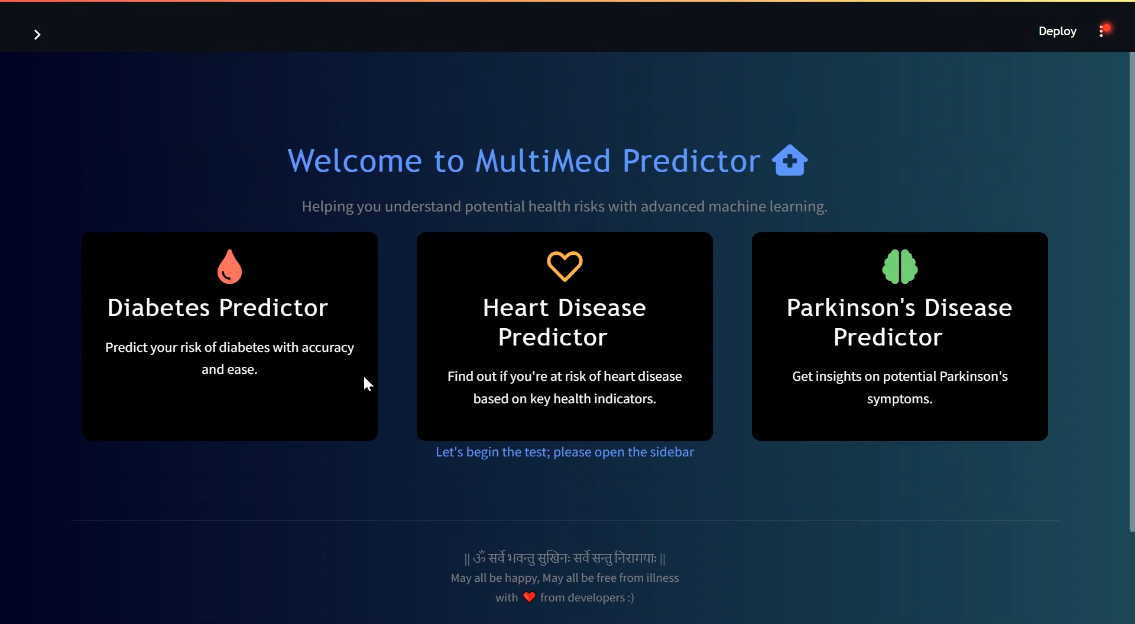
**MultiMed Predictor**

****

*(A multiple disease predictor System)*

**Submitted for**

**Statistical Machine Learning CSET211**

**Submitted by:**

**E23CSEU1865 SOURAV**

**E23CSEU1878 AESHNI**

**E23CSEU1888 VAISHNAVI**

**Submitted to**

**DR. SAKSHI SHARMA**

**July-Dec 2024**

**SCHOOL OF COMPUTER SCIENCE AND ENGINEERING**

A close-up of a logo

Description automatically generated

1. **Abstract**
2. **Introduction**
3. **Related Work**
4. **Methodology**
5. **Hardware/Software Required**
6. **Experimental Results**
7. **Conclusions**
8. **Future Scope**
9. **GitHub Link**

(https://github.com/souravhun27/MutiMed-Predictor)

**Abstract**

The "**Mutimed Predictor**" project aims to predict three diseases— ***Diabetes* *Disease***

***Heart Disease***

***Parkinson's disease***

The project includes ***data preprocessing***, ***visualization***, and ***classification******models***, which are implemented in separate modules. The system integrates ***Streamlit*** *for deployment* and features a user-friendly interface with visually engaging animations. ***The goal is to enable early disease prediction, supporting proactive health management.***

**Introduction**

Health prediction is crucial for early diagnosis and management of chronic conditions.

*The project leverages machine learning models to predict diabetes, heart disease, and Parkinson's, providing users with a quick, reliable method to identify potential risks*.

This work addresses the need for accessible, multi-disease prediction tools that streamline complex predictive analytics into an intuitive application for everyday use.

**Related Work**

Here are a few areas of related research that align with your "Mutimed Predictor" project:

### **(i) Multi-Disease Prediction Models**

Research focuses on developing models capable of predicting multiple diseases with a single framework. Transfer learning is a key method, where a model trained for one disease aids in predicting others, potentially enhancing accuracy and adaptability across different diseases.

