Original Article

An Improved Efficiency in Envisioning the Personage Traits over Online Social Media based on Indian Metrics during Pandemic using Novel Naive Bayes Classifier Algorithm Comparing with Logistic Regression Algorithm

V. Sai Ram Kumar[1], Shri Vindhya A[2]

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai,Tamil Nadu,India. Pincode: 602105.
[2]Project Guide, Corresponding Author, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu,India. Pincode:602105.

Abstract

Aim: The primary aim of this research is to increase the intensity percentage of personage traits detection to reveal the impact of coronavirus on Twitter users by utilizing machine learning classifier algorithms by comparing Novel Naive Bayes Classifier algorithm and Logistic Regression algorithm. **Materials and Methods:** Naive Bayes Classifier algorithm with test size=10 and Logistic Regression algorithm with test size=10 was estimated several times to envision the efficiency percentage with confidence interval of 95% and G-power (value=0.8). Naive Bayes classifier compares whether a specific feature in a class is unrelated to another feature. A logistic regression model predicts the probability of an item belonging to one group or another. **Results and Discussion:** Naive Bayes algorithm has greater efficiency (86%) when compared to Logistic Regression efficiency (60%). The results achieved with significance value p=0.169 (p>0.05) shows that two groups are statistically insignificant. **Conclusion:** Naive Bayes Algorithm executes remarkably greater than the Logistic Regression algorithm.

Keywords: Novel Naive Bayes Classifier, Logistic Regression, Twitter, Covid, Pandemic, Personage Traits Detection.

DOI: 10.47750/pnr.2022.13.S03.081

INTRODUCTION

The aim of this research is to predict Personage Traits Detection during a pandemic by sentiment analysis on social media (Twitter) using a novel Naive Bayes algorithm and to compare the proposed algorithm with Logistic Regression Algorithm. In this case, the investigation aims to enhance the rate of efficiency in detecting personage traits. Coronavirus also known as COVID-19 is a contagious illness that took shape as lung syndrome in December 2019. Dry cough, fever, and exhaustion are the most prevalent symptoms of coronavirus (Sarkodie and Owusu 2020). Governments have implemented a variety of precautionary steps to combat the virus's spread, including social isolation, business closures, and educational institution closures etc. As a result, most people communicate via social media platforms. Twitter and other social media sources are real-time communication tools that enable social media users to communicate and interact with many people at the same time (Chan et al. 2020). This platform helps people to express their feelings on pandemics (Rosenberg, Syed, and Rezaie 2020). An answer to detect user nature on pandemic using Naive Bayes algorithm is aimed in this research. With the huge amount of tweets on COVID-19, different research has been proposed. This research helps for Personage Traits Detection on pandemic using sentiment analysis. The proposed approach is used to identify user nature on pandemic and to calculate the efficiency of each algorithm used. A related research concludes that the novel Naive Bayes algorithm has better detection efficiency and faster detection time (Chirawichitchai 2013). The application of this approach helps in detecting personage traits and visualizing them with sentiment analysis (Ribeiro et al. 2020; Park, Park, and Chong 2020; Rajput, Grover, and Rathi 2020; Dai and Charnigo 2018; T. Wang et al. 2020; Pastor 2020; Dubey 2020)).

Sentiment analysis is used by many researchers for various applications like analyzing customer opinions and brand monitoring etc. On sentiment analysis, 79 papers were published in IEEE Explore and 263 publications in Google Scholar. Yum (Yum 2020) researched in the USA (United States of America) to identify the important

