CSE/ECE 343 : Machine Learning Project Proposal
Title : Diabetes Prediction using Machine Learning

Pragathi Vuppu
vuppu22604@iiitd.ac.in

Bhanu Shandilya
bhanu22134@iiitd.ac.in

Kanak Yadav
kanak22611@iiitd.ac.in

Neelu
neelu22318@iiitd.ac.in

1. Motivation

The motivation for this project stems from the increasing prevalence of diabetes and its severe health complications, including heart disease, kidney failure, and limb amputations. With early detection being critical for effective management and prevention, we recognized a pressing need for more accurate and accessible diagnostic tools. This realization led us to explore the potential of machine learning to improve early diabetes detection, building on existing techniques to develop a solution that can make a significant impact in preventive healthcare.

2. Related Work

1.Tasin, I. , Nabil, T.U. , Islam, S. , Khan, R. : Diabetes prediction using machine learning and explainable AI techniques. Healthc. Technol. Lett. 10, 1–10 (2023). 10.1049/htl2.12039.They were able to achieve 88.8% accuracy using Ensemble(XGBoost).

2. Aishwarya Mujumdar, Dr. Vaidehi V et al. (2019) Diabetes prediction using machine learning. They were able to achieve 96% accuracy using Logistic Regression. Further, this work can be extended to find how likely non-diabetic people are to have diabetes in the next few years.

3.Abdulhadi N, Al-Mousa A (2021) Diabetes detection using machine learning classification methods. In: 2021 International Conference on Information Technology (ICIT). IEEE, p 350–354. They were able to achieve 82% accuracy with the Random Forest.

3. Timeline

A Tentative 12 week timeline:

Week 1-2	Exploring the domain
Week 3	Data Collection
Week 4	Pre-processing and Data Visualization
Week 5	Logistic Regression, K-Nearest Neighbors
Week 6	Decision Trees, Random Forest
Week 7	Analysis & performance of models
Week 8-9	Hyperparameter Tuning, Check for Overfitting, Underfitting
Week 10	Model Deployment
Week 11	Report Writing.
Week 12	Buffer

4. Individual Tasks

The individual task have been distributed in following manner -

Data Collection	**Bhanu,Kanak,Neelu,Pragathi**
Data Visualisation , Pre Processing	**Kanak,Neelu**
Logistic Regression	**Pragathi**
Decision tress & Random forest	**Kanak,Neelu**
Analysis and Performance of model	**Bhanu,Kanak,Neelu,Pragathi**
K-Nearest Neighbors	**Bhanu**
Hyperparameter Tuning	**Kanak,Neelu,**
Check for Overfitting, underfitting	**Bhanu,Pragath**
Model Deployement	**Bhanu,Kanak,Neelu,Pragathi**
Report Writing	**Bhanu,Kanak,Neelu,Pragathi**

5. Final Outcome

As a group, we expect to gain a strong foundational understanding of machine learning by creating a basic model for diabetes prediction. Our primary goal is to apply what we've learned in a practical setting, enhancing our skills in data analysis, model development, and teamwork. Although the model may not be ready for industry use, we aim to contribute by demonstrating how ML can be used to address real-world healthcare challenges, specifically in early diabetes detection. Through this project, we hope to set the stage for more complex and impactful work in the future.