Project Submission Document

Project Title: AI Powered Diabetes Prediction System

Diabetes is a chronic disease that affects millions of people worldwide.
It is characterized by high blood sugar levels, which can lead to serious health complications such as heart disease, stroke, and blindness. Early detection and management of diabetes are essential for preventing these complications.

In this project, we propose to develop an AI powered diabetes prediction system using innovative techniques such as ensemble methods and deep learning architectures. Our goal is to improve the accuracy and robustness of the prediction system, so that it can be used to reliably identify individuals at risk of developing diabetes.

Background

Diabetes is a complex disease with multiple risk factors. Some of the most common risk factors include:

Age
Family history of diabetes
Obesity
Physical inactivity
Gestational diabetes
Polycystic ovary syndrome (PCOS)
Traditional methods of diabetes prediction rely on a combination of these risk factors. However, these methods are not always accurate, as they do not take into account all of the factors that can contribute to the development of diabetes.

About Dataset

Context

This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective is to predict based on diagnostic measurements whether a patient has diabetes.

Content

Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage.

- Pregnancies: Number of times pregnant
- Glucose: Plasma glucose concentration a 2 hours in an oral glucose tolerance test
- BloodPressure: Diastolic blood pressure (mm Hg)
- SkinThickness: Triceps skin fold thickness (mm)
- Insulin: 2-Hour serum insulin (mu U/ml)
- BMI: Body mass index (weight in kg/(height in m)^2)
- DiabetesPedigreeFunction: Diabetes pedigree function
- Age: Age (years)
- Outcome: Class variable (0 or 1)

Sources:

(a) Original owners: National Institute of Diabetes and Digestive and
Kidney Diseases
(b) Donor of database: Vincent Sigillito (vgs@aplcen.apl.jhu.edu)
Research Center, RMI Group Leader
Applied Physics Laboratory
The Johns Hopkins University
Johns Hopkins Road
Laurel, MD 20707
(301) 953-6231
(c) Date received: 9 May 1990

Past Usage:

```
1. Smith,~J.~W., Everhart,~J.~E., Dickson,~W.~C., Knowler,~W.~C., \&
```

```
   Johannes,~R.~S. (1988). Using the ADAP learning algorithm to
forecast
   the onset of diabetes mellitus.  In {\it Proceedings of the
Symposium
   on Computer Applications and Medical Care} (pp. 261--265).  IEEE
   Computer Society Press.

   The diagnostic, binary-valued variable investigated is whether the
   patient shows signs of diabetes according to World Health
Organization
   criteria (i.e., if the 2 hour post-load plasma glucose was at least
   200 mg/dl at any survey  examination or if found during routine
medical
   care).   The population lives near Phoenix, Arizona, USA.

   Results: Their ADAP algorithm makes a real-valued prediction between
   0 and 1.  This was transformed into a binary decision using a cutoff
of
   0.448.  Using 576 training instances, the sensitivity and
specificity
   of their algorithm was 76% on the remaining 192 instances.
```

Relevant Information:

```
 Several constraints were placed on the selection of these instances
from
  a larger database.  In particular, all patients here are females at
  least 21 years old of Pima Indian heritage.  ADAP is an adaptive
learning
  routine that generates and executes digital analogs of
perceptron-like
  devices.  It is a unique algorithm; see the paper for details.
```

Number of Instances: 768

Number of Attributes: 8 plus class

For Each Attribute: (all numeric-valued)

1. Number of times pregnant
2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test
3. Diastolic blood pressure (mm Hg)
4. Triceps skin fold thickness (mm)
5. 2-Hour serum insulin (mu U/ml)
6. Body mass index (weight in kg/(height in m)^2)
7. Diabetes pedigree function
8. Age (years)
9. Class variable (0 or 1)

Missing Attribute Values: Yes

Class Distribution: (class value 1 is interpreted as "tested positive for

diabetes")

Proposed Approach

Our proposed AI powered diabetes prediction system will use a variety of innovative techniques to improve the accuracy and robustness of the prediction. These techniques include:

Ensemble methods:

Ensemble methods combine the predictions of multiple machine learning models to produce a more accurate prediction.
Deep learning architectures: Deep learning architectures are a type of machine learning that can learn complex patterns from data.
We plan to train our system on a large dataset of electronic health records (EHRs) from individuals with and without diabetes. The EHRs will contain a variety of data, including demographic information, medical history, lab results, and prescription medications.

Once the system is trained, we will evaluate its performance on a held-out test set. We will also compare the performance of our system to other state-of-the-art diabetes prediction systems.

Expected Outcomes

We expect that our proposed AI powered diabetes prediction system will be more accurate and robust than traditional methods of diabetes prediction. We believe that our system can be used to reliably identify individuals at risk of developing diabetes, so that they can take steps to prevent or manage the disease.

Innovative Techniques

The following are some of the innovative techniques that we plan to use in our proposed diabetes prediction system:

Ensemble methods: We plan to use ensemble methods to combine the predictions of multiple machine learning models, such as logistic regression, support vector machines, and random forests. This will help to improve the accuracy and robustness of the prediction system.
Deep learning architectures: We plan to use deep learning architectures, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to learn complex patterns from the EHR data. This will help the system to identify subtle patterns in the data that may be associated with an increased risk of developing diabetes.
Conclusion

We believe that our proposed AI powered diabetes prediction system has the potential to make a significant impact on the prevention and management of diabetes. We are committed to developing a system that is accurate, robust, and accessible to everyone.