**Breast Cancer Survival Prediction (Questions and Answers)**

Madhuri Basava

Bellevue University

Applied Data Science DSC680

Amirfarrokh Iranitalab

Aug 5thth, 2024

**Research Questions & Answers:**

1. Which features are most relevant for predicting Breast Cancer survival?

The below chart from the XG Boost Classifier shows that ‘Tumour Stage III’, Surgery type Other, ER2 Status\_Negative, Surgery\_type\_Mastectomy, and ‘Tumour Stage I’ features are important.

A graph with text on it

Description automatically generated

1. Which models are suitable for Breast Cancer survival prediction?

Random Forest, Decision Tree Classifier, XG Boost Classifier, Support Vector Classification, and Logistic Regression algorithms are suitable for modeling. Ensemble methods like Random Forest and Gradient Boosting perform well due to their ability to handle complex interactions between various features.

1. What criteria should be used to evaluate the performance of the models?

Metrics like accuracy, precision, recall, F1-score, and confusion Matrix are used to evaluate the performance of the models.

1. How will missing data be handled in the dataset?

Missing data is handled using the imputation technique by removing the rows with null values.

1. What steps will be taken to ensure data quality and integrity?

Data quality will be ensured through rigorous data cleaning, validation checks, consistency checks, and by setting up a data governance framework to monitor and maintain data integrity.

1. How will data privacy be ensured, given the sensitivity of medical data?

The anonymization of patient data ensures data privacy. Here, in the first column, Patient\_Id’s values in the data set are anonymized.

1. How will the ethical implications of survival predictions be communicated to patients and healthcare providers?

Ethical implications will be communicated through clear, understandable reports and consultations with healthcare providers, ensuring patients are fully informed about their prognosis and treatment options.

1. How will the models be updated with new data and advancements in medical research?

Models will be periodically retrained with new data, and a continuous learning framework will be established to incorporate the latest medical research findings.

1. How will real-time data be incorporated into the prediction models?

Real-time data will be incorporated through integration with IoT platforms and wearable devices, enabling continuous monitoring and dynamic updates to predictions.

1. How will the success of the project be measured in terms of improving patient outcomes and healthcare efficiency?

Success will be measured through metrics such as improved survival rates, patient satisfaction, reduced treatment costs, and enhanced healthcare provider efficiency.