to identify and anticipate pandemics on a massive scale. The entire population has been devastated by the COVID-19 epidemic, which started in Wuhan, China, and has put an excessive strain on the world's most developed healthcare institutions. Approximately 282,244 fatalities and over 4,063,525 confirmed cases have been reported globally as of May 11th, 2020, according to European Centre for Disease Prevention and Control organization. However, the present exponential and rapid expansion in the number of individuals has made accurate and timely prediction of an infected patient's potential outcomes for the purpose of effective treatment necessary. Real-time medical data collection and processing are necessary due to the size and scope of the healthcare sector. Additionally, the challenge of data management, which necessitates real-time prediction and information distribution to practitioners for prompt medical intervention, is at the heart of this sector. Major players in this sector, including doctors, vendors, hospitals, and health-related businesses, have made an effort to gather, manage, and resurrect data with the goal of using it to further technology innovation and medical

procedures. However, dealing with healthcare data has recently turned into a challenging undertaking because of the data's enormous amount, security concerns, shoddy wireless network applications, and the rate at which it is growing.

The Impact of the COVID-19 Pandemic on Mental Health of Children and Adolescents: After being originally identified as an epidemic in Wuhan, China, the coronavirus disease 2019 (COVID-19) outbreak was subsequently believed to have expanded to other countries and become a worldwide pandemic. The coronavirus had spread to more than 216 nations as of August 18th, 2020, with at least 21,756,357 confirmed cases and 771,635 fatalities worldwide. Millions of people were placed under lockdown when this epidemic was proclaimed a national emergency in several nations. The world society, particularly youngsters, have been severely disrupted by this unexpectedly enforced social seclusion. The majority of schools shuttered, cancelled lessons, and switched to home-based or online learning as one of the steps taken to stop the spread of the virus in order to promote and follow social distance rules. The coronavirus is affecting 67.6% of students' education and learning on a global scale in 143 different nations. Children's mental health may be at danger due to the enormous disruption caused by the move away from physical education programmes in the lives of kids and their families. For children's growing brains, this sudden change in the learning environment and a lack of social connections and activities presented an uncommon circumstance. Since many mental health illnesses first appear in youngsters, it is crucial and mandated that the scientific community and healthcare professionals evaluate the psychological effects of the coronavirus epidemic on children and adolescents. Worldwide, including in the US, nations are grappling with the challenge of coming up with effective methods for kids to lessen the psychological effects of coronavirus.

III. METHODOLOGY

In this methodology we are mentioning the concept which is going to implement in this project. That we called as proposed method is mentioned below.

Proposed system:

In our proposed model, we have created a system that which classifies the sentiment and the fake news as well with the help of some machine learning classification techniques. The block diagram of the proposed is shown in below diagram.

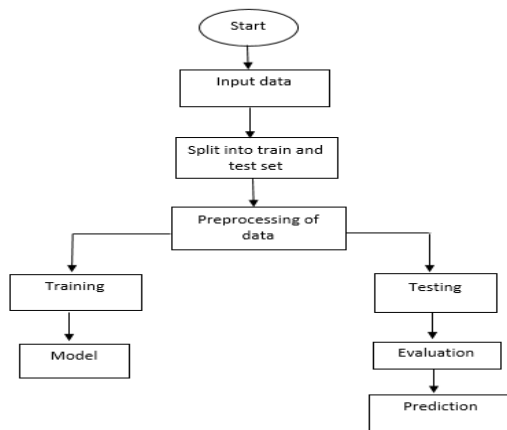


Figure 1 : Block diagram of proposed method

IV. Implementation

The project implementation contains the below listed algorithms.

1. Decision Tree Classifier:

A tree has many similarities in everyday life, and it turns out that it has affected both classification and regression in a broad range of machine learning. A decision tree can be used in decision analysis to formally and graphically reflect decisions and decision-making. It employs a decision-tree-like approach, as the name suggests. Although frequently used in data mining to generate a plan of action to accomplish a certain objective.

An upside-down decision tree is depicted, with the root at the top. In the left figure, the bold writing in black denotes an internal node or condition upon which the tree's branches and edges are built. The choice or leaf, in this example whether the passenger lived or died, lies at the end of the branch that doesn't divide anymore and is represented by red and green letters, respectively.

