FINAL PROJECT PROPOSAL

Course: Machine Learning C Knowledge Representation and Reasoning (Collaborative Project) **Academic Term: _____Group Members (Names, Year/Program):**

1. Kristel N. Ramos (3rd Year, Computer Science)

2. Johann Francois P. Tanyag (3rd Year, Computer Science)

3. Althea Maxene B. Viar (3rd Year, Computer Science)

1.Project Title

Write a clear and concise title:

"VitaSenseAI: A Machine Learning Model for Lifestyle and Health Risk Prediction"

2.Selected SDG Goal

Indicate the specific UN Sustainable Development Goal (SDG) your project addresses:
> SDG No. _____3_____| SDG Title: Good Health and Well-Being

Rationale (Why this SDG?): > The research tackles healthcare as the model aims to predict the health and lifestyle risk of an individual. This research aims to integrate and implement Machine Learning to encourage individuals to monitor their health and seek immediate professional help if their life is at risk. With a synthetic dataset, machine learning can provide a consistent and early diagnosis to individuals based on their input data. Additionally, the program aims to make healthcare accessible by allowing more individuals to have regular checkups without needing to go to the hospital.

3.Problem Statement

What real-world problem related to the SDG are you trying to solve?

> Regular health checks aim to identify early signs of health issues. Finding problems early helps individuals find treatments that can reduce the risk of having a fatal health issue. However, a concerning trend has been observed in the Philippines, where regular health checkups are often neglected. A recent study revealed that 40% of their chosen respondents undergo annual medical check-ups (Capstone Intel-Corporation, 2023). This raises a significant issue in the Philippines, particularly in the context of healthcare. Moreover, the research revealed that 32% of the respondents with a monthly salary of less than P10 000 only seek medical assistance when unwell, while 50% of those with a monthly income of ₱65,000 or more have an annual health checkup. According to Cabalza (2023), almost 800 million people in the Western Pacific region, including the Philippines, do not have full access to basic health services. Thus, to lessen the risk and improve accessibility of healthcare, the researchers have formulated to create a Machine Learning model that can predict the health and lifestyle risk of an individual. The research aims to allow individuals to monitor their health remotely without going to the hospital. With this innovation, the researchers believe that using a predictive model can improve the accessibility of healthcare and help individuals reduce the risk of having a severe health issue.

4.Machine Learning Component

Describe the ML task (classification, prediction, recommendation, clustering, etc.) and planned dataset:

ML Objective: Build a supervised classification model that predicts an individual's health risk level (e.g., Low / Medium / High or At-Risk / Not-At-Risk) from lifestyle, demographic, socioeconomic, and clinical features in the Kaggle "Lifestyle and Health Risk Prediction" dataset to enable remote, early-risk screening and prioritization of preventive care.

Possible Dataset Source: Kaggle — "Lifestyle and Health Risk Prediction" dataset

5.KRR (Knowledge Representations Reasoning) Component

Describe the rule-based or ontology-based reasoning you will implement:

Representation Type (Rules/Ontology/Semantic Web):

- **Rule-Based Reasoning (Using IF–THEN health and lifestyle rules):**
    - The system will incorporate expert-derived heuristic rules from health guidelines (e.g., WHO, DOH) to complement the machine learning model's predictions. These rules serve as interpretable logic for explaining why the model assigns a health risk level and provide additional reasoning that ML alone cannot express.

Example Rule / Knowledge Representation Idea:

- **BMI Risk Rule:**
  - IF BMI > 30
  - THEN Risk_Level = "High"
  - BECAUSE obesity is a known predictor of cardiovascular and metabolic diseases.
- **Smoking Frequency Rule:**
  - IF Smoking_Status = "Yes" AND Exercise_Frequency = "Low"
  - THEN Risk_Level = "High"
- **Combined Lifestyle Risk Rule:**
  - IF (Smoking_Status = "Yes") AND (Alcohol_Intake = "Yes") AND (Sleep_Hours < 6)
  - THEN Lifestyle_Risk = "High"

6.Expected Output / System Overview

Describe what the user can do and what the system will return:

> The system will allow users to input their lifestyle and health-related information, such as age, BMI, sleep duration, smoking status, alcohol consumption, physical activity, and other relevant personal habits. After submitting their data, the system will process the inputs through a trained machine learning model to analyze potential lifestyle-related health risks.

