Disease Prediction Using Machine Learning Algorithms

Presented by

Lalit Mahajan.

Santosh Wakade.

Data Analytics

Project Guide
Prabhjyot Singh
Data Analytics
Edubrigde Technology

Table of contents

- Abstract
- Introduction and relevance
- Literature survey
- Block diagram
- Database
- Problem definition
- Application of project
- Software and hardware resources
- Methodologies and algorithms
- Objectives
- Outcome of the project
- System design
- Conclusion
- Future scope
- References

Abstract

- The development and exploitation of several prominent Data mining techniques in numerous real-world application areas (e.g. Industry, Healthcare and Bio science) has led to the utilization of such techniques in machine learning environments, in order to extract useful pieces of information of the specified data in healthcare communities, biomedical fields etc.

- The accurate analysis of medical database benefits in early disease prediction. The aim of developing proposed system using machine learning algorithms is to immensely help to solve the health-related issues by assisting the physicians to predict and diagnose diseases at an early stage.

- A Sample data of 4920 patients' records diagnosed with 41 diseases was selected for analysis. A dependent variable was composed of 41 diseases. 95 of 132 independent variables(symptoms) closely related to diseases were selected and optimized. The proposed system is developing by using Machine learning algorithms such as Decision Tree classifier, Random forest classifier, and Naïve Bayes classifier.

Introduction and Relevance

- The proposed system predicts the disease of the patient based on the information or the symptoms he/she enter into the proposed system and provides the accurate results based on that information.

- 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 and also it is useful for the patients in case he/she doesn't want to go to the hospital or any other clinics, so just by entering the symptoms the patients/user can get to know the disease he/she is suffering From.

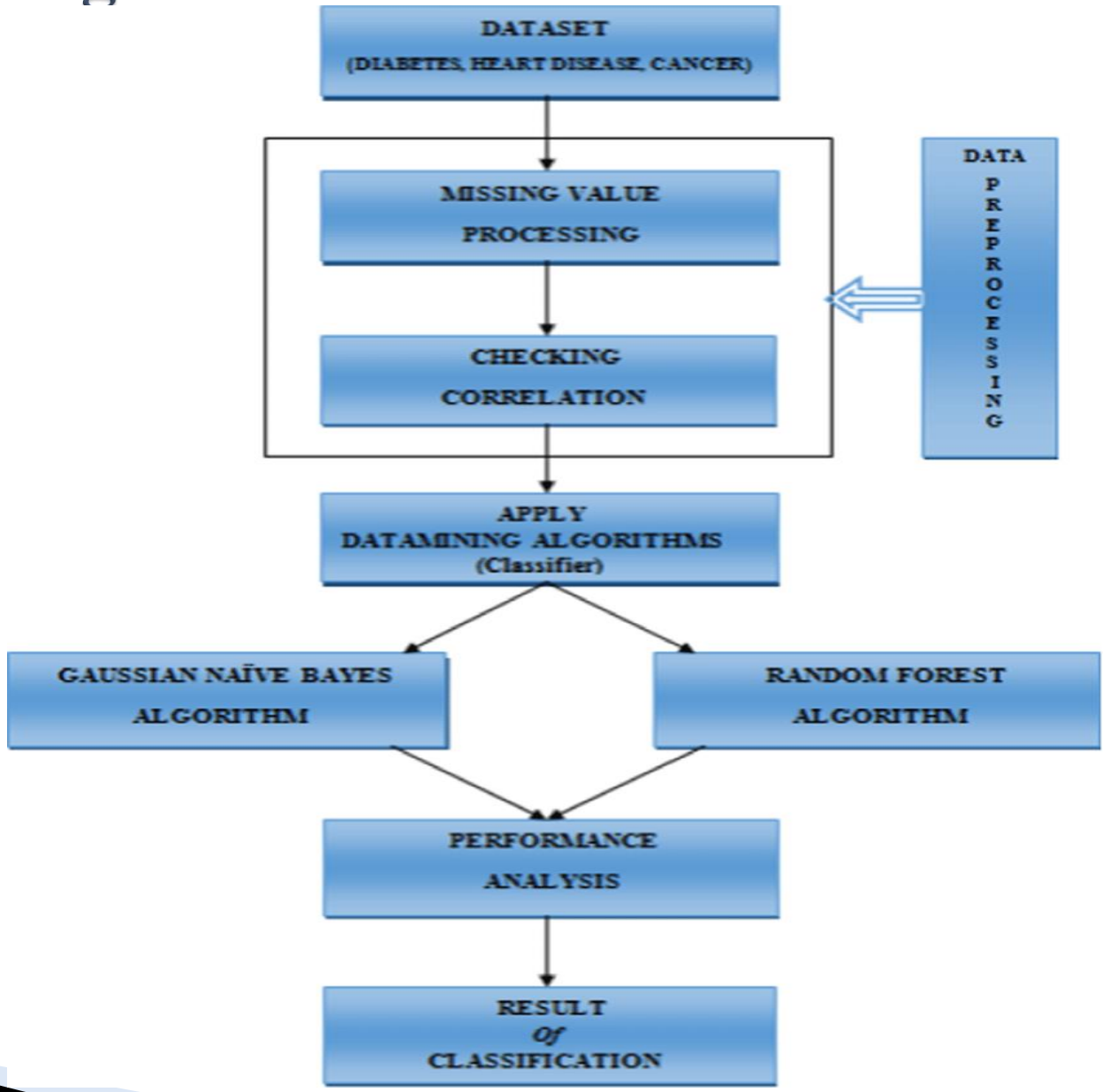
- The proposed system is doing with the help of Machine Learning Algorithms and Python Programming language with Tkinter , Numpy, Pandas, Sklearn libraries.

