

# Diabetes Prediction Project

Capstone

Justin Mpanga



# Purpose

- Purpose: To develop classification models that can predict whether a person is likely to have diabetes or not based on certain risk factors



# Data and Quality

- ✦ 2021 Behavioral Risk Factor Surveillance System (BRFSS) data.
- ✦ BRFSS data is available on the Centers for Disease Control and Prevention (CDC) website
- ✦ Contains information on a large sample of adults through telephone interviews.
- ✦ Information such as health behaviors, chronic conditions, preventative services, demographics and access to healthcare.



# Variables

- ✦ Using existing literature on diabetes and its risks
- ✦ Feature: age, sex, race, education, income, smoking status, alcohol consumption, body mass index (BMI), general health, physical health, mental health, hypertension, and cholesterol levels
- ✦ Outcome: diabetes diagnosis



# The Impact of Diabetes

- ✦ According to the CDC, over 34 million Americans have diabetes.
  - ✦ Associated complication are heart diseases, and kidney failure
- ✦ And high prevalence puts a significant burden on the healthcare system and the economy.
  - ✦ The American Diabetes Association estimates more than \$327 billion dollars/ year.



# Importance of Diabetes Prediction

- ✦ The prediction model(s) will play a crucial role in:
  - ✦ identifying individuals at risk,
  - ✦ promoting preventive measures,
  - ✦ optimizing resource allocation,
  - ✦ supporting personalized medicine,
  - ✦ and contributing to research and public health initiatives.
- ✦ BRFSS data will give the model(s) the potential to make a significant impact on diabetes prevention and management strategies.



# Jupyter Notebooks

- Proceed...