Name- Kalpita Vilas Birhade

Internship Domain- Data Science

Batch-June 2022- August 2022

Date of Submission- 15/08/2022

Fastfindfirm, Thane,

Maharashtra, India 400610

Multiple Disease Prediction System

A case study submitted for the mandate of

fastfindfirm

for internship work completed during

Data Science internship

By

Kalpita Vilas Birhade

Multiple Disease Prediction System

Kalpita Vilas Birhade
Data Science
Fastfindfirm
Thane, Maharashtra
kalpitabirhade28@gmail.com

***Abstract*— The project implements 3 linear models and one deep learning model: Naïve Bayes, Support Vector Machine, K-Nearest Neighbors network to investigate their performance on diabetes, heart and Parkinson's disease datasets obtained from the UCI data repository. In addition to the comparison of the algorithms, each algorithm has been integrated into a prediction engine. The project also includes a web platform to facilitate collaboration among researchers and doctors. As the results show, our prediction engine is capable of recognizing the presence of the disease and also predict it accurately. Performance improvements could also be achieved by using complex deep learning methods.**

Keywords—disease, diabetes, heart, Parkinson's, data science, machine learning, web app.

I. INTRODUCTION

In this digital world, data is an asset, and enormous data was generated in all the fields. Data in the healthcare industry consists of all the information related to patients. Here a general architecture has been proposed for predicting the disease in the healthcare industry. Many of the existing models are concentrating on one disease per analysis. Like one analysis for diabetes analysis, one for heart analysis, one for parisons diseases like that. There is no common system present that can analyze more than one disease at a time. Thus, we are concentrating on providing immediate and accurate disease predictions to the users about the symptoms they enter along with the disease predicted. So, we are proposing a system which used to predict multiple diseases by using python web app. In this system, we are going to analyze Diabetes, Heart, and parisons disease analysis. Later many more diseases can be included. To implement multiple disease prediction system, we are going to use machine learning algorithms, and deep learning. Python pickling is used to save the behavior of the model. The importance of this system analysis is that while analyzing the diseases all the parameters which cause the disease is included so it is possible to detect the disease efficiently and more accurately. The final model's behavior will be saved as a python pickle file.

A. Description

A lot of analysis over existing systems in the health care industry considered only one disease at a time. For example, one system is used to analyze diabetes, another is used to analyze diabetes retinopathy, and another system is used to predict heart disease. Maximum systems focus on a particular disease. When an organization wants to analyze their patient's, health reports then they have to deploy many models. The approach in the existing system is useful to analyze only particular diseases. In multiple diseases prediction system, a user can analyze more than one disease on a single website. The user doesn't need to traverse different places in order to predict whether he/she has a particular disease or not. In multiple diseases prediction system, the user needs to select the name of the particular disease, enter its parameters and just click on submit. The corresponding machine learning model will be invoked and it would predict the output and display it on the screen

B. Problem statement

Many of the existing machine learning models for health care analysis are concentrating on one disease per analysis. For example, first is for liver analysis, one for cancer analysis, one for lung diseases like that. If a user wants to predict more than one disease, he/she has to go through different sites. There is no common system where one analysis can perform more than one disease prediction. Some of the models have lower accuracy which can seriously affect patients' health. When an organization wants to analyze their patient's health reports, they have to deploy many models which in turn increases the cost as well as time Some of the existing systems consider very few parameters which can yield false results.

C. Proposed system

In multiple disease prediction, it is possible to predict more than one disease at a time. So, the user doesn't need to traverse different sites in order to predict the diseases. We are taking three diseases that are parisons, Diabetes, and Heart. As all the three diseases are correlated to each other. To implement multiple disease analyses we are going to use machine learning algorithms. When the user is accessing this data sets, the user has to send the parameters of the disease along with the disease name.

II. LITERATURE REVIEW

1. According to the paper focuses about as diabetes is one of the dangerous diseases in the world, it can cause many varieties of disorders which includes blindness etc. In this paper they have used machine learning techniques to find out diabetes disease as it is easy and flexible to forecast whether

the patient has illness or not. Their aim of this analysis was to invent a system that can help the patient to detect the diabetes disease of the patient with accurate results. Here they used mainly 4 main algorithms Decision Tree, Naïve Bayes, and SVM algorithms and compared their accuracy which is 85%,77%, 77.3% respectively. They also used ANN algorithm after the training process to see the reactions of the network which states whether the disease is classified properly or not. Here they compared the precision recall and F1 score support and accuracy of all the models.

2. The main aim of the paper is, as heart plays an important role in living organisms. So, the diagnosis and prediction of heart related disease should be perfect and correct because it is very crucial which can cause death cases related to heart. So, Machine learning and Artificial Intelligence supports in predicting any kind of natural events. So, in this paper they calculate accuracy of machine learning for predicting heart disease using k-nearest neighbor ,decision tree, linear regression and SVM by using UCI repositor dataset for training and testing . They also compared the algorithm and their accuracy SVM 83 %, Decision tree 79%, Linear regression 78%, k-nearest neighbors 87%.

3. The system defines that liver diseases are causing high number of deaths in India and is also considered as a life threating disease in the world. As it is difficult to detect the liver disease at early stage. So, using automated program using machine learning algorithms we can detect the liver disease accurately. They used and compared SVM, Decision Tree and Random Forest algorithm and measures precision, accuracy and recall metrics for quantitative measurement. The accuracy is 95%,87%,92% respectively.

III. SYSTEM ANALYSIS

A. *Functional requirement*

- The system allows the patient to predict the disease
- The user adds the input for the particular disease and based on the trained model of the user input the output will be displayed.

B. *Nonfunctional requirement*

- The website will provide range of the values during the prediction of the disease.
- The website should be reliable and consistent.

Implementation

In the system diabetes disease prediction model used knn algorithm, heart disease uses the xgboost algorithm and liver uses the random forest algorithm as these gave the best accuracy accordingly. There when the patient adds the parameter according to the disease it will show whether the patient has a disease or not according to the disease selected. The parameters will show the range of the values needed and if the value is not between the range or is not valid or is empty it will show the warning sign that add a correct value

ACCURACY FOR EACH DISEASE:

