# PROJECT REPORT

### On

Multiple Disease Prediction system

**(CSE V Semester Mini project) 2023-2024**



**Submitted to: Submitted by:**

Ms. Himadri Vaidya Mr. Shivam Rawat

(CC-CSE-K-V-Sem) Roll. No:2119185 CSE-k-V-Sem

Session: 2023-2024

**DEPARTMENT OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY**

**GRAPHIC ERA HILL UNVERSITY, DEHRADUN**

# CERTIFICATE

Certified that Mr. Shivam Rawat (Roll No.- 2119185) has developed mini project on “Multiple Disease Prediction System” for the CS V Semester Mini Project Lab in Graphic Era Hill University, Dehradun. The project carried out by Students is their own work as best of my knowledge.

Date:

Ms. Himadri Vaidya **Class Co-ordinator CSE-k-V-Sem**

(CSE Department) GEHU Dehradun

**ACKNOWLEDGMENT**

We would like to thank particularly our project Class Coordinator Ms. Himadri Vaidya and encouragement throughout the completion of this project and having faith in us.

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 far in our studies.

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

### Mr. Shivam Rawat

### Roll No.- 2119185

### CSE-K-V-Sem

**Session: 2023-2024 GEHU, Dehradun**

### TABLE OF CONTENTS

**CHAPTER NO. TITLE PAGE NO.**

|  |  |  |
| --- | --- | --- |
| **1.** | **Introduction**   * 1. **What is Multiple Disease Prediction System?**   **1.2 What Disease Model can predict?** | **6** |
| **2.** | **Methodology**   * 1. **Data collection**   2. **Data Preprocessing**   3. **Data Visualization** | **7** |
| **3.** | **Algorithm Used**    **3.1 Support Vector Machine**  **3.2 Logistic Regression** | **10** |
| **4.** | **Conclusion** | **14** |
| **5.** | **References** | **15** |

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

1. **INTRODUCTION**

#### **Multiple Disease Prediction System**

Building a multiple disease prediction model using machine learning involves developing a system that can predict the likelihood of different diseases based on various input features.

A multiple disease prediction model is a valuable tool that contributes to a paradigm shift from reactive to proactive healthcare. By leveraging data and predictive analytics, healthcare systems can focus on prevention, early detection, and personalized care, ultimately leading to better health outcomes for individuals and populations.

#### What Disease Model can predict?

A multiple disease prediction model can be designed to predict various diseases based on the available data and features. The diseases that can be predicted depend on the scope and focus of the model, as well as the availability of relevant data.

The following disease it can predict :

1.Diabetes Disease

2.Heart Disease

3.Parkinson Disease

#### 

## METHODOLOGY

The methodology for developing a multiple disease prediction model involves a systematic approach, combining elements of data preparation, model building, evaluation, and deployment.

The methodology for developing a multiple disease prediction model is iterative, requiring collaboration between data scientists, healthcare professionals, and other stakeholders. Regular validation, updates, and ongoing monitoring are essential to ensure the model remains accurate and relevant over time.