aspects of coronavirus by using Twitter data streams. In addition to that, Jain and Sinha (Jain and Sinha 2020) have recognized the influential users on social media using the tweets dataset. Furthermore, during the COVID-19 epidemic, Schild et al. (Tahmasbi et al. 2021) and his colleagues investigated the exposure of emotions on Twitter. Chen and Wang, (Chen, Wang, and Wu 2021) have presented an automated online aggressive conduct interpretation system to interpret tweets about COVID-19. The next part of research focused on using data mining techniques to analyze social network data like tweets. These include social network analysis, e.g., (Ribeiro et al. 2020; Park, Park, and Chong 2020; Rajput, Grover, and Rathi 2020; Dai and Charnigo 2018; T. Wang et al. 2020; Pastor 2020; Dubey 2020)), and disinformation detection, modeling, and forecasting, e.g., (Al-Rakhami and Al-Amri 2020; Kouzy et al. 2020; Pourghomi, Dordevic, and Safieddine 2018; Safieddine, Dordevic, and Pourghomi 2017). Further researches were to identify COVID-19 patients using deep and machine learning algorithms that overcomes detection failures and improved efficiency, e.g., (Ozturk et al. 2020; Oh, Park, and Ye 2020; Nour, Cömert, and Polat 2020; Minaee et al. 2020; Sethy et al. 2020). Overall, Selmi and Al-Shargabi (Al-Shargabi and Selmi 2021) visualizes Arabic user tweets during the COVID-19 pandemic by machine learning algorithms and social network analysis which was the best method and produced accurate results compared to other studies.
Our team has extensive knowledge and research experience that has translate into high quality publications(Bhansali et al. 2021; Jayanth et al. 2021; Sudhakar, Ravel, and Perumal 2021; Sathiyamoorthi et al. 2021; Deepanraj et al. 2021; Raju et al. 2021; Arun Prakash et al. 2020; Kamath et al. 2020; Shanmugam et al. 2021; Rajasekaran et al. 2020; Adhinarayanan et al. 2020; Rajesh et al. 2020; Aurtherson et al. 2021). The drawback of the existing Personage traits Detection system is less efficient in predicting user nature especially when there is a huge set of data. The main aim of our proposed system is to improve efficiency in predicting user nature by sentiment analysis using a novel Naive Bayes Classifier algorithm.

Materials and Methods

This research work was carried out at Cyber Forensic Laboratory, Saveetha School of Engineering, SIMATS (Saveetha Institute of Medical and Technical Sciences). The proposed work contains two groups. Group 1 is taken as Naive Bayes Classifier and group 2 as Logistic Regression Classifier. The Naive Bayes algorithm and Logistic Regression algorithm were executed and evaluated a different number of times with a sample size of 40 (Omer 2015) with a confidence interval of 95%, and with pretest power of 80% and maximum accepted error is fixed as 0.05.
After data collection, the invalid values, and independent content in the datasets were separated by pre-processing and data cleaning steps. After data cleaning and preprocessing the data, a perfect input for the Personage traits detection model is created, which are refined into the detection model by python libraries, and efficiency of both novel Naive Bayes Classifier algorithm and Logistic Regression algorithm is calculated. The studying process of Naive Bayes Classifier and Logistic Regression algorithms are given below.

Naive Bayes Classifier Algorithm
Naive Bayes classifiers are a group of classifier algorithms built by Bayes' Theorem. It is a set of algorithms that share a similar premise, namely that each pair of qualities being classified is separate from the others (Ali et al. 2022). Figure 1 shows the algorithm for the Naive Bayes Classifier algorithm from dataset processing to output and efficiency generation.

Logistic Regression Classifier Algorithm
Logistic Regression algorithm is one of the supervised machine learning algorithms which is used to calculate the probability of an outcome variable. As the dependent variable used in Logistic Regression is binary in nature there are only two possible outcomes for the Logistic Regression algorithm (Y.-H. Wang et al. 2019). Figure 2 shows the algorithm for the Logistic Regression algorithm from dataset processing to output and efficiency generation.

Procedure for Personage Traits detection model:
Step-1: Data gathering
First and foremost, we require the data that will be analyzed later. Scraping tools, APIs, customers' data feeds, and other methods can be used to collect data from social media, specifically Twitter. We can also collect information from consumer reviews on sites such as Google and Yelp. We'll be on the lookout for any mentions of the firm or brand throughout a particular time period. This is a frequent technique in all types of social media listening. I gathered a twitter dataset about covid-19 using different IEEE xplore papers.
Step-2: Analyze the data
It's critical to preprocess data to achieve high-quality results. Data preparation is separated into four steps to make the process easier: data cleaning, data reduction and data transformation.

- Outliers are removed, missing values are replaced, noisy data is smoothed, and inconsistent data is corrected using data cleaning techniques. Each of these activities is carried out using a variety of ways, each of which is tailored to the user's preferences or issue set.
- The goal of data reduction is to provide a condensed version of the data set that is less in size while keeping the original data set's integrity. As a consequence, efficient but equivalent outcomes are obtained.
- Transforming the data into a format suitable for data modeling is the final phase in data preparation.

Data preprocessing was done by using python libraries like “punkt” and “wordnet”

Step-3: Personage Detection Model

After preprocessing, the data was divided into two groups: train and test. The Naive Bayes method is a data classification technique that divides data into classes or categories (Ali et al. 2022). We need to identify people into good and negative persons based on their tweets on Covid-19 in personage detection. As a result, the Naive Bayes Algorithm is used to divide the dataset into four categories: positive, negative, very positive, and very negative. The python sklearn package was used to implement the Naive Bayes method. Based on past observations of a data set, logistic regression is a statistical analytic approach for predicting a binary result, such as yes or no (Y.-H. Wang et al. 2019). To predict a dependent data variable, a logistic regression model examines the relationship between one or more existing independent variables. The dataset must be classified as either negative or positive. As a result, I utilized the "sklearn" module to discover personality characteristics using the logistic regression approach.

