**Project Proposal**

**Team Members:** Group 3: Alex Hunt, Oluchi Ejehu, Zainab Iyiola

**Project Title:** Supervised Classification Algorithms for Early Detection of Diabetes

**Description of the problem context:** Diabetes is a chronic condition that affects over 37 million people in the United States, with an estimated 8.5 million cases going undiagnosed, according to the Centers for Disease Control and Prevention (CDC). Early detection is critical for effective management and prevention of severe complications, such as heart disease, kidney failure, and nerve damage. This project seeks to develop a supervised classification machine learning model to accurately detect diabetes based on key patient health data, including glucose levels, BMI, age, and family history. By analyzing these features, the model will aim to identify patterns and risk factors associated with diabetes, providing a reliable tool for healthcare professionals to diagnose the disease early. The goal is to improve early detection rates and facilitate timely interventions, ultimately reducing the burden of diabetes-related complications in the U.S. healthcare system.

**Type of the problem:** Supervised Learning - The problem at hand is a supervised classification problem. The dataset includes labeled examples where each patient's health data is associated with a binary outcome (0 for non-diabetic and 1 for diabetic). This makes it suitable for classification algorithms that can be learned from historical data to predict whether new patients are likely to develop diabetes.

**Initial thoughts on techniques that might be used:** We plan to employ various supervised classification algorithms to predict the likelihood of diabetes and compare these algorithms. Potential models include Logistic Regression, Random Forest, Support Vector Machines (SVM), and Gradient Boosting. Hyperparameter tuning and cross-validation techniques will be used to optimize model performance. We will also explore feature selection methods to identify the most important predictors of diabetes.

**Hyperlink of the problem description:** <https://www.kaggle.com/datasets/akshaydattatraykhare/diabetes-dataset>