**Flight Fare Prediction System**

**ABSTRACT:**

The Flight Fare Prediction System is a comprehensive tool designed to predict flight ticket costs with accuracy and give travellers useful information for more informed planning and selection. Predicting flight costs has gotten more difficult due to the airline industry's exponential growth and the complexity of pricing structures that is growing. To produce precise fare estimates, this method makes use of past flight data and machine learning techniques. The method makes use of an extensive dataset that contains historical flight prices together with other pertinent characteristics including departure times, airlines, travel dates, and destinations. Through the application of sophisticated machine learning techniques to analyse this data, the system picks up patterns and associations that allow it to estimate flight fares with confidence in the future. Combining several regression techniques is how the Flight Fare Prediction System operates.

**Key words:** Prediction, Accuracy, and Regression algorithm.

## I. INTRODUCTION

The Flight Fare Prediction System represents a machine learning initiative dedicated to estimating aircraft ticket costs using relevant features and historical data. This approach serves travellers, travel firms, and airlines, aiding them in projecting trip expenses for planning, budgeting, and making informed choices.

The project's objective lies in constructing a dependable machine-learning model for forecasting flight expenses by encompassing various factors such as travel class, airline, departure and arrival destinations, travel dates, and other relevant details. To train the algorithm effectively, an extensive dataset comprising historical flight data, including ticket pricing and related attributes, will be utilized.

Users of the Flight Fare Prediction System will gain access to a user-friendly interface where they can input their travel information and receive an estimated flight fare. The system will meticulously analyse input data and generate precise predictions through feature engineering, data preprocessing, and machine learning methodologies. Rigorous evaluation criteria will be applied to ensure the accuracy and reliability of the model and its associated attributes.

While the accuracy of predictions heavily relies on the quality of the training and prediction data, the project places a premium on data quality and integrity. Data preprocessing techniques such as data cleaning, handling missing values, and feature scaling will be employed to validate the legitimacy and trustworthiness of the data utilized for training and prediction purposes.

The Flight Fare Prediction System holds the potential to assist consumers in planning their travel budgets, aid travel agencies in offering competitive pricing to their clientele, and support airlines in devising effective pricing strategies and revenue management tactics. By leveraging machine learning, the system aims to offer valuable insights and advantages to the travel industry, precisely estimating airline fares and enhancing decision-making processes.

In conclusion, the Flight Fare Prediction System signifies a machine learning endeavor aimed at developing a system capable of accurately predicting trip expenses based on historical data and relevant attributes. This technology harbors the potential to enhance decision-making within the travel industry and offer valuable insights and benefits to travelers, travel agencies, and airlines alike

## II. RELATED WORK

In this Section, we will look at a few studies that demonstrate how machine learning is implemented in flight fare prediction systems.

A. "Airline Ticket Price Prediction: A Machine Learning Approach" by M. L. Ahirrao, et al. (2018): This research paper proposes a flight fare prediction model using machine learning techniques such as regression algorithms and time-series analysis. The study explores various factors influencing ticket prices and compares the performance of different algorithms in predicting fare trends.

Flight Fare Prediction using Historical Data and Machine Learning Techniques" by A. Kumar, et al. (2019): The paper presents a flight fare prediction system that combines historical flight data and machine learning algorithms to forecast ticket prices. It analyzes factors such as departure time, travel duration, and airline popularity to generate accurate fare predictions. The study compares the performance of different algorithms and discusses the potential for improving prediction accuracy.

B. "Airline Fare Prediction Using Machine Learning" by A. L. Rodrigues, et al. (2020): This work focuses on predicting airline fares using machine learning techniques. The study considers various parameters, including airline popularity, route distance, and historical fare data, to train a predictive model. The authors explore the performance of different algorithms and discuss the implications of their findings for fare prediction accuracy.

C. "Predicting Airfare Using Machine Learning Techniques" by S. Aruna, et al. (2020): The paper presents a comparative analysis of different machine learning algorithms for predicting airfare. The study considers factors such as seasonality, time of booking, and flight class to develop a prediction model. The authors evaluate the performance of regression algorithms, including linear regression, support vector regression, and random forest regression.

D. "Flight Fare Prediction Using Ensemble Learning Techniques” by M.Sharma, et al. (2021): This research focuses on the application of ensemble learning techniques for flight fare prediction. The study combines multiple machine learning models, including decision trees, random forests, and gradient boosting, to improve prediction accuracy. The authors compare the performance of individual models and ensemble methods to identify the most effective approach.