Although a real dataset will have many more attributes and this will just be one branch of a much larger tree, the simplicity of this technique cannot be overlooked. The significance of the trait is obvious, and relationships are simple to see. This process is more usually referred to as learning decision trees from data, and the tree above is known as a classification tree because the goal is to categorise passengers as having survived or having passed away. The only difference is that regression trees forecast continuous quantities, such as the price of a house. CART, or Classification and Regression Trees, is the common abbreviation for decision tree algorithms.

2. Random Forest Classifier:

A random forest is a machine learning method for tackling classification and regression issues. It makes use of ensemble learning, a method for solving complicated issues by combining a number of classifiers.

In a random forest algorithm, there are many different decision trees. The random forest algorithm creates a "forest" that is trained via bagging or bootstrap aggregation. The accuracy of machine learning algorithms is improved by the ensemble meta-algorithm known as bagging.

Based on the predictions of the decision trees, the (random forest) algorithm determines the result. It makes predictions by averaging or averaging out the

results from different trees. The accuracy of the result grows as the number of trees increases.

The decision tree algorithm's shortcomings are eliminated with a random forest. It improves precision and decreases dataset overfitting. Without requiring several settings in packages, it makes.

Features of a Random Forest Algorithm:

- It's more accurate than the decision tree algorithm.
- It provides an effective way of handling missing data.
- It can produce a reasonable prediction without hyper-parameter tuning.
- It solves the issue of over fitting in decision trees.
- In every random forest tree, a subset of features is selected randomly at the node's splitting point.

3. SUPPORT VECTOR MACHINES:

Support Vector Machine(SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well its best suited for classification. The objective of SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three.

4. SVM Kernel:

The SVM kernel is a function that takes low dimensional input space and transforms it into higher-dimensional space, ie it converts not separable problem to separable problem. It is mostly useful in

non-linear separation problems. Simply put the kernel, it does some extremely complex data transformations then finds out the process to separate the data based on the labels or outputs defined.

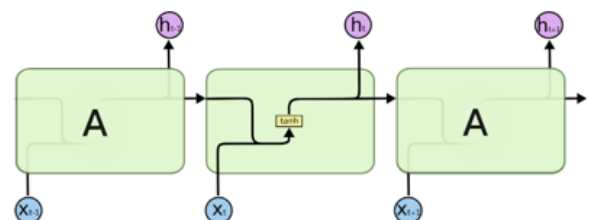
Advantages of SVM:

- Effective in high dimensional cases
- Its memory efficient as it uses a subset of training points in the decision function called support vectors
- Different kernel functions can be specified for the decision functions and its possible to specify custom kernels

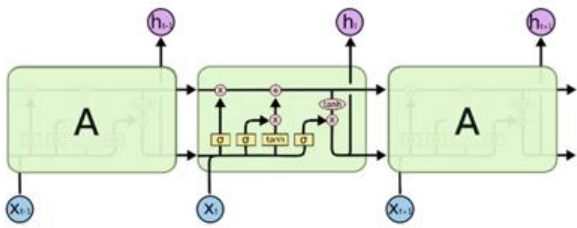
5. LSTM:

Long Short Term Memory Networks, most often referred to as "LSTMs," are a unique class of RNN that can recognize long-term dependencies. They were first presented by Hoch Reiter & Schmidhuber (1997), and several authors developed and popularized them in subsequent works. They are currently frequently utilised and perform incredibly well when applied to a wide range of issues.

Intentionally, LSTMs are created to prevent the long-term reliance issue. They don't strain to learn; rather, remembering knowledge for extended periods of time is basically their default habit. All recurrent neural networks have the shape of a series of neural network modules that repeat. This repeating module in conventional RNNs will have a fairly straightforward structure, such as a single tan h layer.



LSTMs also have this chain like structure, but the repeating module has a different structure. Instead of having a single neural network layer, there are four, interacting in a very special way.



Each line in the aforementioned figure conveys a full vector from one node's output to another's input. The yellow boxes are learnt neural network layers, and the pink circles represent point-wise operations like vector addition. Concatenation is shown by lines merging, whereas lines forking indicate that their content has been replicated and is being sent to other destinations.