Step- 4 : Data Visualization

The interpretation of data into a graphical representation is known as data visualization. Using visual components like charts, graphs, and maps, data visualization tools make it simple to explore and grasp trends, outliers, and patterns in data. I used python "numpy" and "panda" libraries to visualize personage traits.

The detection model follows the above procedure. Table 1 gives the source of the dataset and its properties. The dataset is processed using the visualization and NLP (Natural Language Processing) libraries and the data set is processed into a data frame. It is represented in Fig. 3, dependent variables are selected from the data frame and visualized using the matplotlib library. A train and a test set of data are created and implemented using two algorithms to predict user nature. Python programming language was used to implement this work.

Hardware specifications are concerned with the system resource settings allocated for specific devices. The following are minimum hardware requirements to execute this model processor: intel i3, RAM 4GB, 250 GB HDD storage.

Software specifications are concerned with the resources that must be installed in the target system to get an application to work. The minimal software requirements for this model to work are windows operating system version 7/8/10, python programming language version 3 or above, Jupyter Notebook, or Google Collab.

Statistical Analysis

IBM SPSS v26 is used for statistical analysis (George and Mallery 2019). The independent variable is tweet_id, date, keyword, user_id and the dependent variable is user_nature and text. The independent T-test analysis is performed.

Results

Table 1 represents the details and source of the dataset. Table 2 shows the simulated efficiency analysis of novel Naive Bayes Classifier and Logistic Regression algorithms. Table 3 represents group statistical analysis with the mean value of 86.10 and 60.10, the standard deviation of 5.174 and 6.154 for novel Naive Bayes Classifier and Logistic Regression algorithms respectively. Table 4 represents the independent T-test analysis of both the groups with significance value p= 0.169 (p>0.05) states that both groups are statistically insignificant.

Figure 4 shows the bar graph analysis based on the efficiencies of two algorithms. The mean efficiencies of novel Naive Bayes Algorithm and Logistic Regression are 86% and 60% respectively. From the results obtained it is inferred that the novel Naive Bayes Personage Traits Detection algorithm is more efficient than the Logistic Regression algorithm.

Discussion

In this research work, Naive Bayes Classifier and Logistic Regression were executed for predicting the efficiency of Personage Traits Detection of Twitter users. After validating the two models using the same dataset it was noticed that the Naive Bayes algorithm has better performance than the Logistic Regression algorithm. The novel Naive Bayes classifier model for predicting the user nature of Twitter users was developed, which makes use of

NLB (Natural Language Processing) and visualization libraries to process the user nature by text (user tweet). The proposed model predicts the user's nature using Naive Bayes and displays it by using a bar graph. The datasets from different publications assisted in improving the efficiency percentage.

The research affects less development of efficiency in predicting user sentiment on Twitter (Babu et al. 2017). A similar work presents on sentiment classification (Singh 2016) using the novel Naive Bayes algorithm. The results achieved after all iterations on each dataset showed a constant 86% efficiency. The model proposed resulted in reaching more than a 26% rise in efficiency compared to the existing model (Perski et al. 2021). Similar research carried out is about negative sentiment detection which is the best option for future researchers interested in sentiment analysis (Cheeti 2021; Kang, Song, and Ju 2022). There are no such opposite findings regarding existing personage traits detection for predicting user nature.

Although our proposed system is faster than Logistic Regression in predicting user nature, it is generally extracting only limited features from the tweets and another limitation of this research work is, it is limited to testing only twitter data. Currently, it is not programmed to embed with other social media (Huang et al. 2022). Further, this research work can be improved by deploying a model that analyzes more features from tweets in less time so that wait will be less and it can be embedded with other social media as in this research.

Conclusion

In this research work, prediction of efficiency percentage for Personage Traits Detection using Naive Bayes algorithm appears to have enhanced efficiency (86%) when compared to Logistic Regression algorithm (60%). Personage Traits Detection has been successfully employed for predicting twitter's user behavior. The results reveal the maximum number of true positives compared to true negatives from all the observations.

DECLARATION

Conflict of Interest

The author declares no conflict of interest.

Authors Contribution

Author VSRK was involved in data collection, data analysis, and manuscript writing. Author SVA was involved in conceptualization, data validation, and critical review of the manuscript.