#### Data Collection

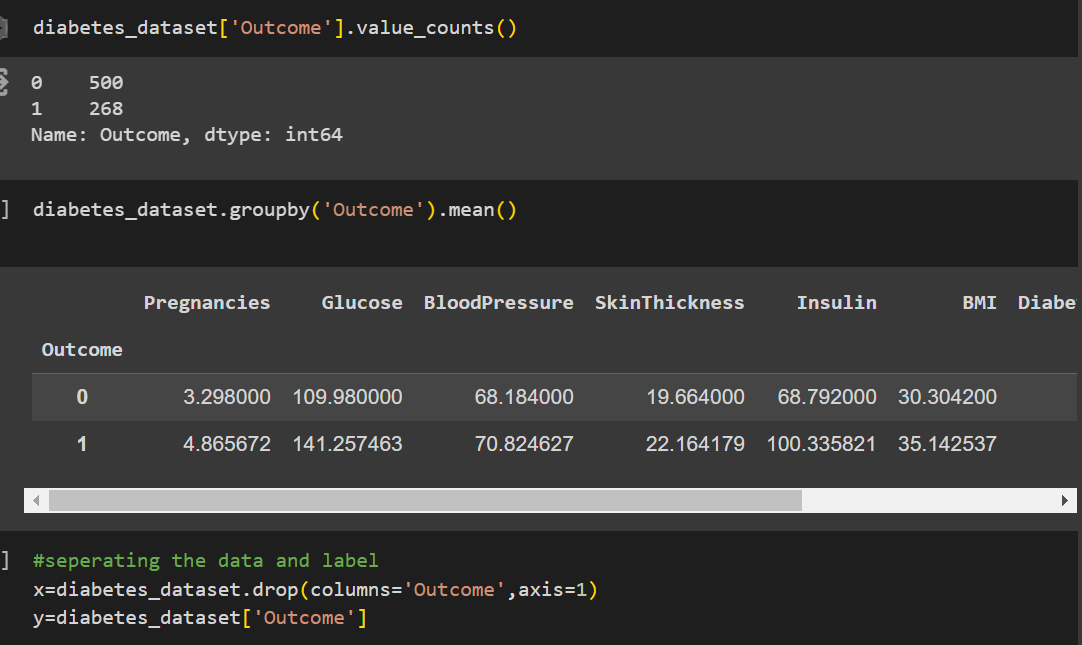
* Gather a comprehensive dataset that includes information on patients, their medical history, lifestyle factors, and relevant health metrics.
* Ensure that the dataset is diverse and representative of the population you want to make predictions for.

#### Data Preprocessing:

* Handle missing data by imputing values or removing instances with incomplete information.
* Normalize or standardize numerical features.
* Encode categorical variables.
* Address any data imbalance issues if present.

#### Data Visualization

In the realm of Twitter sentiment analysis, effective data visualization is crucial for gaining insights, identifying patterns, and presenting findings in a comprehensible manner. In this section, we discuss the data visualization techniques employed to enhance the understanding of the sentiment analysis results.

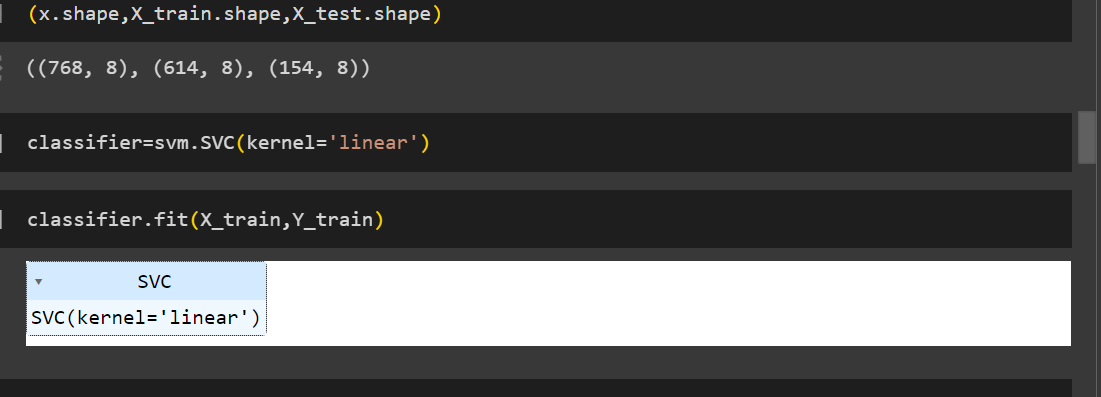


### 2.3 Feature Selection

* Identify the most relevant features for disease prediction.
* Use techniques like feature importance analysis, recursive feature elimination, or dimensionality reduction.