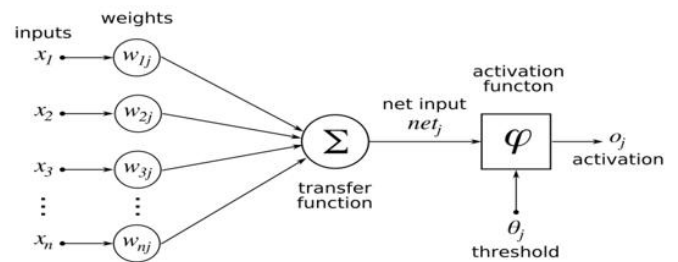
6. Convolutional neural network (CNN):

A component of a computing system called an Convolutional neural network (CNN) is made to mimic how the human brain evaluates and processes information. It serves as the cornerstone of artificial intelligence (AI) and resolves issues that, by human or statistical standards, would be impossible or challenging. Because CNNs are self-learning, they can provide better outcomes as more data becomes available.

A CNN consists of hundreds or thousands of processing units, which are artificial neurons connected by nodes. Input and output units make up these processing units. The neural network makes an effort to learn about the information supplied to generate one output report from the input units, which receive diverse forms and structures of information based on an internal weighting system. Back propagation, which stands for backward propagation of error, is a set of learning rules that CNNs use to refine their output results, much as people use rules and guidelines to produce a result or output.

A CNN first goes through a period of training when it learns to identify patterns in data, whether they be textual, auditory, or visual. The network compares its actual output with the output it was intended to

produce—the expected output—during this supervised phase. The back propagation method is used to correct the discrepancy between the two results. This implies that the network adjusts the weight of its connections between the units backward, from the output unit to the input units, until the disparity between the actual and anticipated outcome generates the lowest possible error.



7. Logistic Regression:

Logistic Regression was used in the biological sciences in early twentieth century. It was then used in many social science applications. Logistic Regression is used when the dependent variable (target) is categorical.

For example,

To predict whether an email is spam (1) or (0)

Whether the tumor is malignant (1) or not (0)

Consider a scenario where we need to classify whether an email is spam or not. If we use linear regression for this problem, there is a need for setting up a threshold based on which classification can be done. Say if the actual class is malignant, predicted continuous value 0.4 and the threshold value is 0.5, the data point will be classified as not malignant which can lead to serious consequence in real time.

From this example, it can be inferred that linear regression is not suitable for classification problem. Linear regression is unbounded, and this brings logistic regression into picture. Their value strictly ranges from 0 to 1.

Purpose and examples of logistic regression:

Logistic regression is one of the most commonly used machine learning algorithms for binary classification problems, which are problems with two class values,

including predictions such as “this or that,” “yes or no” and “A or B.”

The purpose of logistic regression is to estimate the probabilities of events, including determining a relationship between features and the probabilities of particular outcomes.

One example of this is predicting if a student will pass or fail an exam when the number of hours spent studying is provided as a feature and the variables for the response has two values: pass and fail.

Organizations can use insights from logistic regression outputs to enhance their business strategies so they can achieve their business goals, including reducing expenses or losses and increasing ROI in marketing campaigns, for example.

An e-commerce company that mails expensive promotional offers to customers would like to know whether a particular customer is likely to respond to the offers or not. For example, they’ll want to know whether that consumer will be a “responder” or a “non responder.” In marketing, this is called propensity to respond modeling.

Likewise, a credit card company develops a model to decide whether to issue a credit card to a customer or not will try to predict whether the customer is going to default or not on the credit card based on such characteristics as annual income, monthly credit card payments and number of defaults. In banking parlance, this is known as default propensity modeling.

Uses of logistic regression:

Logistic regression has become particularly popular in online advertising, enabling marketers to predict the likelihood of specific website users who will click on particular advertisements as a yes or no percentage.

- Logistic regression can also be used in:
- Healthcare to identify risk factors for diseases and plan preventive measures.
- Weather forecasting apps to predict snowfall and weather conditions.

- Voting apps to determine if voters will vote for a particular candidate.

