Final Project Proposal

Course: Machine Learning & Knowledge Representation and Reasoning

(Collaborative Project)

Academic Term: 2025-2026

Group Members:

1. Alberto, Marc Justin G. (BSCS 3A)
2. Aquino, Jose V. III (BSCS 3A)
3. Garin, Jeremy M. (BSCS 3A)

I. Project Title

"Predictive System for Diabetes Risk Using Clinical, Laboratory, and Lifestyle Indicators"

II. Selected SDG Goal

This project intends to produce a resolution plan through software engineering in subject the following Sustainable Development Goals from United Nations (2025):

SDG 3: Good Health and Well-Being

TThis study contributes to SDG 3 by enabling early detection of diabetes risk through data-driven health analytics. Early identification of individuals at risk helps prevent complications, reduce hospital admissions, and lower long-term healthcare costs. By supporting timely screening and intervention, the system reinforces global targets on reducing non-communicable diseases and improving access to quality healthcare.

SDG 11: Sustainable Cities and Communities

This table of predictor and outcome variables also supports SDG 11 (Sustainable Cities and Communities) by linking diabetes risk to urban factors such as food accessibility, walkability, physical activity opportunities, pollution exposure, and local healthcare availability. Integrating demographic, lifestyle, clinical, and environmental data allows city planners and public health officials to identify vulnerable populations, improve resource allocation, and design community-based preventive programs. This approach helps build healthier, more resilient, and sustainable urban environments by reducing the burden of diabetes.

III. Problem Statement

Diabetes is a rising global health concern and a major contributor to illness, disability, and premature death. According to the World Health Organization (2024), despite the availability of routine clinical assessments, many individuals remain undiagnosed until the disease has already progressed, resulting in complications that could have been prevented with early detection. This gap highlights the need for more accessible and data-driven methods to identify at-risk individuals before symptoms become severe.

In urban settings, the risk of developing diabetes is further influenced by environmental and lifestyle factors such as limited access to healthy food, low walkability, pollution exposure, and uneven distribution of healthcare services (Banday et.al., 2020). Traditional screening approaches often overlook these contextual determinants, preventing a holistic understanding of diabetes risk across different communities (Dulyapach et.al, 2022). As a result, vulnerable populations

may remain underserved, and disease burden continues to rise in densely populated areas.

This study addresses these challenges by developing a diabetes risk prediction model that integrates demographic, lifestyle, clinical, and environmental data. The goal is to enable earlier identification of individuals at risk, support preventive healthcare interventions, and guide urban planning decisions that promote healthier living conditions. By aligning with SDG 3 (Good Health and Well-Being) and SDG 11 (Sustainable Cities and Communities), the project aims to reduce diabetes-related health disparities and contribute to more resilient and health-supportive urban communities.

IV. Machine Learning Component

This study aims to develop a machine learning model for predicting diabetes risk using structured, numeric clinical and demographic data. The model will leverage patient characteristics such as fitness activities, laboratory results, and comorbidities as indicators to predict the likelihood of diabetes or its severity (Centers for Disease Control and Prevention, 2024).

Type	Category	Indicator
Input/Independent Variable	Comorbidities	● Heart Attack/Disease ● Stroke
	Laboratory Results	● High Blood Pressure ● High Cholesterol
	Lifestyle	● Physical Activity ● Smoker ● Cholesterol Checkup

Output/Dependent Variable	Result	• Diabetes Diagnosis • Severity Classification • Intervention recommendation

Table 1. Predictor and Outcome Variables for Diabetes Diagnosis and Severity Prediction

Moreover, this research aims to test the ML integrated program by producing data on the following metrics to evaluate optimality, usability, and efficiency of the developed model.

Category	Metrics
Model	• Random Forest
Statistical Baseline	• Logistic Regression
Evaluation Metrics	• Accuracy • Precision • Recall • F1 Score

Table 2. Modeling Approach for Program Development

The dataset used to develop and train the machine learning model in this study is sourced from Teboul, A. (2021), which provides comprehensive health indicators related to diabetes.

V. KRR (Knowledge Representation & Reasoning) Component

The Diabetes risk assessment system leverages several Knowledge Representation and Reasoning (KRR) concepts to encode expert knowledge, standardize clinical information, and ensure interpretable and clinically safe predictions. The key KRR concepts applied include:

1. *Forward-chain Reasoning*. Diagnosis and risk probability interpretation will be analyzed through Forward-chain reasoning rule primarily in the assessment of the given metrics provided on section IV. This induces a conclusion from the patient data in comparison to the trained dataset.

