Mini Project Report

on

Multiple Disease Prediction

(BTech CSE VI Semester) 2023-2024

Submitted to:

Dr. Susheela

(CC- Section H- VI-Sem)

Submitted by:

Aikansh Shridhar

Roll No.: 11

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GRAPHIC ERA HILL UNIVERSITY, DEHRADUN

CERTIFICATE

Certified that Mr.Aikansh Shridhar (Roll No.- 11) has completed Mini Project with title “Multiple Disease Detection” for the fulfilment of **Course- BTech CSE** VI Semester in Graphic Era Hill University, Dehradun. Student has successfully completed this course as best of myknowledge.

Date: 12-07-2024

(Dr. Susheela)

Class Coordinator

CC-Section H

(VI-Sem)

CSE Department GEHU, Dehradun

ACKNOWLEDGMENT

I would like to express our gratitude to the Almighty, the most Beneficent and the most Merciful Dr. Susheela, for Successful completion of MINI PROJECT.

I wish to thank our parents for their continuing support and encouragement. We also wish to thank them for providing us with the opportunity to reach this far in our studies.

I would like to thank particularly my External Supervisor for his patience, support and encouragement throughout the completion of this Course. At last but not the least I greatly indebted to all other persons who directly or indirectly helped me during this course.

Aikansh Shridhar

Roll No.- 11

Course : BTech CSE -H-VI-Sem

Session: 2023-2024

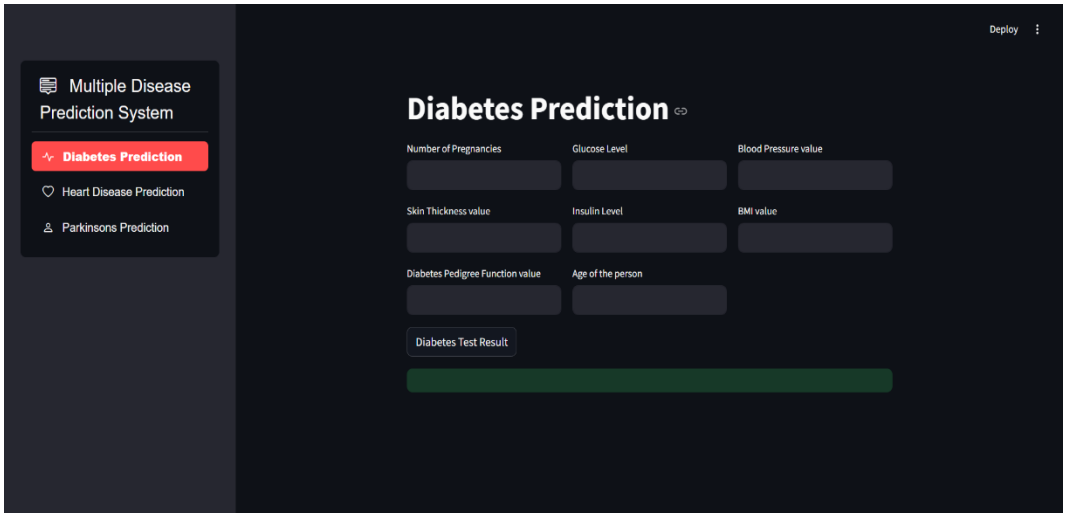
GEHU, Dehradun

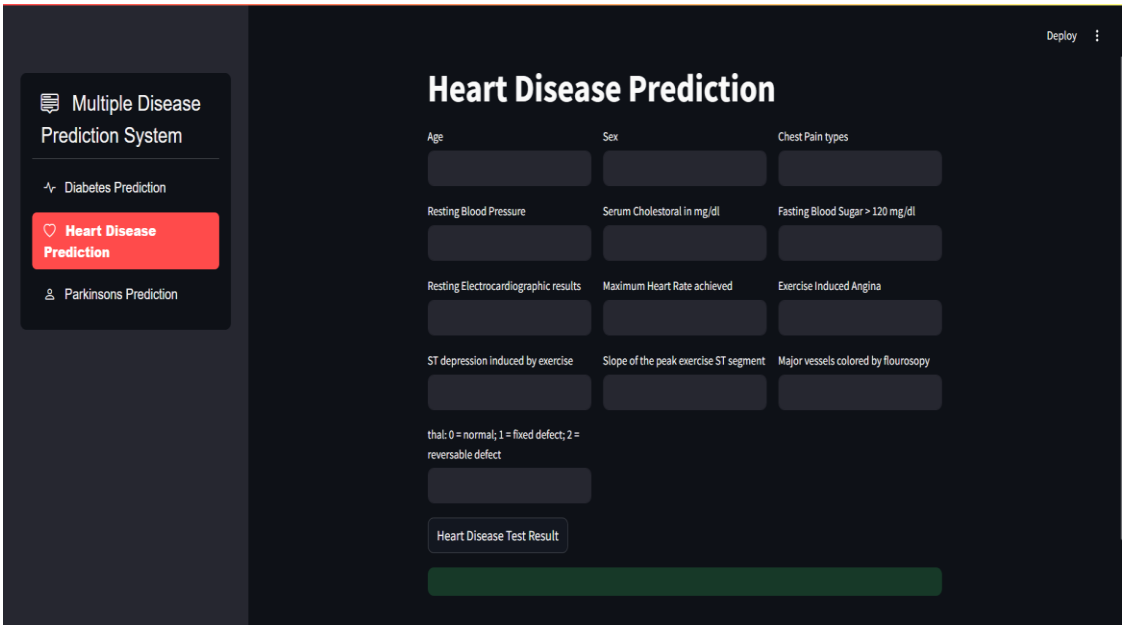
TABLE OF CONTENTS

1. INTRODUCTION

The Title of my Mini Project is “Multiple Disease Prediction”. As the name suggests my project is on detecting any disease like diabetes , Parkinson and heart related disease.

It shows that if the user has the disease or not.





2. RESOURCES USED

2.1 JUPYTER NOTEBOOK

Jupyter Notebook, now known as JupyterLab, is an open-source web-based interactive computing environment that allows you to create and share documents containing live code, visualizations, explanatory text, and more. It supports various programming languages, including Python, R, Julia, and others.

2.2 DATASET

Dataset used in this project was downloaded from Kaggle.com. There are 3 datasets for each type of disease prediction and each dataset has different set of columns which will help in training the model for further prediction.

2.3 VS CODE

Visual Studio Code (VS Code) is a free, open-source code editor developed by Microsoft. It is widely popular among developers due to its versatility, ease of use, and rich set of features. VS Code supports multiple programming languages through built-in functionality and an extensive extension marketplace. Key features include intelligent code completion, debugging tools, integrated Git control, and customizable themes. Its lightweight yet powerful design makes it suitable for various development tasks, from web development to complex software engineering projects, providing a streamlined and efficient coding environment.