Insurance to predict the chances that a policy holder will die before the term of the policy expires based on certain criteria, such as gender, age and physical examination.

Banking to predict the chances that a loan applicant will default on a loan or not, based on annual income, past defaults and past debts.

Logistic regression vs. linear regression:

The main difference between logistic regression and linear regression is that logistic regression provides a constant output, while linear regression provides a continuous output.

In logistic regression, the outcome, such as a dependent variable, only has a limited number of possible values. However, in linear regression, the outcome is continuous, which means that it can have any one of an infinite number of possible values.

Logistic regression is used when the response variable is categorical, such as yes/no, true/false and pass/fail. Linear regression is used when the response variable is continuous, such as number of hours, height and weight.

For example, given data on the time a student spent studying and that student’s exam scores, logistic regression and linear regression can predict different things.

With logistic regression predictions, only specific values or categories are allowed. Therefore, logistic regression can predict whether the student passed or failed. Since linear regression predictions are continuous, such as numbers in a range, it can predict the student’s test score on a scale of 0 -100.

8. Naïve Bayes:

A Naive Bayes classifier is a probabilistic machine learning model that’s used for classification task. The crux of the classifier is based on the Bayes theorem.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Using Bayes theorem, we can find the probability of **A** happening, given that **B** has occurred. Here, **B** is the evidence and **A** is the hypothesis. The assumption made here is that the predictors/features are independent. That is presence of one particular feature does not affect the other. Hence it is called naive.

Let's understand it using an example. Below I have a training data set of weather and corresponding target variable 'Play' (suggesting possibilities of playing). Now, we need to classify whether players will play or not based on weather condition. Let's follow the below steps to perform it.

Step 1: Convert the data set into a frequency table

Step 2: Create Likelihood table by finding the probabilities like Overcast probability = 0.29 and probability of playing is 0.64.

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table		
Weather	No	Yes
Overcast	3	4
Rainy	3	2
Sunny	2	3
Grand Total	5	9

Likelihood table			
Weather	No	Yes	
Overcast	3	4	=4/14 0.29
Rainy	3	2	=5/14 0.36
Sunny	2	3	=5/14 0.36
All	5	9	
	=5/14	=9/14	
	0.36	0.64	

Step 3: Now, use Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.

Problem: Players will play if weather is sunny. Is this statement is correct?

We can solve it using above discussed method of posterior probability.

$$P(\text{Yes} | \text{Sunny}) = P(\text{Sunny} | \text{Yes}) * P(\text{Yes}) / P(\text{Sunny})$$

$$\text{Here we have } P(\text{Sunny} | \text{Yes}) = 3/9 = 0.33, P(\text{Sunny}) = 5/14 = 0.36, P(\text{Yes}) = 9/14 = 0.64$$

$$\text{Now, } P(\text{Yes} | \text{Sunny}) = 0.33 * 0.64 / 0.36 = 0.60, \text{ which has higher probability.}$$

Naive Bayes uses a similar method to predict the probability of different class based on various attributes. This algorithm is mostly used in text

classification and with problems having multiple classes.

- It is easy and fast to predict class of test data set. It also perform well in multi class prediction
- When assumption of independence holds, a Naive Bayes classifier performs better compare to other models like logistic regression and you need less training data.
- It perform well in case of categorical input variables compared to numerical variable(s). For numerical variable, normal distribution is assumed (bell curve, which is a strong assumption).

Applications of Naive Bayes Algorithms:

Real time Prediction: Naive Bayes is an eager learning classifier and it is sure fast. Thus, it could be used for making predictions in real time.

Multi class Prediction: This algorithm is also well known for multi class prediction feature. Here we can predict the probability of multiple classes of target variable.

Text classification/ Spam Filtering/ Sentiment Analysis: Naive Bayes classifiers mostly used in text classification (due to better result in multi class problems and independence rule) have higher success rate as compared to other algorithms. As a result, it is widely used in Spam filtering (identify spam e-mail) and Sentiment Analysis (in social media analysis, to identify positive and negative customer sentiments).

9. Gradient Boosting Classifier: Gradient boosting algorithm is one of the most powerful algorithms in the field of machine learning. As we know that the errors in machine learning algorithms are broadly classified into two categories i.e. Bias Error and Variance Error. As gradient boosting is one of the boosting algorithms it is used to minimize bias error of the model.

