

Problem statement formation:

What are the information, symptoms, or combinations of symptoms needed by patients and doctors to determine and diagnose early signs of diabetes?

Context:

In 2021, it was reported that “537 million adults (20-79 year)” or “1 in 10” adults across the globe live with diabetes each year, while it’s further reported that diabetes accounts for “6.7 million deaths,” and costs in diabetes expenditure result in “USD 966 billion dollars” (diabetesatlas.org). In addition, “3 in 4 adults with diabetes live in low and middle income countries” (diabetesatlas.org). Diabetes is a disease that not only afflicts significant financial costs to both patients and healthcare facilities, but also can be prevented. With roughly 75% of the world’s diabetes population living in low and middle income countries, both detection and prevention of diabetes could improve the quality of life for those impacted by the disease.

Criteria for success:

- An accurate indicator to identify both early signs and likelihood of developing diabetes.

Scope of solution space:

- Identify predominant symptoms and/or potential patient demographic information to most accurately predict both the likelihood and early signs of contracting diabetes.
- Develop a model that most accurately informs both patients and healthcare professionals on which symptoms will show the strongest indication of developing early stages of diabetes.
- The model will be developed using following machine learning tasks: Binary Classification, Regression, Clustering, and Feature Selection.
 - The following machine learning tasks could save time, resources, and provide better diagnoses of diabetes for both patients and healthcare providers.

Constraints:

- The patient sample size may not represent the actual population size of the world’s diabetes population.
- The data will not contain patient genetic diabetes information.

Stakeholders:

- Patients
- Healthcare professionals

Data sources:

- UC Irvine Machine Learning Depository:
<https://archive.ics.uci.edu/dataset/529/early+stage+diabetes+risk+prediction+datas+et>