International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Machine Learning-Based Flight Price Prediction System Using Linear Regression and GUI Integration

VOLIPILLI GOWRI MALLIKA

PG Scholar. Department of MCA, DNR College, Bhimavaram, Andhra Pradesh

A. Naga Raju

(Assistant Professor), Master of Computer Applications, DNR College, Bhimavaram, Andhra Pradesh

ABSTRACT

The rapid growth of the aviation industry has led to dynamic and highly volatile flight pricing mechanisms influenced by multiple factors such as demand, seasonality, airline competition, and route characteristics. Predicting flight ticket prices accurately has become a crucial requirement for both customers seeking affordable travel and airlines aiming to optimize revenue. This project presents a machine learning-based Flight Price Prediction System that leverages historical flight data to estimate ticket prices efficiently. The proposed system uses Linear Regression, a supervised learning algorithm, to model the relationship between various input features and the flight price. The dataset is preprocessed using techniques such as label encoding for categorical variables and feature scaling for numerical consistency. Important features such as airline, source, destination, journey date, duration, and number of stops are extracted and transformed into meaningful numerical representations. These features are then used to train the regression model. To enhance usability, the system integrates a Graphical User Interface (GUI) built using Tkinter. This interface allows users to input flight details through dropdown menus and text fields, making the system interactive and user-friendly. Upon submission, the system processes the input data, applies the trained model, and displays the predicted price instantly.

The system demonstrates how machine learning can be effectively applied to real-world problems involving dynamic pricing. While Linear Regression provides a simple and interpretable baseline model, the system architecture allows for future enhancements using advanced algorithms such as Random Forests or Gradient Boosting. Overall, this project highlights the importance of predictive analytics in the travel domain and provides a scalable solution that can be extended with real-time data and improved models. The integration of machine learning with a GUI ensures accessibility for non-technical users, making the solution practical and impactful.

KEYWORDS: Flight Price Prediction, Machine Learning, Linear Regression, Tkinter GUI, Data Preprocessing, Feature Engineering, Airline Pricing, Predictive Analytics

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

I. INTRODUCTION

In recent years, the airline industry has witnessed significant advancements driven by data analytics and intelligent systems. One of the most challenging aspects of this industry is the fluctuation of flight ticket prices, which depend on various dynamic factors such as booking time, travel season, route popularity, and airline policies. These fluctuations make it difficult for passengers to determine the optimal time to purchase tickets. Flight price prediction has emerged as an important application of machine learning, enabling users to make informed decisions and airlines to implement efficient pricing strategies. Traditional methods rely heavily on manual analysis or static pricing models, which fail to capture the complexity and dynamic nature of pricing trends. Therefore, there is a need for automated systems that can analyze large datasets and generate accurate predictions. This project focuses on developing a Flight Price Prediction System using Linear Regression, a widely used machine learning algorithm for regression problems. The system utilizes historical flight data and extracts relevant features such as airline type, source and destination locations, journey date, flight duration, and number of stops. These features are processed and fed into the model to learn patterns and relationships that influence ticket prices.

In addition to predictive modeling, the system emphasizes usability through a graphical interface developed using Tkinter. This interface allows users to interact with the system without requiring technical expertise. By simply selecting options and entering basic details, users can obtain an estimated flight price instantly. The significance of this project lies in its practical application and scalability. It demonstrates how machine learning techniques can be integrated with user-friendly interfaces to create intelligent systems. Furthermore, it lays the foundation for future enhancements such as incorporating real-time data, improving model accuracy, and deploying the system as a web or mobile application. In summary, this project aims to bridge the gap between complex data-driven models and end-user accessibility, providing a reliable and efficient solution for flight price prediction.