Unlike, Ada boosting algorithm, the base estimator in the gradient boosting algorithm cannot be mentioned by us. The base estimator for the Gradient Boost algorithm is fixed and i.e. Decision Stump. Like, AdaBoost, we can tune the $n_estimator$ of the gradient boosting algorithm. However, if we do not mention the value of $n_estimator$, the default value of $n_estimator$ for this algorithm is 100.

Gradient boosting algorithm can be used for predicting not only continuous target variable (as a Regressor) but also categorical target variable (as a Classifier). When it is used as a regressor, the cost function is Mean Square Error (MSE) and when it is used as a classifier then the cost function is Log loss.

Let us now understand the working of the Gradient Boosting Algorithm with the help of one example. In the following example, Age is the Target variable whereas Likes Exercising, GotoGym, Drives Car are independent variables. As in this example, the target variable is continuous, Gradient Boosting Regressor is used here.

Let us now find out the estimator-2. Unlike AdaBoost, in the Gradient boosting algorithm, residues ($age_i - \mu$) of the first estimator are taken as root nodes as shown below. Let us suppose for this estimator another dependent variable is used for prediction. So, the records with False GotoGym.

4. Results and Discussion:

The following images will visually depict the process of our project.

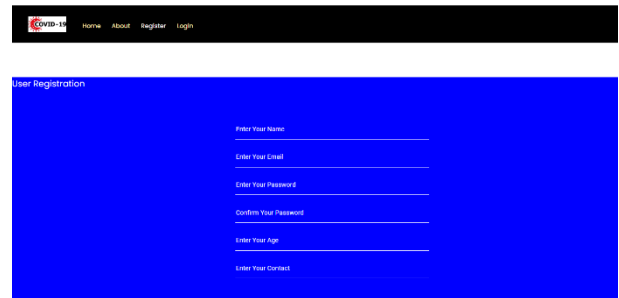
Home page: This is the home page of the web application.



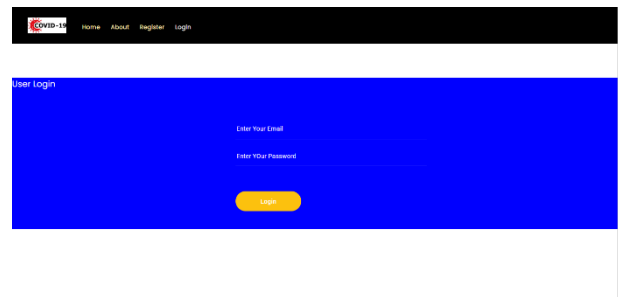
ABOUT: Here the user can see the about page of the web application.



Registration: In this page user can register him/herself by entering the details.



Login: Here the user can log in with the valid credentials.



User home: After successful login user will enter into the user home page.



View: Here we can see the sentiment data.

ID	user_name	user_location	user_description
14622210170989	Yashraj GPT	Los Angeles, California, CA	Appreciate of Japan's medical services, especially the excellent COVID-19 vaccine
13338164439204243	JBaer1 Long	San Francisco, CA	Modeling shows both peak, heavy rainfall & 5th major jet flow, forecasted for meteorology and LR
1073333333333333	ahhha	Your Best	Real, high

Sentiment prediction: Here we can predict the sentiment of the tweet data on covid 19 vaccination.

The Entered Text Has Positive Sentiment

Enter Your Text to Classify

Predict

View fake data: Here we see our fake news data.

