HealthCare Analytics (Patient length of stay (LOS))

The analysis provided a comprehensive overview of the HealthCare Analytics dataset for predicting patient length of stay (LOS). Here's a breakdown of the key steps and findings:

1. **Data Pre-processing**: The initial steps included importing necessary libraries, reading the dataset, and performing basic data pre-processing tasks such as handling missing values and dropping unnecessary columns.
2. **Exploratory Data Analysis (EDA):** Various visualizations were generated to understand the distribution of categorical variables and their relationships with the target variable (Stay). Chi-square tests were also conducted to assess the significance of these categorical variables in predicting patient length of stay.
3. **Label Encoding:** Categorical variables were encoded using one-hot encoding and label encoding techniques to prepare the data for machine learning models.
4. **Machine Learning Models:** Several machine learning models were trained and evaluated using the pre-processed data. These models include Logistic Regression, Random Forest, Gradient Boosting Machines (GBM), XGBoost, Catboost, Naive Bayes, Fully Connected Layer (ANN), and Convolutional Neural Network (CNN).
5. **Evaluation Metrics:** The models were evaluated based on both ML model scores and F1 scores to assess their performance in predicting patient length of stay.
6. **Model Comparison:** A summary table was created to compare the performance of each model based on the evaluation metrics. Catboost emerged as the top-performing model, followed by Random Forest and GBM. However, Gaussian Naive Bayes had a lower ML model score but a competitive F1 score, indicating a balance between accuracy and precision-recall.
7. **Conclusion:** The analysis concluded that Catboost was the most effective model for predicting patient length of stay in this dataset. However, the choice of the best model depends on specific requirements such as interpretability, computational efficiency, and the importance of precision-recall balance.

Overall, the analysis provided valuable insights into predicting patient length of stay in healthcare settings, highlighting the importance of machine learning techniques in optimizing resource allocation and improving healthcare management efficiency.