# Prediction of Diabetes

Diabetes, a chronic metabolic disorder, is one of the fastest growing diseases regardless of geographic, racial, or ethnic context. There are two types of disbetes: Type I and Type II.

- Type I diabetes happens when the pancreatic beta cells produce either very little insulin or no insulin.

- Type II dibetes happens when body become insulin resistance.

The common symptoms for Type I diabetes are sudden weight loss, Dehydration, delayed healing, blurred vision and Type II have similar symptoms or no symptoms at all depends on body reaction. [1, 2]. Type I diabetes strikes both kids and adults at any age.There is no cure for Diabetes but it can be managed if it diagnose in early stage. Type I diabetes can be managed by medication and Type II can be managed by balanced diet, excercise and a healthy lifestyle. In this project, we will be dealing exclusively with data about Type II diabetic patients.

>>>>>>> 051c4b03614a5c3d5bc0ade979055c8392453882

## Why we have chosen "Prediction of Diabetes" as a topic

8.7% of males and 7.6% of females in Canada, equaling roughly three million individuals, had Type II diabetes in the year 2014 [4]. Nationally, in the year 2014, the incidence of new diabetes cases in the population was 5.9 out of every 1000 people [4]. Total health care costs resulting from diabetes in Canada is expected to increase to over $16.9 billion (CAD) anually by 2020. We would like to create a machine learning model that can correlationally predict whether or not an individual is at risk for Type II Diabetes based on a range of health features, in order to aid clinicians in preventing and treating the illness.

![Diabetes Canada](https://github.com/asadca4u/Final\_Project\_Group\_Five/blob/Presentation/Diabetes%20Incidence%20Canada.png)

## Description of Data

The dataset contains a record of 16 attributes about each of 520 patients, which may be correlated with their diabetic status, or "class". The data was collected from the patients in the Sylhet Diabetes Hospital of Sylhet, Bangladesh.

![Table 1](https://user-images.githubusercontent.com/67460581/98498587-6551b800-2204-11eb-9d2d-b425945e8966.PNG)

![Table 2](https://user-images.githubusercontent.com/67460581/98498616-73073d80-2204-11eb-9160-cbed55bb1a2e.PNG)

## Questions to Consider

To what extent can a machine-learning model predict whether or not a patient is at risk of diabetes based on 16 different health and demographic attributes? Is such a model useful for classifying patients in a clinical setting so that preventative and if nescessary, therapeutic measures may be implemented?

## exploring the dataset

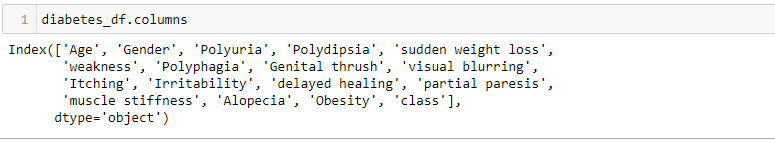
definition: data exploration is the initial steps in data analysis in which, used data visulization and techniques to describe data characteristics, such as, size, data type, rows and columns, in order to understand the nature of data

data exploration tools:

we used manual exploration to explore diabetes dataset. Jupyter notebook was used to write the script with help of Pandas and Python.

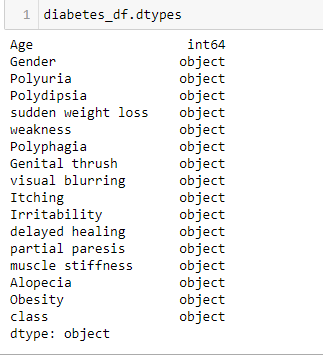
what data is available?

Diabetes\_df.columns to explore the columns in data set



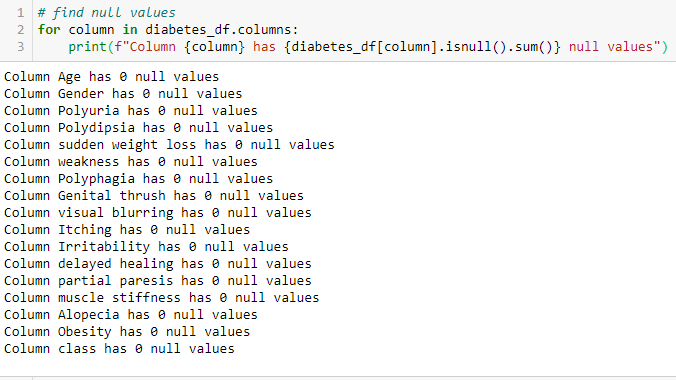
what type of data is available?

Diabetes\_df.dtypes to display the type of data for each column



what data is missing?

pandas "isnull" method to check if any data is missing. Used print statement to get sum of null values for each column and “diabetes\_df.dropna()” to delete all null values if any existed.



what data can be removed?

Panadas ".dropna()" method can be used to remove null data between rows and columns.



Duplicate data in rows and columns can be removed with Pandas ".drop\_duplicates" method.

Data Processing:

The diabetes dataset was obtained as a result of question and answers from patients which are classified with “age” and “gender”. Our dataset consists of “yes” and “no” answers which need to transform into binary data for our supervised machine learning model.

We transformed data with “LabelEncoder” from “sklearn.preprocessing” library and commit to a for loop to transform all the entries in data set one by one.

A picture containing text

Description automatically generated

Data type was checked “df2.dtypes” after converting data.

Text

Description automatically generated

## References:

1. https://www.diabetes.ca/en-CA/about-diabetes

2. https://www.albertadiabetesfoundation.com/type-2-diabetes?gclid=CjwKCAiAqJn9BRB0EiwAJ1SztWQsJPkJ0cSmIA5\_mo6FY-YJaOfQ3yL8lVYJY2vqKnNk5ba-Uz8mpRoC2YkQAvD\_BwE

3. https://www.jdrf.ca/who-we-are/type-1-diabetes/

<<<<<<< HEAD

4. Computer Vision and Machine Intelligence in Medical Image Analysis

=======

4. https://www.canada.ca/en/public-health/services/publications/diseases-conditions/diabetes-canada-highlights-chronic-disease-surveillance-system.html

>>>>>>> 051c4b03614a5c3d5bc0ade979055c8392453882