Task Performed: Model Selection and Training

Flight fare prediction using machine learning involves selecting an appropriate model and training it on a dataset of flight information and fare prices. In this section, we will discuss the model selection and training process based on the reference article "A Comparison Between Machine Learning Models for Air ticket Price Prediction".

Model Selection:

The authors of the reference article compared the performance of several regression models, including linear regression, decision tree, random forest, and support vector regression (SVR). They evaluated the models using mean absolute error (MAE), root mean squared error (RMSE), and R-squared values. Based on their evaluation, they found that the random forest model outperformed the other models in terms of accuracy, with the lowest MAE, RMSE, and the highest R-squared value. Therefore, they selected the random forest model for predicting air ticket prices.

Training:

The authors used a dataset that combined flight information and fare prices from three different sources. The dataset required modifications as the project proceeded further. The authors performed feature selection and hyperparameter tuning to improve the performance of the random forest model.

Feature Selection:

The authors used the recursive feature elimination (RFE) method to rank the attributes based on their importance and select the top k features that provide the best predictions. They found that the combination of departure date, departure airport, and arrival airport is the most important set of features for predicting air ticket prices.

Hyperparameter Tuning:

The authors also performed hyperparameter tuning to optimize the performance of the random forest model. They tuned the hyperparameters, such as the number of trees in the forest, the maximum depth of the tree, and the minimum number of samples required to split a node. They used a grid search method to find the best hyperparameter values that minimize the prediction error.

After selecting the model and performing feature selection and hyperparameter tuning, the authors trained the random forest model on the modified dataset. They used the training dataset to fit the model and then evaluated its performance on the testing dataset using MAE, RMSE, and R-squared values. The authors found that the optimized random forest model provided more accurate predictions than the other models and was the best choice for air ticket price prediction.