Acknowledgment

The Authors would like to convey their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (previously known as Saveetha University) for providing the required infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the research.

1. Cyclotron Technologies, Chennai, Tamilnadu.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences (SIMATS).
4. Saveetha School of Engineering.

References

1. Adhinarayanan, Rajesh, Aravindh Ramakrishnan, Gopal Kaliyaperumal, Melvinvíctor De Poures, Rajesh Kumar Babu, and Damodharan Dillikannan. 2020. "Comparative Analysis on the Effect of 1-Decanol and Di-N-Butyl Ether as Additive with diesel/LDPE Blends in Compression Ignition Engine." Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, June, 1–18.
2. Ali, Ashraf, Weam Samara, Doaa Alhaddad, Andrew Ware, and Omar A. Saraereh. 2022. "Human Activity and Motion Pattern Recognition within Indoor Environment Using Convolutional Neural Networks Clustering and Naive Bayes Classification Algorithms." Sensors 22 (3). https://doi.org/10.3390/s22031016.
3. Al-Rakhami, Mabrook S., and Atif M. Al-Amri. 2020. "Lies Kill, Facts Save: Detecting COVID-19 Misinformation in Twitter." IEEE Access : Practical Innovations, Open Solutions 8 (August): 155961–70.
4. Al-Shargabi, Amal A., and Afef Selmi. 2021. "Social Network Analysis and Visualization of Arabic Tweets During the COVID-19 Pandemic." IEEE Access. https://doi.org/10.1109/access.2021.3091537.
5. Arun Prakash, V. R., J. Francis Xavier, G. Ramesh, T. Maridurai, K. Siva Kumar, and R. Blessing Sam Raj. 2020. "Mechanical, Thermal and Fatigue Behaviour of Surface-Treated Novel Caryota Urens Fibre–reinforced Epoxy Composite." Biomass Conversion and Biorefinery, August. https://doi.org/10.1007/s13399-020-00938-0.
6. Aurtherson, P. Babu, Bhanu Teja Nalla, Karthikeyan Srinivasan, Kulmani Mehar, and Yuvarajan Devarajan. 2021. "Biofuel Production from Novel Prunus Domestica Kernel Oil: Process Optimization Technique." Biomass Conversion and Biorefinery, May. https://doi.org/10.1007/s13399-021-01551-5.
7. Babu, Aarabhi, Mtech Student, Department of Computer Science and Engineering, Sahrdaya College Of Engineering and Technology Kodakara, Kerala, India., Vince Paul, et al. 2017. "Comparitive Study on Sentiment Analysis Techniques and User Behavior Prediction on Twitter Data." International Journal Of Engineering And Computer Science. https://doi.org/10.18535/ijecs/v6i2.01.
8. Bhansali, Karan J., Kamlesh R. Balinge, Subodh U. Raut, Shubham A. Deshmukh, M. Senthil Kumar, C. Ramesh Kumar, and Pundlik

