International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-12 Issue-1, May 2023

Multiple Disease Prediction Using ML

Alok Katiyar, Sajid Ali, Sameer Ray

Abstract: ***Accurate and on-time analysis of any health-related drawback is vital for the interference and treatment of the sickness. The normal method of diagnosing might not be sufficient within the case of a significant illness. Developing a medical diagnosing system supported machine learning (ML) algorithms for prediction of any unwellness will facilitate during a lot of correct diagnosis than the standard methodology. We've designed a disease prediction system using ML. Disease Prediction System using Machine Learning could be a system that predicts the sickness supported data or symptoms that he/she enter into the system and gives correct results supported that data. This predictive disease using Machine Learning is completed entirely with the assistance of Learning Machines and Python programing language with its Flask Interface and mistreatment antecedently offered databases with hospitals that use that we'll predict the unwellness.***

Keywords: Machine Learning, Disease Prediction, Decision Tree, Random Forest, Symptoms.

I. INTRODUCTION

Disease Prediction using Machine Learning is a system that predicts that disease is supported to the information provided by the user. It also predicts a patient's illness or the user based on the information or symbols that enter the system as well provides accurate results based on that information. If the patient is not so bad and the user just wants to know the type of disease, go through it. It's a way of system that provides the user with tips and tricks to maintain a healthy user also provides a way to diagnose the disease using this prediction. This Disease Prediction Using Machine Learning is fully developed with the help of Python Programming Learning and Language Equipment with its Flask Interface and using the previously available database by Hospitals that use that will predict the disease (Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., & Lo, B. et al. (2017). Deep Learning for Health Informatics. IEEE Journal of Biomedical and Health Informatics, 21(1), 4-21.) [1].

Manuscript received on 27 March 2023 | Revised Manuscript received on 04 April 2023 | Manuscript Accepted on 15 May 2023 | Manuscript published on 30 May 2023.
*Correspondence Author(s)
Dr. Alok Katiyar, Professor, Department of CSE, Galgotias University Greater Noida, Gautam Buddha Nagar, Uttar Pradesh, India ORCID ID: https://orcid.org/0000-0002-1645-6585
Sajid Ali, Student, Department of CSE, Galgotias University Greater Noida, Gautam Buddha Nagar, Uttar Pradesh, India. ORCID ID: https://orcid.org/0009-0007-5961-3070
Sameer Ray*, Student, Department of CSE, Galgotias University Greater Noida, Gautam Buddha Nagar, Uttar Pradesh, India. E-mail: sameer.ray.official@gmail.com, ORCID ID: https://orcid.org/0009-0006-7319-2332

Now day doctors use many scientific technologies and how to do both identification and diagnosis not only common diseases, but also many deadly diseases. Effective treatment is permanent it results from an accurate and accurate diagnosis. Doctors can sometimes fail to take the right one decision while diagnosing a patient's disease, so disease planning systems which uses machine learning skills that help in such situations to obtain accurate results.

II. LITERATURE REVIEW

A Bayesian Inference Naive Bayes classifiers are a type of simple probabilistic classifier based on applying Bayes' theorem to characteristics with strong (naive) independence assumptions. They're one of the most basic Bayesian network models, but when combined with kernel density estimation, they can achieve higher levels of accuracy. The number of parameters required by Naive Bayes classifiers is linear in the number of variables (features/predictors) in a learning problem. Instead of expensive iterative approximation, which is used for many other types of classifiers, maximum likelihood training can be done simply evaluating a closed form expression, which requires linear time.

A. Logistic regression:

Logistic regression is a type of analytical modeling. It's used to examine a dataset in which one or more independent variables influence the outcome. With a random state of 0, Logistic Regression was imported. The training model was then fitted. The accuracy of the testing was 87.09 percent.

