Team 2

DATS 6103: Summary Report

Professor Dr. Ning Rui

Dec 11, 2024

**Diabetes Risk Prediction**

**Introduction**

Diabetes represents one of the most pressing global health challenges of the 21st century, characterized by its chronic nature and profound impact on individual and population health. The World Health Organization estimates that diabetes affects approximately 537 million adults worldwide, with projections suggesting this number could rise to 783 million by 2045. Beyond its immediate health implications, diabetes contributes significantly to cardiovascular diseases, kidney dysfunction, neurological complications, and reduced quality of life.

In an era of advancing medical technology and data-driven healthcare, predictive analytics emerges as a transformative approach to understanding and mitigating complex health risks. This study leverages a comprehensive dataset of 100,000 patient records to explore the intricate relationships between multiple health indicators and diabetes likelihood, representing a critical step toward personalized preventive medicine.

Smart Questions:

* What are the primary risk factors (e.g., blood glucose level, BMI, age, hypertension) for diabetes in this population, and how accurately can they predict the likelihood of a diabetes diagnosis?
* Are there any notable differences in diabetes prevalence based on gender, age, or smoking history? For example, does smoking history combined with a high blood glucose level increase the risk?
* How does the distribution of BMI and HbA1c levels differ between those with and without diabetes?
* How sensitive are the models to changes in certain variables (e.g., slight increases in blood glucose or BMI)? Can we identify an actionable threshold for intervention?

By answering these smart questions, we hope to contribute to the early detection and prevention of diabetes, ultimately improving the quality of life for at-risk individuals.

**Literature Review**

Diabetes is a chronic metabolic disorder characterized by elevated blood glucose levels. It is a significant global health concern, affecting millions of people worldwide. Early detection and timely intervention are crucial for managing diabetes and preventing complications.

In recent years, machine learning techniques have emerged as powerful tools for predicting diseases, including diabetes. Several studies have explored the use of machine learning algorithms to identify individuals at risk of developing diabetes.

**Traditional Risk Factors:**

* **Age:** Older individuals are at a higher risk of developing type 2 diabetes.
* **Family history:** A family history of diabetes is a significant risk factor.
* **Obesity:** High BMI and body fat percentage are associated with increased risk.
* **Physical inactivity:** Lack of physical activity is a major contributing factor.

**Emerging Risk Factors and Machine Learning:**

* **Metabolic Syndrome:** A cluster of conditions that increase the risk of heart disease, stroke, and type 2 diabetes. Machine learning models can identify individuals with metabolic syndrome based on various biomarkers.
* **Genetic Factors:** Specific genetic variants have been linked to increased diabetes risk. Machine learning can analyze genetic data to identify individuals with a higher predisposition.
* **Lifestyle Factors:** Factors like diet, sleep patterns, and stress levels can influence diabetes risk. Machine learning models can incorporate these factors to improve prediction accuracy.
* **Description Of Data**
* **Exploratory Data Analysis**
* **Smart Questions**
  + **Smart Question 1**
  + **Smart Question 2**
  + **Smart Question 3**
  + **Smart Question 4**
* **Limitations:**
* **Conclusion:**