R. Bhagat. 2021. "Visible Light Assisted Sulfonic Acid-Functionalized Porphyrin Comprising Benzimidazolium Moiety for Photocatalytic Transesterification of Castor Oil." Fuel 304 (November): 121490.
9. Chan, Teresa M., Kristina Dzara, Sara Paradise Dimeo, Anuja Bhalerao, and Lauren A. Maggio. 2020. "Social Media in Knowledge Translation and Education for Physicians and Trainees: A Scoping Review." Perspectives on Medical Education 9 (1): 20–30.
10. Cheeti, Swetha Sree. 2021. Twitter Based Sentiment Analysis of Impact of COVID-19 on Education Globally.
11. Chen, Toly, Yu-Cheng Wang, and Hsin-Chieh Wu. 2021. "Analyzing the Impact of Vaccine Availability on Alternative Supplier Selection Amid the COVID-19 Pandemic: A cFGM-FTOPSIS-FWI Approach." Healthcare. https://doi.org/10.3390/healthcare9010071.
12. Chirawichitchai, Nivet. 2013. "Sentiment Classification by a Hybrid Method of Greedy Search and Multinomial Naïve Bayes Algorithm." In 2013 Eleventh International Conference on ICT and Knowledge Engineering. IEEE. https://doi.org/10.1109/ictke.2013.6756285.
13. Dai, Hongying, and Richard Charnigo. 2018. "A SENTIMENT ANALYSIS OF MERS-CoV OUTBREAK THROUGH TWITTER SOCIAL MEDIA MONITORING." JP Journal of Biostatistics 15 (2): 107–25.
14. Deepanraj, B., N. Senthilkumar, D. Mala, and A. Sathiamourthy. 2021. "Cashew Nut Shell Liquid as Alternate Fuel for CI Engine—optimization Approach for Performance Improvement." Biomass Conversion and Biorefinery, February. https://doi.org/10.1007/s13399-021-01312-4.
15. Dubey, Akash Dutt. 2020. "Twitter Sentiment Analysis during COVID19 Outbreak." SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3572023.
16. George, Darren, and Paul Mallery. 2019. IBM SPSS Statistics 26 Step by Step: A Simple Guide and Reference. Routledge.
17. Huang, Zhilian, Evonne Tay, Dillon Wee, Huiling Guo, Hannah Yeefen Lim, and Angela Chow. 2022. "Public Perception on the Use of Digital Contact Tracing Tools Post COVID-19 Lockdown: Sentiment Analysis and Opinion Mining." JMIR Formative Research, January. https://doi.org/10.2196/33314.
18. Jain, Somya, and Adwitiya Sinha. 2020. "Identification of Influential Users on Twitter: A Novel Weighted Correlated Influence Measure for Covid-19." Chaos, Solitons & Fractals. https://doi.org/10.1016/j.chaos.2020.110037.
19. Jayanth, Bellappu Venkat, Melvin Victor Depoures, Gopal Kaliyaperumal, Damodharan Dillikannan, Dilipsingh Jawahar, Kumaran Palani, and Ganesha Prasad Meravanigee Shivappa. 2021. "A Comprehensive Study on the Effects of Multiple Injection Strategies and Exhaust Gas Recirculation on Diesel Engine Characteristics That Utilize Waste High Density Polyethylene Oil." Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, June, 1–18.
20. Kamath, Manjunath, Subha Krishna Rao, Jaison, Sridhar, Kasthuri, Gopinath, Sivaperumal, and Shantanu Patil. 2020. "Melatonin Delivery from PCL Scaffold Enhances Glycosaminoglycans Deposition in Human Chondrocytes – Bioactive Scaffold Model for Cartilage Regeneration." Process Biochemistry 99 (December): 36–47.
21. Kang, Eunkyo, Narae Song, and Hyorim Ju. 2022. "Contents and Sentiment Analysis of Newspaper Articles and Comments on Telemedicine in Korea: Before and after of COVID-19 Outbreak." Health Informatics Journal 28 (1): 14604582221075549.
22. Kouzy, Ramez, Joseph Abi Jaoude, Afif Kraitem, Molly B. El Alam, Basil Karam, Elio Adib, Jabra Zarka, Cindy Traboulsi, Elie W. Akl, and Khalil Baddour. 2020. "Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter." Cureus 12 (3): e7255.
23. Minaee, Shervin, Rahele Kafieh, Milan Sonka, Shakib Yazdani, and Ghazaleh Jamalipour Soufi. 2020. "Deep-COVID: Predicting COVID-19 from Chest X-Ray Images Using Deep Transfer Learning." Medical Image Analysis 65 (October): 101794.
24. Nour, Majid, Zafer Cömert, and Kemal Polat. 2020. "A Novel Medical Diagnosis Model for COVID-19 Infection Detection Based on Deep Features and Bayesian Optimization." Applied Soft Computing 97 (December): 106580.
25. Oh, Yujin, Sangjoon Park, and Jong Chul Ye. 2020. "Deep Learning COVID-19 Features on CXR Using Limited Training Data Sets." IEEE Transactions on Medical Imaging 39 (8): 2688–2700.
26. Omer, Ahmed Abdelkarim Eldud. 2015. Prediction of Protein Secondary Structure Using Binary Classificationtrees, Naive Bayes Classifiers and the Logistic Regression Classifier.
27. Ozturk, Tulin, Muhammed Talo, Eylul Azra Yildirim, Ulas Baran Baloglu, Ozal Yildirim, and U. Rajendra Acharya. 2020. "Automated Detection of COVID-19 Cases Using Deep Neural Networks with X-Ray Images." Computers in Biology and Medicine 121 (June): 103792.
28. Park, Han Woo, Sejung Park, and Miyoung Chong. 2020. "Conversations and Medical News Frames on Twitter: Infodemiological Study on COVID-19 in South Korea." Journal of Medical Internet Research 22 (5): e18897.
29. Pastor, Cherish Kay. 2020. "Sentiment Analysis of Filipinos and Effects of Extreme Community Quarantine due to Coronavirus (COVID-19) Pandemic." SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3574385.
30. Perski, Olga, Noreen L. Watson, Kristin E. Mull, and Jonathan B. Bricker. 2021. "Identifying Content-Based Engagement Patterns in a Smoking Cessation Website and Associations With User Characteristics and Cessation Outcomes: A Sequence and Cluster Analysis." Nicotine & Tobacco Research: Official Journal of the Society for Research on Nicotine and Tobacco 23 (7): 1103–12.
31. Pourghomi, Pardis, Milan Dordevic, and Fadi Safieddine. 2018. "The Spreading of Misinformation Online: 3D Simulation." In 2018 5th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE). IEEE. https://doi.org/10.1109/icitacee.2018.8576937.
32. Rajasekaran, S., D. Damodharan, K. Gopal, B. Rajesh Kumar, and Melvin Victor De Poures. 2020. "Collective Influence of 1-Decanol Addition, Injection Pressure and EGR on Diesel Engine Characteristics Fueled with diesel/LDPE Oil Blends." Fuel 277 (October): 118166.
33. Rajesh, A., K. Gopal, De Poures Melvin Victor, B. Rajesh Kumar, A. P. Sathiyagnanam, and D. Damodharan. 2020. "Effect of Anisole Addition to Waste Cooking Oil Methyl Ester on Combustion, Emission and Performance Characteristics of a DI Diesel Engine without Any Modifications." Fuel 278 (October): 118315.
34. Rajput, Nikhil Kumar, Bhavya Ahuja Grover, and Vipin Kumar Rathi. 2020. "Word Frequency and Sentiment Analysis of Twitter Messages during Coronavirus Pandemic." http://arxiv.org/abs/2004.03925.
35. Raju, P., K. Raja, K. Lingadurai, T. Maridurai, and S. C. Prasanna. 2021. "Glass/Caryota Urens Hybridized Fibre-Reinforced nanoclay/SiC Toughened Epoxy Hybrid Composite: Mechanical, Drop Load Impact, Hydrophobicity and Fatigue Behaviour." Biomass Conversion and Biorefinery, March. https://doi.org/10.1007/s13399-021-01427-8.
36. Ribeiro, Haroldo V., Andre S. Sunahara, Jack Sutton, Matjaž Perc, and Quentin S. Hanley. 2020. "City Size and the Spreading of COVID-19 in Brazil." PloS One 15 (9): e0239699.
37. Rosenberg, Hans, Shahbaz Syed, and Salim Rezaie. 2020. "The Twitter Pandemic: The Critical Role of Twitter in the Dissemination of Medical Information and Misinformation during the COVID-19 Pandemic." CJEM 22 (4): 418–21.
38. Safieddine, Fadi, Milan Dordevic, and Pardis Pourghomi. 2017. "Spread of Misinformation Online: Simulation Impact of Social Media Newsgroups." In 2017 Computing Conference. IEEE. https://doi.org/10.1109/sai.2017.8252201.
39. Sarkodie, Samuel Asumadu, and Phebe Asantewaa Owusu. 2020. "Investigating the Cases of Novel Coronavirus Disease (COVID-