### **(ii) Ensemble Learning in Healthcare**

Ensemble methods, such as Random Forest and XGBoost, are used to combine different machine learning algorithms to improve prediction accuracy for diseases like diabetes, cardiovascular issues, and neurological conditions. These models are particularly useful in handling complex healthcare data.  
**paper**: "Ensemble methods in medical diagnostics"( https://arxiv.org/abs/2304.05367)

### **(iii) Feature Engineering for Disease Prediction**

Methods like genetic algorithms and recursive feature elimination are explored to enhance model efficiency, especially in multi-disease scenarios that involve diverse data types (e.g., genetic, biometric, lifestyle).  
 **paper:** "Feature Selection and Feature Engineering for Healthcare" (https://www.sciencedirect.com/science/article/abs/pii/S0010482519302525 )

### **(iv) Explainable AI in Health Predictions**

Explainable AI (XAI) methods, such as SHAP and LIME, are employed to interpret and explain predictions in healthcare models, making them more transparent and understandable for healthcare professionals.  
 **paper**: "Explainable Artificial Intelligence for Healthcare Decision Support" (https://www.researchgate.net/publication/375672036\_Explainable\_AI\_for\_Healthcare\_Decision\_Support\_Systems)

### **(v) Predictive Modeling with Real-Time Data Integration**

Studies are investigating the integration of real-time data from wearable devices (e.g., smartwatches, fitness trackers) for chronic disease prediction, particularly for diabetes and heart diseases.

**paper:** "Real-time Healthcare Prediction with IoT Devices” (https://www.researchgate.net/publication/351413253\_Real\_Time\_Health\_Monitoring\_System\_Using\_IoT).

### **(vi) Early Detection of Parkinson’s Using Voice and Motor Skills Analysis**

Research is being conducted on using voice and motor skill to detect early signs of Parkinson’s disease

**paper:** "Early Detection of Parkinson’s Disease using Speech and Motor Skills" (https://arxiv.org/html/2406.02608v1).

### **(vii) Cross-Domain Transfer Learning**

Cross-domain transfer learning explores applying models trained on one disease to predict other related conditions, especially useful in scenarios where data for rare diseases is limited. This method has potential in expanding disease prediction capabilities without requiring large, disease-specific datasets.  
 **paper**: "Transfer Learning for Disease Prediction” (https://www.researchgate.net/publication/319870314\_Disease\_Prediction\_Based\_on\_Transfer\_Learning\_in\_Individual\_Healthcare)

**Methodology**

**1. Data Collection from Kaggle:**

* Datasets for Diabetes (https://www.kaggle.com/datasets/mathchi/diabetes-data-set)
* Heart Disease (https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset)
* Parkinson’s Disease (https://www.kaggle.com/datasets/vikasukani/parkinsons-disease-data-set)

**2. Data Preprocessing:**

Implemented through a preprocessing script that handles missing values, feature scaling, and encoding.

**3. Data Visualization:**

By sklearn library*- seaborn, matplotlib* are used for plotting different types of graphs for data visualization to understand data.

**4. feature selection:**

Recursive Feature Elimination (RFE) was used for feature selection, ranking features by importance to understand their impact on results. The top features were retained, while the least important ones were set to their central tendency by default, allowing users the option to input custom values for improved accuracy.

**5. Model Selection:**

Classification algorithms like *Logistic Regression*, *SVM*, were evaluated to choose the best-performing model for each disease.

**6. Model Training and Testing:**

Each model was trained on respective disease data, with test sets used for validation. Metrics such as accuracy and F1-score were utilized to assess performance.

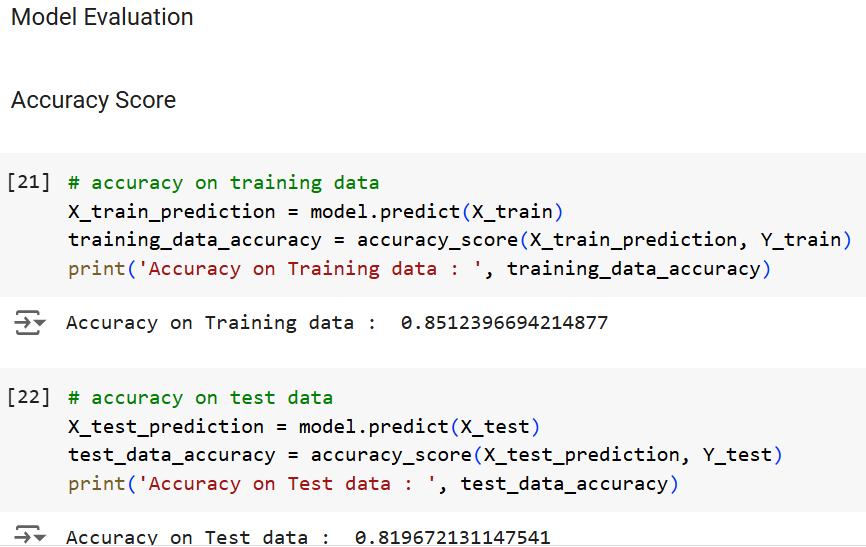
**7. Software Required**

* Python: Core programming language.
* Streamlit: For creating the web interface.
* Pandas: For data manipulation and analysis.
* Scikit-learn: For machine learning algorithms and feature selection.
* Matplotlib: For data visualization.
* Streamlit-Lottie: For adding animations.
* CSS: For interface design.
* Git: For version control.
* Spider IDE and Anaconda Navigator: For development environment setup.
* Google Colab: For collaborative coding and running experiments.

