**Churn Prediction Model**

**1. Introduction**

The primary objective of this project was to develop predictive models for identifying customers at risk of churning in a telecommunications company. Two approaches were employed: traditional machine learning algorithms (Gradient Boosting, Logistic Regression, and Random Forest) and a neural network using deep learning.

**2. Data Collection and Pre-processing**

The dataset, obtained from Kaggle, contained customer information such as demographics, service usage, and churn status. Data pre-processing involved handling missing values, encoding categorical variables, and creating new features like 'TotalCharges' as the product of 'tenure' and 'MonthlyCharges’.

**3. Exploratory Data Analysis (EDA)**

EDA provided valuable insights into customer behavior and factors influencing churn. Key findings included an imbalance in the dataset, with significantly more instances of 'No' churn compared to 'Yes.' Visualizations helped reveal patterns and understand feature distributions.

**4. Feature Engineering**

Feature engineering aimed to enhance the predictive power of the models. Besides 'TotalCharges,' additional features were explored to capture potential indicators of churn, such as interactions between existing variables.

**5. Machine Learning Algorithms**

**5.1. Gradient Boosting**

The Gradient Boosting model was fine-tuned using grid search to optimize hyperparameters. The best model achieved the following results on the validation set:

* Accuracy: 0.8097
* Precision: 0.6732
* Recall: 0.5508
* F1 Score: 0.6059

**5.2. Logistic Regression**

The Logistic Regression model achieved the following results on the validation set:

* Accuracy: 0.7287
* Precision: 0.4933
* Recall: 0.7861
* F1 Score: 0.6062

**5.3. Random Forest**

The Random Forest model achieved the following results on the validation set:

* Accuracy: 0.7855
* Precision: 0.6169
* Recall: 0.5080
* F1 Score: 0.5572

**6. Neural Network**

The neural network model, implemented using Keras with TensorFlow backend, achieved the following results:

