**Draft Proposal- Diabetics Prediction**

**Team Members:**

**Lakshmi Kavya Ravi**

**Omkar Katkade**

**1.What is your issue of interest (provide sufficient background information)?**

The motivation behind this project was to gain hands-on experience in developing a complete machine learning project, as well as to explore deployment platforms such as Heroku. Additionally, the growing health issue of diabetes, caused by sedentary lifestyles, inspired the project. Early detection of diabetes through proper medical treatment can prevent severe health complications. By utilizing technology, such as machine learning, a predictive model can be built to determine if a patient has diabetes. This project not only provides a learning opportunity but also has the potential to benefit society.

**2. Why is this issue important to you and/or to others?**

Diabetes is a serious health issue that affects millions of people worldwide. It is a chronic disease characterized by high levels of sugar (glucose) in the blood, and it can lead to serious health complications if not properly managed. The following are some reasons why diabetes is an important issue:

**Prevalence**: Diabetes is a growing health concern, with the World Health Organization estimating that there are over 400 million people with diabetes worldwide.

**Health complications**: Diabetes can lead to a range of health complications, including heart disease, stroke, kidney failure, blindness, and amputations.

**Economic impact**: Diabetes is a significant burden on healthcare systems and the economy, with costs related to the disease estimated to reach over $1 trillion globally by 2040.

**Disproportionate impact on vulnerable populations**: Diabetes disproportionately affects vulnerable populations, such as the elderly and low-income communities, making it an important social issue as well as a medical one.

**Lack of effective treatments**: Despite advances in diabetes management and treatment, there is still a need for more effective strategies to manage the disease and prevent its complications.

**3. Where do you get the data to analyze and help answer your questions (credibility of source, quality of data, size of data, attributes of data. etc.)**

**Source :** National Institute of Diabetes and Digestive and Kidney Diseases (NIH).

**Attributes of Data:** Pregnancy, BMI, Skin thickness, Glucose, Blood pressure, Insulin, Age, Pedigree Function, Outcome.

**Quality of Data:** We didn't find any null values in the dataset.

**4. What will be your unit of analysis (for example, patient, organization, or country)?**

**Roughly how many units (observations) do you expect to analyze?**

We are gonna finally analyze whether a person has diabetes or not.

**5. What kinds of techniques/models do you plan to use (for example, clustering,**

We want to use logistic regression, decision tree classifier and Random forest classifier. And find which model provides the best accuracy.

**6.What outcomes do you have (better understanding of problems, tools to help solve problems, predictive analytics with practical applications, etc)?**

We plan to deploy our web application which users can use to enter a few details and predict if they have diabetes or not.

**7. Process:**

**Data collection:** Collecting data from various sources like medical records, surveys, and public health databases.

**Data cleaning:** Cleaning and preprocessing the collected data to remove any missing or incorrect values.

**Exploratory data analysis (EDA):** Conducting exploratory data analysis to understand the distribution of the variables and identify any patterns or trends in the data.

**Feature engineering:** Creating new features from the existing data to increase the predictive power of the model.

**Model selection:** Selecting the appropriate machine learning model for the project, such as linear regression, logistic regression, decision trees, etc.

**Model training:** Training the selected model using the cleaned and preprocessed data.

**Model evaluation:** Evaluating the performance of the trained model using appropriate evaluation metrics, such as accuracy, precision, recall, etc.

**Model tuning:** Tuning the model parameters to improve its performance.

**Results interpretation:** Interpreting the results of the model to draw meaningful insights and conclusions about the diabetes data.

**Visualization:** Creating interactive visualizations to present the results in an intuitive and accessible way.

**8. Tasks Distribution:**

**Kavya Ravi :**Data collection**,** Exploratory data analysis, Model selection, Model Evaluation, Results Interpretation, Visualization**.**

**Omkar Katkade:**Data Cleaning, Feature Engineering, VisualizationModel Training, Model Tuning,Web Application Development.