- This proposed system uses Machine Learning algorithms , For predicting diseases Naïve Bayes algorithm, for clustering KNN algorithm, is used.

Literature Review (previous investigations)

NAME OF CONFERE/ JOURNAL	Author name, / Year	Paper title	Proposed method	Advantages of the system	Limitations
2020 International Conference for Emerging Technology (INCET)	1. Sneha Grampurohit . 2. Chetan Sagarnal, / Jun 5-7, 2020.	Disease Prediction using Machine Learning Algorithms	Decision tree classifier	The accuracy in terms of percentage: 95.12 percentage.	**It led to Overfitted**
Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019)	1. Dhiraj Dahiwade. 2. Prof. Gajanan Patle. 3. Prof. Ektaa Meshram.	Designing Disease Prediction Model Using Machine Learning Approach	K-Nearest Neighbor (KNN) and Convolutional neural network (CNN)	Low time consumption and minimal cost possible for disease prediction and risk prediction	**Limitation of this paper they could not consider large dataset**

Block Diagram



Database

- CSV (comma-separated values) FILE IN USE

A CSV (**comma-separated values**) file is a text file that has a specific format which allows data to be saved in a table structured format.

Problem Definition

- Now a days in Health Industry there are various problems related to machines or devices which will give wrong or unaccepted results, so to avoid those results and get the correct and desired results we are building a program or project which will give the accurate predictions based on information provided by the user and also based on the datasets that are available in that machine.

- The health industry in information yet and knowledge poor and this industry is very vast industry which has lot of work to be done. So, with the help of all those algorithms, techniques and methodologies we have done this project which will help the peoples who are in the need.

- So the problem here is that many people goes to hospitals or clinic to know how is their health and how much they are improving in the given days, but they have to travel to get to know there answers and sometimes the patients may or may not get the results based on various factors such as doctor might be on leave or some whether problem so he might not have come to the hospital and many more reasons will be there so to avoid all those reasons and confusion we are making a project which willhelp all those person's and all the patients who are in need to know the condition of their health, and at sometimes if the person has been observing few symptoms and he/she is notsure about the disease he/she is encountered with so this will lead to various diseases in future.

- So, to avoid that and get to know the disease in early stages of the symptoms this disease prediction will help a lot to the various people's ranging from children to teenagersto adults and also the senior citizens.

Application of project

- Mostly used in Hospital, Rural area, medical store etc.
- It is minorly used in Army sector.
- Anyone can easily use that application.
- After understanding or predict a disease, take action accordingly.