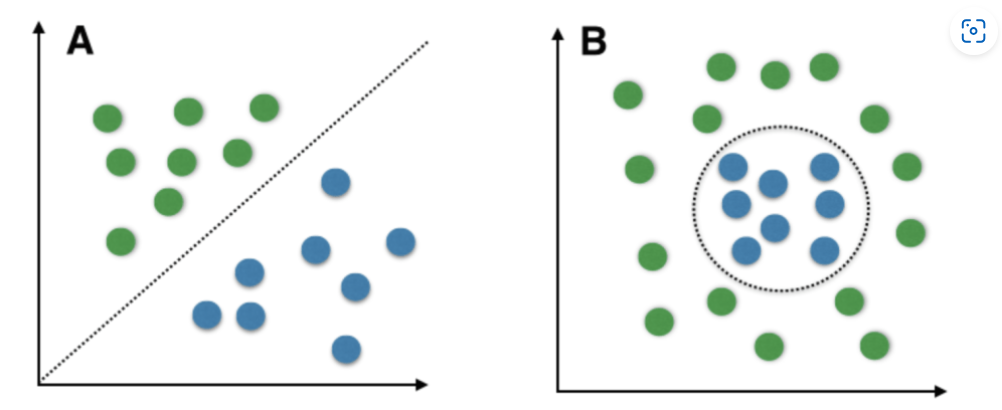
2.4**Model Selection**

* Choose appropriate machine learning algorithms for your task. Common algorithms include:
  + Logistic Regression
  + Support Vector Machines



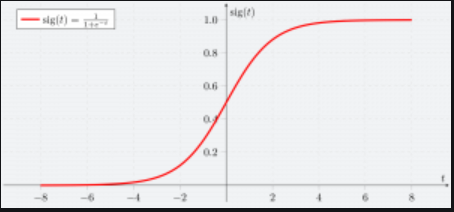
SUPPORT VECTOR MACHINE

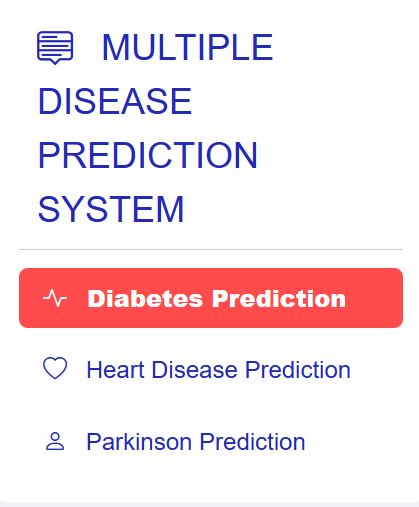
Support Vector Machine (SVM) is a [supervised machine learning](https://www.geeksforgeeks.org/supervised-unsupervised-learning/) algorithm used for both classification and regression. Though we say regression problems as well it’s best suited for classification. The main objective of the SVM algorithm is to find the optimal [hyperplane](https://www.geeksforgeeks.org/separating-hyperplanes-in-svm/) in an N-dimensional space that can separate the data points in different classes in the feature space. The hyperplane tries that the margin between the closest points of different classes should be as maximum as possible. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three.



## **Logistic Regression**

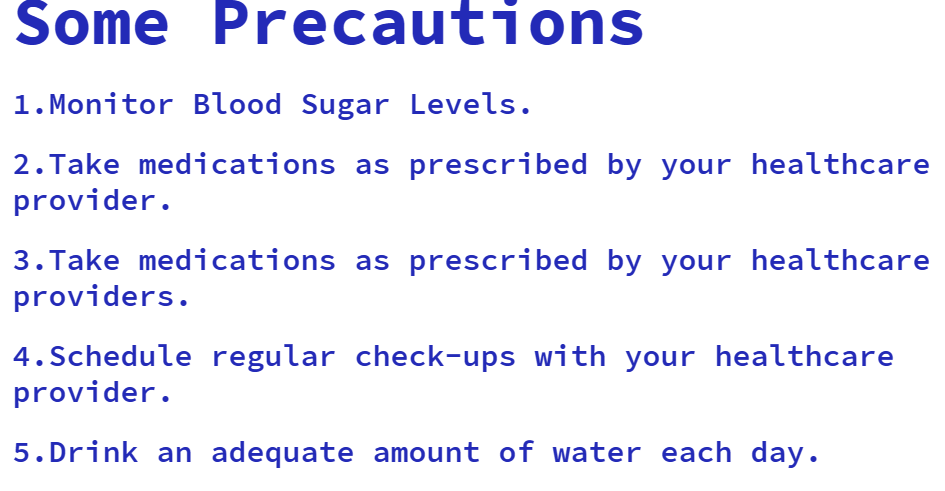
Logistic regression is a [supervised machine learning](https://www.geeksforgeeks.org/supervised-unsupervised-learning/) algorithm mainly used for binary [classification](https://www.geeksforgeeks.org/getting-started-with-classification/) where we use a logistic function, also known as a sigmoid function that takes input as independent variables and produces a probability value between 0 and 1. For example, we have two classes Class 0 and Class 1 if the value of the logistic function for an input is greater than 0.5 (threshold value) then it belongs to Class 1 it belongs to Class 0. It’s referred to as regression because it is the extension of linear regression but is mainly used for classification problems. The difference between [linear regression](https://www.geeksforgeeks.org/ml-linear-regression/) and logistic regression is that linear regression output is the continuous value that can be anything while logistic regression predicts the probability that an instance belongs to a given class or not.









#### **CONCLUSION**

#### Developing a multiple disease prediction model is a complex yet impactful endeavor that brings together the realms of data science and healthcare. As technology advances, these models have the potential to revolutionize healthcare delivery by shifting the focus from reactive to proactive and preventive care.

#### In summary, multiple disease prediction models represent a promising frontier in healthcare. By harnessing the power of data and machine learning, these models contribute to a paradigm shift toward more effective, patient-centered, and sustainable healthcare practices. Continued research, collaboration, and advancements in technology will play pivotal roles in shaping the future of predictive healthcare models

**REFERENCES**

[Diabetes Dataset (kaggle.com)](https://www.kaggle.com/datasets/mathchi/diabetes-data-set)

[Parkinson’s Disease Data Set (kaggle.com)](https://www.kaggle.com/datasets/vikasukani/parkinsons-disease-data-set)

[Heart Disease Dataset (kaggle.com)](https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset)

[Logistic Regression in Machine Learning - GeeksforGeeks](https://www.geeksforgeeks.org/understanding-logistic-regression/)

[Introduction to Support Vector Machines (SVM) - GeeksforGeeks](https://www.geeksforgeeks.org/introduction-to-support-vector-machines-svm/)

[Parkinson's disease - Symptoms and causes - Mayo Clinic](https://www.mayoclinic.org/diseases-conditions/parkinsons-disease/symptoms-causes/syc-20376055)

1. **CONCLUSION**

Twitter sentiment analysis stands as a pivotal means of extracting and comprehending public sentiments within the dynamic landscape of social media. Through the utilization of machine learning algorithms, meticulous data preprocessing, and insightful data visualization techniques, this project aspires to furnish valuable insights into the prevailing sentiment trends. The overarching goal is to offer a comprehensive understanding of user opinions, thereby contributing to informed decision-making processes across various spheres, including business strategies, political landscapes, and societal awareness.

1. <https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-pyt>
2. [https://www.analyticsvidhya.com/blog/2021/06/twitter-sentiment-](https://www.analyticsvidhya.com/blog/2021/06/twitter-sentiment-analysis-a-nlp-use-case-for-beginners/) [analysis-a-nlp-use-case-for-beginners/](https://www.analyticsvidhya.com/blog/2021/06/twitter-sentiment-analysis-a-nlp-use-case-for-beginners/)
3. [https://medium.com/nerd-for-tech/twitter-sentiment-analysis-using-](https://medium.com/nerd-for-tech/twitter-sentiment-analysis-using-logistic-regression-ff9944982c67) [logistic-regression-ff9944982c67](https://medium.com/nerd-for-tech/twitter-sentiment-analysis-using-logistic-regression-ff9944982c67)
4. .