The system will return a predicted **health risk level** (e.g., Low / Medium / High or At-Risk / Not-At-Risk) along with brief insights indicating the key factors that contributed to the prediction. Additionally, the system may provide optional recommendations or reminders encouraging users to improve their health habits based on the identified risk. This enables users to remotely assess their health status without the need for a physical medical check-up, supporting preventive health monitoring and awareness.

7.Project Timeline (Weeks 1–6)

Provide a brief week-by-week plan:

Week	Task
Week 1–2	Proposal + Literature Review

Week 3–4	Data cleaning + Baseline ML model
Week 5–6	Rule base + Inference integration
Week 7	Web deployment
Week 8	Testing + Refinement
Week 9	Final report & defense

8. Tools Technology Stack

- **Front-End:** HTML, CSS, JavaScript
- **Back-End:** Python, Flask
- **Algorithms:** Linear Regression, K-Means Clustering, Train-Test Split, Logistic Regression, Cross-Validation, Confusion Matrix, Learning Curve
- **Library:** Scikit-learn

9. Anticipated Challenges

- Limited Medical Validity
    - Predictions cannot replace real medical diagnosis, so the system must be clearly framed as a screening tool, not a clinical device.
- Scalability Issues
    - If multiple users access the system simultaneously, the server may experience increased load, potentially slowing down the prediction process.
- Feature Misinterpretation by Users
    - Users may input incorrect or unrealistic values (impossible BMI, Incorrect Lifestyle Entries. etc.), affecting the reliability of predictions.
- Rule Conflict Between ML and KRR System
    - ML predictions may contradict expert-defined rules (model may say “Low Risk” but rule flags is “High Risk”), requiring a conflict-resolution strategy.
- Overfitting or Poor Model Generalization
    - The model may perform well during training but poorly on unseen data, requiring careful validation, tuning, and regularization.
- Computational Resource Limitations

- Training and evaluating models, especially if multiple algorithms are tested, may be slow on devices with limited RAM or CPU power.

10. References (If any)

- **Kaggle dataset**
  MIADUL. (n.d.). *Lifestyle and Health Risk Prediction* [Data set]. Kaggle. https://www.kaggle.com/datasets/miadul/lifestyle-and-health-risk-prediction
- **Capstone-Intel article**
  Capstone-Intel Corporation. (2023, September 4). *New Study Reveals Only 40% of Filipinos Get Yearly Checkups, an Issue That Can Be Addressed by UHC Law* [Web page]. Capstone-Intel. https://www.capstone-intel.com/new-study-reveals-only-40-of-filipinos-get-yearly-checkups-an-issue-that-can-be-addressed-by-uhc-law-atty-conti Capstone Intel
- **Inquirer Globalnation article**
  Inquirer Global Nation. (2024, April 8). *WHO: Basic health services lacking in PH, Western Pacific*. Philippine Daily Inquirer. https://globalnation.inquirer.net/230863/who-basic-health-services-lacking-in-ph-western-pacific globalnation.inquirer.net
- **Better Health Victoria page**
  Victorian Department of Health. (n.d.). *Health checks*. Better Health Channel. https://www.betterhealth.vic.gov.au/healthyliving/health-checks

Prepared by:

1. Kristel N. Ramos (3rd Year, Computer Science)
2. Johann Francois P. Tanyag (3rd Year, Computer Science)
3. Althea Maxene B. Viar (3rd Year, Computer Science)

Date: 11/16/2025

Approved by Instructor: _____Date: