Practicum Sprint 2

Weicong Duan, Sena King, Wan Lam, Yara Seif
wduan35@gatech.edu, sking300@gatech.edu, wlam42@gatech.edu,
yaraseif@gatech.edu

1 PROJECT RESEARCH

1.1 Background

According to the Center for Disease Control and Prevention (2022), Diabetes is a chronic health condition that affects how someone's body manages sugar in the bloodstream. In order for the human body to benefit from sugar, the pancreas normally releases insulin into the bloodstream to help turn sugar into energy for the cells to use. For a person with diabetes, either the body doesn't produce enough insulin or doesn't use insulin as well as it should. Statistically speaking, more than 37 million adults in the United States are diabetic and diabetes is the 7th leading cause of death in the U. S. ,(Disease Control and Prevention, 2022). More disturbingly, the rate of having diabetes in adults in the US has doubled in the last 20 years.

There are three types of diabetes: type 1, type 2, and gestational diabetes. People with type 1 diabetes often have diabetes since childhood/adolescence and this type is thought to be an autoimmune reaction where the body stops producing insulin. However, in type 2 diabetes, the body does produce insulin, but it does not effectively use it well enough to manage healthy sugar levels in the bloodstream. Gestational diabetes happens only to women who are pregnant, and it could go away after giving birth. While gestational diabetes may go away after birth, women who had gestational diabetes are more likely to develop type 2 diabetes later on.

While type 1 diabetes cannot be prevented, type 2 and gestational diabetes can be prevented given risk factors for each type (Center for Disease Control and Prevention, 2022). Therefore, it is important to understand whether a person is at risk of developing diabetes so he/she can follow a healthier, more careful lifestyle.

1.2 Justification

Considering that diabetes is a one of the leading causes of death in the United States, machine learning models must be used to predict that chances of someone acquiring diabetes in order to prevent it as much as possible. According to a

literature survey by Saxena et al., "Analysis of diabetes data disease is quite challenging because most of the data in the medical field are nonlinear, nonnormal, correlation structured, and complex in nature" (2022). There are multiple methods used by researchers to predict diabetes, such as "naïve Bayes, support vector machine, decision trees, random forest, k-nearest neighbour, multilayer perceptron, and logistic regression" (Saxena et al., 2022). Implementing a modular web service that interacts with FHIR-compliant resources could help in discovering which of these methods is most accurate at predicting diabetes. Patient health data can be retrieved from public FHIR servers, and FHIR resources provide a standardized data model to utilize machine learning models.

1.3 Solution

In order to predict diabetic risk, we first compare and use the ML models trained on the data in the Pima indians. By building a web service/user interface for machine learning model application, we can predict the diabetic risk by making API calls. Our web application which contains patient data will call the web service to find out the diabetic risk for a particular patient.

2 TECHNICAL DESIGN

2.1 Tools/Technology

- Python
- Python fhir.resources library
- Heroku
- ReactJS
- Typescript
- SMART on FHIR
- MongoDB
- Flask
- SMART Health IT FHIR Sandbox
- Three clarification models - KNN, Random Forest, and Support Vector Classifier -to be compared in predicting the onset of diabetes

2.2 Data/Data Sources

- FHIR Public Database
- Pre-trained diabetes dataset from Kaggle: https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