Software and hardware resources

- Software resources.
  - Operating System : Windows 7, 10 or Higher Versions
  - Platform : Jupiter Notebook
  - Front End : Python Tkinter
  - Back End : Python and Files
  - Programming Lang : Python

- Hardware resources.
  - Laptop with better configuration
  - System : Pentium 4, Intel Core i3, i5, i7 and 2 GHz Minimum
  - RAM : 512Mb or above
  - Hard Disk : 10 GB or above
  - Input Device : Keyboard and Mouse
  - Output Device: Monitor or PC

Methodologies and algorithms

- Decision Tree Algorithm :
  - The decision tree splits the nodes on all available variables and then selects the split which results in most homogeneous sub-nodes. The ID3 algorithm builds decision trees using a top-down greedy search approach through the space of possible branches with no backtracking.
- RANDOM FOREST ALGORITHM :
  - Random forest algorithms have three main hyperparameters, which need to be set before training. These include node size, the number of trees, and the number of features sampled. From there, the random forest classifier can be used to solve for regression or classification problems.
- NAÏVE BAYES ALGORITHM :
  - Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset.

Objectives:

The purpose of making this proposed system called “Disease Prediction Using Machine Learning algorithms” is to predict the accurate disease of the patient using all their symptoms.

- If this Prediction is done at the early stages of the disease with the help of this proposed system and all other necessary measure the disease can be cured and in general this prediction system can also be very useful in health industry.

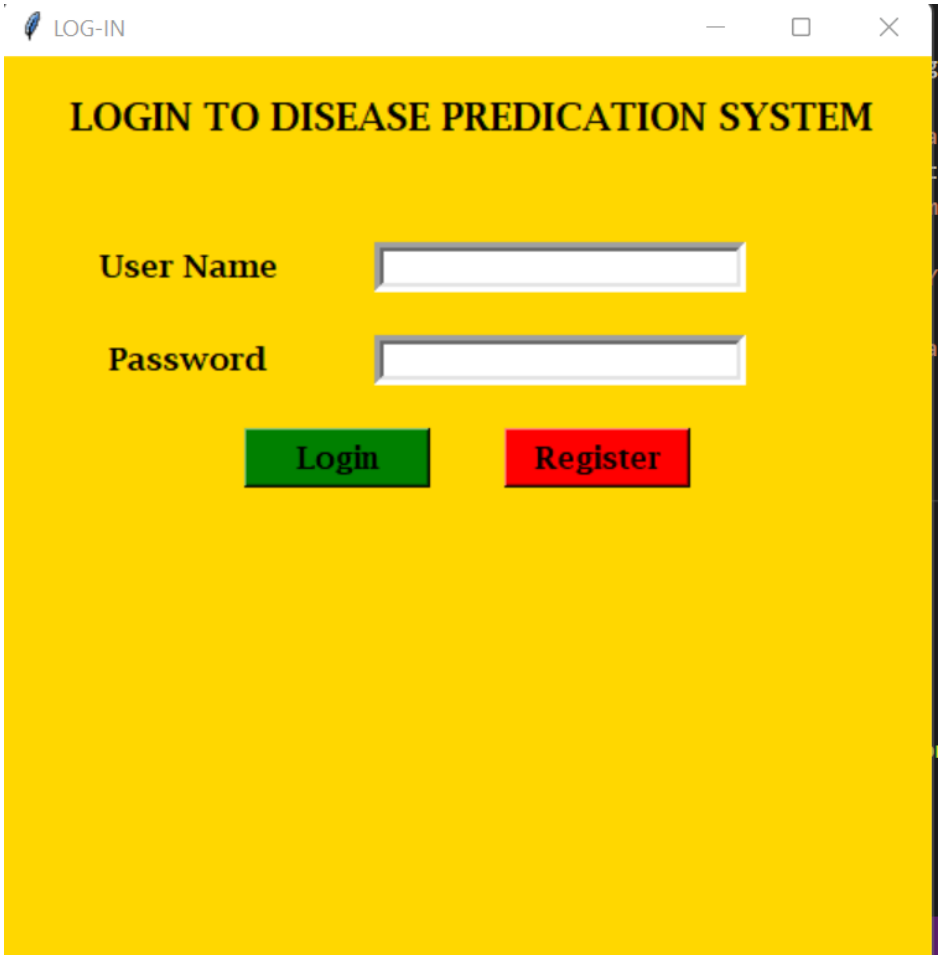
- If health industry adopts this project then the work of the doctors can be reduced and they can easily predict the disease of the patient.