II. LITERATURE SURVEY (WITH EXISTING METHODS)

Flight price prediction has been widely studied in the field of machine learning and data analytics due to its practical importance and complexity. Several researchers have explored different techniques to model and predict airline ticket prices with varying degrees of success. Early approaches relied on statistical methods such as linear regression and time-series analysis. These methods provided a baseline understanding of pricing trends but were limited in capturing nonlinear relationships between variables. Linear regression, despite its simplicity, has been widely used due to its interpretability and efficiency in handling structured datasets. With the advancement of machine learning, more sophisticated models such as Decision Trees, Random Forests, and

Support Vector Machines (SVM) have been applied to this problem. Random Forest, in particular, has shown significant improvements in prediction accuracy by combining multiple decision trees and reducing overfitting. Similarly, Gradient Boosting algorithms like XGBoost have been used to achieve high performance by iteratively improving prediction errors.

Deep learning approaches, including Artificial Neural Networks (ANNs), have also been explored for flight price prediction. These models are capable of capturing complex nonlinear patterns and interactions among features. However, they require large datasets and significant computational resources, making them less suitable for simpler implementations. Feature engineering plays a crucial role in improving model performance. Researchers have emphasized the importance of extracting meaningful features such as travel date, booking window, airline popularity, and route demand. Encoding categorical variables and scaling numerical features are essential preprocessing steps that enhance model efficiency. Recent studies have also focused on integrating prediction systems with user interfaces and real-time data sources. Web-based and mobile applications have been developed to provide users with instant price predictions and recommendations. These systems often use APIs to fetch live data and continuously update models. Despite these advancements, challenges remain in achieving high accuracy due to the dynamic nature of airline pricing. Factors such as sudden demand changes, weather conditions, and airline policies introduce unpredictability. The current project builds upon these existing methods by implementing a Linear Regression model combined with effective preprocessing techniques and a user-friendly GUI. While it may not achieve the highest accuracy compared to advanced models, it provides a simple, interpretable, and efficient solution suitable for demonstration and practical use.

III. EXISTING SYSTEM

The existing systems for flight price prediction primarily rely on manual analysis or basic online tools that provide limited insights into ticket pricing trends. Most travel websites display current prices without offering predictive capabilities, leaving users uncertain about whether prices will increase or decrease in the future. Some platforms use rule-based systems or simple heuristics to suggest booking times. These systems often consider factors such as seasonal demand or historical averages but lack the ability to adapt to complex and dynamic pricing patterns. As a result, their predictions are often inaccurate and unreliable. In recent years, machine learning-based systems have been introduced to address these limitations. These systems use algorithms such as Random Forests, Support Vector Machines, and Neural Networks to predict prices based on historical data. While they provide improved accuracy, they often require large datasets, high computational power, and complex implementation processes.

Another limitation of existing systems is the lack of user-friendly interfaces. Many machine learning models are developed as backend systems without proper integration into interactive applications. This makes them less accessible to non-technical users. Additionally, existing systems may not handle data preprocessing effectively. Issues such as missing values, inconsistent formats, and unscaled features can significantly impact model performance. The proposed system addresses these challenges by combining a simple yet effective Linear Regression model with proper data preprocessing techniques and an intuitive GUI. This approach ensures ease of use, faster computation, and reasonable prediction accuracy, making it suitable for practical applications and academic purposes.

IV. PROPOSED METHOD

The proposed system is a machine learning-based Flight Price Prediction System designed to estimate airline ticket prices based on user-provided inputs. The system leverages historical flight data and applies regression techniques to identify patterns and relationships between various influencing factors such as airline, source, destination, journey date, duration, and number of stops. The core of the system is built using the Linear Regression algorithm, which models the relationship between dependent and independent variables. The dataset undergoes preprocessing steps including handling categorical variables using label encoding and normalizing numerical values using feature scaling techniques. These preprocessing steps improve model performance and ensure consistency in predictions. A key feature of the proposed system is the integration of a Graphical User Interface (GUI) using Tkinter. The GUI allows users to interact with the system easily by selecting options from dropdown menus and entering relevant flight details. Once the user inputs the data, the system processes it, applies the trained model, and outputs the predicted flight price instantly.