2.3 Architecture Diagram

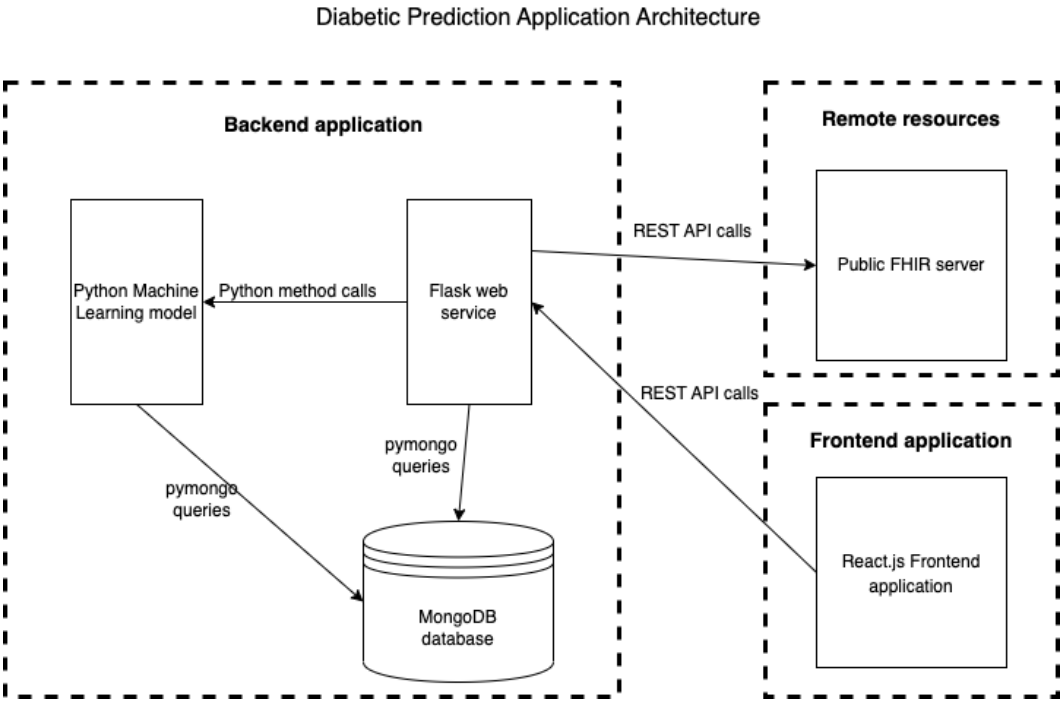

Figure 1—Architecture Diagram

2.4 Screen Mockup

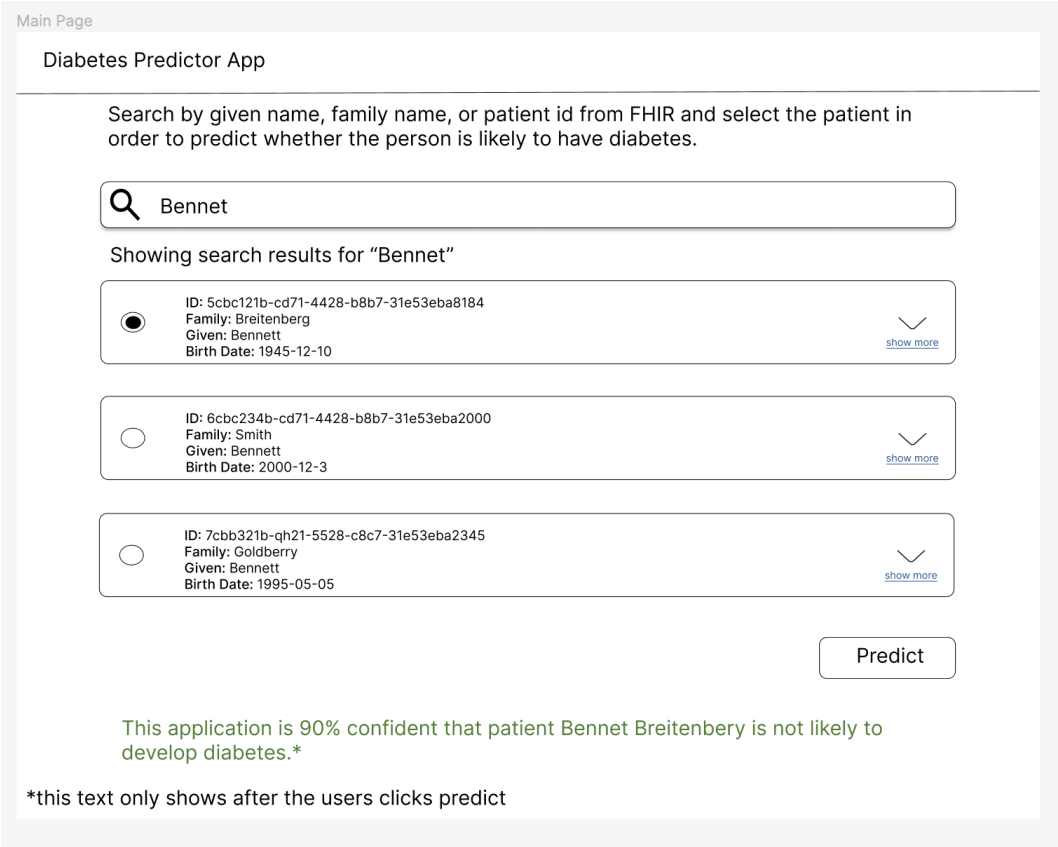

Figure 2—Wireframe of the UI's main page

Link to the figma: https://www.figma.com/file/U8KJK7KD1j24ny3COQnH74/Diabetes-Prediction?node-id=0%3A1t=n6B8KJRruFVXCdTZ-0

3 IMPLEMENTATION PLAN

3.1 Project Tasks

Project tasks can be found on this Idea Board: https://ideaboardz.com/for/Diabetics%20Prediction/4851360

3.2 Project Timeline

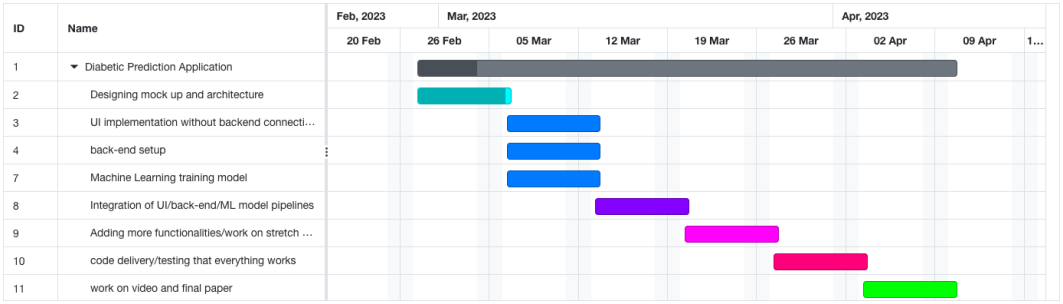

Figure 3—Predicted Project Timeline

3.3 Needs/Risks

- The training data for the model is diabetes data that includes parameters like numbers of pregnancies which are not always available in the patient data. For some of the information like insulin (2 hour serum insulin), we do not know if a 2 hour range is a general practice for all data.
- The information about pregnancy in the training data seems to indicate that it is about gestational diabetes of women. But since we have not located a special dataset concerning this group, we might have to work with the general patient data that is available for now.
- The other risk is about the fitting of the models. Three models have different accuracies, and one might be slightly better than the other ones for the training data. It might be different for different training data.
- Our team does not have experience with the Flask web server and there may be some difficulties and extra time needed in order to integrate Flask with ReactJS.

4 REFERENCES

Centers for Disease Control and Prevention. (2022, April 5). Diabetes risk factors. Centers for Disease Control and Prevention. Retrieved March 4, 2023, from https://www.cdc.gov/diabetes/basics/risk-factors.html

Centers for Disease Control and Prevention. (2022, July 7). What is diabetes? Centers for Disease Control and Prevention. Retrieved March 4, 2023, from https://www.cdc.gov/diabetes/basics/diabetes.html

Saxena, R., Sharma, S. K., Gupta, M., Sampada, G. C. (2022, April 12). A comprehensive review of various diabetic prediction models: A literature survey. Journal of healthcare engineering. Retrieved March 4, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9018179/