2. *Ontology-based Representation.* Clinical concepts can be standardized using SNOMED CT or ICD-10 codes, enabling semantic interoperability across datasets and healthcare systems.

This hybrid approach enhances model interpretability and clinical safety, allowing clinicians to understand and trust automated pneumonia risk assessments.

VI. Expected Output/System Overview

The study aims to produce the following outputs as a sustainable contribution for the development of the community:

1. Literature review (local & foreign)
2. Cleaned dataset and ML notebook
3. Web application with ML + KRR
4. Reflection and SDG alignment report

Conceptual Framework

Analyzing and engineering the proposed project intends to follow the conceptual framework below for architecture guidelines and development structure:

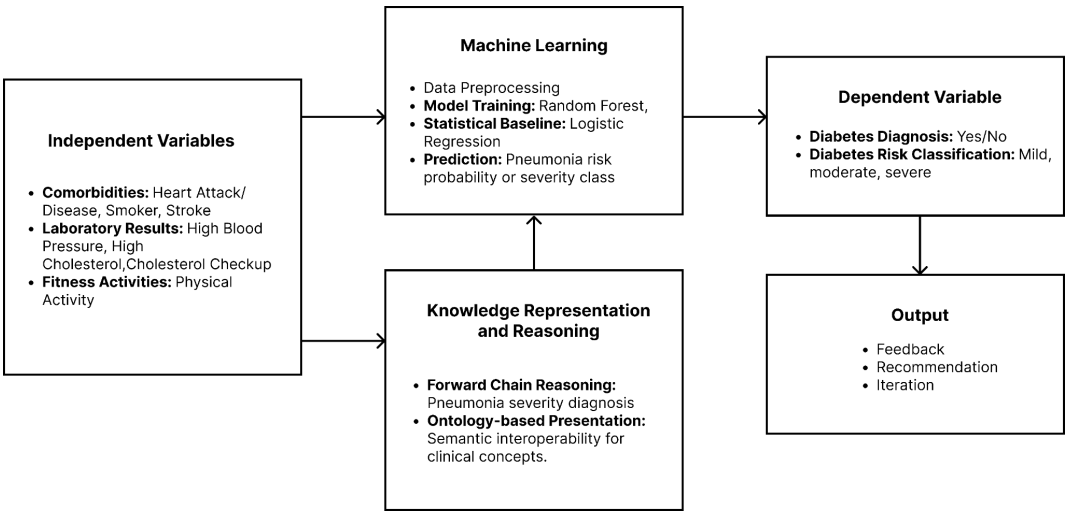

Figure 1. Conceptual Framework

VII. Project Timeline

The project proposal and study intends to follow the timeline below:

Week Number	Objective
Week 1	Project Concept Proposal
Week 2	Initial Dataset Collection
Week 3	Dataset Consultation
Week 4	KRR and ML Model Development
Week 5	Start of Prototype Development
Week 6	Document Consultation
Week 7	Prototype Consultation
Week 8	Refining of Prototype
Week 9	Project Presentation

Table 3. Project Development Timeline

Bibliography

Banday, M. Z., Sameer, A. S., & Nissar, S. (2020). Pathophysiology of diabetes: An overview. Avicenna Journal of Medicine, 10(4), 174-188. https://doi.org/10.4103/ajm.ajm_53_20

Centers for Disease Control and Prevention. (2024, May 15). Symptoms of diabetes. https://www.cdc.gov/diabetes/signs-symptoms/index.html

Dulyapach, K., Ngamchaliew, P., Vichitkunakorn, P., Sornsenee, P., & Choomalee, K. (2022, November 28). Prevalence and associated factors of delayed diagnosis of type 2 diabetes mellitus in a tertiary hospital: A retrospective cohort study. International Journal of Public Health, 67, 1605039. https://doi.org/10.3389/ijph.2022.1605039

Teboul, A. (2023). Diabetes Health Indicators Dataset [Dataset]. UCI Machine Learning Repository. DOI: 10.24432/C53919

World Health Organization. (2024, November 14). Diabetes [Fact sheet]. https://www.who.int/news-room/fact-sheets/detail/diabetes