B. Random forest:

Random forest classifier may be a powerful supervised classification tool. RF generates a forest of classification trees from a given dataset, instead of a single classification tree. every of those trees produces a classification for a given set of attributes. From 'sklearn ensemble', 'Random Forest Classifier' was imported (Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32) [2]. The estimators are unbroken at ten and random state at 0. Then the coaching model was fitted. The testing accuracy was 90.32%. call Tree The testing accuracy was 90.32%. a choice tree is a tree shape-like diagram, wherever the internal nodes represent a check on Associate in Nursing attribute, every branch denotes the outcome of the test, each leaf node denotes a category label. call Tree was foreign wherever the random state was unbroken as zero and so the coaching model was fitted. The testing accuracy was 83.87%. 6. Results Amongst all classification techniques, testing accuracy was best within the case of the random forest and SVM approach with an accuracy of 90.32%.

III. PROPOSED SYSTEM

We have blended based and unstructured facts inside the Healthcare fields to decide ailment danger on this venture. The Use of a latent aspect model to recreate missing statistics in scientific Information acquired from on line sources. We may also verify the Most important persistent sicknesses in a particular area and populace the usage of Statistical facts. We consult sanatorium specialists to learn Approximately useful capabilities while coping with dependent facts. Inside the Case of unstructured textual content files, we use the random forest Algorithm to robotically pick out capabilities.

A. Data collection

Data collection has been done from the web to spot the disease here the important symptoms of the sickness area unit collected i.e. no dummy values area unit entered. The symptoms of the sickness area unit collected from completely different health connected websites.

Data Preprocessing

Before feeding the information into the Prediction model, following data cleaning and preprocessing steps area unit performed.

- Checking null values and filling mistreatment forward fill technique
- changing information into completely different cases
- Standardizing the information mistreatment mean and variance
- ripping the dataset into coaching and testing sets

B. Building model

Many methods are used to perform data mining. Machine learning is one of the approaches. Random forest Machine learning strategies include grouping, clustering, summarization, and many others. Since classification techniques are used in this project, classification is one of the data mining processes in the Phase of categorial data classification.

C. Prediction

Prediction using random forest:

Prediction done by Random Forest model using Flask frame work model trained but training chronic disease dataset.

IV. THE DATABASE AND THE METHODS

A. Set of data

The dataset we utilize is the Kaggle dataset, which is freely available Kaggle. (n.d.). Chronic Kidney Disease Dataset Retrieved from https://www.kaggle.com/mansoordaku/ckdisease [3]. The dataset was retrieved from its repository. It comprises 402-samples. Numerous types of classes Eleven of the twenty-five qualities are numerical, and thirteen are non-numerical, one is a class attribute while the other is a nominal. The data collection includes the total number of missing values the dataset's information is available here data from the patient, such as age, blood pressure, and specific symptoms albumin, sugar, red blood cells, and so forth.

B. Data Gathering

We use a dataset from the Kaggle Machine Learning Repository in this experiment (Mehta, N., Pandey, S., & Verma, S. (2017). Disease Prediction System using Machine Learning over Cloud. 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS), 331-336.) [4]. Furthermore, the initial dataset was gathered. There are 583 liver patients in this dataset, with 75.64 percent of male patients and 24.36 percent of female patients. There were 11 distinct parameters in this sample, whereas

Select ten parameters for further investigation and one parameter as a target class. as an example,

- Patient's age (in years)
- Gender: The Patients' Gender
- TB stands for total bilirubin.
- DB: Bilirubin Direct
- Alkphos: Phosphotase Alkaline
- Alamine Aminotransferase (Sgpt)
- Asparatate Aminotransferase (Sgot)
- Total Proteins (TP)

Fig. 1 Web applications interface

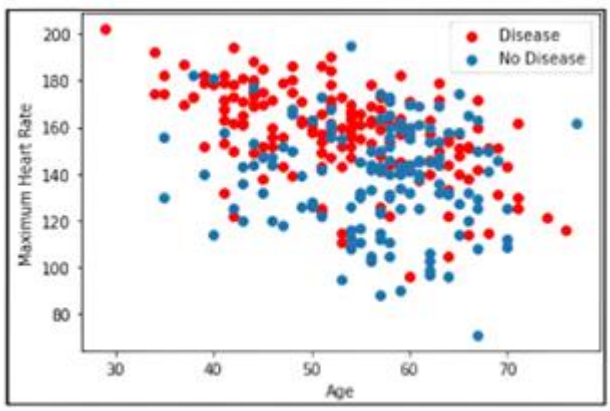

Fig. 2 Positive and negative case

V. RELATED WORK

Many researchers have used machine learning techniques like KNN, Naïve Bayes and Decision trees to develop disease Prediction strategies.