3. INFORMATION ABOUT PROJECT

1- We have used 3 datasets having the features as Insulin level , skin depth etc.

2- We have used Linear Regression Algorithm for Heart Disease Detection and Support Vector Machine(SVM) for Diabetes and Parkinson Disease .

3- Linear Regression Algorithm :- Linear regression is a popular algorithm used in statistical modeling and machine learning to establish a relationship between a dependent variable (target variable) and one or more independent variables (features). It assumes a linear relationship between the input variables and the target variable and aims to find the best-fit line that minimizes the difference between the predicted and actual values.

   Support Vector Machine :- Support Vector Machine (SVM) is a powerful supervised learning algorithm commonly used for classification and regression tasks. Developed by Vladimir Vapnik and his colleagues, SVM works by finding the optimal hyperplane that maximally separates different classes in the feature space. It is effective in high-dimensional spaces and can handle both linear and non-linear data through the use of kernel functions. SVM aims to create the widest possible margin between the data points of different classes, enhancing the model's generalization ability. Its robustness and versatility make SVM a popular choice for various applications, including text categorization, image recognition, and bioinformatics.

4- We have to select multiple inputs like insulin level , number of pregnancies etc.

4. CONCLUSION

In conclusion, the multiple disease prediction mini-project successfully demonstrates the capability of machine learning models to accurately detect diseases such as diabetes, Parkinson's, and heart disease. By utilizing Support Vector Machines (SVM) and preprocessing techniques, we developed robust classifiers for each disease, achieving commendable accuracy levels. The project underscores the importance of data preprocessing, feature scaling, and model selection in building effective predictive models. The integration of these models into a unified system showcases a practical approach to early disease detection, which can significantly aid in timely medical intervention and personalized healthcare. Future enhancements could include incorporating more advanced algorithms, larger and more diverse datasets, and real-time prediction capabilities to further improve the system's accuracy and applicability in real-world scenarios.