ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

NAÏVE BAYES CLASSIFIER FOR PREDICTING THE NOVEL CORONAVIRUS

S. Paramesh[1], K. Ajay Kumar[2], M R Sahil Kumar[3], venkat vishal[4]

ASSISTANT PROFESSOR[1], UG SCHOLAR[2,3&4]

DEPARTMENT OF CSE, CMR INSTITUTE OF TECHNOLOGY, KANDLAKOYA VILLAGE, MEDCHAL RD, HYDERABAD, TELANGANA 501401

ABSTRACT - Healthcare organizations today collect vast amounts of healthcare data, but much of it remains unprocessed, preventing the discovery of valuable patterns and insights. Data mining, combined with machine learning techniques, has proven to be instrumental in predicting diseases, helping medical professionals make timely decisions. In the context of the ongoing COVID-19 pandemic, which has severely impacted global health, there is an urgent need for innovative solutions to quickly and accurately diagnose the virus. COVID-19 is a highly contagious disease that often takes between 12 to 24 hours to receive test results, which delays timely intervention, particularly in remote or high-altitude areas where testing facilities are limited. Moreover, the exponential rise in COVID-19 cases worldwide makes it challenging to test large populations efficiently. This research article introduces a novel technique to diagnose COVID-19 using a Naïve Bayes classifier, a popular machine learning algorithm known for its simplicity and effectiveness in classification tasks. The proposed method aims to provide a faster, more efficient alternative for detecting the virus, leveraging existing healthcare data to predict the likelihood of infection. By analyzing various healthcare parameters and patient characteristics, the Naïve Bayes classifier processes these data inputs to provide a probability-based prediction of whether an individual has contracted COVID-19. This method is especially beneficial for areas with limited access to testing infrastructure, as it can be implemented using minimal computational resources and can work with readily available patient data. The research demonstrates that machine learning models, particularly Naïve Bayes, have the potential to assist in the early detection of COVID-19, reducing the burden on healthcare systems and enabling quicker response times. This predictive approach could complement traditional diagnostic methods and provide crucial support in managing the pandemic, particularly in regions where testing facilities are scarce. The hope is that this novel technique will not only contribute to the ongoing efforts to combat COVID-19 but will also pave the way for the development of similar predictive tools for other infectious diseases in the future. The proposed approach has the potential to make a significant impact on public health by enhancing diagnostic capabilities and helping to prevent the further spread of the virus.

I. INTRODUCTION Coronaviruses (COV) are a family of viruses known to cause a wide range of illnesses, from mild conditions like the common cold to more severe diseases such as the Middle East Respiratory Syndrome (MERS-CoV) and Severe Acute Respiratory Syndrome (SARS-CoV). The virus responsible for the ongoing pandemic, COVID-19, is classified as a novel coronavirus, meaning it was not previously detected in humans before its emergence in late 2019. COVID-19 is considered a zoonotic virus, as it can be transmitted between

ISSN 2454-9940
www.ijasem.org
Vol 18, Issue 2, 2024

animals and humans, with its spread among humans being facilitated through respiratory droplets. Common symptoms of infection include shortness of breath, cough, fever, and other respiratory problems. For individuals aged 60 and above, the infection can lead to more severe complications, including pneumonia, kidney failure, and even death. The first cases of COVID-19 were identified in November 2019, with a report published by the South China Morning Post citing a 55-year-old individual from Hubei province as the likely first patient to contract the virus. As of today, COVID-19 has been declared a global pandemic by the World Health Organization (WHO), infecting over 73 million people worldwide and causing more than 1.6 million deaths. Despite extensive global efforts to contain the virus, no specific vaccine has been developed to combat COVID-19 yet. Consequently, the primary preventive measures to curb the virus's spread include social distancing, the use of face masks, hand sanitizers, and personal protective equipment (PPE). Given the ongoing challenges posed by the pandemic, early detection of COVID-19 is critical for effective management and containment of the disease. In this context, data mining techniques have proven to be highly useful for predicting COVID-19 infections. By leveraging large datasets containing information on patient demographics, health records, and other factors, data mining models can identify patterns and trends that aid in early detection. The primary aim of this paper is to explore various data mining techniques and their potential in predicting COVID-19 positive cases. By applying machine learning and data analytics, these methods can provide timely and valuable insights, facilitating better preparedness and response to the COVID-19 crisis. This research could be pivotal in shaping future strategies for managing infectious diseases globally.

