**Topic Area: Health sector**

**Project:** Predict the likelihood of chronic diseases such as diabetes, depression, dementia, cardiovascular diseases in Australian populations.

*Note: A specific disease will be chosen based on the quality of dataset we get.*

**Model:** Use machine learning models like Logistic Regression or Deep Neural Networks.

**General Research Questions**

1. *Prediction Accuracy:* How accurately can we predict the onset of chronic diseases using demographic, lifestyle, and medical data from Australian populations?
2. *Feature Importance:* Which features (e.g., age, gender, lifestyle factors, medical history) are the most important predictors for each chronic disease?
3. *Model Performance:* How do different machine learning models (e.g., Logistic Regression, Decision Trees, Neural Networks) compare in terms of prediction accuracy and robustness?

Proposed Target for prediction accuracy: **80 percent.**

**Disease-Specific Research Questions**

**Diabetes Prediction:**

1. What is the relative importance of factors like BMI, physical activity, and family history in predicting the likelihood of developing diabetes?
2. How well do models perform in predicting diabetes onset in different age groups and ethnicities within the Australian population?

**Depression Prediction:**

1. Can we identify early warning signs of depression based on socio-economic status, lifestyle, and medical history?
2. How does the inclusion of mental health survey data improve the prediction accuracy for depression?

**Dementia Prediction:**

1. What lifestyle and genetic factors are most strongly associated with the risk of developing dementia?
2. How early can we predict the onset of dementia with reasonable accuracy?

**Cardiovascular Disease Prediction:**

1. Which factors (e.g., cholesterol levels, blood pressure, smoking status) are the strongest predictors of cardiovascular disease?
2. How do lifestyle changes (e.g., diet, exercise) over time impact the risk prediction for cardiovascular diseases?

**Intervention and Prevention Research Questions**

**Early Intervention:**

1. Can early prediction models identify individuals who would benefit most from preventive interventions?
2. What is the impact of early lifestyle interventions on reducing the risk of chronic diseases?

**Healthcare Utilization:**

1. How does the utilization of healthcare services (e.g., regular check-ups, preventive screenings) impact the prediction and management of chronic diseases?
2. Can prediction models help optimize healthcare resource allocation for chronic disease management?

**Personalized Medicine:**

1. How can machine learning models be used to develop personalized prevention and treatment plans for individuals at high risk of chronic diseases?
2. What role does genetic data play in enhancing the accuracy and personalization of chronic disease predictions?

**Data source:** Use datasets from Keggle, Australian Institute of Health and Welfare (AIHW), Health Data Australia, ABS, or state health departments.