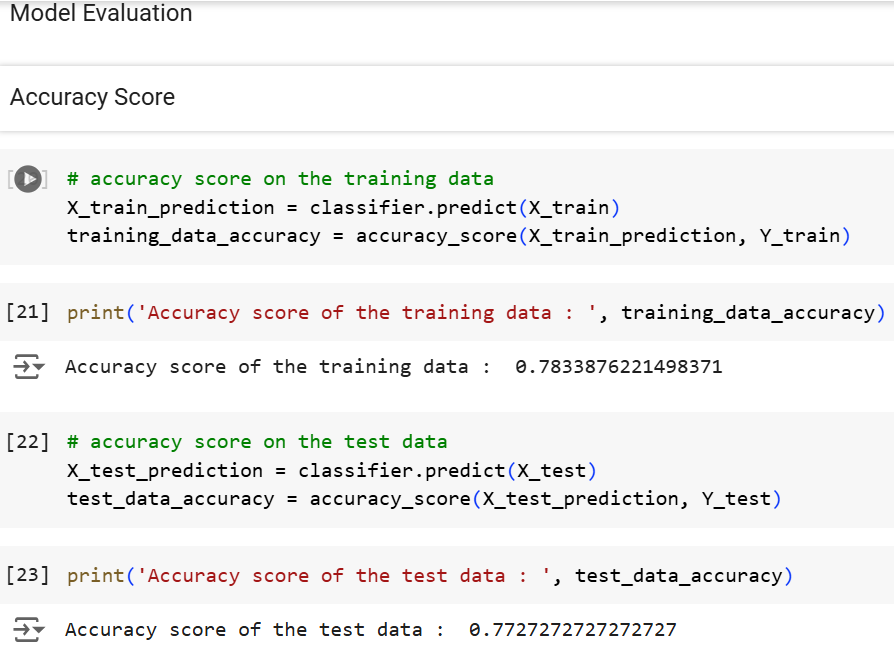
**Experimental Results**

The system achieved notable accuracy in predictions across the three diseases with –

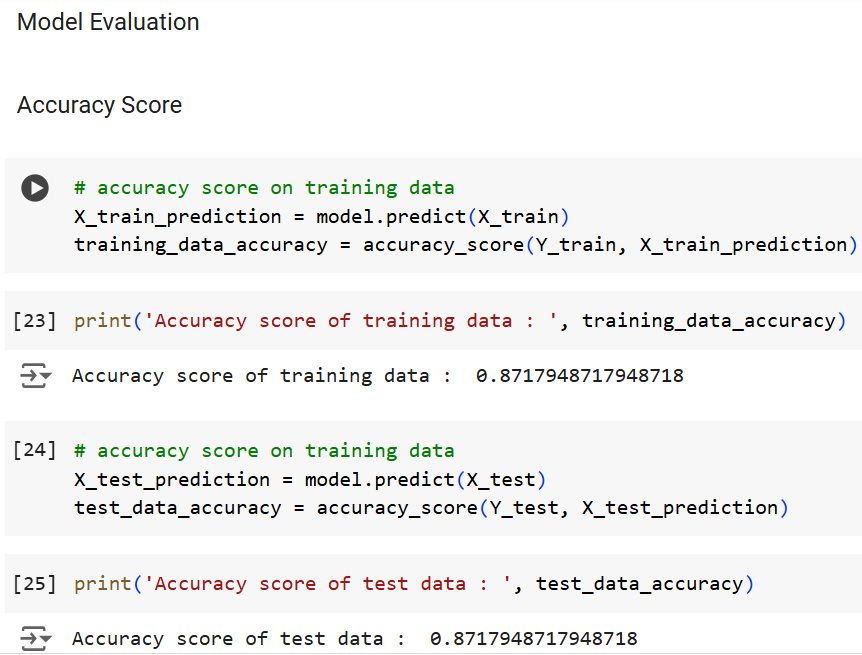
* *Heart Disease predictor trained on SVM:*



* *Logistic Regression provided efficient predictions for Diabetes*



* *SVM provided efficient predictions for Parkinson's*



**Conclusions**

The project successfully integrates multi-disease prediction within a single application. It demonstrates the potential of machine learning in health prediction, enabling users to assess risks for multiple chronic diseases. The tool provides accurate predictions, supporting early detection and proactive health care.

**Future Scope**

Future improvements could include integrating real-time data from wearable devices and enhancing explainability for better user comprehension. Expansion to additional diseases and the inclusion of neural networks could also boost prediction accuracy and broaden the system’s utility.

**Check out my project using GitHub Link**

*[GitHub Repository for Mutimed-Predictor]*

(https://github.com/souravhun27/MutiMed-Predictor)