Retrieval Number: 100.1/ijrte.A75680512123
DOI: 10.35940/ijrte.A7568.0512123
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Satyabhama Balasubramanian, Balaji Subramani discussed the system to reduces the multiple diseases showing the similar symptoms problem and it will increase the accuracy of such diagnosis (Balasubramanian, S., & Subramani, B. (2019). A system for reducing the multiple diseases showing the similar symptoms problem using machine learning techniques. Journal of Medical Systems, 43(7), 157.) [5]. It has received 71.53% accuracy. Aditya Arya, Sudhanshu, Rohan Agarwal, attempted to show and visualized the result of our study and this project By comparing with other techniques, it scores accuracy of 68.5% (Arya, A., Sudhanshu, & Agarwal, R. (2018). Disease prediction using machine learning algorithms. International Journal of Advanced Research in Computer Science, 9(5), 415-419.) [6]. Iqra Anjum, Mohammed Afreed, Mohammed Kalam has developed a system which predicts the disease based on the information or the symptoms he/she enter into the system and provides the accurate results based on that information (Anjum, I., Afreed, M., & Kalam, M. (2020). Disease prediction using machine learning. International Journal of Advanced Science and Technology, 29(5), 1592-1601.) [7]. Raj H. Chauhan, Daksh N. Naik, Rinal A. Halpati, Sagarkumar J. Patel, Mr. A.D. Prajapati developed a system analyzes the symptoms provided by the user as input and gives the probability of the disease as an output Disease Prediction is done by implementing the Decision tree Classifier (Chauhan, R. H., Naik, D. N., Halpati, R. A., Patel, S. J., & Prajapati, A. D. (2021). Disease prediction using decision tree classifier. In Proceedings of the 2021 3rd International Conference on Advances in Electronics, Computers and Communications (pp. 148-152). IEEE) [8]. Decision tree Classifier calculates the probability of the disease. With big data growth in biomedical and health care communities, and it also provides the probability estimation of the system such it shows the assorted probability like how the system behaves when there are n number of predictions are done and it also does the recommendations for the patients from their final result and also from their symptoms prefer it can show what to use and what to not use from the given datasets and therefore the final results. Here we've got combined the structure and unstructured form to data for the general risk analysis that's required for doing the prediction of the disease. Using the structured analysis, we will identify the chronic kinds of disease in an exceedingly particular region and particular community. In unstructured analysis we select the features automatically with the help of algorithms and techniques. This technique takes symptoms from the user and predicts the disease accordingly supported the symptoms that it takes and also from the previous datasets, it also helps in continuous evaluation of viral diseases, heart rate, blood pressure, sugar level and far more which is within the system and along with other external symptoms its predicts the acceptable and accurate disease and it gives the prescription details for that disease. And also, the data entered by the user are stored within the created database.

VI. EVALUATING THE MODEL % RESULT

The result obtained from our model are summarized in the following table:

Model	Accuracy
Liver disease	77.97%
Kidney disease	90%
Heart disease	85%

VII. CONCLUSION

So, Finally I conclude by saying that, this project Disease prediction using machine learning is very much useful in everyone's day to day life and it is mainly more Important for the healthcare sector, because they are the one that daily uses these systems to predict the diseases of the patients based on their general information and there symptoms that they are been through. Now a day's health industry plays major role in curing the diseases of the patients so this is also some kind of help for the health industry to tell the user and also it is useful for the user in case he/she doesn't want to go to the hospital or any other clinics, so just by entering the symptoms and all other useful information the user can get to know the disease he/she is suffering from and the health industry can also get benefit from this system by just asking the symptoms from the user and entering in the system and in just few seconds they can tell the exact and up to some extent the accurate diseases. If health industry adopts this project then the work of the doctors can be reduced and they can easily predict the disease of the patient. The Disease prediction is to provide prediction for the various and generally occurring diseases that when unchecked and sometimes ignored can turns into fatal disease and cause lot of problem to the patient and as well as their family members.

