Thesis Proposal

Research Objective and Significance

In the airline sector, forecasting flight prices is a crucial problem. Customers may find the best ticket prices at various times with the use of accurate forecasts, which can also help airlines optimize their revenue management and pricing strategies. Big data-based flight price prediction has emerged as a significant study subject in recent years due to the development of data analysis tools. In order to identify the main determinants of flight costs and forecast future costs, this study will use linear regression techniques.

The simplicity, interpretability, and capacity to demonstrate the linear correlations between different factors and flight prices are the driving forces behind the selection of linear regression. We can measure the impact of several variables, including time, route, airline, and seat class, using this model. The study has a great deal of practical significance since it can help consumers make wise selections and airlines with their pricing strategies.

Literature Review

Predicting flight prices has drawn a lot of interest from both industry and academia. Numerous research have examined the variables influencing airfare and put forth different forecasting techniques. The current body of literature falls into the following categories:

1. Statistical Techniques: For predicting flight prices, conventional regression techniques like ridge regression and linear regression have been employed. The relationship between flight pricing and variables including time, destination, and airline has been examined in a number of studies using linear models. For example, studies have revealed that the time of booking, travel distance, and seasonal variations all have a major impact on ticket costs.

2. Techniques for Machine Learning: Numerous research have used machine learning methods such decision trees, support vector machines, and random forests in addition to linear regression. Compared to conventional regression techniques, these models can increase prediction accuracy and are more suited to capture more intricate, nonlinear interactions. To improve prediction performance, for instance, some research has employed gradient boosting techniques and random forests.

3. Neural networks and deep learning: Neural networks have been used in several research to anticipate airline fares as deep learning technologies have grown in popularity. Neural networks have a high computational complexity and need a lot of data for training, even though they may make strong predictions in big datasets.

4. Hybrid Approaches: To improve prediction accuracy, some research has tried to create hybrid models by fusing machine learning techniques with conventional regression models.

There are still issues, like managing seasonal variations, flight frequency fluctuations, and enhancing model interpretability, despite the range of approaches investigated. Because of its ease of use and interpretability, linear regression is still a popular option for predicting airline prices.

Main Research Content, Expected Goals, and Innovation

Principal Information:
1. Data collection and feature selection: identify and evaluate factors such airline, route distance, flight time, departure and destination cities, and flight time that impact trip costs.
2. Building a Linear Regression Model: Use linear regression to construct a model that predicts flight prices and investigate the linear relationship between various characteristics and costs.
3. Model Evaluation and Optimization: Examine possible enhancements and assess the model using metrics like Mean Squared Error (MSE) and R2.
4. Analysis of the Results and Conclusion: Examine the model's output to determine the main variables affecting airfare costs and talk about potential enhancements.

Anticipated Objectives:
1. Develop a linear regression-based model for predicting flight prices and assess its effectiveness using actual datasets.
2. Determine the main determinants of flight costs and offer airlines and customers useful information.

Innovation: 1. This study offers a straightforward and understandable solution to a challenging issue by predicting flight prices using linear regression.
2. In order to provide useful suggestions for airline pricing strategies, it will investigate the linear correlations between various elements and their impact on travel prices.

Research Approach, Methods, and Plan

Research Approach:

1. Data Collection: Gather flight data containing relevant information such as flight date, departure and destination cities, airline, and seat class.
2. Feature Engineering and Data Processing: Clean and preprocess the data, handle missing values, encode categorical variables, and standardize numerical features.
3. Model Training and Validation: Train the linear regression model and evaluate its performance

using cross-validation and split testing.
4. Result Analysis and Conclusion: Analyze the model’s coefficients and performance to identify significant factors affecting flight prices.

Research Methods:

1. Linear Regression: To estimate the coefficients and forecast flight prices, create a conventional linear regression model using the least squares approach.
2. Model Evaluation: To evaluate the model's performance and prevent overfitting, use evaluation measures like MSE, R2, and cross-validation.
3. Analysis of the Results: Examine the regression coefficients to ascertain which variables most significantly affect airfare.

Research Plan:

1. Months 1: Review the literature, collect flight data, preprocess the data, and choose features.
2. Months 2-3: Develop and train the linear regression model, assess its effectiveness, and refine it.
3. Month 4: Compose the final report, provide an overview of the results, and make recommendations for further research.

References

[1] Abdella J A, Zaki N M, Shuaib K, et al. Airline ticket price and demand prediction: A survey[J]. Journal of King Saud University-Computer and Information Sciences, 2021, 33(4): 375-391.
[2] Panigrahi A, Sharma R, Chakravarty S, et al. Flight Price Prediction Using Machine Learning[C]//ACI@ ISIC. 2022: 172-178.
[3] Samunderu E, Farrugia M. Predicting customer purpose of travel in a low-cost travel environment—A Machine Learning Approach[J]. Machine Learning with Applications, 2022, 9: 100379.
[4] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.