

Justin Ku

Capstone Two - Project Proposal

**Problem Statement:** *(What is the business problem you are investigating)*

- Can a machine learning model be created, using a Diabetes Health Indicators dataset, to predict if an individual has Diabetes or is at high risk of developing Diabetes?

**Context:** *(Why are you working on this problem)*

- Diabetes is a serious chronic disease that plagues the United States impacting millions of lives each year. It's a disease which renders individuals' ability to effectively regulate glucose levels in their blood a futile effort.
- Food is broken down through digestion, and sugars are released into the bloodstream. The pancreas then secretes a hormone called insulin which acts as a key to open cell doors and allow glucose to move from the bloodstream into cells throughout the body.
- Diabetes can be characterized by either the body being unable to use the insulin it produces effectively or the body not making enough insulin. Complications such as heart damage which lead to heart attacks/strokes, permanent vision loss, kidney failure, and lower extremity amputations occur when high levels of sugar remain in the bloodstream of those with diabetes.
- Early identification of diabetes or risk of diabetes can prevent individuals from a life-long struggle of health complications.

**Criteria for Success:** *(What is the key criteria that will deem this work successful)*

- The model is able to predict if an individual already has Diabetes or is at high risk of Diabetes with high accuracy.

**Scope of Solution Space:** *(What is the focus of this business initiative? I.e. What are your specific items you will focus on exclusively?)*

- The specific items that will be focused on are the health indicators provided in the Diabetes Health Indicators dataset.

**Constraints within Solution Space:** *(What constraints exist that may prevent this business initiative from succeeding?)*

- The last time this dataset was updated was 3 years ago (2021). There may be new developments in Diabetes research that this dataset does not account for. There are physical symptoms of Diabetes such as 'Extreme Thirst', 'Slow-healing Wounds', and 'Blurred vision' that are not outwardly stated in the dataset.

**Stakeholders to Provide Key Insight:** *(Who are the key stakeholders that need to be involved in this project? Where will you source your data from and who will you present your recommendation to once you have identified a solution?)*

- Medical Professionals - Doctors
- Patients
- Kaggle - Provided Diabetes Health Indicators Dataset

**Key Data Sources:** *(What are the key pieces of data you need to answer the questions related to the problem you are trying to solve?)*

- 'Diabetes Health Indicators Dataset' - Sourced from Kaggle