Outcome of the project

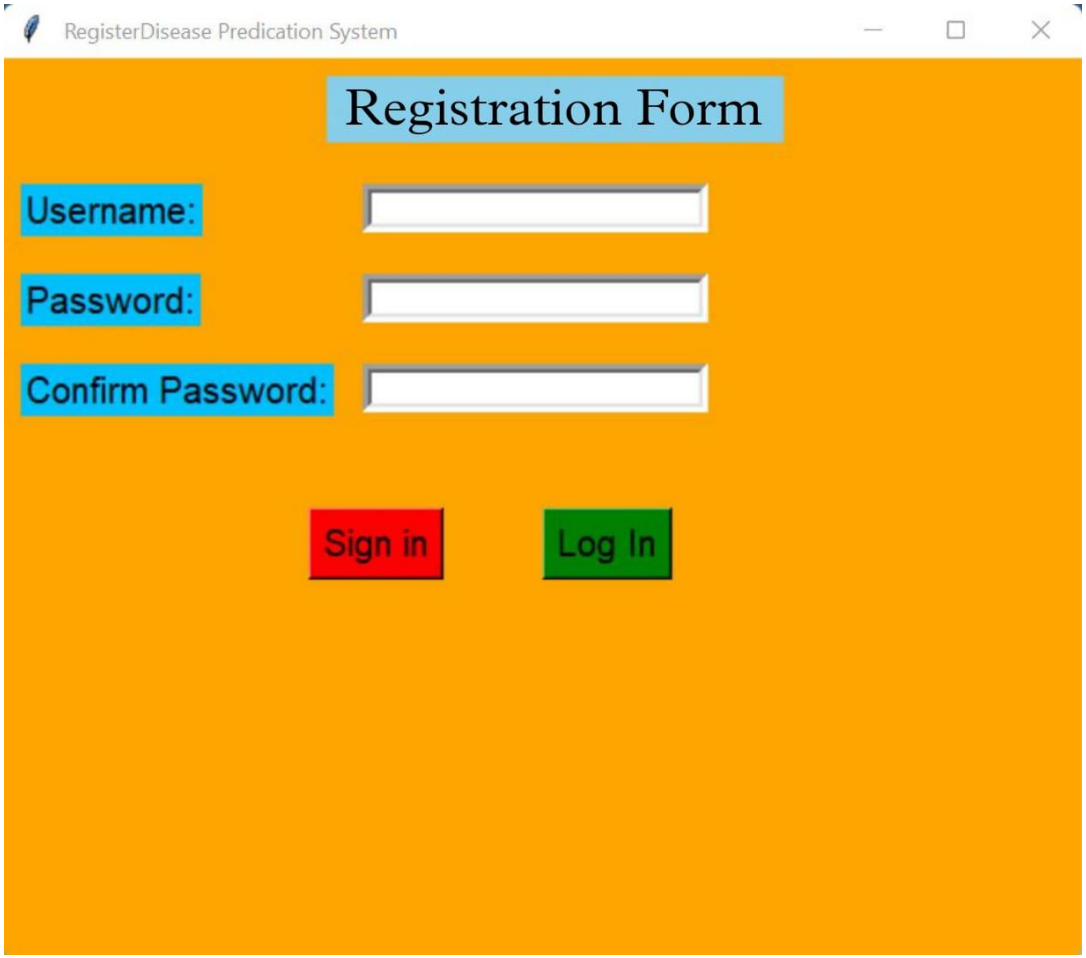
1. It is difficult to handle the huge amount of data of the patients. It is easier to handle this data through Big Data Analytics.

2. Developing a medical diagnosis system based on machine learning (ML) algorithms for prediction of any disease can help in a more accurate diagnosis

