<u>Task Performed: Week-1</u>

In first week, the reference article was thoroughly reviewed and analysis for the attributes used for justifying the prediction-based learning of different machine learning models was completed and on basis of the review, the feature selection is done.

The next task will be to justify the data collection and detailed feature analysis. The data included in the shared Kaggle link has been combined from 3 different sources based on the initial problem definition. The dataset will require the modifications as the project proceeds further.

The article review:

The paper "A Comparison Between Machine Learning Models for Air ticket Price Prediction" compares the performance of different machine learning models for predicting air ticket prices. The authors of the paper present a comprehensive study of various regression models, including linear regression, decision tree, random forest, and support vector regression (SVR), and evaluate their performance using mean absolute error (MAE), root mean squared error (RMSE), and R-squared values.

The results of the study indicate that the random forest model outperforms the other models in terms of accuracy, with the lowest MAE, RMSE, and the highest R-squared value. The authors also show that the random forest model is more robust to outliers compared to the other models.

Additionally, the authors perform feature selection and hyperparameter tuning to further improve the performance of the random forest model. They conclude that the optimized random forest model provides more accurate predictions than the other models and is the best choice for air ticket price prediction.

The study provides valuable insights for researchers and practitioners in the field of air ticket price prediction and demonstrates the importance of model selection, feature selection, and hyperparameter tuning in achieving accurate predictions. The authors also suggest future research directions, including the integration of external factors such as weather conditions and economic indicators, to further enhance the accuracy of air ticket price predictions.

Overall, the "A Comparison Between Machine Learning Models for Air ticket Price Prediction" is a well-designed and well-executed study that contributes to the literature in the field of air ticket price prediction. The results of the study can be useful for airlines, travel agencies, and other stakeholders in the aviation industry to make informed decisions about air ticket pricing.

Feature Analysis based on reference article:

The paper used several attributes to make predictions about air ticket prices. The authors performed a comprehensive feature analysis to determine the importance of these attributes in predicting air ticket prices.

The attributes used in the study include:

1.   Departure date: The date on which the flight departs was used as a feature in the study. The authors found that the departure date has a significant impact on air ticket prices and is an important attribute to consider.

2.   Arrival date: The date on which the flight arrives was also used as a feature. The authors found that the arrival date has a moderate impact on air ticket prices and is an important attribute to consider.

3.   Departure airport: The airport from which the flight departs was used as a feature. The authors found that the departure airport has a moderate impact on air ticket prices and is an important attribute to consider.

4.   Arrival airport: The airport to which the flight arrives was used as a feature. The authors found that the arrival airport has a moderate impact on air ticket prices and is an important attribute to consider.

5.   Flight duration: The duration of the flight was used as a feature. The authors found that flight duration has a moderate impact on air ticket prices and is an important attribute to consider.

6.   Number of stops: The number of stops on the flight was used as a feature. The authors found that the number of stops has a moderate impact on air ticket prices and is an important attribute to consider.

7.   Airline: The airline operating the flight was used as a feature. The authors found that the airline has a moderate impact on air ticket prices and is an important attribute to consider.

The authors also performed feature selection to determine the most important attributes and reduce the dimensionality of the dataset. They 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. The authors found that the combination of departure date, departure airport, and arrival airport is the most important set of features for predicting air ticket prices.

In conclusion, the authors found that several attributes are important in predicting air ticket prices, including departure date, departure airport, arrival airport, flight duration, number of stops, and airline. The results of the feature analysis provide valuable insights into the factors that influence air ticket prices and can be useful for airlines, travel agencies, and other stakeholders in the aviation industry to make informed decisions about air ticket pricing.