19) in China Using Dynamic Statistical Techniques." Heliyon 6 (4): e03747.
40. Sathiyamoorthi, Ramalingam, Gomathinayakam Sankaranarayanan, Dinesh Babu Munuswamy, and Yuvarajan Devarajan. 2021. "Experimental Study of Spray Analysis for Palmarosa Biodiesel-diesel Blends in a Constant Volume Chamber." Environmental Progress & Sustainable Energy 40 (6). https://doi.org/10.1002/ep.13696.
41. Sethy, Prabira Kumar, Santi Kumari Behera, Pradyumna Kumar Ratha, and Preesat Biswas. 2020. "Detection of Coronavirus Disease (COVID-19) Based on Deep Features and Support Vector Machine." International Journal of Mathematical, Engineering and Management Sciences 5 (4): 643–51.
42. Shanmugam, Rajasekaran, Damodharan Dillikannan, Gopal Kaliyaperumal, Melvin Victor De Poures, and Rajesh Kumar Babu. 2021. "A Comprehensive Study on the Effects of 1-Decanol, Compression Ratio and Exhaust Gas Recirculation on Diesel Engine Characteristics Powered with Low Density Polyethylene Oil." Energy Sources, Part A: Recovery, Utilization, and Environmental Effects 43 (23): 3064–81.
43. Singh, Jagmeet. 2016. "Analysis on Hinglish Opinion Using Multinomial Naive Bayes Algorithm." IOSR Journal of Computer Engineering. https://doi.org/10.9790/0661-15010020273-82.
44. Sudhakar, M. P., Merlyn Ravel, and K. Perumal. 2021. "Pretreatment and Process Optimization of Bioethanol Production from Spent Biomass of Ganoderma Lucidum Using Saccharomyces Cerevisiae." Fuel 306 (December): 121680.
45. Tahmasbi, Fatemeh, Leonard Schild, Chen Ling, Jeremy Blackburn, Gianluca Stringhini, Yang Zhang, and Savvas Zannettou. 2021. "'Go Eat a Bat, Chang!': On the Emergence of Sinophobic Behavior on Web Communities in the Face of COVID-19." Proceedings of the Web Conference 2021. https://doi.org/10.1145/3442381.3450024.
46. Wang, Tianyi, Ke Lu, Kam Pui Chow, and Qing Zhu. 2020. "COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model." IEEE Access: Practical Innovations, Open Solutions 8: 138162–69.
47. Wang, Yi-Han, Yang Ou, Xu-Dong Deng, Lu-Ran Zhao, and Chao-Yu Zhang. 2019. "The Ship Collision Accidents Based on Logistic Regression and Big Data." In 2019 Chinese Control And Decision Conference (CCDC). IEEE. https://doi.org/10.1109/ccdc.2019.8832686.
48. Yum, Seungil. 2020. "Social Network Analysis for Coronavirus (COVID-19) in the United States." Social Science Quarterly, May. https://doi.org/10.1111/ssqu.12808.