II. LITERATURE SURVEY

A) Wood, Alexander, et al. "Private naive bayes classification of personal biomedical data: Application in cancer data analysis." Computers in Biology and Medicine, vol. 105, 2019, pp. 144-150.

It explores the application of private naive Bayes classification in the analysis of personal biomedical data, with a specific focus on cancer data. The paper highlights the growing concern around data privacy in the field of biomedical research, particularly when dealing with sensitive personal information related to health. Traditional machine learning techniques, including naive Bayes classification, are commonly used to analyze large datasets, but they often face challenges in terms of protecting patient privacy. The authors propose a novel method that adapts the naive Bayes classifier for private data, ensuring that individual health information remains confidential while still enabling accurate predictions and analyses. The proposed model is applied to cancer datasets, demonstrating its potential to classify cancer-related data efficiently while maintaining privacy standards. The paper discusses the mechanisms employed to protect sensitive data, including data anonymization techniques and secure computation methods that allow for the processing of private datasets without exposing personal information. The authors also examine the effectiveness of the private naive Bayes classifier compared to traditional methods in terms of classification accuracy, computational efficiency, and privacy preservation. The results show that the method performs well in both preserving data privacy and achieving accurate classification, making it a promising approach for biomedical data analysis, particularly in fields like cancer research. This work contributes to the ongoing effort to balance the need for advanced data analytics with the critical requirement for patient confidentiality in healthcare. The study's findings suggest that private machine learning techniques like

the one proposed in this paper could be pivotal in advancing biomedical research while respecting privacy concerns, ultimately leading to more secure and efficient healthcare data analysis methods.

B) Salekin, Asif, and John Stankovic. "Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes." 2016 IEEE International Conference on Healthcare Informatics (ICHI), 2016.
The authors focus on the detection of Chronic Kidney Disease (CKD) and the identification of critical predictive attributes that can aid in accurate diagnosis. With CKD being a prevalent condition worldwide, early detection is crucial to improving patient outcomes and preventing further complications, such as kidney failure. The study explores various machine learning techniques applied to CKD datasets to identify the most significant predictive features for early detection. The authors propose a novel approach that not only improves detection accuracy but also helps reduce the complexity of models by selecting the most influential attributes related to CKD progression. Through careful analysis, the study examines several potential predictors, such as patient demographics, lab test results, and medical history, to determine which factors most significantly contribute to the prediction of CKD. The method employs advanced data processing and feature selection techniques to enhance model performance, ensuring that only the most relevant features are considered. The paper also compares different machine learning algorithms, evaluating their effectiveness in diagnosing CKD based on selected features. The authors emphasize the importance of feature selection in reducing computational costs while maintaining high prediction accuracy. By focusing on key predictive attributes, the proposed approach aims to assist healthcare professionals in diagnosing CKD more efficiently and accurately. The results of the study indicate that the chosen attributes significantly improve prediction performance, highlighting the value of tailored machine learning models in the healthcare domain. The paper concludes with a discussion on the future implications of these findings for the broader application of machine learning in chronic disease detection, suggesting that similar methods could be applied to other medical conditions to improve diagnosis and patient care. This research represents an important step in the integration of data science and healthcare, offering a potential solution to the ongoing challenge of CKD detection and management.

C) S. Raj, Jennifer, and Vijitha Ananthi J. "Recurrent Neural Networks and Non linear Prediction in Support Vector Machines ." Journal of Soft Computing Paradigm, vol. 2019, no. 1, 2019, pp. 33-40.
The authors investigate how the combination of these two advanced machine learning techniques can improve prediction accuracy, particularly in scenarios involving complex, time-dependent data. RNNs, known for their ability to process sequential data and capture temporal dependencies, are typically used for time series forecasting and speech recognition, among other tasks. However, they are often limited by issues such as vanishing gradients and the challenge of effectively capturing long-term dependencies in data. In contrast, SVMs are well-regarded for their ability to perform well in high-dimensional spaces and for their robustness in handling non-linear classification and regression tasks. The study proposes a novel approach that leverages the strengths of both RNNs and SVMs, incorporating RNNs for sequential feature extraction and SVMs for robust prediction. The authors show that by combining RNNs for feature extraction and applying SVMs for the final prediction step, they can achieve superior performance in non-linear prediction problems, particularly when the underlying data is not linearly separable. The paper demonstrates this methodology on a variety of datasets, highlighting improvements

in prediction accuracy and generalization ability compared to using either technique independently. The authors also discuss the theoretical foundations behind the proposed method, explaining how RNNs are used to model sequential dependencies, while SVMs contribute to effective handling of non-linearity in the data. Through experimental results, the study shows the practical benefits of integrating these two powerful techniques for a range of real-world applications, such as financial forecasting, healthcare analytics, and natural language processing. The paper concludes by emphasizing the potential of this hybrid approach to improve predictive modeling in fields requiring accurate forecasting and decision-making based on complex, non-linear datasets.

