NAME : Hemanth Naga Pavan Kumar
UBID : hswarna
UBnumber: 50559715

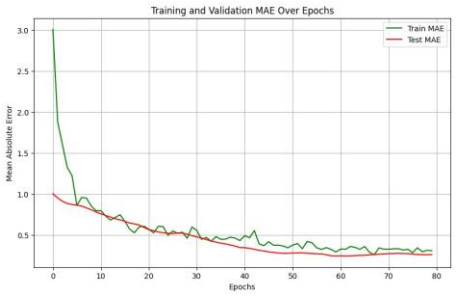
HOSPITAL LENGTH OF STAY PRDICTION

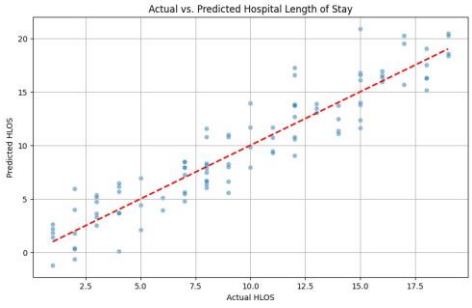
In this project, aimed to predict hospital length of stay (HLOS) using a deep learning model called LSTM (Long Short-Term Memory) with an attention mechanism. The dataset had information about patients' demographics, medical conditions, and clinical features recorded during their stay. I merged different datasets together to include important features like age, sex, vital signs, and lab results. The target variable, "adjusted HLOS," was calculated by subtracting the current day from the total hospital length of stay. I only kept the rows where adjusted HLOS was greater than zero for model training.

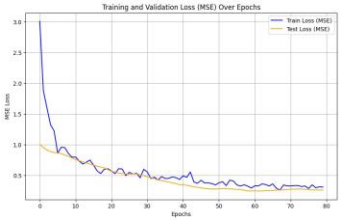
For data cleaning, I dropped columns with too many missing values and filled the missing numerical data with medians, and categorical data with the most common value. I used LabelEncoder to encode categorical features like sex and comorbidities, and scaled the numerical features using MinMaxScaler. After this, the data was split into training and testing sets.

I built the LSTM model with an attention mechanism that helps the model focus on the most important parts of the sequence. LSTM is good for sequential data like medical records taken over time. The attention mechanism makes the model pay more attention to important features, which is useful when predicting HLOS because certain medical conditions may have more effect on the length of stay.

To improve the model's performance, I used a random search to tune hyperparameters like dropout rate, learning rate, and batch size. I also applied early stopping and learning rate scheduling to avoid overfitting and make the model perform better. Cross-validation was used to check how the model works on different parts of the data.In the end, the model got a mean squared error (MSE) of 0.184 and an R^2 score of 0.85, showing that it performed well. This model can be used in hospitals to predict how long patients will stay and help in resource management







NAME : Hemanth Naga Pavan Kumar
UBID : hswarna
UBnumber: 50559715

This project focuses on predicting hospital length of stay (HLOS) using a given dataset containing patient admission details and clinical measurements. The data was pre-processed by handling missing values through median imputation for numerical variables and KNN imputation for specific features. Categorical features, such as 'sex' and 'intubation status,' were one-hot encoded. Comobidities were processed using Word2Vec embeddings to capture latent information from textual data, followed by encoding specific comorbidity categories.

In this project, I developed a model using the Keras Sequential API with three hidden layers. To avoid overfitting of the model, I applied dropout and batch normalization. I also performed a grid search to find the best learning rates and dropout rates. The training process included early stopping to help the model perform better better.

Model Performance: Final R2 score (Test): 0.6724

The final model achieved a mean squared error (MSE) of 0.2604 and an R^2 score of 0.6724 on the test data, which shows that the model performed pretty well. To evaluate the performance, I used various visualizations, including distribution plots for hospital length of stay (HLOS), correlation heatmaps, actual vs. predicted graphs, and learning curves. These helped me understand how well the model was doing. Overall, the deep learning model did a good job predicting hospital length of stay, and I identified the top 15 features that are most closely related to HLOS. This information can be useful for making better clinical decisions in the future.