Unlike traditional systems that rely on static pricing or manual analysis, this system provides dynamic predictions based on learned data patterns. It also ensures faster computation and accessibility for non-technical users. The modular design of the system allows for future enhancements such as integration with real-time APIs, advanced machine learning models like Random Forest or Gradient Boosting, and deployment as a web-based application. The proposed system aims to provide an efficient, accurate, and user-friendly solution for predicting flight prices, thereby assisting users in making informed travel decisions.

V. IMPLEMENTATION

The implementation of the Flight Price Prediction System involves several stages, including data collection, preprocessing, model training, and GUI integration. Initially, the dataset is loaded using the Pandas library. The dataset contains important attributes such as airline, source, destination, date of journey, duration, and price. Data preprocessing is performed to clean and transform the dataset into a suitable format for model training. Categorical variables such as airline, source, and destination are encoded into numerical values using LabelEncoder. Additionally, the date of journey is split into separate features such as journey day and month to extract meaningful temporal information. After preprocessing, the dataset is divided into input features (X) and target variable (y), where the target variable represents the flight price. The dataset is further split into training and testing sets using train_test_split to evaluate the model's performance effectively. Feature scaling is applied using StandardScaler to normalize the data, ensuring that all features contribute equally to the model. This step is particularly important for regression models as it improves convergence and accuracy.

The Linear Regression model is then trained using the training dataset. The model learns the relationship between input features and the target variable by minimizing the error between predicted and actual values. Once trained, the model is tested using the test dataset to evaluate its performance. The next phase involves developing a user-friendly GUI using Tkinter. The GUI includes dropdown menus for selecting airline, source, and destination, as well as input fields for duration, journey day, month, and number of stops. A "Predict" button triggers the prediction process. When the user inputs the required data and clicks the button, the system collects the input values, preprocesses them (including scaling), and feeds them into the trained model. The predicted price is then displayed on the interface in a formatted manner. Error handling mechanisms are included to ensure that invalid inputs are managed effectively. The system displays appropriate error messages when incorrect values are entered. The implementation demonstrates a complete pipeline from data preprocessing to model deployment in a GUI environment, making it a practical and user-centric application.

VI. ALGORITHMS

The primary algorithm used in this system is Linear Regression, a supervised machine learning technique used for predicting continuous values. Linear Regression establishes a linear relationship between dependent and independent variables by fitting a line that minimizes the error between predicted and actual values.

The mathematical representation of Linear Regression is:

Y=b0+b1X1+b2X2+...+bnXnY = b_0 + b_1X_1 + b_2X_2 + ... + b_nX_nY=b0+b1X1 +b2X2+...+bnXn

where YYY is the predicted value, b0b_0b0 is the intercept, b1,b2,...,bnb_1, b_2, ..., b_nb1,b2,...,bn are coefficients, and X1,X2,...,XnX_1, X_2, ..., X_nX1,X2,...,Xn are input features.

In this system, features such as airline, source, destination, duration, journey day, month, and stops are used as independent variables, while the flight price is the dependent variable. The algorithm works by minimizing the Mean Squared Error (MSE), which measures the difference between actual and predicted values. The objective is to find the optimal coefficients that reduce this error. In addition to Linear Regression, preprocessing algorithms play a crucial role. Label Encoding is used to convert categorical variables into numerical form, enabling the model to process them effectively. StandardScaler is used to normalize feature values so that they have a mean of zero and standard deviation of one.

