# Diabetes Prediction

**Team Members:**             Vedang Patel - 2022565             Abhirup Das - 2022019

**Problem:-** According to researched data, undiagnosed diabetes is increasing rapidly. Early detection of diabetes for managing and preventing any further health complications is important. While the traditional methods work, they require costly tests and time from the patient's end. There is also a good chance that people with diabetes don't know they have diabetes and will never take these tests. The project proposes a machine learning approach to predict diabetes using readily available data, potentially enabling earlier diagnosis of diabetes. The prediction can be integrated into smartwatches that monitor the target user's health data.

**Method:-** The project will involve using various machine learning algorithms to predict the presence of diabetes. The process will ideally involve data acquisition, data processing, model training and model testing. We will be preprocessing the data to convert some of the data which is in words into numbers as machine learning algorithms cannot work directly with words. Some of the algorithms that could be used to train the algorithm are logistic regression, naive bayes, support vector machines and neural networks. Depending on the scope of the course and future analysis of the project we may increase or decrease the set of tools that might use for the project.

**Dataset:-** We will be using a diabetes prediction dataset on Kaggle - [Diabetes prediction dataset](). The following dataset provides 100 thousand data entries. Each entry has the information: gender, age, hypertension, heart_disease, smoking_history, bmi, HbA1c_level, blood_glucose_level, diabetes. While not all of this data is readily available in most cases, the data provided is relevant to predicting diabetes.

**Results (anticipated):-** Confusion matrices of prediction accuracy to evaluate how accurately the algorithm is able to predict diabetes. Performance metrics across different models in order to see whether it is reasonable to process data across users. Identification of factors that lead to diabetes is also something we expect to figure out.

**Literature reviewed:-**
1) [Prediction of Diabetes Empowered With Fused Machine Learning | IEEE Journals & Magazine]()
2) [IDF diabetes Atlas: Global estimates of undiagnosed diabetes in adults for 2021 - ScienceDirect]()
3) [Trends in Prevalence of Diabetes and Control of Risk Factors in Diabetes Among US Adults, 1999-2018 | Cardiology | JAMA]()