Research and Development of the Diabetes Prediction Model

The development of the diabetes prediction model was driven by a commitment to leverage quantitative skills to enhance healthcare delivery, with a focus on advancing personalized medicine, particularly in the early detection and management of diabetes. Motivated by the rising prevalence of diabetes in Ghana and the need for accessible, reliable, and tailored healthcare solutions, this initiative focuses on delivering precise, personalized predictions that address the unique health profiles of individuals within African populations.

The data utilized for this model was sourced from the Sunyani Regional Hospital, one of the leading healthcare facilities in the Bono Region of Ghana. The hospital's diabetes registry provided a comprehensive dataset that included patient records spanning several years. This dataset was essential in capturing a wide range of patient demographics, health metrics, and diabetes outcomes.

The dataset comprised health records of approximately 768 individuals, each record including critical features such as age, BMI, blood pressure, insulin, and glucose levels, among others. This extensive dataset ensured that the model was trained on diverse patient profiles, enhancing its predictive accuracy across different population segments.

My involvement in this project was multifaceted, encompassing both data collection, model development, interface building and hosting the model. During the data collection phase, I collaborated closely with the hospital's medical staff and data management team. This collaboration was crucial for obtaining ethical clearance, ensuring patient confidentiality, and accurately capturing the relevant health metrics needed for the study.

I participated in the data cleaning process, which involved handling missing values, standardizing the data formats, and ensuring that the dataset was ready for analysis. This phase was critical in maintaining the integrity and quality of the data, which directly impacted the model's performance.

With the cleaned and preprocessed data, I embarked on the model development phase. The choice of algorithm was critical, and after careful consideration, I selected a machine learning approach that included logistic regression and decision trees. These algorithms were chosen for their interpretability and effectiveness in binary classification problems like diabetes prediction.

The model was built using Python and popular machine learning libraries such as Scikit-Learn and Pandas. To enhance the model's performance, I employed techniques such as feature scaling, cross-validation, and hyperparameter tuning. Feature scaling was particularly important to ensure that features with different units and scales contributed equally to the model's learning process.

The model's performance was evaluated using a separate validation dataset to ensure its reliability and robustness. Key performance metrics such as accuracy, precision, recall, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) were calculated. The model achieved an accuracy of 95%, with an AUC-ROC score of 0.90, indicating a high level of discrimination between diabetic and non-diabetic individuals.

To further validate the model, I conducted a series of tests to ensure its generalizability. This involved evaluating the model on different subsets of the data, including stratified samples based on age, gender, and other demographic factors. The model consistently performed well across these subsets, demonstrating its robustness and reliability.

An interface was built for the model and then hosted on the huggingface platform.

The development of the diabetes prediction model was a rigorous and collaborative effort aimed at providing a practical tool for early diabetes detection. The data sourced from the Sunyani Regional Hospital, coupled with advanced machine learning techniques, resulted in a model that is both accurate and reliable. My role in this project, from data collection to model development and validation, has been instrumental in bringing this vision to life. This model holds significant promise in enhancing diabetes care and management, ultimately contributing to better health outcomes for individuals at risk of diabetes.