

ABSTRACT

Diabetes Classification Using Machine Learning Techniques

Diabetes mellitus is a chronic metabolic disorder characterized by high blood glucose levels due to the body's inability to produce or effectively use insulin. Accurate classification of diabetes is crucial for effective disease management and treatment. This study explores various machine learning techniques for the classification of diabetes, focusing on predicting whether an individual has diabetes based on specific medical and demographic features. We employed a dataset containing medical records, including attributes such as number of pregnancies, glucose concentration, blood pressure, skin thickness, insulin levels, BMI, diabetes pedigree function, age.

The study compared the performance of several classification algorithms, including Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), and k-Nearest Neighbors (k-NN). Model performance was evaluated using metrics such as accuracy, precision, recall, F1-score, and the area under the Receiver Operating Characteristic (ROC) curve. The results demonstrated that ensemble methods like Random Forest provided higher accuracy and better generalization to unseen data compared to single classifiers. Feature importance analysis revealed that glucose level, BMI, and age were the most significant predictors of diabetes.

In conclusion, the study highlights the potential of machine learning in improving diabetes diagnosis and emphasizes the importance of selecting the appropriate algorithm and features for optimal classification. Future work will focus on enhancing the model's predictive power by incorporating more complex features and exploring deep learning approaches.

Technology Stack: HTML, CSS, JavaScript, Flask, MongoDB Database, Python, Machine Learning Technology.

Project Members:

Takbide Shubhangi_253

Kotlwar Tejas_218

Mustare Mayuri_228

Shaikh Fardin_212