E. “Flight Fare Prediction using Machine Learning Techniques” by K. Kumar and team (2017). This study compares the performance of various machine learning techniques, including decision trees, support vector machines, k-nearest neighbors, and random forests for flight fare prediction. This study also employs features engineering techniques to extract relevant features from flight data and evaluates the models using metrics such as mean squared error (MSE) and R-squared.

**III. METHODOLOGY**

## The following steps are involved in the proposed architecture of our project.

***3.1. Data Collection***

The act of obtaining, acquiring, and combining the data that will be used to develop, test, and verify a machine learning model is known as data collection in machine learning. This step plays crucial role in implementation. Here data is collected from flight fare dataset which is imported from Kaggle. The dataset consists of both categorical data and numerical data. The categorical data includes source, destination, type of airline, additional info and numerical data includes arrival and departure dates, number of Stops. There are 11 columns (each represents a feature) and 10683 rows in this large dataset.

***3.2. Data Preprocessing***

Data preprocessing means nothing but cleaning data, which can be used for model training and testing. By this step we can make our data useful for model training purpose. Data preprocessing involves cleaning, transforming, and preparing the data for data analysis. The sub steps involved in the data preprocessing are:

**Data Cleaning:** In this step the null values are removed, missing values are removed and if any duplicates are present that are also removed.

**Feature selection and engineering:** In this step the features of our model are extracted and all the relevant features are used for model training. In dataset it contains date of journey, arrival date, departure date columns and all the numerical values are extracted as Departure hour, departure minutes, arrival hour, arrival minutes, journey day, journey month. As dataset contains both categorical and numerical features, by using 'On hot encoding' method for nominal categorical data and 'label encoding' for ordinal categorical data was used to convert the categorical values to numerical values. The dataset consists of categorical variables like airline, source, destination, route, total number of stops and additional info.

***3.3 Data Splitting***

This step involves splitting our data into two parts for training and testing purpose. For model training 80 percent of data was used by using Random Forest regressor model was trained. The machine learning algorithms are:

**LGBM Regressor**

LGBM stands for Light Gradient Boosted Machine. It is a gradient boosting framework based on decision trees that can be used for various machine learning tasks such as regression, classification and ranking. LGBM Regressor is a class in lightgbm package that can be used to train and predict regression models.

**Decision Tree Regressor**

Decision tree regressor is a class in Sklearn tree module that can be used to train and predict regression models. It is a decision tree-based algorithm that recursively partitions the input data based on the values of the input features, forming a tree-like structure. It initially chooses independent variable from dataset as decision nodes for decision making and then it divides the entire dataset into sub-sections and when test data is passed to the model the output is decided by checking the data point belongs to the decision tree will give output as the average value of all the datapoints in the sub-section.

**Randon Forest Regressor**

Random Forest regressor uses multiple decision trees to perform regression tasks. It is an example of ensemble learning. Random forest is a Supervised Learning algorithm which uses ensemble learning approach for classification and regression. Decision trees are sensitive to the specific data on which they are trained. If the training data is changed the resulting decision tree can be quite different and in turn the predictions can be quite different. Also, Decision trees are computationally expensive to train, carry a big risk of overfitting, and tend to find local optimal because they can’t go back after they have made a split to address these weaknesses, we turn to Random Forest.

***3.4 Model evaluation***

This is an important step in our project, as it helps us to measure the performance and accuracy of our model. Test data is used for model evaluation. Here, we employed Cross-validation for model evaluation. This method divides the data into k-subsets, called folds. the model is trained on k-1 folds and tested on the remaining fold. this process is repeated k times, so that each fold is used as a test set once. The average performance across all k-folds is reported as the final result. The metrics that are used for model evaluation purpose are:

**Root Mean Squared Error (RSME)**: It gives the root of the average squared difference between the actual values and the predicted values for a regression problem.

**Mean Absolute Error (MAE)**: It gives the absolute difference between the actual values and predicted values. The higher negative mean values indicate the better performance of model.

**R-Squared**: This metric measures how well the regression model fits the data, by comparing it to a baseline model that always predict the mean value. It shows how much variation in the data is explained by the model.

***3.5 Model Architecture***

The architecture is crafted with essential components to optimize prediction. Our proposed system analyses historical data to identify patterns, seasonal trends and additional information that influence flight fares.

## IV. RESULT AND FUTURE WORKS

### The developed machine learning model demonstrates high accuracy in predicting flight fares, with low mean absolute error (MAE) and root mean squared error (RMSE) values compared to baseline models. The user-friendly interface allows travellers to input their travel details and receive estimated flight fares quickly and efficiently. The flight fare prediction system provides tangible benefits such as cost savings, improved decision-making, and enhanced customer satisfaction for travellers, travel agencies, and airlines. Future enhancements may include integrating real-time data feeds, expanding coverage to include additional features, and ongoing maintenance to ensure continued effectiveness and relevance in the travel industry.

