

Machine Learning Algorithm to Identify Hesitancy Towards Covid-19 Vaccination Among the Rural Citizens Under E-Governance Initiatives

M.Bhuvana¹, K.Priya², Suresh.V³, Jyothi Palukiri⁴, Premkumar. B⁵

¹Assistant Professor, Faculty of Management
SRM Institute of Science & Technology (SRMIST)
Kattankulathur, India
bhuvitejal@gmail.com

²Assistant Professor, Faculty of Management
SRM Institute of Science & Technology (SRMIST)
Ramapuram
priyak1@srmist.edu.in

³Assistant Professor, Faculty of Management
SRM Institute of Science & Technology (SRMIST)
Kattankulathur, India
sureshv@srmist.edu.in

⁴Associate Professor, MBA Department
Saveetha Engineering College
jyothip@saveetha.ac.in

⁵Associate Professor, Faculty of Management
SRM Institute of Science & Technology (SRMIST)
Kattankulathur, India
premkumb@srmist.edu.in

Abstract— The Indian Government has taken massive initiatives to control the pandemic virus of COVID-19. Recently, the Government has decided to begin the process of a large-scale vaccination programme to create an end to the crisis due to COVID-19. However, vaccination is found to be the major solution as suggested by various health care experts to stop the widespread of this deadly virus, hesitancy towards getting vaccinated is found to be more in the rural villages of India. The process of getting vaccinated was not made compulsory by the government, but still there exists skepticism in minds of rural citizens towards COVID-19 vaccination. Text Analytics has been implemented to identify the exact emotions among rural citizens towards COVID Vaccinations. Hence the present research study has identified the major reasons for hesitancy towards getting vaccinated for COVID-19 using machine learning techniques. The researchers of the current study have also measured the emotions and signs of sentiments of the rural citizens on COVID-19 vaccination.

Keywords- COVID-19; E-Governance; Emotions; Vaccination; Rural Citizens.

I. INTRODUCTION

On 31st December 2019, in China, the WHO (World Health Organization) has identified the new pneumonia deadly virus called “Novel Coronavirus 2019”. A cluster of cases has been reported by developed and developing countries all over the world. The entire country across the globe has been under complete lockdown for more than eight months to control the spread of this deadly virus. This virus directly infects the respiratory tract of the human body and it can be easily transferred between humans [4]. Several guidelines and safety measures have been suggested by WHO to all the countries to safeguard their people. Maintaining social distancing, wearing

face masks, and frequently sanitizing the hands are the suggestion posted by the WHO to control the spread of the virus from one person to another person. The Government of India has taken multiple initiatives during this pandemic situation of COVID-19 to protect the lives of the people especially in remote villages as the maximum space of the entire population is occupied by the rural population [6].

II. INITIATIVES TAKEN BY THE INDIAN GOVERNMENT TO CONTROL COVID-19

To control the large spread of Coronavirus across the states, a large number of coronavirus vaccinations have been investigated globally under several technologies and platforms.

Vaccinations for COVID-19 have been facilitated with two dosages to the citizens. In India, about 183 million dosages of vaccines have been provided and about 40 million citizens have obtained their second dosages of COVID vaccination. Covaxin, Covishield, and Sputnik V are the three different vaccinations approved by the Health Ministry of India [10].

A. E-GOVERNANCE INITIATIVES FOR VACCINE REGISTRATION

The online platform has been constructed by the e-governance system of India under Government to Citizen services for registering the COVID-19 vaccination. Co-WIN, Aarogya Setu Application, and UMANG are the three different platforms developed under the Ministry of Health and Family Welfare India to reach COVID-19 vaccination to every citizen in the country.

- **Co-WIN Application:** It is an application developed for registering the COVID-19 vaccination. The citizens can access this application and confirm their appointment and choose their nearby health care centers for vaccination through their registered mobile number. The one-time password will be sent to the registered mobile number for confirming the registration [12].
- **Aarogya Setu Application:** It is a mobile application developed by the Ministry of Electronics and Information Technology. This app detects and indicates the nearby areas which were affected by COVID-19. This application also has a special feature that directs the URL towards the Co-WIN application for vaccination registration [1].
- **UMANG Application:** This application is defined as the Unified Mobile Application for New-Age Governance that was structured by the Ministry of Electronics and Information Technology and the National e-governance division. This application also directs the URL towards the Co-WIN application through their Aadhar Number [14].