ID	text	label
1	The CDC currently reports 1600 deaths. It generated the discrepancies in death counts between different sources are small and explainable. The death toll stands at roughly 16000 people today.	fake
2	States reported 152 deaths is small rise from last Tuesday. Southern states reported 148 of those deaths https://twitter.com/ABC7News	fake
3	Politically Correct Women (PCW) have been denied as excuse not to leave Plastic Bag https://www.youtube.com/watch?v=... #PCW	fake
4	Wendy's California has been found #COVID testing laboratory in India and so on 20th August 2022. 1000000 tests have been done. @WendysCA @WendysCA #COVID19 #WendysCA #WendysCA #WendysCA #WendysCA	fake
5	Phishing sites are generated to get more counts but if you look at the new cases per million today's similar status are coming more cases per million than California or Texas. AL, HI, ID, IL, IN, IA, MI, MN, MO, ND, SD, NE, KS, OK, TX, VA, WV, DC, HI, AK, PR, GU, VI, AS, FM, MK, PW, FJ, TO, NU, NC, NZ, PG, PY, RE, TF, TG, TT, TV, UM, VG, VI, WF, WS, YE, ZM, ZW	fake
6	COVID-19 has found 1% coverage each person in India with COVID-19 in finding 11 other sources. Data or news that the infection growth rate has declined over time. This factors in the observed increase in order and other variables such as the spread of the disease. https://www.who.int/news-room/press-releases/2022/08/18/covid-19-in-india	fake
7	If you believe positive for #COVID19 and have no symptoms, stay home or at least from other people. Learn more about CDC's recommendation about when you can be around others after COVID-19 infection.	fake

News Prediction: Here we can predict by entering a news regarding covid whether it is a fake or real news.

Detecting Detecting Fake News During Covid-19

This is A Real News

Enter Your Text to Classify

Predict

V. CONCLUSION

Sentiment analysis on twitter dataset is done on the recent issue of demonetization. The analysis shows largely a positive opinion on the move though negative opinion is also present. Various machine learning algorithms for classification are used for the results comparison. For our dataset, LSTM shows the maximum accuracy. The major disadvantage of our

approach is that many tweets are written in local languages like Hindi. The Hindi words are written in English language and we didn't include any such words in the positive and negative corpus. Because of this all tweets in Hindi, written in English are classified as neutral. In future those words can also be included and accuracy can be increased for analysis. As future work similar techniques like can be used to reduce the neutral count.

VI. REFERENCES

- [1]. Jianlong Zhou, Yang Shuiqiao, Xiao Chun, Chen Fang. Examination of community sentiment dynamics due to COVID-19 pandemic: a case study from a state in Australia. SN Comput Sci 2021;2(3):1–11.
- [2]. Ewen Callaway, Ledford Heidi. How bad is omicron? What scientists know so far. Nature 2021;600(7888):197–9.
- [3]. Women's Aid, UK. The Impact of COVID-19 on Women and Children Experiencing Domestic Abuse, and the Life-Saving Services that Support Them. Women's Aid UK; 2020.
- [4]. Livingston E, Bucher K. Coronavirus disease 2019 (COVID-19) in Italy. JAMA. (2020) 323:1335–5. doi: 10.1001/jama.2020.4344
- [5]. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, et al. COVID-19 patient health prediction using boosted random forest algorithm. Front Public Health. (2020) 8:357. doi: 10.3389/fpubh.2020.00357
- [6]. Srivastava G. The Impact of the COVID-19 Pandemic on Mental Health of Children and Adolescents. (2020). Available online at: <http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-414191>
- [7]. Jiahua Du, Michalska Sandra, Subramani Sudha, Wang Hua, Zhang Yanchun. Neural attention with character embeddings for hay fever detection from twitter. Health Inf Sci Syst 2019;7(1):1–7.

- [8]. Rubina Sarki, Ahmed Khandakar, Wang Hua, Zhang Yanchun. Automated detection of mild and multi-class diabetic eye diseases using deep learning. Health Inf Sci Syst 2020;8(1):1–9.
- [9]. Wang Xinjue, Deng Ke, Li Jianxin, Yu Jeffery Xu, Jensen Christian S, Yang Xiaochun. Efficient targeted influence minimization in big social networks. World Wide Web 2020;23(4):2323–40.
- [10]. Qiao Tian, Li Jianxin, Chen Lu, Deng Ke, Li Rong-hua, Reynolds Mark, et al. Evidence-driven dubious decision making in online shopping. World Wide Web 2019;22(6):2883–99

Cite this article as :

Deepak Nakka, Pavan Kalyan Nakka, Bhargav Pulipati, "Twitter Network Sentimental Analysis on Vaccination", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 6, pp. 493-504, November-December 2022.

Journal URL : <https://ijsrcseit.com/CSEIT228674>