Topic: Investigating the Generalization Performance of Machine Learning Models for Diabetes Prediction across Diverse Population Demographics

[1]Author: MD IRSHAD ANWAR, [1*]Second Author: MD TAHSEEN EQUBAL
*MD IRSHAD ANWAR, [1] 23f1000996@ds.study.iitm.ac.in , IIT MADRAS , INDIA
[2] JAMIA HAMDARD UNIVERSITY, INDIA

A. Abstract:

Diabetes prediction remains a critical area of research in public health, aiming to identify individuals at risk and implement preventive measures. In this study, we delve into the assessment of the generalization performance of machine learning models for diabetes prediction across different population demographics. Specifically, we employ Logistic Regression, Random Forest, Decision Tree, and K-Nearest Neighbors (KNN) algorithms to develop prediction models leveraging a comprehensive dataset encompassing pregnancies, glucose levels, blood pressure, skin thickness, insulin levels, BMI, diabetes pedigree function, and age. Furthermore, we incorporate hyperparameter tuning techniques to optimize model parameters and enhance predictive accuracy.

Our research focuses on evaluating the robustness and efficacy of these models across diverse population groups, considering factors such as age, gender, ethnicity, and socioeconomic status. By systematically analyzing model performance metrics, including accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC), we aim to identify potential biases and disparities in diabetes prediction across different demographic cohorts.

Through rigorous experimentation and cross-validation, we aim to provide insights into the effectiveness of machine learning algorithms in accurately predicting diabetes onset in varied population groups. Additionally, we intend to highlight the importance of model generalization and the need for tailored approaches to address demographic heterogeneity in diabetes prediction.

By contributing to a deeper understanding of the generalization capabilities of machine learning models in diabetes prediction, our research endeavors to inform the development of more equitable and effective preventive healthcare strategies tailored to diverse population demographics.

B. Justification with Atmanirbhar Bharat

Our research employs machine learning for diabetes prediction, aligning with Atmanirbhar Bharat's vision by fostering innovation in healthcare technology and promoting self-reliance