SPRINGBOARD

CAPSTONE 2 PROJECT PROPOSAL

EARLY STAGE DIABETES RISK PREDICTION

Student: Igor Verevkin

Mentor: A.J. Sanchez

10/17/2020

1. Problem Statement:

- Is it possible to predict early stage diabetes among the patients based on their symptoms?
- The goal of this project is to build predictive models that will estimate the probability that a patient is diabetic, as a function of available features.

2. Context:

- Diabetes is a group of metabolic disorders characterized by a high blood sugar level over a prolonged period of time. If left untreated, diabetes can cause many complications, such as diabetic ketoacidosis, cardiovascular disease, stroke, chronic kidney disease, damage to the nerves, cognitive impairment or even death. About 422 million people worldwide have diabetes, the majority living in low-and middle-income countries, and 1.6 million deaths are directly attributed to diabetes each year. Both the number of cases and the prevalence of diabetes have been steadily increasing over the past few decades.[1]

3. Criteria For Success:

Several models will be built and compared against appropriate performance metrics. Measurable

[1] https://en.wikipedia.org/wiki/Diabetes

connections between features and target will be explored with respect to multiple models.

4. Scope Of Solution Space:

- The models will be limited to the data that is available.

5. Constraints within solution space:

- The patients in the dataset are all from one town in Bangladesh. It is unknown if there is any influence of their local nutrition habits on higher/lower chance of having diabetes.
- The dataset is unbalanced since there are significantly more negative target values than positive ones.

6. Stakeholders and Key Data Sources:

- A dataset[2] collected using direct questioning of patients of Sylhet Diabetes Hospital in Sylhet, Bangladesh by

  i. M Faniqul Islam

  ii. Rahatara Ferdousi

  iii. Sadikur Rahman

  iv. Yasmin Bushra

  From Metropolitan University Sylhet, Bangladesh

7. Deliverables:

- Jupyter notebooks
- Presentation
- Project report

[2] https://archive.ics.uci.edu/ml/datasets/Early+stage+diabetes+risk+prediction+dataset.