Project Proposal

- **Topic:**

  Creating a Diabetes prediction application using k-Nearest Neighbors and Decision Trees as classification models.

- **Data Set:**

  https://www.kaggle.com/uciml/pima-indians-diabetes-database

- **Why this topic:**

  Diabetes is one of the most prevalent disease among people. It occurs when the pancreas reduce the production of insulin which is a major enzyme required to break down glucose in our body. This results in increased glucose levels that have damaging effects resulting in health problems.
  I wanted to create an application that would help users get an estimate prediction whether they could have diabetes or not. This would greatly help people in making some lifestyle changes to combat the disease.
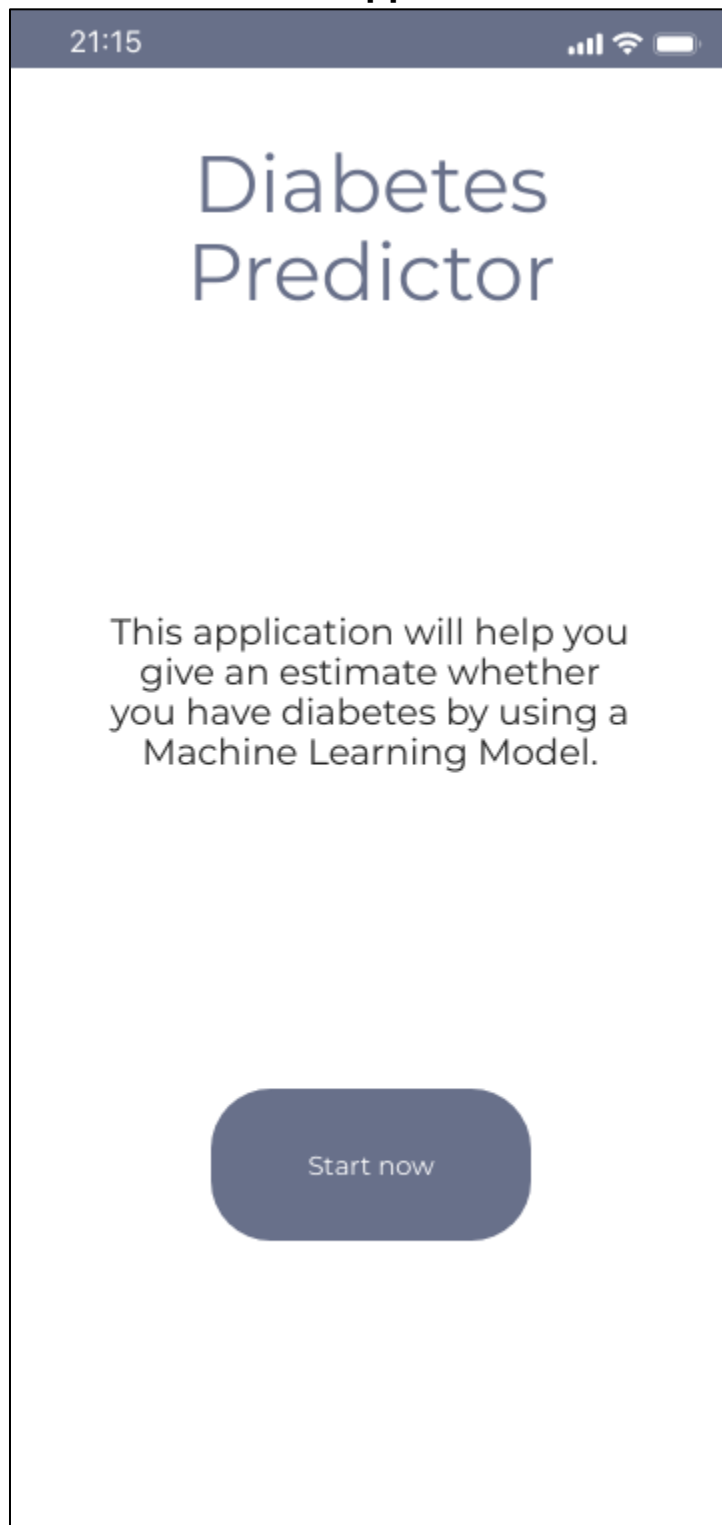
- **Features of the application:**
  - It will use the medical history of the person such as :
    - Age
    - Blood Glucose Level
    - Insulin Level
    - Body Mass Index (Weight to Height ratio)
    - Diabetic Pedigree Function (History of diabetes in the family)
    - Skin Thickness
  - The application will then use a series of classification models to predict whether the person has diabetes or not. It will be using k-Nearest Neighbors model and Decision Tree to classify the features

- **Features of the Classification Models:**
  - It will use two sampling techniques called Principal Component Analysis which is used to reduce the dimensionality of the data set thus reduce the calculations required and SMOTE(Synthetic Minority Oversampling Technique) which is artificially adding samples to the data set to reduce imbalance between classes.
  - The models will incorporate Grid Search 5-fold Cross Validation Technique to estimate and tune the hyper-parameters such as k value(Number of Neighbors) and weight(weights of the features) in k-Nearest Neighbors and Entropy and GINI index values to decide on the purity of the nodes of the Decision Tree.

- **Sketch of UI of the application with features:**

21:15

Diabetes Predictor

This application will help you give an estimate whether you have diabetes by using a Machine Learning Model.

Start now

Enter Details

Age

Blood Pressure (mm/Hg)

Gender ⌄

Insulin Level (mu U/ml)

Body Mass Index (kg/m2)

Skin Thickness (mm)

Predict

72%

There is a 72%
chance that you
have diabetes

- **Similar applications on the market:**
  1. **Cardiogram:** Application that can detect beginnings of diabetes by analyzing heart rate data. Link: https://cardiogram.com/payers/
  2. **Diabits App:** Predictive monitoring of glycemia levels to check for diabetes. Link: https://diabetes.jmir.org/2020/3/e18660/
  3. **Diabetes Prediction**: Takes medical history input to predict diabetes.Link:https://play.google.com/store/apps/details?id=com.ihsinc.ipredictdiabetes&hl=en&gl=US