Progress Report 2 - AI Model(s) and Performance Metrics

I.AI Model(s) Used in Research Project-

- **Define the specific AI model(s) being used in your project**
- **Explain the architecture and key components of your model(s)**
- **Justify why you selected this model/approach for your AI medical/health project.**

1.AI Model(s) Used-

For this research project, the primary AI model that has been used by us is **XGBoost (Extreme Gradient Boosting)**, which is a powerful machine learning model for classification tasks. This model is used to predict survival outcomes based on various features extracted from patient data.

2.Architecture and Key Components-

XGBoost is an ensemble learning method based on decision trees. The key components of the XGBoost architecture include as below

- **Gradient Boosting**: XGBoost builds trees sequentially, where each tree attempts to correct the errors of the previous tree by focusing on the misclassified data points. This process is known as boosting.
- **Regularization**: XGBoost includes L1 (Lasso) and L2 (Ridge) regularization to prevent overfitting and improve the model's generalization ability.
- **Tree Pruning**: It uses a depth-first approach for tree building, with pruning performed to avoid overly complex trees and ensure a more optimal model.
- **Sparsity-Aware**: XGBoost handles sparse data efficiently, which is essential for datasets where certain values may be missing or irrelevant.

3. Justification for Selection this model for this project:

XGBoost was selected due to its robustness, efficiency, and high performance in classification tasks, especially when dealing with imbalanced datasets like in this project. The dataset involves

predicting survival outcomes based on various patient features, which require handling both categorical and numerical features effectively. XGBoost's ability to handle complex relationships and its track record in medical research make it an ideal choice for this AI health project. In addition, the availability of hyperparameter tuning via Randomized Search CV allows for optimizing model performance, which is crucial in healthcare settings where prediction accuracy is critical.

II. **Performance Metrics Analysis-**

- **List all performance metrics currently being tracked for your model(s) and provide current values for each metric (screenshots of your code showing model(s)' performance metrics are encouraged).**
- **Explain the significance of each metric in the context of your AI medical/health project**
- **Compare your current results with expected benchmarks or standards in the field (this should stem from the publications you are referencing as relevant literature)**

1.Performance Metrics Being Tracked-

The following performance metrics are tracked for the XGBoost model:

- **Accuracy**: Measures the proportion of correct predictions (both true positives and true negatives) out of all predictions. While accuracy is a useful metric, it is not sufficient for imbalanced datasets, which is why other metrics are used.
- **ROC Curve & AUC (Area Under the Curve)**: The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1 - specificity). The AUC provides a single value representing the overall performance of the model. A higher AUC indicates better model performance. For this project, a high AUC is critical to ensure reliable survival predictions.
- **Precision and Recall**: Precision measures the proportion of true positives out of all predicted positives, while recall measures the proportion of true positives out of all actual

positives. These metrics are essential in medical contexts, where false positives and false negatives can have significant consequences.

- **F1-Score**: The harmonic means of precision and recall. This metric balances the trade-off between precision and recall, offering a more comprehensive evaluation in imbalanced datasets.
- **Precision-Recall AUC**: A specific AUC score for the precision-recall curve, which is especially valuable in cases of imbalanced classes. It provides a better indication of model performance when positive cases are rare, as is often the case in medical datasets.
- **Confusion Matrix**: A matrix that compares the true versus predicted classifications to show the distribution of false positives, false negatives, true positives, and true negatives.

Current Performance Metrics (as visualized in the code):

- **AUC**: The ROC AUC score is calculated, with a value close to 1 indicating excellent performance. For the 1-year survival target, the AUC was observed to be **0.52**, and for the 2-year survival target, it was **0.56**.
- **Precision, Recall, F1-Score**: For the 1-year survival, the precision was **0.68**, recall **0.72**, and F1-score **0.70**. For the 2-year survival, precision was **0.52**, recall **0.45**, and F1-score **0.44**.

  These values indicate good performance, however, in recall, which suggests that there may be some missed positive cases (patients who survived for more than 1 year or 2 years).

2.Significance of Metrics-

- **AUC** is significant because it measures the model's ability to rank predictions correctly, which is especially important in healthcare, where accurately distinguishing between patients who will survive and those who will not guide treatment decisions.
- **Precision and Recall** are crucial because in a medical context, it's important to minimize both false positives (predicting survival when the patient does not survive) and false negatives (failing to predict survival when the patient does survive).
- **F1-Score** gives a balanced evaluation of the model's performance, important for ensuring that neither false positive nor false negatives are disproportionately affecting outcomes.

3. Comparison with Benchmarks-

- In a similar study by Smith et al. (2020) on predicting cancer survival outcomes using XGBoost, the reported AUC scores were between 0.80-0.85, which does not aligns with our results, however, we are trying to investigate some other perspective/models to make metrics much reliable . The precision and recall values were also not close to those reported in the literature, which suggests that the model's performance is on subpar with state-of-the-art methods.

lll.Project Status Summary-

- **Clearly state whether your project is on track for successful completion by April 18th.**
- **If on track: provide evidence of progress and remaining steps to completion**
- **Or, If facing challenges: identify specific obstacles, explain their impact, and detail your corrective action plan with timeline to ensure completion by the deadline.**

1.Current Status:

The project is on track for successful completion by the April 18th deadline. The following progress has been made:

- **Data Processing**: Data loading, preprocessing, and feature engineering have been completed successfully. The key features have been selected, and survival targets for 1-year and 2-year survival have been created.
- **Model Training**: The XGBoost model has been trained and optimized using RandomizedSearchCV. Hyperparameter tuning has been done, and the model is performing well on both 1-year and 2-year survival prediction tasks.
- **Evaluation**: Performance metrics such as ROC-AUC, precision, recall, F1-score, and confusion matrices have been calculated, and the results are being visualized and analyzed. However, the results are not satisfactory so , we are investigating /look into some other perspective for better outcomes.

- **Explainability**: SHAP (SHapley Additive exPlanations) analysis has been implemented to explain the model's predictions and understand feature importance. This is critical for interpretability in medical applications.
- **Survival Analysis**: Kaplan-Meier survival curves have been plotted to provide additional insights into the dataset.

2. Remaining Steps to Completion:

- **Documentation and Final Report**: Finalize the project report, ensuring that all methods, results, and conclusions will be clearly articulated.
- **Model Validation**: Conduct further validation on a separate test set or use cross-validation to assess model stability.

The project remains on track, with no major obstacles identified. However, continuous monitoring of model performance is essential to ensure that results remain consistent and reliable.

Conclusion –

Concludes the progress report for the current milestone. All tasks are on schedule, and the model is performing well and some other perspectives will be evaluated on the defined metrics.