# Leading University
## Department of Computer Science & Engineering
## CSE - 4800


# Proposal for Thesis

# Research on Diabetes Prediction using Machine Learning Algorithms


| | |
|---|---|
| Ahsanur Rahman | ID: 1912020074 |
| Symum Hasan | ID: 1912020080 |
| Ismat Jerin Arpa | ID: 1912020038 |


**Submitted To**

Shafkat Kibria
Assistant Professor & Head (Acting)
Department of Computer Science and Engineering

2nd April, 2022

# Abstract

Diabetes mellitus is a common human disease characterized by a series of metabolic abnormalities in which blood sugar levels are abnormally high for an extended length of time. It has a variety of effects. It damages a huge number of people in the circulatory system of the organism, particularly the blood veins and nerves. Early diagnosis and treatment of such diseases can be helpful. We will conduct research by collecting datasets and implementing various machine learning algorithms on them for a potential early prediction. It will help us investigate the numerous risk factors associated with this condition through the use of machine learning techniques to develop an efficient result for extracting knowledge models for predicting outcomes from local diagnostic medical datasets available in our region. Diabetic patients are the source of this information.

# Methodology

In order to achieve the goal, our study methodology might comprise a few stages, which are likely accumulation of diabetes dataset with the relevant attributes of the patients, preprocessing the numeric value attributes, to apply various machine learning classification techniques and corresponding predictive analysis utilizing such data. First, we have to identify the problem. Then, we need to collect the necessary data from Diagnostic data storages. Then, we need to train and test the datasets using machine learning and applying it eventually to get the results. We have studied various research papers related to this topic and also wrote three Literature Reviews. And from that experience, in our research, we are primarily planning to use various well-known machine learning algorithms, such as Support Vector Machine (SVM), Naive Bayes (NB), K-Nearest Neighbor (KNN) and C4.5 decision tree, on adult population data to predict diabetic mellitus.

# Motivation

Diabetes is a major chronic syndrome caused by a series of metabolic abnormalities in which blood glucose levels are abnormally high for an indeterminate amount of time. It influences various organs in the human body, resulting in a variety of complex diseases such as stroke, renal disease, pulmonary embolism, eyesight, and so on. Diabetes Disorders (DD) are presently one of the healthcare top causes of mortality. Predictive analytics in the healthcare system is a huge obstacle, but if accurate early prediction is achieved, the potential risk and degree of diabetes may be significantly decreased. We are using machine learning methods and our local datasets to predict early diabetes. We are eager to learn about machine learning and solve a problem using this technology. Also this thesis will help us learn about research work and how to approach it.

# Conclusion

In our research, we have to analyze the early prediction of diabetes by taking into account various risk factors related to this disease using machine learning techniques. Extracting knowledge from real health care dataset can be useful to predict diabetic patients. We believe that our experimental study could be helpful to make a control plan for diabetes mellitus in future.