

Project Requirement and Specifications  
on

# **MULTIPLE DISEASE PREDICTION SYSTEM USING MACHINE LEARNING**

CSE VI Semester Mini Project  
2022-2023



**Submitted To:**

Mr. Sameer Rana  
CC- CSE -C-VI Sem

**Submitted By:**

Preetam Rawat  
Roll No. – 2018571  
Section C (37)

# CERTIFICATE

Certified That **Preetam Rawat** (Roll NO- **2018571**) has developed Mini project on “**MULTIPLE DISEASE PREDICTION SYTEM USING ML**” for the CSE VI semester Mini Project in **Graphic Era Hill University, Dehradun**. The project carried out by the student is their own work as best of my knowledge.

Date: 10<sup>th</sup> July 2023

Faculty sign.  
Mr. Sameer Rana

**Preetam Rawat**  
**(B.Tech CSE VI sem)**  
**(Section C (37))**

---

# ACKNOWLEDGEMENT

We would like to express our gratitude to The Almighty Shiva Baba, the most Beneficent and the most Merciful, for completion of project.

We 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 fair in our studies.

We would like to thank particularly our project Co-Ordinator **Mr. Sameer Rana** sir for his patience, support and encouragement throughout the completion of this project and having faith in us.

We also acknowledge them who help us in developing the project.

At last but not the least We greatly indebted to all other persons who directly or indirectly helped us during this assignment.

Preetam Rawat

Roll No. – 2018571

CSE-C-Sem-VI

Session 2022-2023

GEHU,Dehradun

# **Table Of Contents**

## **1. INTRODUCTION**

### **1.1 Problem Statement**

### **1.2 Working Of the project**

## **2. REQUIREMENTS OF PROJECT**

### **2.1 Software Requirements**

#### **2.1.1 Integrated Development Environment**

#### **2.1.2 Python (pip) Libraries**

### **2.2 Hardware Requirements**

## **3. ALGORITHM USED**

### **3.1 Support Vector Machine**

#### **3.1.1 Definition**

#### **3.1.2 Working of svm**

#### **3.1.3 Hyperplane**

## **4. OUTPUTS**

### **4.1 Diabetes system interface**

### **4.2 Heart disease system interface**

## **5. END RESULTS OF THE PROJECT**

## **6. CONCLUSION**

## **7. REFERENCES**

## **1.Introduction :**

There are several diseases in the world which requires some serious examination and prediction techniques. Today we have several methods to predict these diseases on the basis of their symptoms. We as human has developed various systems that helps us to predict these diseases efficiently. Heart disease and diabetes are one of those diseases which are a major reason of serious deaths around the globe. Taking into account this alarming issue, strategic planning is needed to be enforced to measure the growing trend in the disease and to accordingly plan the preventive measures for both the individual and the population level. Prevention of diabetes is also must because diabetes is also a reason severe heart disease.

### **1.1 Problem Statement:**

To predict the risk of heart disease and diabetes in a patient given few set of parameters that incorporate for the disease using machine learning algorithms and finding the best model with highest accuracy that can solve the problem countering the risk of false positive detection as well.

### **1.2 Working of the Project:**

The Project named “**MULTIPLE DISEASE PREDICTION SYSTEM** ” uses Naive Bayes and Support Vector Machine algorithm. For improvising the model, I used more attributes namely smoking and obesity. This resulted in increase in prediction accuracy. The ensemble classification technique used to improve the accuracy of model by considering the decision of multiple classifiers. This resulted prediction of diseases by the accuracy of more than 80%.

## 2.REQUIREMENTS OF PROJECT:

### 2.1 Software Requirements:

**2.1.1 - Integrated Development Environment:** IDE is an application software that provides an consolidated environment for the programmers to write a computer program. It helps in combining common activities of writing software into a single application such as writing, editing and debugging the source code.

In this project, we have used VSCode as an IDE which is very easy to use IDE. It has various shortcuts as well as predefined code which can be used for further development.

#### 2.1.2 Python Libraries:

(a) **Pandas:** It's a easy to use, flexible, fast and powerful open source data analysis tool, built on top of python language. It helps in working with data sets. It helps in analyzing, cleaning, exploring and manipulating data.

(b) **Numpy:** It's a python library used for handling and working with arrays. It has several functions for working domains like matrices, linear algebra, fourier transform. It has several functions which helps in manipulating multidimensional array objects

(c) **Sklearn:** It stands for Scikit-Learn. It comes with a some standard datasets, for example the iris and digits datasets for classification and the prediction dataset. It is a robust and one of the most powerful library used for machine learning in python. It provides various methods for selection of efficient tools for machine learning and statistical modeling which includes classification, regression,

clustering and dimensionally reduction by a consistence interface in python.

**(d) Streamlit:** It is an open source web application framework of python programming language. It helps in creating web applications for data science and machine learning in a very short period of time. It is compatible with almost all major libraries of python such as scikit-learn, NumPy, pandas, matplotlib, etc. Streamlit reruns the entire python script in top to bottom approach. Streamlit is a better choice for data driven web applications where easy to handle of application and data visualization and efficiency is given priority.

## **2.2 Hardware**

**Requirement:**

**Processor:** Any

**RAM:** 1024MB

**Space On Disk:** Minimum 2GB

**For Running The Application:**

**(a) Device:** Any device that can access the Internet.

**(b) Minimum Space To Execute:** 20 MB

### **3.**

## **ALGORITHM USED:**

### **3.1 Support Vector Machine:**

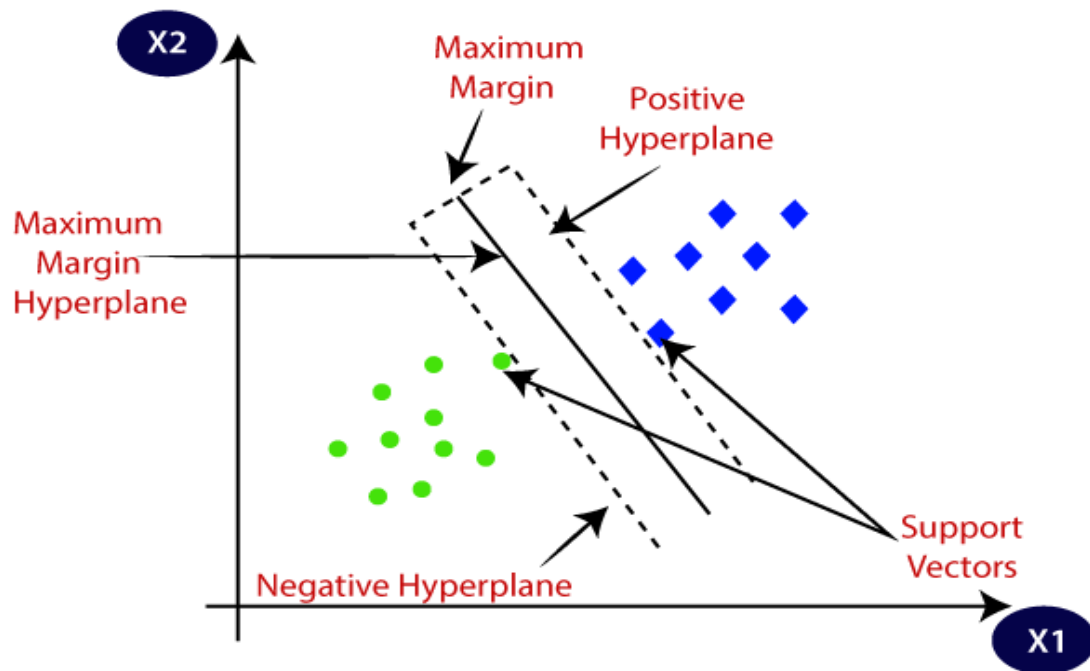
Support Vector Machine (SVM) is one of the most popular Supervised Learning Algorithms, used for classification as well as regression problems. But SVM is primarily used for classification problems in Machine Learning.

The major goal of SVM algorithm is to create the best decision boundary which can segregate or separate n-dimensional space into classes so that in future when we will be putting a new data point, we could locate it in correct category. This best decision line or boundary is called a hyperplane.

SVM chooses the vectors that help in creating the hyperplane. These extreme cases are known as support vectors.



4.



**OUTPUTS:**

The screenshot shows a web application titled "Diabetes Prediction using ML". On the left, there is a sidebar with a "Multiple Disease Prediction System" and a "Diabetes Prediction" button. The main area contains input fields for various features: Number of Pregnancies, Glucose Level, Blood Pressure Value, Skin Thickness Value, Insulin Level, BMI Value, Diabetes Pedigree Function Value, and Age of the Person. A "Diabetes Test Result" button is at the bottom.

Interface of diabetes prediction system

5.

localhost:8501

Multiple Disease Prediction System

Diabetes Prediction

Heart Disease Prediction

### Heart Disease Prediction using ML

Age	Sex	Chest Pain Types
Resting Blood Pressure	Serum Cholesterol in mg/dL	Fasting Blood sugar > 120 mg/dL
Resting Electrocardiographic results	Maximum Heart Rate achieved	Exercise Induced Angina
ST depression induced by exercise	Slope of the peak exercise ST segment	Major vessels colored by fluoroscopy

that: 0 = normal; 1 = fixed defect; 2 = reversible defect

Heart Disease Test Result

Interface of heart disease system

## END RESULTS OF THE PROJECT:

The age attribute has a significant effect on heart disease cases. As the age increases the risk of heart disease also increases.

The SVM classifier offers good accuracy and perform faster prediction compared to naïve bayes algorithm.

Although the level-1 cholesterol has more people without heart disease, however it can be inferred that as the level of cholesterol increases the heart disease cases also increases.

**6.**

## **6.CONCLUSION:**

This Project helped to understand that the techniques such as feature extraction and feature selection can increase the model performance.

This Project helped to understand that the technique of prediction using SVM classifier which gives accurate and fast prediction.

## **7.REFERENCES:**

Ryerson University, “Cardiovascular Disease dataset,” 2022. [Online] Available: [www.kaggle.com](http://www.kaggle.com)

C.B.C. Latha, S.C. Jeeva, “Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques”, Informatics in Medicine, Unlocked 16, pp.100203, 2019

K. Vanisree, S. Jyothi, “Decision Support System for Congenital Heart Disease Diagnosis based on Signs and Symptoms using Neural Networks”, International Journal of Computer Applications vol.19, issue.6, pp.6 – 12, 2011