DECLARATION

Funding/ Grants/ Financial Support	No, We did not receive.
Conflicts of Interest/ Competing Interests	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval and consent to participate with evidence
Availability of Data and Material/ Data Access Statement	Data sets are taken from the Kaggle community.
Authors Contributions	All this project is completed under the guidance (Author) Dr Alok katiyar. All documentation work and somewhat research has been done by (Author) Sajid Ali. Complete Web applications and Research has been done by (Author) Sameer Ray.

REFRENCES

1. Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., & Lo, B. et al. (2017). Deep Learning for Health Informatics. IEEE Journal of Biomedical and Health Informatics, 21(1), 4-21. [CrossRef]
2. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. [CrossRef]
3. Kaggle. (n.d.). Chronic Kidney Disease Dataset. Retrieved from https://www.kaggle.com/mansoordaku/ckdisease
4. Mehta, N., Pandey, S., & Verma, S. (2017). Disease Prediction System using Machine Learning over Cloud. 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS), 331-336.

5. Balasubramanian, S., & Subramani, B. (2019). A system for reducing the multiple diseases showing the similar symptoms problem using machine learning techniques. Journal of Medical Systems, 43(7), 157.
6. Arya, A., Sudhanshu, & Agarwal, R. (2018). Disease prediction using machine learning algorithms. International Journal of Advanced Research in Computer Science, 9(5), 415-419.
7. Anjum, I., Afreed, M., & Kalam, M. (2020). Disease prediction using machine learning. International Journal of Advanced Science and Technology, 29(5), 1592-1601.
8. Chauhan, R. H., Naik, D. N., Halpati, R. A., Patel, S. J., & Prajapati, A. D. (2021). Disease prediction using decision tree classifier. In Proceedings of the 2021 3rd International Conference on Advances in Electronics, Computers and Communications (pp. 148-152). IEEE.

AUTHORS PROFILE

Dr. Alok Katiyar, Professor, Department of CSE (Galgotias University Greater Noida, Gautam Buddha Nagar, Uttar Pradesh, India **About**: Dr. Alok Katiyar is a highly respected professor in the Department of Computer Science and Engineering at Galgotias University, located in Greater Noida, Gautam Buddha Nagar, Uttar Pradesh, India. With years of experience in academia and industry, Dr. Katiyar is known for his expertise in areas such as data analytics, machine learning, and artificial intelligence. He is an accomplished researcher and has published numerous papers in top-tier conferences and journals. Dr. Katiyar is also a sought-after speaker, having delivered talks at various national and international conferences. His dedication to teaching and research has earned him the respect and admiration of his students, colleagues, and peers in the academic community.

Sajid Ali, Student, Department of CSE (Galgotias University Greater Noida, Gautam Buddha Nagar, Uttar Pradesh, India. **About:** Sajid Ali is a dedicated student pursuing his education in the Department of Computer Science and Engineering at Galgotias University in Greater Noida, Uttar Pradesh, India. With a passion for technology, he is constantly exploring new areas within the field of computer science and working to expand his knowledge and skills. Sajid is a hardworking individual with a keen eye for detail and a commitment to excellence. He is actively involved in various extracurricular activities, such as coding contests and hackathons, which allow him to apply his skills and gain practical experience. Sajid's enthusiasm for computer science and his drive to succeed make him a valuable member of the Galgotias University community.

Sameer Ray, Student, Department of CSE (Galgotias University Greater Noida, Gautam Buddha Nagar, Uttar Pradesh, India. **About:** Sameer Ray is a dedicated and enthusiastic student currently pursuing his studies in the Department of Computer Science and Engineering at Galgotias University, located in Greater Noida, Uttar Pradesh, India. With a passion for technology and a desire to innovate, Sameer has already begun to establish himself as a promising young talent within the field. He is known for his strong work ethic and commitment to learning, as well as his ability to work well within a team. With his drive and determination, Sameer is sure to achieve great things both academically and professionally in the years to come.