Capstone Two -Project Proposal

Problem Identification

Problem Statement Formation:
My goal is to develop a classification model that predicts the likelihood of breast cancer recurrence or death based on patient information and tumor characteristics.

Context:
Breast cancer is a significant health concern, and timely prediction of patient outcomes is crucial for effective treatment planning and care.
The dataset at hand contains information on breast cancer patients, including age, menopausal status, tumor size, grade, lymph node involvement, hormone receptor levels, hormonal therapy, and survival outcomes.

Criteria for Success:
The success of this project will be determined by the accuracy, precision, and interpretability of the classification model.
The model should provide valuable insights into the influence of various factors on breast cancer recurrence and death.

Scope of Solution Space:
My solution will involve building a classification model, focusing on supervised learning techniques.
The model will predict two outcomes: 0 for patients who are alive without recurrence and 1 for patients with recurrence or death.
The solution space includes data preprocessing, feature selection, model development, hyperparameter tuning, and evaluation.

Constraints:
Ethical considerations and patient data privacy are paramount constraints. We must handle sensitive medical data with care and ensure compliance with relevant regulations.
Limited computational resources may impose constraints on the complexity of the model and hyperparameter optimization.

Stakeholders:
Stakeholders include medical professionals, researchers, and healthcare institutions interested in predicting breast cancer outcomes.
Patients are indirect stakeholders as the model's predictions can influence their treatment plans and well-being.

Data Sources:
The primary data source is the breast cancer dataset from Kaggle, which includes patient information, tumor characteristics, and survival outcomes.
Additional data sources, if necessary, may be consulted for feature enrichment and validation.