III. COVID VACCINATION ACCEPTANCE AMONG THE RURAL CITIZENS

Refusal and Hesitancy for COVID-19 vaccination have been identified among the rural citizens. People from the slum areas and remote villages have very low literacy levels, lack of confidence and awareness for getting vaccinated to fight against the Coronavirus [2]. In India, the major space of population in India is occupied by remote villages. Hence the Indian government must encourage and educate the rural citizens to get vaccinated and safeguarding their lives. Researchers have stated that proper knowledge and information turn the negative

attitude of rural people towards COVID vaccination. Lack of technology is found to be another obstacle faced by the rural people for accessing the e-governance system for vaccination registration. Moreover, the information generated in the application developed for COVID vaccination must be in the local language for making the rural people understand and follow the instructions[3]. Poor knowledge and misinformation spread through social media make the rural people get a fear getting vaccinated for COVID-19 [9]. Hence in the present study, the researcher aims to identify the primary reasons for hesitancy towards COVID-19 vaccination among rural citizens.

IV. RESEARCH METHODS

The major objective of this study is to analyze the reasons among rural citizens for avoiding the COVID-19 vaccination. The current research study is exploratory in creation. The researchers have maintained a qualitative approach for conducting the study. Authors have stated that a qualitative study has been made in natural settings and attempting to interpret the meaning of people's life. It has been further argued that qualitative research includes the collection and use of empirical materials such as interviews and personal experiences (Denzin & Lincoln, 1998). The researchers have selected the 500 citizens (245 males and 255 females) who belong to the 18 to 60 years of the age group who are not taken COVID-19 vaccination in the rural villages of the Kanchipuram district. The rural respondents of the research study have been selected by adopting the purposive non-probability sampling technique (Miller & Chandler, 2002). An in-depth face-to-face interview has been done among the respondents. The respondents have been interviewed to give the five major reasons for not taking the COVID-19 vaccination. The interview has been connected between 5th March 2021 to 2nd April 2021. The Choice of the researchers for opting rural areas of Kanchipuram district has been rationalized with the literacy level (76%) of Kanchipuram district. It is found to be lesser than the average literacy rate (80%) of the state of Tamil Nadu (Census, 2011). About 10-15 minutes have been taken to complete the interview. The handwritten notes have been taken and recorded by the researchers for further analysis.

V. DATA ANALYSIS

The researchers have used Text Analytics in r studio for analyzing the data gathered by the respondents. The imported data from the rural respondents have been converted into a text file. This file has been given as the input for performing sentimental analysis in r studio. Sentimental analysis is defined as an automatic method of analyzing and extracting the subjective judgments on various aspects of an entity or an item [13]. It is a process of executing the machine learning algorithm that determines the usage of Natural Language Processing

(NLP) to find the emotions expressed through words mentioned in the given text [8]. Word Associations, Word Cloud, Emotional Classifications, and Sentimental Scores have been generated through sentimental analysis for the data gathered through feedback, customer reviews, survey responses, and comments from the social media posts [11]. In the present study, the gathered data from the interview has been converted into the text file and given as an input for sentimental analysis in r studio to identify the emotions and factors that leads to the hesitancy of rural citizens to get vaccinated for COVID-19.

A. Word Frequency

Table 1 displays the frequency of words in the text file. Fear, death, health disorder, and side effects are the words that have been mentioned in the text file.

Table 1: Word Frequency

word	freq
fear	155
death	146
health	94
disord	93
effect	63
side	63
vaccin	31
allerg	30
reaction	30
lack	28
clariti	23
clinic	23
trial	23
trust	5
enforc	5
inform	3
media	3
social	3
condit	1
poor	1

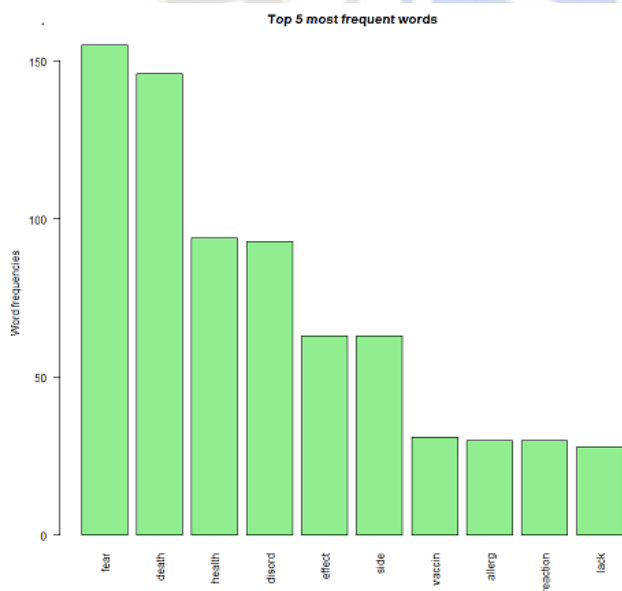


Figure 1: Top 5 Most Frequent Words

Figure 1 displays the Bar graph of the most frequent words in the text file. From the graph, it is clear that the words “fear”, “death”, “health”, “disorder”, “effect”, and “side” are the top five words that have to be analyzed deeply.

B. Word Cloud Generation

A Word Cloud is the best method to analyze and visualize qualitative research data. It is represented as an image associated with keywords displayed in the body of text, whereas the size of all the words represents its frequency present in the body of text. Figure 2 represents the generation of Word Cloud for the text given as the input file. Figure 2 indicates that the words like “health, disorders, fear, side, effects, allergic, and reactions” have maximum frequencies in the input file.



Figure 2: Word Cloud Generation

C. Word Association

Word Association is described as the statistical technique that defines the degree of strongly paired variables from the survey responses. This methodology helps to identify and measure the words that occur often in association with other frequently used words in the survey responses. Table 2 displays the association of words found in the text file given as an input. Table 2 highlights that “Health Disorders”, Fear of Death” and “Side Effects” are the strongly associated words represented in the input file.

Table 2: Word Association

SNO	Text
1.	health disord 0.99
2.	fear death 0.95
3.	effect side 1

D. Sentiment Scores

Sentiment Scores represent the Positive, Neural, or Negative comments highlighted in the input file. It is expressed in terms of numerical values, to explore the negative or positive strength of the sentiment represented in the body of text. Figure 3 represents the Syuzhet Vector of input text. The values of Syuzhet Vector of input text show that the value generated by the first element of the input text is found to be -0.75. The sentiment scores using the Syuzhet Vector method are represented using the scale ranges between -1 to +1. The first element of the input file displays a score of -0.75. Hence it indicates the negative responses have been indicated in the text file given as the input for sentimental analysis. Moreover, the median score generated in the text file is found to be -0.25 which below zero and it is interpreted as the overall sentimental average score value for all the responses is found to be negative.

```
> syuzhet_vector <- get_sentiment(text, method="syuzhet")
> head(syuzhet_vector)
[1] -0.75 -0.25 -0.25 -0.25 0.00 -0.25
> summary(syuzhet_vector)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-1.5000 -1.5000 -0.2500 -0.5984  0.0000  0.6000
```

Figure 3: Syuzhet vector

Figure 4 represents the sentiment score calculated through the bing vector and afinn vector. From figure 4 it is clear that the median values evaluated from the bing vector and afinn vector are found to be negative and hence the is negative responses have been given by the rural citizens towards COVID-19 Vaccination.

```
> bing_vector <- get_sentiment(text, method="bing")
> head(bing_vector)
[1] -1 0 0 0 0 0
> summary(bing_vector)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-2.0000 -2.0000  0.0000 -0.7981  0.0000  0.0000
> afinn_vector <- get_sentiment(text, method="afinn")
> head(bing_vector)
[1] -1 0 0 0 0 0
> afinn_vector <- get_sentiment(text, method="afinn")
> head(afinn_vector)
[1] -2 -1 -1 -1 0 -1
> summary(afinn_vector)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-4.000 -4.000 -2.000 -2.055  0.000  0.000
```

Figure 4: Bing Vector and Afinn Vector

Figure 5 displays the comparison of all the three vectors (Syuzhet, Bing, and Afinn). From figure 5 it can be interpreted that the first element of each row represents the vector value of -1. Hence from all three methods, it is proved that negative sentimental scores have been generated for the text given as an input file.

```
> rbind(
+   sign(head(syuzhet_vector)),
+   sign(head(bing_vector)),
+   sign(head(afinn_vector))
+ )
      [,1] [,2] [,3] [,4] [,5] [,6]
[1,]  -1  -1  -1  -1  -1  0  -1
[2,]  -1  0  0  0  0  0  0
[3,]  -1  -1  -1  -1  -1  0  -1
```

Figure 5: Comparison of Syuzhet, Bing, and Afinn Vectors

E. Emotional Classification

The Emotional Classification in the text file has been analyzed by the “NRC Emotional Lexicon”. It represents the list of words in the English Language and defines its relations with eight different basic emotions namely “Disgust, Joy, Sadness, Surprise, Trust, Anticipation, Fear, and Anger”. Also, Positive and Negative are the two different sentiments represented by the NRC Emotional Lexicon. Figure 6 represent the output of emotional classification generated through get_nrc_sentiment().

```
> d<-get_nrc_sentiment(text)
> head(d,10)
      anger anticipation disgust fear joy sadness surprise trust negative positive
1         1             0         0         1         0         0         0         0         1         0
2         0             0         0         0         0         0         0         0         1         0
3         0             0         0         0         0         0         0         0         1         0
4         0             0         0         0         0         0         0         0         1         0
5         0             0         0         0         0         0         0         0         0         0
6         0             0         0         0         0         0         0         1         1         0
7         0             0         0         0         0         0         0         0         0         0
8         0             0         0         0         0         0         0         0         0         1
9         0             0         0         0         0         0         0         0         0         0
10        0             0         0         0         0         0         0         0         0         0
```

Figure 6: Emotional Classification

From Figure 6, it is clear that there exists a “0” occurrences in anticipation, disgust, joy, sadness, and surprise. Occurrences of “1” in anger, fear, trust, negatives, and positives. Figure 7 displays the bar graph that represents the association of words towards emotions. From figure 6 it is proved that the negative emotions “Anger”, and “Fear” have occurred maximum times in the input file. Figure 8 represents the percentage of words in the text file associated with emotions. From figure 8 it is found that about 80 % of the space in the text file has been occupied by the emotions “Anger, and Fear”.

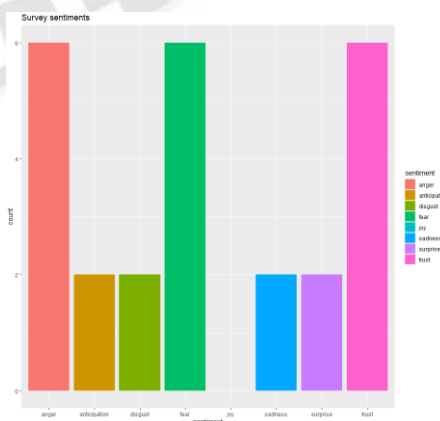


Figure 7: Association of Words with Emotions

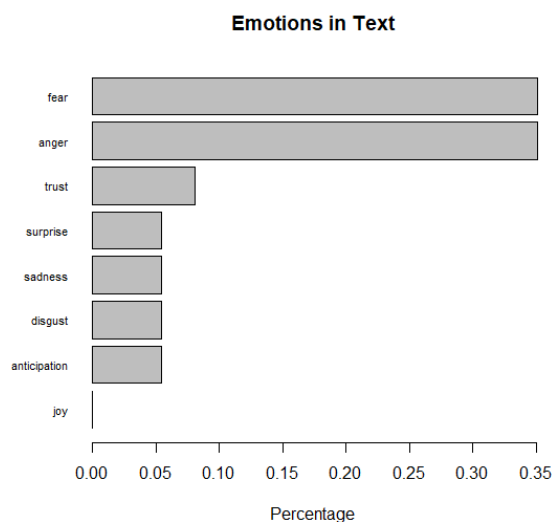


Figure 8: Percentage of Association of Words with Emotion.

VI. FINDINGS AND DISCUSSION

From Text Analytics, the researchers have adopted the sentimental analysis to identify the major reasons among the rural citizens for the hesitancy towards COVID-19 vaccination. From the sentimental analysis, it is found that the “Health Disorders”, “Fear of Death” and “Side effects are the major reasons highlighted by the rural citizens for their hesitancy towards getting vaccinated for COVID-19. Poor knowledge and misinformation spread through social media make the rural people get a fear getting vaccinated for COVID-19 (Saiful Islam, et.al, 2021). The emotional classification generated from the text analysis proves that “Anger and Fear” are highly associated emotions of the rural citizens towards their reasons mentioned for avoiding COVID-19 vaccination. Lack of technology is found to be another obstacle faced by the rural people for accessing the e-governance system for vaccination registration.

Moreover, the information generated in the application developed for COVID vaccination must be in the local language for making the rural people understand and follow the instructions [3]. The sentimental scores proved that the rural citizens have given negative comments on COVID-19 vaccines. People from the slum areas and remote villages have very low literacy levels, lack of confidence and awareness for getting vaccinated to fight against the Coronavirus [2]. Hence the Government of India should focus on the factors explored by the rural citizens for getting COVID-19 vaccination. The government officials should address the fear of the rural citizens and take immediate and appropriate actions for the spreading misinformation to the rural citizens regarding COVID-19 vaccination.

VII. RECOMMENDATIONS FOR FURTHER RESEARCH AND CONCLUSION

The present study has recognized the primary reasons highlighted by the rural citizens for hesitancy in taking COVID-19 Vaccination using Machine Learning techniques. The authors of the current study have also analyzed the signs of sentiments and emotions of the rural citizens for taking vaccination for COVID-19. Further research study could be done by the researchers for analysing the emotions and sentiments among the citizens who have taken two dosages of COVID-19 Vaccination and Booster dosage as well. Citizens’ improvement in health and the status of the immune system in their body during post-pandemic situations can also be identified by adopting the similar methodology of the present study.