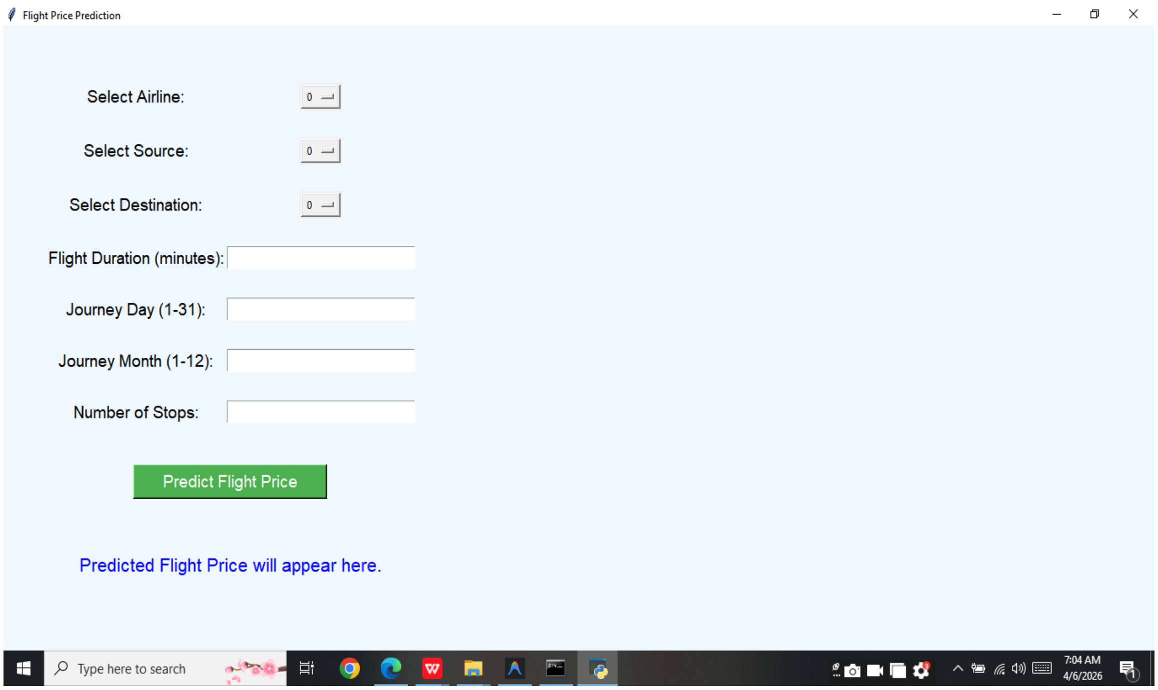
Other algorithms that can be used for improvement include Decision Trees, Random Forest, and Gradient Boosting, which can capture nonlinear relationships more effectively. Studies show that advanced models like Random Forest and deep learning techniques such as LSTM and GRU can significantly improve prediction accuracy in complex datasets. Thus, while Linear Regression provides a simple and interpretable baseline, the system can be extended with more sophisticated algorithms for better performance.