### V. REFERENCES

[1]. "Predicting Airline Ticket Prices Using Historical Data" Authors: R. B. Patterson and R. M. Patel Published in: International Journal of Computer Applications, Vol. 52, No. 5, 2012.

[2]. Flight Fare Prediction using Historical Data and Machine Learning Techniques" Authors: A. Kumar, et al. Published in: Proceedings of the 3rd International Conference on Computer, Communication, and Signal Processing, 2019.

[3]. "Flight Fare Prediction Using Machine Learning Techniques" Authors: S. G. Sonawane and A. N. Kadam Published in: International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 7, Issue 7, 2017.

[4]. "Airline Fare Prediction Using Machine Learning" Authors: A. L. Rodrigues, et al. Published in: Proceedings of the International Conference on Data Engineering and Communication Technology, 2020.

[5]. "Flight Fare Prediction Using Machine Learning Techniques" Authors: R. N. Sahoo, B. Mishra, and S. P. Dash Published in: Proceedings of the 3rd International Conference on Computational Intelligence in Data Science (ICCIDS), 2019. DOI: 10.1109/ICCIDS.2019.9010314

[6]. "Predicting Airfare using Machine Learning Algorithms" Authors: K. Gupta and S. Agarwal Published in: International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 9, Issue 3, 2019.

[7]. "Airfare Prediction with Deep Learning" Authors: A. Bansal, M. Shukla, S. Parthiban, and B. Sundarraj Published in: 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)

[8]. "Flight Price Prediction using Machine Learning Techniques" Authors: S. Chatterjee, A. Saini, and P. Sharma Published in: International Journal of Computer Applications, Vol. 172, No. 8, 2017. DOI: 10.5120/ijca2017913795.

[9]. "Predicting Flight Prices using Machine Learning Algorithms" Authors: S. Verma and A. Gupta Published in: International Journal of Computer Applications, Vol. 117, No. 15, 2015. DOI: 10.5120/ijca2015907272

[10]. "Airfare Prediction using Machine Learning Techniques" Authors: S. Kumar, V. Bajaj, and S. Joshi Published in: International Journal of Computer Applications, Vol. 91, No. 8, 2014. DOI: 10.5120/16084-9437

[11]. "Airline Ticket Fare Prediction using Machine Learning Techniques" Authors: A. Sharma and S. Verma Published in: 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) DOI: 10.1109/ICCCNT45670.2020.9226029

[12]. "Flight Fare Prediction using Machine Learning Techniques and Feature Selection" Authors: R. R. Saravana Kumar, M. S. Siva, and S. P. Shantharajah Published in: 2020 International Conference on Inventive Research in Computing Applications (ICIRCA) DOI:

10.1109/ICIRCA48642.2020.9169227

[13]. "Predicting Airfare using Machine Learning Algorithms" Authors: R. Varghese, S. Jacob, and S. J. Aby Published in: 2019 International Conference on Communication and Signal Processing (ICCSP) DOI: 10.1109/ICCSP.2019.8698027

[14]. "Flight Fare Prediction using Machine Learning Algorithms" Authors: M. M. Aboul Hassan, A. M. El-Sayed, and A. A. Fergany Published in: 2018 9th Cairo International Biomedical Engineering Conference (CIBEC) DOI: 10.1109/CIBEC.2018.8743157

[15]. "Airfare Prediction: A Deep Learning Approach" Authors: H. Nguyen, A. Nguyen, and K. Nguyen Published in: 2018 International Conference on Advanced Computing and Applications (ACOMP) DOI: 10.1109/ACOMP.2018.00008

[16]. "Airfare Prediction System using Machine Learning Techniques" Authors: N. Arifin, S. Hartono, and N. P. D. Simamora Published in: 2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE) DOI: 10.1109/ICITISEE.2017.8256920

[17]. "Predicting Flight Fares using Machine Learning Techniques" Authors: P. Kaushik and M. Kaur Published in: 2016 International Conference on Computing, Communication and Automation (ICCCA) DOI: 10.1109/CCAA.2016.7813906

[18]. "Machine Learning Approaches for Airfare Prediction" Authors: M. Gupta and A. Goyal Published in: 2015 International Conference on

Green Computing and Internet of Things (ICGCIoT) DOI: 10.1109/ICGCIoT.2015.7380513