Final Project Proposal

I. Overview

Although patients suffering from diabetes approximates to millions worldwide, people are still not aware of the early-symptoms of diabetes. Early detection of diabetes can lead to controlling the seriousness, save huge amount of hospital costs and in some cases prevent the disease from developing faster. In this analysis, a regression model is built to predict the odds of diabetes from its early symptoms.

II. Research Questions

Main questions of interest are: (1) what are the critical early-symptoms of developing diabetes, (2) is itching a significant predictor, (3) are there any interesting interaction terms with itching.

Using this dataset, Islam, MM Faniqul, et al., has published a research paper in 2020. The link to the research paper is stated below:

Islam, MM Faniqul, et al. 'Likelihood prediction of diabetes at early stage using data mining techniques.' Computer Vision and Machine Intelligence in Medical Image Analysis. Springer, Singapore, 2020. 113-125. https://link.springer.com/chapter/10.1007/978-981-13-8798-2_12

III. Data

The data was collected by asking direct questions to the patients of Sylhet Diabetes Hospital in Sylhet, Bangladesh, and was approved by a doctor. The data is multivariate, having 17 attributes and 520 instances in total. There are no missing values.

Codebook:

	Attributes	Value		Attributes	Value
1	Age	20 – 65 years	**10**	Itching	Yes:1, No:0
2	Sex	Male:1, Female:0	**11**	Irritability	Yes:1, No:0
3	Polyuria	Yes:1, No:0	**12**	Delayed healing	Yes:1, No:0
4	Polydipsia	Yes:1, No:0	**13**	Partial paresis	Yes:1, No:0
5	Sudden weight loss	Yes:1, No:0	**14**	Muscle stiffness	Yes:1, No:0
6	Weakness	Yes:1, No:0	**15**	Alopecia	Yes:1, No:0
7	Polyphagia	Yes:1, No:0	**16**	Obesity	Yes:1, No:0
8	Genital thrush	Yes:1, No:0	**17**	Classification	Positive:1, Negative:0
9	Visual blurring	Yes:1, No:0			

Link to the dataset is below:
https://archive.ics.uci.edu/ml/datasets/Early+stage+diabetes+risk+prediction+dataset.

IV. Project Plan

The plan for this project is outlined as below:

(1) Oct 19th – 22th: Finish analysis
(2) Oct 26th – 30th: Finalize analysis and write reports
(3) Oct 30th – Nov.: Finalize presentation