

# FINAL PROJECT:

## Predicting the Risk of Diabetes

### PROJECT BRAINSTORM

Group3: BRAINTTRAIN; Mia Metni, Betty Paniagua, Timothy O'Donoghue, Seifeddine Mawardi

**Project title:** Behavioral & Health Risk Factor Prediction Model for Diabetes

**Track:** Health

**Sphere of Interest:** Healthcare: Data-Driven Machine-Learning Methods for Risk Prediction

**TOPIC:** Data-Driven Machine-Learning Methods for Predicting Risk of Diabetes in youth and adults.

- BACKGROUND:
  - Diabetes mellitus is a common metabolic disease characterized by high blood glucose levels. In diabetes, the body inefficiently produces little or no insulin. Increased blood sugar (hyperglycemia) and impaired glucose metabolism occur either as a result of decreased insulin secretion or due to decreased sensitivity of the body cells to the action of this hormone (insulin). Depending on the insulin disorder, diabetes can lead to general morbidity and, in some cases, mortality.
- PROBLEM SIGNIFICANCE:
  - Nearly  $\frac{1}{4}$  of all individuals having diabetes in 2021 were undiagnosed. Undiagnosed and untreated individuals having diabetes can lead to increased risk of getting serious problems with your eyes, feet, heart and nerves.
  - Total: 38.4 million people of all ages had diabetes (11.6% of the population) in 2021. 38.1 million were adults ages 18 years or older.
  - Diagnosed: 29.7 million people of all ages had been diagnosed with diabetes (8.9% of the population).
  - *Undiagnosed:* 8.7 million adults ages 18 years or older had diabetes but were undiagnosed (22.8% of adults with diabetes were undiagnosed).
  - (redundant but...) Early diagnosis is one of the most effective ways to minimize the burden caused by diabetes.
- NEEDS STATEMENT:
  - Because the onset of Type 2 is not catastrophic, the disease is more manageable/mitigated earlier on (during the first two phases – insulin resistance and prediabetes). Moreover, in accordance with the World Health Organization (WHO), apart from prevention, **early diagnosis is one of the most effective ways to minimize the burden caused by diabetes.**
  - As such, there is a critical need for accessible tools that have the potential to increase rates of early diagnosis.

- SOLUTION:
  - A behavioral & health risk factor prediction model for diabetes, preemptively imparting health literacy to better equip individuals at risk of/predisposed to diabetes.
  - Model will train an algorithm to recognize disease symptoms and predict if a patient is at risk of/predisposed to Type 2 Diabetes. Model serves as an early self-assessment, intended to encourage proactive care. Model does NOT supersede medical consultation BUT, does serve as a preemptive step/preventative care with low barrier to entry (i.e. at no/low cost and improved accessibility).
- DATA:
  - Our prediction dataset will be a collection of medical and demographic data from patients, along with their diabetes status (0,1). This data will include features such as age, gender, body mass index (BMI), hypertension, heart disease, smoking history, HbA1c level, and blood glucose level.
  - DATA SOURCES BEING CONSIDERED:
    - Diabetes Prediction Dataset:  
<https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset>
    - Diabetes Health Indicator Dataset:  
[https://www.kaggle.com/datasets/julnazz/diabetes-health-indicators-dataset?select=diabetes\\_012\\_health\\_indicators\\_BRFSS2021.csv](https://www.kaggle.com/datasets/julnazz/diabetes-health-indicators-dataset?select=diabetes_012_health_indicators_BRFSS2021.csv)
- USER/POPULATION/CONSUMER SERVED:
  - Patients who believe they might be predisposed to diabetes can use the model to assess their symptoms and risk levels early on, without medical consultation.