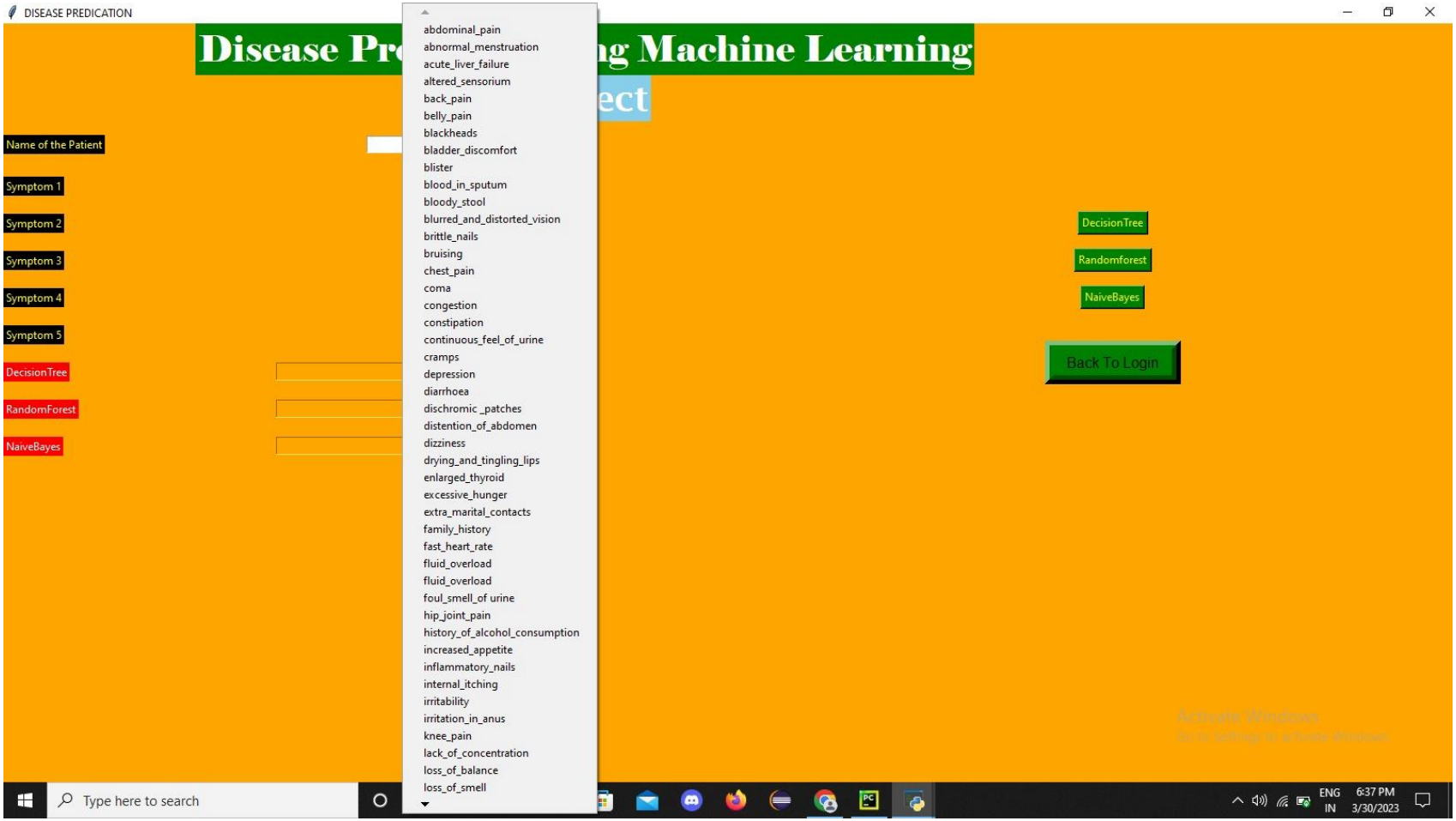
System design



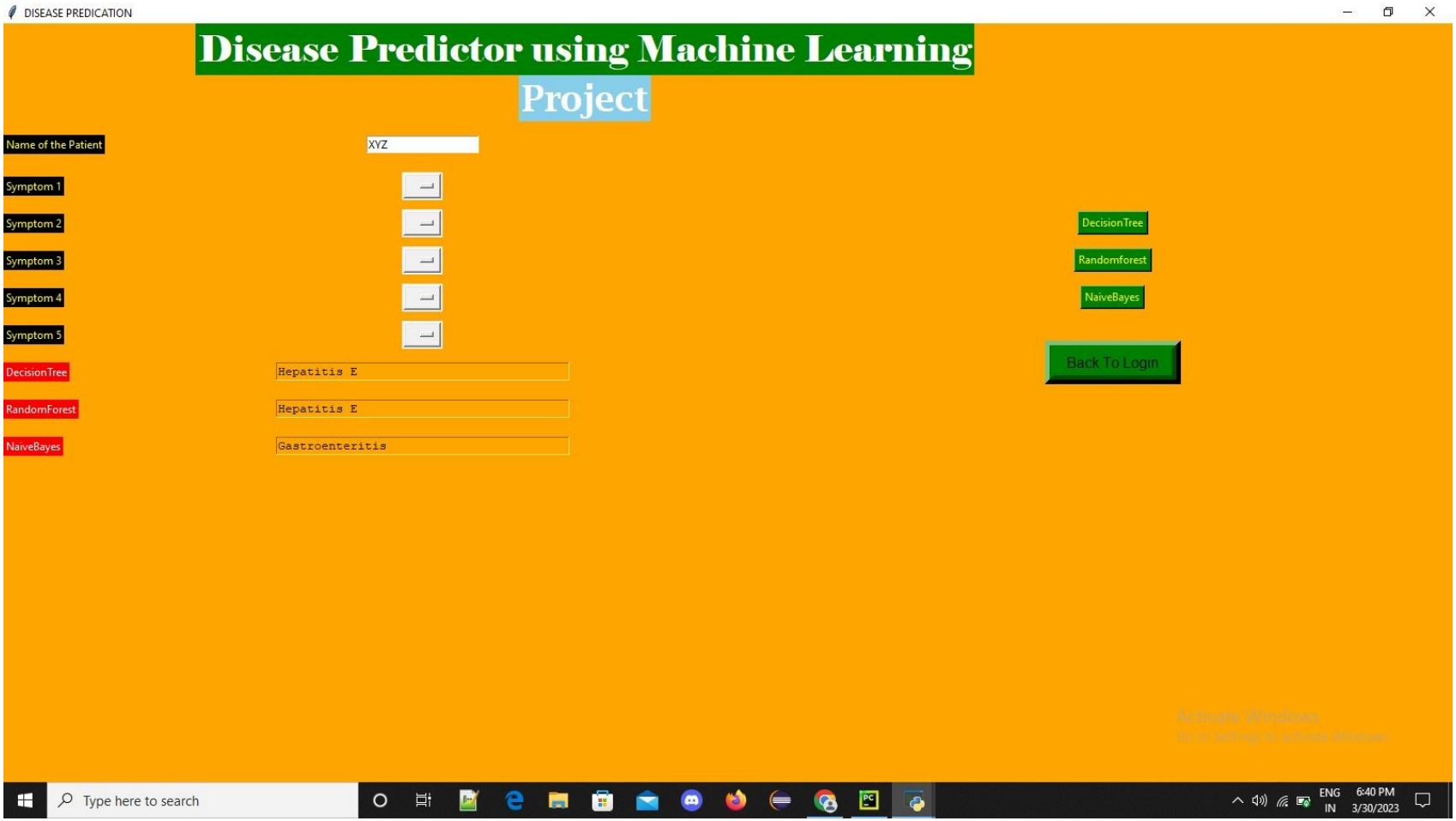
Login Page Window



Registration Form



Project Window



Output Window

References

- Min Chen, Yixue Hao, Kai Hwang, Fellow, IEEE, Lu Wang, and Lin Wang “Disease Prediction by Machine Learning over Big Data from Healthcare Communities” (2017).
- "Disease prediction by machine learning over big data from healthcare communitites" IEEE Access ,2017
- Mr. Chala Beyene, Prof. Pooja Kamat, “Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques” , International Journal of Pure and Applied Mathematics, 2018
- S. Patel and H. Patel, “Survey of data mining techniques used in healthcare domain, ” Int. J. of Inform. Sci. and Tech., Vol. 6, pp. 53-60, March 2016
- V.V. Ramalingam, AyantanDandapath, M Karthik Raja, “Heart disease prediction using machine learning techniques: a survey”, International Journal of Engineering & Technology, 2018, pp. 684–687.

Conclusion

- The use of different Machine Learning algorithms enabled the early detection of many maladies such as heart, kidney, breast, and brain diseases and some common disease

- Throughout the literature, SVM (support vector machine), RF (random forest) and LR (logistic regression) algorithms were the most widely used at prediction.

Future scope

- Useful in Health Care System.
- Useful to design different kiosk to give genuine information to patient.
- This System can be furturistic used in blockchain technology to keep record safely.

THANK YOU