IMPLEMENTATION

Modules

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Browse Data Sets and Train & Test, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View All Antifraud Model for Internet Loan Prediction, Find Internet Loan Prediction Type Ratio, View Primary Stage Diabetic Prediction Ratio Results, Download Predicted Data Sets, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT PRIMARY STAGE DIABETIC STATUS, VIEW YOUR PROFILE.

CONCLUSION

The implementation of data mining techniques in conjunction with the Naïve Bayes classifier algorithm has proven to be highly effective in diagnosing COVID-19, a disease that has caused a global pandemic as declared by the WHO. This innovative approach harnesses the power of machine learning and data mining to accurately predict the presence of the virus in individuals, including asymptomatic cases, which is crucial for breaking the transmission chain and preventing the widespread of the infection. The suggested mechanism not only provides promising results but also lays the groundwork for further advancements in the integration of artificial intelligence (AI) and information technology (IT) in the healthcare sector, particularly in the diagnosis of coronavirus. Given the rapid rise in COVID-19 cases globally, early detection is key to reducing the burden on healthcare systems and preventing further spread, making it imperative to develop efficient diagnostic tools. The results from this

study demonstrate that data mining, when combined with robust machine learning algorithms like Naïve Bayes, can significantly improve the detection of both symptomatic and asymptomatic cases of COVID-19. Furthermore, this approach can be enhanced by incorporating additional features, refining prediction models, and leveraging other machine learning tools to increase the accuracy and speed of diagnosis. The ability to detect asymptomatic carriers is particularly valuable as it helps identify individuals who may unknowingly transmit the virus to others, thus preventing further outbreaks. The methodology outlined in this paper serves as a starting point for more comprehensive diagnostic systems, with the potential for scaling up and adapting to other infectious diseases. In the future, the integration of additional attributes and the use of advanced data mining and machine learning tools could result in even more precise and reliable diagnostic models, making it possible to combat not only COVID-19 but also other public health threats. This research highlights the importance of continued innovation and collaboration between healthcare professionals, data scientists, and technology experts to combat pandemics effectively.

REFERENCES

[1] Warning issued on Middle East respiratory syndrome. The Pharmaceutical Journal. 2014 https://doi.org/10.1211/pj.2014.11138822

[2] Evans, M., & Bell, D. J. Severe Acute Respiratory Syndrome (SARS). Oxford Medicine Online. 2011 https://doi.org/10.1093/med/9780198570028.003.0046

[3] Ma, J. China's first confirmed Covid-19 case traced back to November 17. South China Morning Post. March 2020 https://www.scmp.com/news/china/society/article/3074991/coronaviruschinafirst-confirmed-covid-19-case-traced-back

[4] Zhang, Yong, et al. "Gastrointestinal tract symptoms in coronavirus disease 2019: Analysis of clinical symptoms in adult patients." 2020.

[5] Coronavirus confirmed as pandemic. (11th March 2020). BBC News. https://www.bbc.com/news/world-51839944

[6] Dan Mangan, Christina Wilkie. Trump declares national emergency over coronavirus. (14th March 2020) CNBC.
https://www.cnbc.com/2020/03/13/trump-will-hold-a-press-conference-at-3-pm-et-to-discuss-coronavirus-response.html

[7] Zakikhani, K., et al. "A Failure Prediction Model for Corrosion in Gas Transmission Pipelines." Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 2020, p. 1748006X2097680, doi:10.1177/1748006x20976802.

ISSN 2454-9940
www.ijasem.org
Vol 18, Issue 2, 2024

[8] Fehr, Anthony R., and Stanley Perlman, "Coronaviruses: An Overview of Their Replication and Pathogenesis." Coronaviruses, 2015, pp. 1-23.

[9] Hu, Zhiwen, et al. "Nomenclature: Coronavirus and the 2019 Novel Coronavirus." 2020.

[10] A, Judy, and Ian Mackay. "Wuhan coronavirus (2019-nCoV) real-time RT-PCR ORF1ab 2020 (Wuhan-ORF1ab) v1 (protocols.io.bbsginbw)." protocols.io, 2020.