TABLES AND FIGURES

Table 1. Dataset Name, Extension and Source.

S.NO	DATASET NAME	DATASET EXTENSION	DATASET SOURCE
1	TWEETS DATASET	CSV	IEEE Xplore

Table 2. Efficiency of Naive Bayes Algorithm and Logistic Regression Algorithm.
The Naive Bayes is 26% more efficient than the Logistic Regression algorithm.

ITERATION NO.	Naive Bayes Algorithm NB(%)	Logistic Regression Algorithm LR(%)
1	93	70
2	92	67
3	90	65
4	89	64
5	88	60
6	87	58

7	82	57
8	81	55
9	80	53
10	79	52

Table 3. Group Statistics of NB and LR algorithm with the mean value of 86.10% and 60.10%

GROUP	**N**	**Mean(%)**	**Std.Deviation**	**Std.Error Mean**
Naive Bayes	10	86.10	5.174	1.636
Logistic Regression	10	60.10	6.154	1.946

Table 4. Independent sample T-test is performed for the two groups for significance and standard error determination. The significance value p=0.169 (p>0.05) shows that two groups are statistically insignificant.

		Levene's Test for Equality of Variance		T-test for Equality of Means						
								Std. Error Differ-ence	95% Confidence Interval of the Difference	
	Equal Variance	F	Sig	t	df	Sig (2-tailed)	Mean Difference		Lower	Upper
Efficiency	Assumed	.310	.169	10.226	18	.002	26.000	2.543	20.658	31.342

	Not Assumed			10.226	17.483	.002	26.000	2.543	20.647	31.353

Input:

Tweets dataset

Output:

Accuracy, confusion matrix.

Steps:

1. Remove unwanted variables from the dataset.
2. Load the data(train set, test set).
3. Import Multinomial Naive Bayes model from sklearn.naive_bayes
4. Fit train set and test set to the multi nominal naive bayes model.
5. Do K-fold cross validation on our train set
6. Measure the mean accuracy of K-fold cross validation
7. Predict the user nature of test set
8. Measure the accuracy of the model on test set
9. Plot confusion matrix between predicted test set and original test set
10. Compile model

End

Fig. 1. Pseudocode for novel Naive Bayes Classifier algorithm.

Input:

Tweets dataset

Output:

Accuracy, confusion matrix.

Steps:

1. Remove unwanted variables from the dataset.
2. Load the data(train set, test set).
3. Import logistic Regression model from sklearn.linear_model
4. Fit train set and test set to Logistic Regression model.
5. Do K-fold cross validation on our train set
6. Measure the mean accuracy of K-fold cross validation
7. Predict the user nature of test set
8. Measure the accuracy of the model on test set
9. Plot confusion matrix between predicted test set and original test set
10. Compile model

End

Fig. 2. Pseudocode for Logistic Regression algorithm.