Loss: 0.3341

Accuracy: 0.8596

Validation Loss: 0.4033

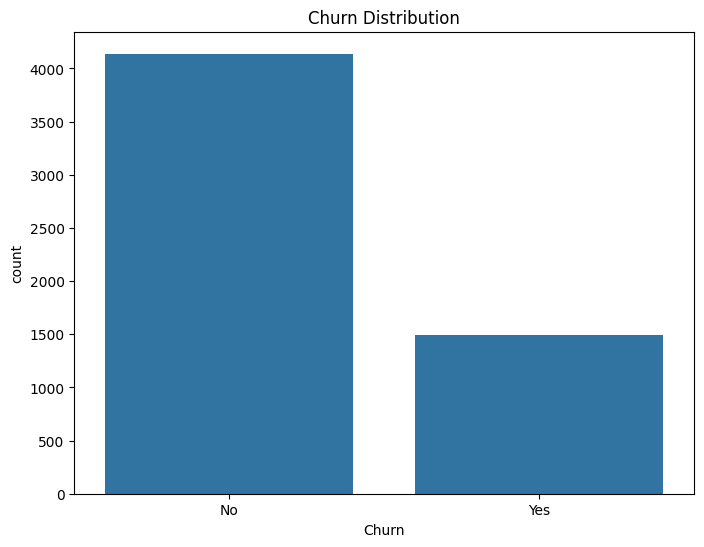
Validation Accuracy: 0.8148

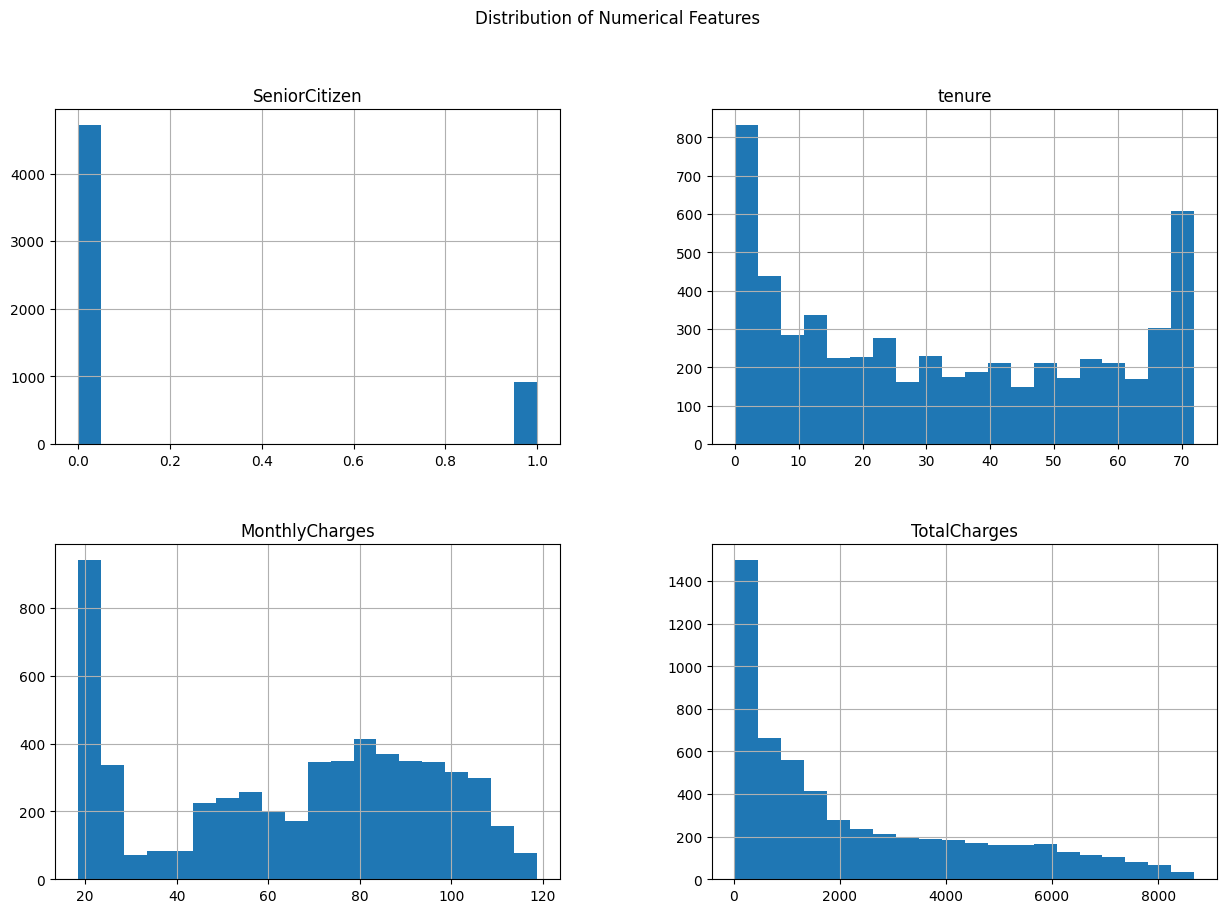
**7. Challenges**

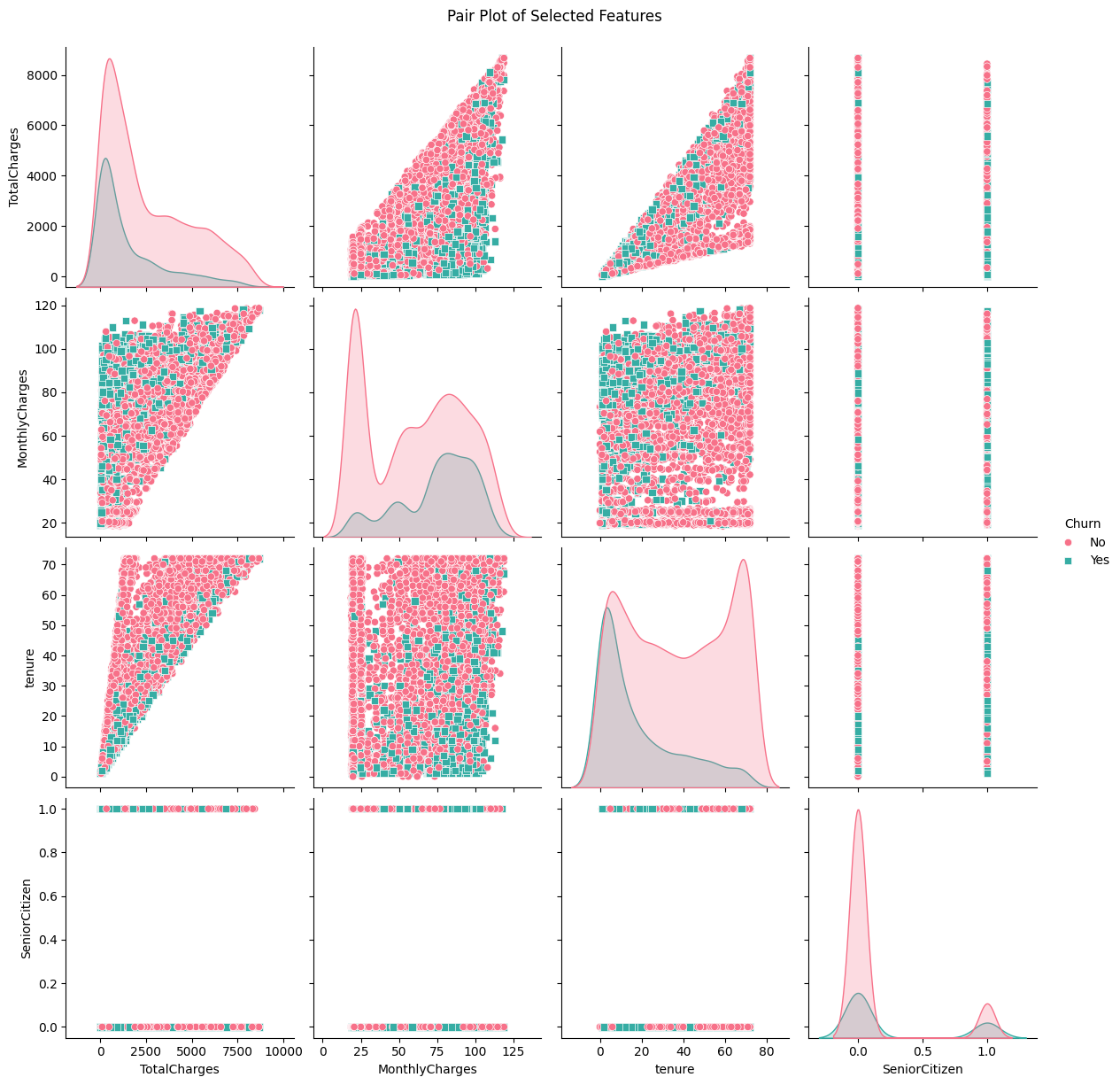
The main challenge encountered was dealing with imbalanced data, particularly in the machine learning models. Techniques like class weights were applied to mitigate this issue and improve model performance.

