Project Report

Topic – Disease Prediction using Machine Learning

Team member –Harsh Bhardwaj and Swayam Arora

INTRODUCTION –

The aim of this project was to develop machine learning models for predicting three different diseases: heart disease, Parkinson's disease, and diabetes. Each disease prediction model was implemented separately using Python libraries such as NumPy, Pandas, Scikit-learn, Stream lit, Pickle, and OS. The project aimed to provide a user-friendly interface for individuals to input relevant medical data and receive predictions regarding their likelihood of having any of these diseases.

METHODOLGY –

1 Data Collection:

- Datasets for heart disease, Parkinson's disease, and diabetes were obtained from reliable sources such as medical research repositories or publicly available datasets.
- The datasets were pre-processed to handle missing values, encode categorical variables if necessary, and normalize/standardize numerical features.

2 Model Development:

- For each disease, a separate machine learning model was developed using Scikit-learn.
- Various algorithms such as Logistic Regression, Random Forest, Support Vector Machine, etc., were experimented with to determine the most suitable model for each disease prediction task.
- Model hyperparameters were fine-tuned using techniques such as Grid Search or Random Search to improve model performance.
- Model evaluation metrics such as accuracy, precision, recall, and F1-score were utilized to assess the performance of each model.

3 Deployment:

- The Stream lit library was employed to create an interactive web application for disease prediction.
- The models were serialized using the Pickle library and saved to disk.
- Upon receiving input from the user through the web interface, the relevant model was loaded, and predictions were generated based on the provided medical data.

IMPLEMENTATION –

1 Feature Engineering:

- Features relevant to each disease were identified and included in the respective datasets.
- Feature engineering techniques such as feature scaling, encoding categorical variables, and feature selection were applied to prepare the data for modelling.

2 Model Training:

- The pre-processed datasets were split into training and testing sets.
- Each disease prediction model was trained using the training data.
- Cross-validation techniques were employed to ensure the robustness of the models.

3 Model Evaluation:

- The trained models were evaluated using the testing data.
- Evaluation metrics such as accuracy, precision, recall, and F1-score were computed to measure the performance of each model.
- Confusion matrices and ROC curves were generated to visualize the model's performance.

4 Web Application Development:

- A user-friendly web interface was created using Stream lit.
- The interface allowed users to input their medical data, select the disease they wanted to predict, and receive instant predictions.
- Proper error handling and validation were implemented to ensure the robustness of the application.

RESULT –

- The developed machine learning models demonstrated promising performance in predicting heart disease, Parkinson's disease, and diabetes.
- The accuracy and reliability of the models were assessed through rigorous evaluation techniques.
- The web application provided an intuitive platform for users to access the prediction models and obtain personalized health insights.

CONCLUSION –

In conclusion, the project successfully developed machine learning models for predicting heart disease, Parkinson's disease, and diabetes. The use of Python libraries such as NumPy, Pandas, Scikit-learn, Stream lit, Pickle, and OS facilitated the entire development process. The interactive web application provides a valuable tool for individuals to assess their risk of developing these diseases based on their medical data. Further improvements and enhancements can be made to the models and the application to enhance usability and accuracy.

FUTURE WORK –

- Incorporate additional features and data sources to further improve the accuracy of the prediction models.
- Explore advanced machine learning algorithms and techniques to enhance model performance.
- Conduct real-world validation studies to assess the practical utility of the prediction models.
- Continuously update and maintain the web application to ensure compatibility with the latest technologies and medical guidelines.

REFRENCES –

Dataset - https://www.kaggle.com

Research Paper –
https://www.sciencedirect.com/science/article/pii/S1877050920300557

https://link.springer.com/article/10.1007/s42979-020-00365-y

https://ieeexplore.ieee.org/abstract/document/6921958/

https://www.sciencedirect.com/science/article/pii/S2214785321052202

Libraries –

- Sklearn
- Stream lit
- NumPy

This project report summarizes the development and implementation of machine learning models for disease prediction using Python. It outlines the methodology, implementation details, results, conclusion, and suggestions for future work.