**Airline Flight Price Prediction - Technical Report**

Business Problem

It is crucial for airlines and passengers to be able to accurately predict flight prices. For airlines it helps them to accurately forecast revenue streams, and for passengers it allows them to plan better and understand price savings.

Data Source and Preprocessing Steps

The analysis uses flight data from Kaggle, the data contains records of flights between Indian cities. The following preprocessing steps were implemented, non-applicable columns removed (index, flight number, source/destination city), a route feature created (combining source/destination city), split data into training and test sets using an 80/20 split and scaling features with the standard scaler function.

Model Selection

Three regression models were used in the analysis linear regression, random forest and lasso regression. Model performance was evaluated using MSE, RMSE, MAE and R2 on the test set.

Key Insights and Recommendations

All the models explain over 90% of the variance in flight prices. The features with the most impact on price are route, airline, class and day left. The recommendation from this analysis is for domestic Indian airlines and passengers to leverage this model to support pricing and purchasing strategies. The R2 value above .90 suggests the model can confidently predict flight prices based off the input features.

Limitations and Future Improvements

The dataset is limited to Indian flights and prices, so the results are not applicable to other flight markets. The model does not account/incorporate other factors such as holidays and promotions etc. For future work other markets could be included or this model could be applied to individual markets. In addition, other models could be used such as XGboost or Ridge. Also, a Boolean feature could be used to indicate holidays/surge pricing.