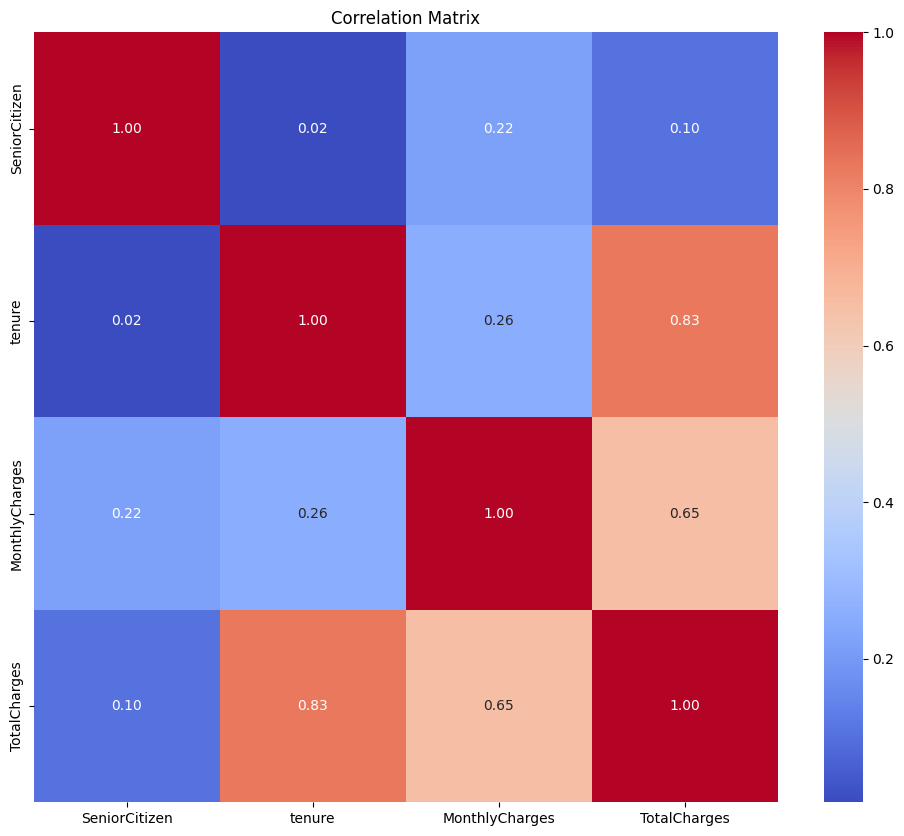
**8. Visualization of insights**

**8.1. EDA**

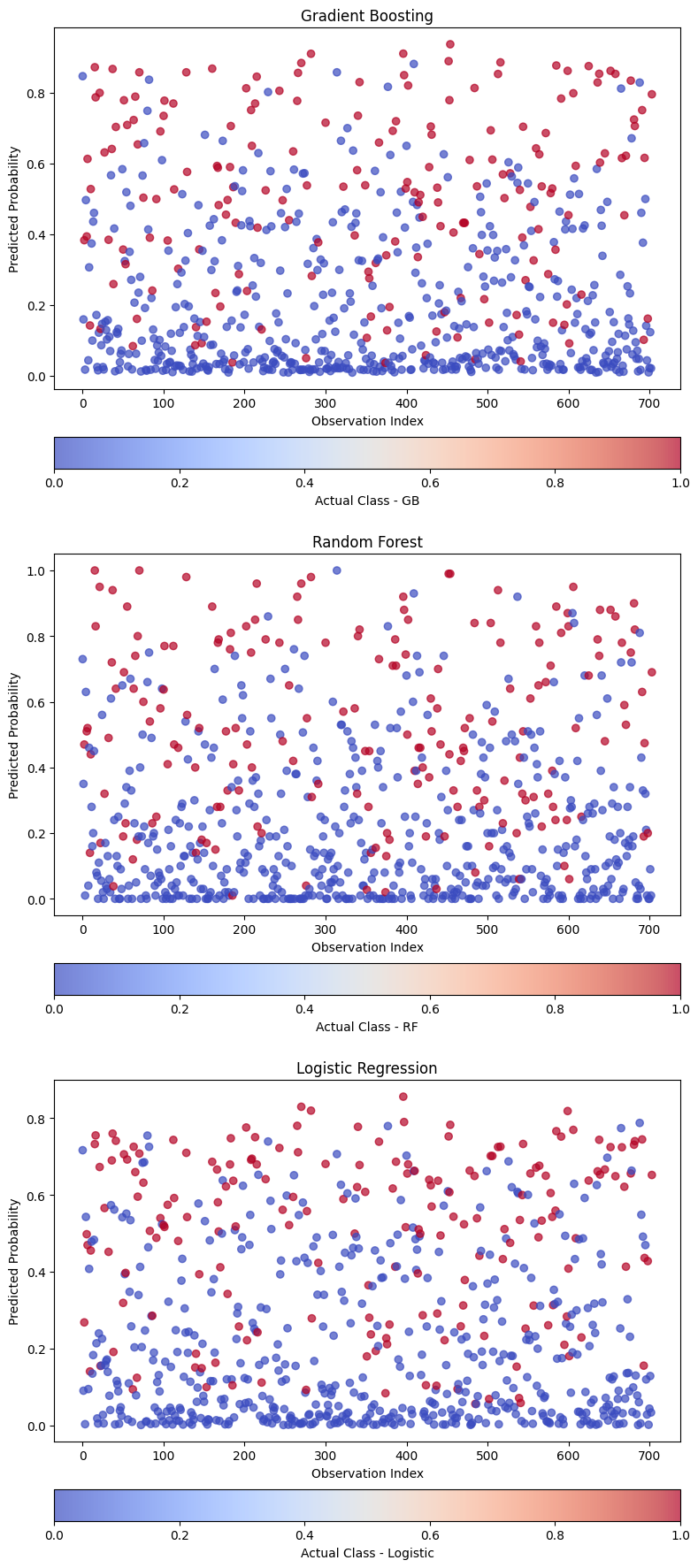
****

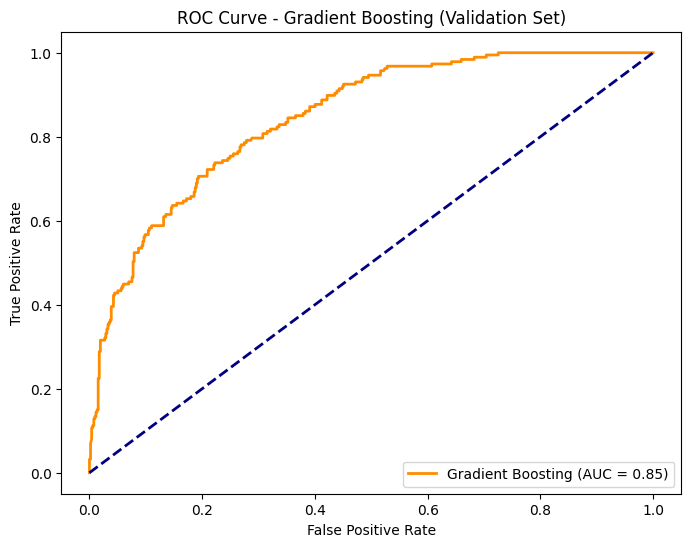
****

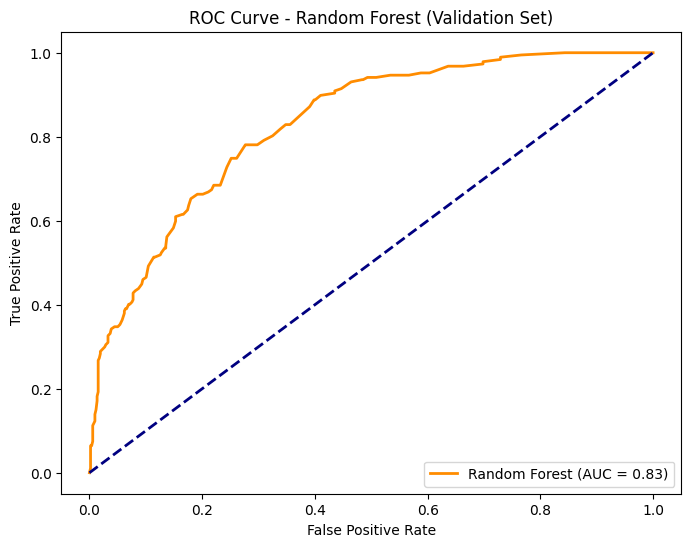


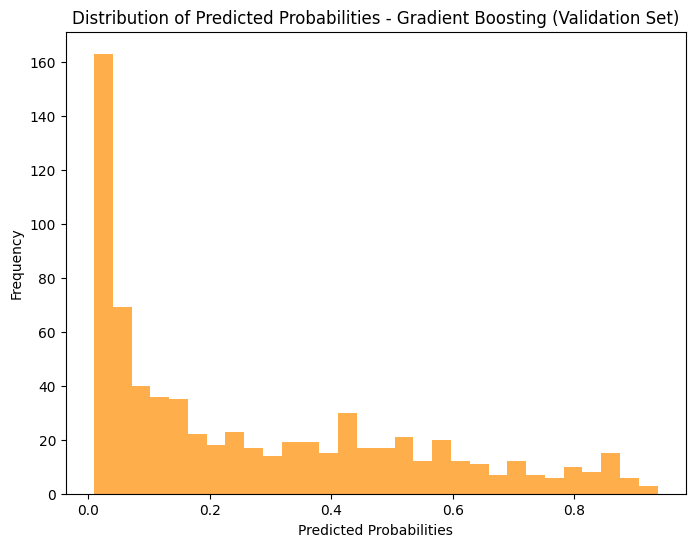
****

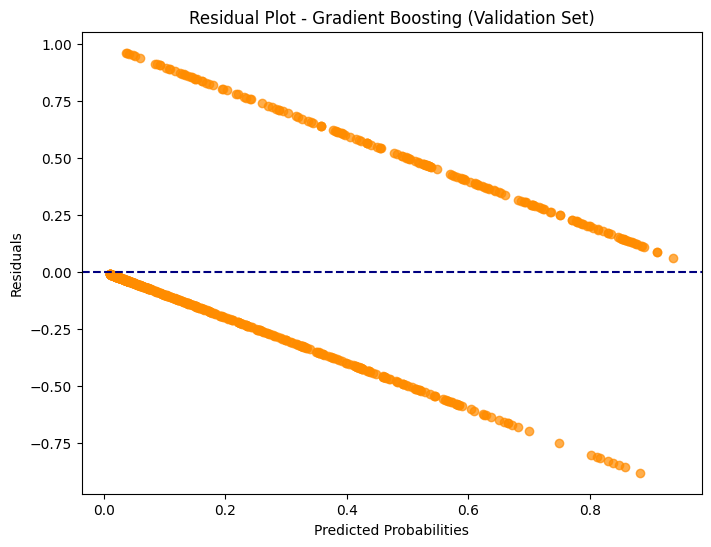
**8.1 ML algorithms**

****

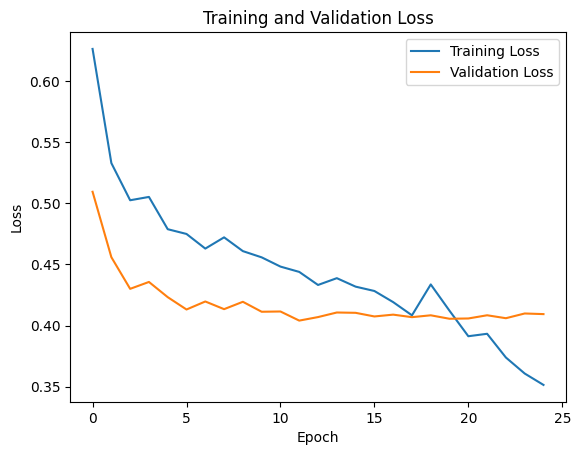
****

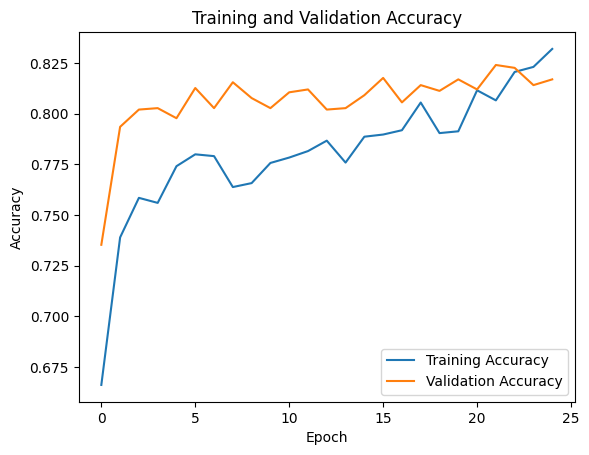
****

****

****

**8.2 Neural Networks**

****

****

**9. Conclusion:**

In this project, we aimed to develop a predictive model for identifying customers at risk of churning in a telecommunications company. The dataset was preprocessed to handle missing values, encode categorical variables, and create relevant features. Exploratory Data Analysis (EDA) provided insights into customer behaviour and factors influencing churn.

The ROC curve, Precision-Recall curve, and confusion matrices provided additional insights into the models' behavior. Further fine-tuning of hyperparameters, feature engineering, and exploring advanced techniques could potentially improve model performance.

Overall, the Gradient Boosting model showed the best balance between precision and recall, making it a promising candidate for deployment in real-world scenarios. However, continuous monitoring and updates to the model will be crucial to adapt to changing customer behaviors.