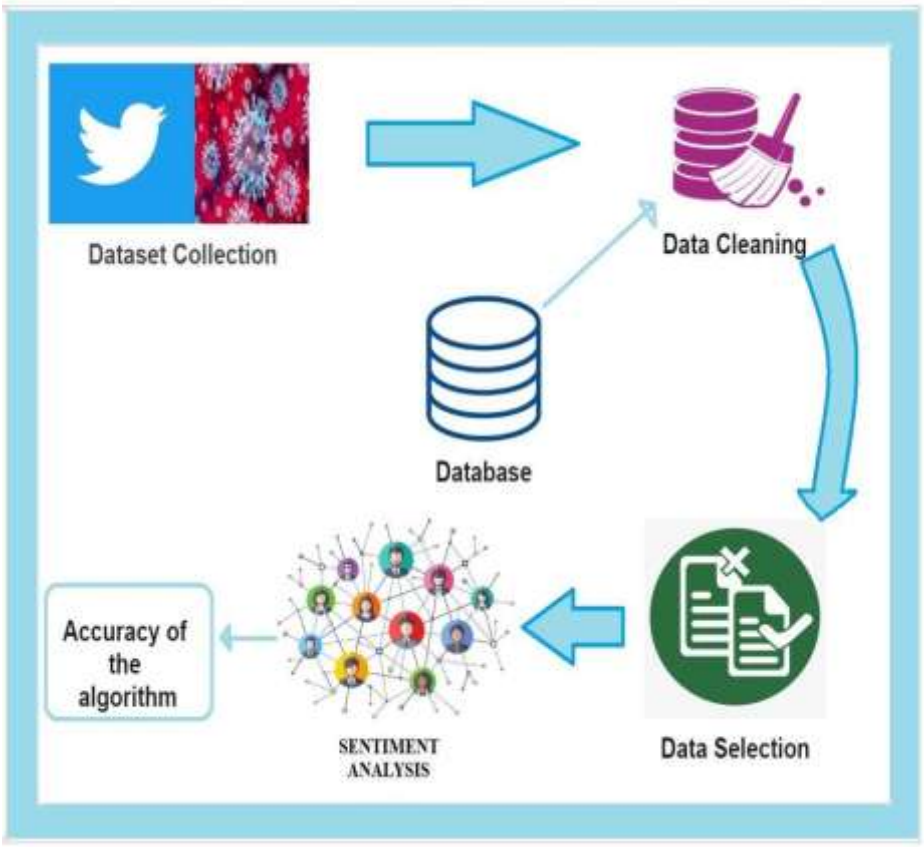

Fig. 3. Architecture for Personage Traits Detection for predicting user nature using novel Naive Bayes algorithm, from dataset collection to accuracy calculation.

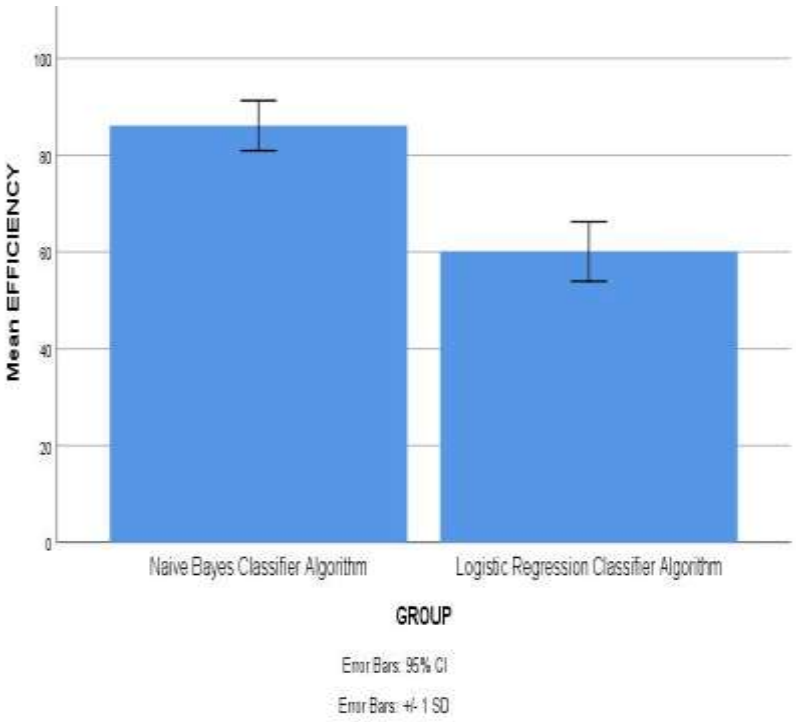

Fig. 4. Bar graph analysis of novel Naive Bayes algorithm and Linear Regression algorithm. Graphical representation shows the mean efficiency of 86% and 60% for the proposed algorithm (Naive Bayes) and Logistic Regression respectively. X-axis : Naive Bayes Classifier Algorithm vs Logistic Regression Classifier Algorithm, Y-axis : Mean precision ± 1 SD.