Table No. 1: Diabetes Disease

Algorithm	diabetes
Random forest	88%
XGBoost	89%

Table No. 2: Heart disease

Algorithm	Heart
KNN	85%
Random forest	77%

Table No.3: Parkinson's disease

Algorithm	Parkinson's
Random forest	73%
XGBoost	68%

Figure No. 1: Diabetes disease input data

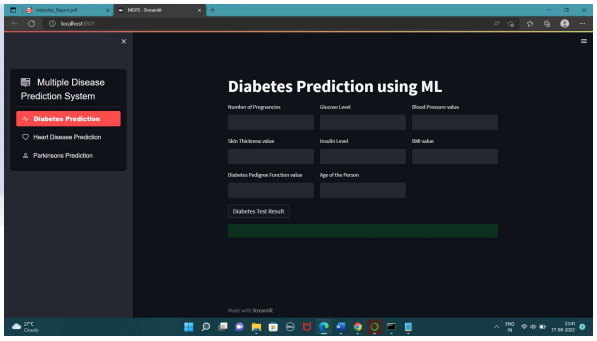


Figure No. 2: Parkinson's disease input data

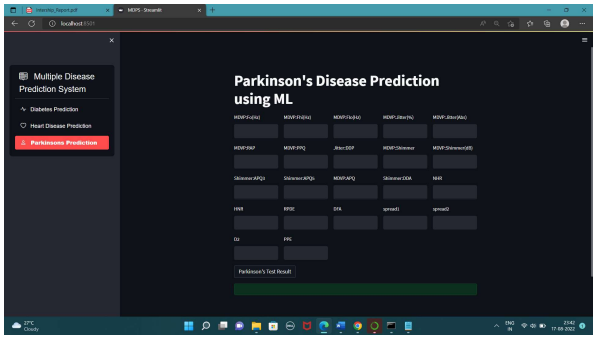


Figure No. 3: Heart disease input data

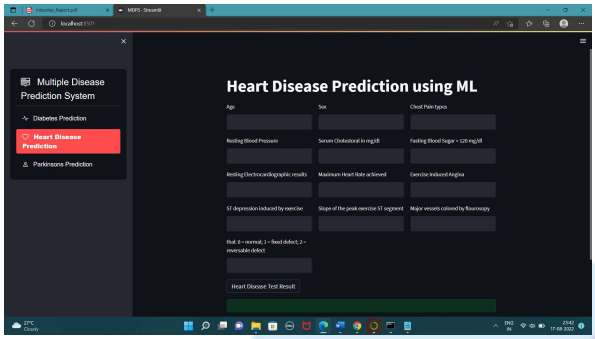
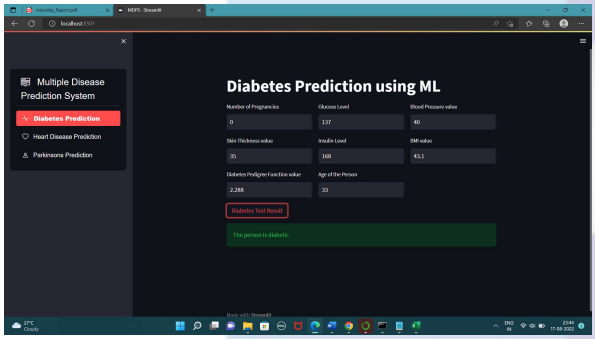


Figure No. 4: Diabetes prediction result



ACKNOWLEDGMENT

We take this opportunity to express our sincere gratitude and respect to FastFindFirm for providing us a platform to pursue our studies and carry out our internship project

We have a great pleasure in expressing our deep sense of gratitude to Krishna sir, FastFindFirm, Mumbai, for his constant encouragement. We would like to thank all team members of FastFindFirm, HR, FastFindFirm, Mumbai, who has been a constant support and encouragement throughout the course of this project. We consider it a privilege and honor to express our sincere gratitude to our guide Krishna sir, for the valuable guidance throughout the tenure of this review.

We also extend our thanks to all the Team of FastFindFirm who directly or indirectly encouraged us. Finally, we would like to thank our parents and friends for all their moral support they have given us during the completion of this work.

REFERENCES

[1] P. Groves, B. Kayyali, D. Knott and S. van Kuiken, The'Big Data'Revolution in Healthcare: Accelerating Value and Innovation, 2016.

[2] *Hong W, Xiong Z, Zheng N and Weng Y 2019 A medical-history-based potential disease prediction algorithm; IEEE Access. 7 131094–131101*

[3] *Haq A Q, Li J P, Memon M H, Khan J, Malik A, Ahmad T, Ali A, Nazir S, Ahad I and Shahid M 2019 Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson's disease using voice recordings; IEEE Access 7: 37718–37734*

[4] *S. Rao Kondapally Seshasai et al., "Diabetes mellitus fasting glucose and risk of cause-specific death", New Engl. J. Med., vol. 364, no. 9, pp. 829-841, Mar. 2011.*

[5] *K. Vembandasamyp, R. R. Sasipriyap and E. Deepap, "Heart Diseases Detection Using Naive Bayes Algorithm", IJISET-International J. Innov. Sci. Eng. Technol, vol. 2, no. 9, 2015, [online] Available: www.ijiset.com.*

[6] *H. Karthikeyan and T Menakadevi, "Multi-disease prediction model using improved SVM-radial bias technique in healthcare monitoring system", Journal of Ambient Intelligence and Humanized Computing, 2020.*