Even though a huge number of innovative initiatives were taken by the Indian Government for COVID-19 vaccination, there are several uncertainties found among the citizens residing in rural villages regarding COVID-19 vaccination. The findings of the research study disclosed that rural citizens still have a feeling of getting distressed about the COVID-19 pandemic that results in rejecting COVID-19 vaccination. The researchers have identified that Fear of Death, Distrust in medicines, Allergic reactions, Side effects, and personal health disorders are the major concerns of rural citizens in avoiding COVID-19 vaccination. The authors of the study like to conclude that misinformation and misinterpretation on COVID-19 vaccination among rural citizens should be addressed by the health care workers, medical experts, and government officials to safeguard their lives from this deadly virus of COVID-19.

REFERENCES

- [1]. Aarogya Setu. (2020). <https://www.mygov.in/aarogya-setu-app/>
- [2]. Abedin M, Islam MA, Rahman FN, Reza HM, Hossain MZ, Hossain MA, et al. (2021) Willingness to vaccinate against COVID-19 among Bangladeshi adults: Understanding the strategies to optimize vaccination coverage. PLOS ONE 16(4): e0250495.
- [3]. Anil Parmar (2021) Taking vaccination to Rural India, <https://www.news18.com/news/lifestyle/health-and-fitness-taking-vaccination-to-rural-india-3750935.html>
- [4]. Civil Services India (2020) <https://www.civilserviceindia.com/subject/General-Studies/notes/e-governance.html>
- [5]. Ms. Ritika Dhabalia, Ms. Kritika Dhabalia. (2012). An Intelligent Auto-Tracking Vehicle. International Journal of New Practices in Management and Engineering, 1(02), 08 - 13. Retrieved from <http://ijnpme.org/index.php/IJNPME/article/view/5>
- [6]. Census (2011). <https://www.census2011.co.in/census/state/tamil+nadu.html>

- [7]. CSC (2020). <http://cschealth.in/Portalclass/index/services>
- [8]. Denzin, N. K., & Lincoln, Y. S. (1998). *The landscape of qualitative research: Theories and issue*. London: Sage Publications.
- [9]. Li J., Hovy E. Springer; Cham. (2017) "Reflections on sentiment/opinion analysis", *A practical guide to sentiment analysis*; pp. 41–59
- [10]. Ruchika, & Chhilar, R. S. . (2023). Novel Approach for Improving Secure Scheduling in Fog Environment and in Context of Smart Homes. *International Journal of Intelligent Systems and Applications in Engineering*, 11(1s), 23–29. Retrieved from <https://ijisae.org/index.php/IJISAE/article/view/2473>
- [11]. Md.Saiful Islam, Abu Bakkar Siddique, Rejina Akter, Rafia Tasnim, Md. Safaet Hossain Sujon, Paul R Ward, Md. Tajuddin Sikder (2021). Knowledge, attitudes and perceptions towards COVID-19 vaccinations: a cross-sectional community survey in Bangladesh, edRxiv 2021.02.16.21251802; doi: <https://doi.org/10.1101/2021.02.16.21251802>
- [12]. Pharmeasy (2021), COVAXIN and COVISHIELD, <https://pharmeasy.in/blog/covaxin-vs-covishield-a-detailed-comparison/>
- [13]. Sanil Mhatre (2020). Text Mining And Sentiment Analysis: Introduction, <https://www.red-gate.com/simple-talk/sql/bi/text-mining-and-sentiment-analysis-introduction/>
- [14]. Jones, D., Taylor, M., García, L., Rodriguez, A., & Fernández, C. Using Machine Learning to Improve Student Performance in Engineering Programs. *Kuwait Journal of Machine Learning*, 1(1). Retrieved from <http://kuwaitjournals.com/index.php/kjml/article/view/101>
- [15]. Siliguri. (2021). Covid-19 vaccine registration now open for all adults: Here's how to register on Co-WIN portal, <https://indianexpress.com/article/technology/tech-news-technology/covid-19-vaccine-registration-now-open-for-all-adults-how-you-can-register-7291196/>
- [16]. Soleymani M., Garcia D., Jou B., Schuller B., Chang S.F., Pantic M. (2017). A survey of multimodal sentiment analysis. *Image Vis Comput*. 65:3–14.
- [17]. UMANG (2021). <http://www.umang.gov.in/>