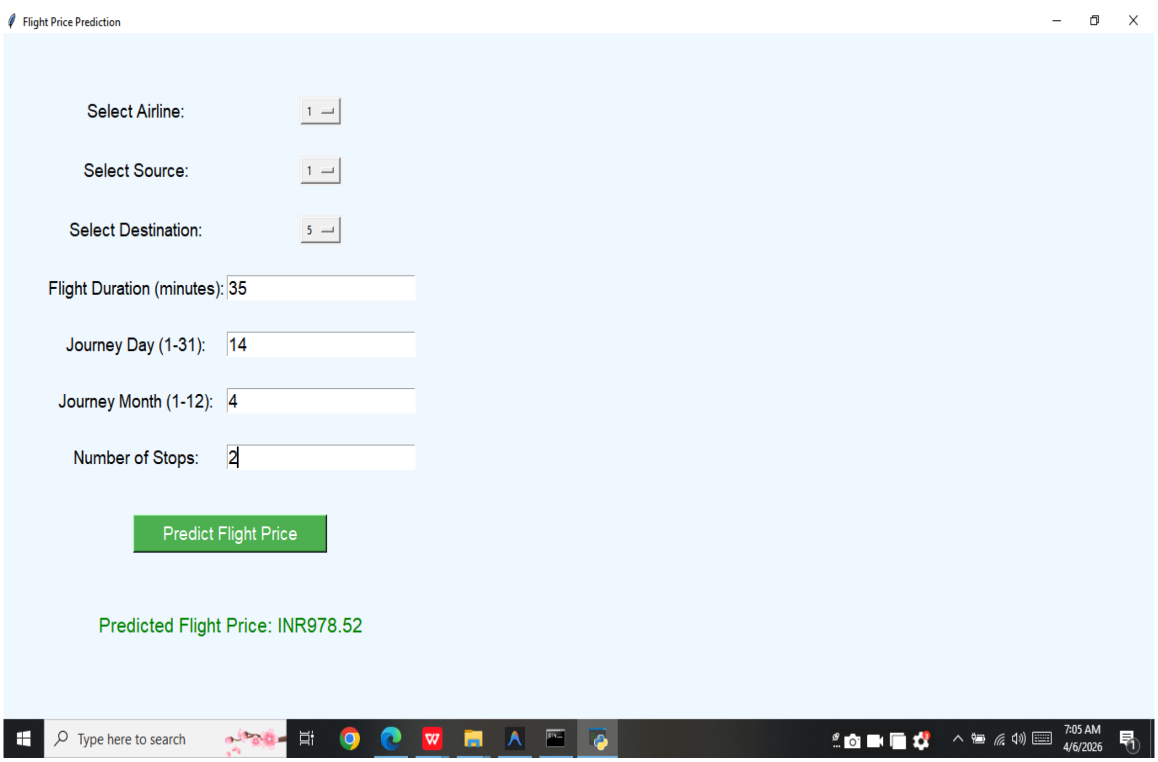
VII. SYSTEM DESIGN

The system design of the Flight Price Prediction System follows a modular architecture consisting of data processing, model training, prediction engine, and user interface components. The first component is the Data Processing Module. This module is responsible for loading the dataset, handling missing values, encoding categorical variables, and extracting relevant features such as journey day and month. Proper preprocessing ensures that the data is clean and suitable for machine learning algorithms. The second component is the Model Training Module. In this module, the processed data is split into training and testing sets. Feature scaling is applied to normalize the data. The Linear Regression model is then trained using the training dataset. The trained model is stored and used for making predictions. The third component is the Prediction Engine. This module takes user input from the GUI, preprocesses it using the same transformations applied during training, and feeds it into the trained model. The model generates a predicted flight price, which is then displayed to the user.

The fourth component is the User Interface Module. The GUI is developed using Tkinter and provides an interactive platform for users to input flight details. The interface includes dropdown menus, text fields, and buttons, making it easy to use even for non-technical users. The system follows a pipeline architecture where data flows sequentially from input to output. The interaction between components ensures efficient processing and accurate predictions. From a design perspective, the system emphasizes simplicity, modularity, and scalability. Each module operates independently, allowing for easy updates and enhancements. For example, the Linear Regression model can be replaced with more advanced algorithms without affecting the GUI. The system design also considers performance and usability. By using efficient preprocessing techniques and lightweight algorithms, the system ensures fast response times. The GUI enhances user experience by providing clear instructions and immediate results. Overall, the system design provides a robust framework for implementing machine learning-based prediction systems in real-world applications.

SYSTEM DESIGN IMAGES





International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

VIII. CONCLUSION

The Flight Price Prediction System demonstrates the practical application of machine learning techniques in solving real-world problems. By leveraging historical flight data and applying Linear Regression, the system provides an efficient method for predicting airline ticket prices based on various influencing factors. One of the key strengths of the system is its simplicity and interpretability. Linear Regression allows users to understand how different features contribute to price prediction. The integration of preprocessing techniques such as label encoding and feature scaling ensures that the model performs effectively and produces reliable results. The inclusion of a user-friendly GUI enhances the usability of the system, making it accessible to users without technical expertise. This feature bridges the gap between complex machine learning models and end users, enabling practical adoption. However, the system also has certain limitations. Linear Regression may not capture complex nonlinear relationships in the data, which can affect prediction accuracy.

Future improvements can include the use of advanced algorithms such as Random Forest, Gradient Boosting, and deep learning models, which have been shown to provide better performance in airfare prediction tasks. Additionally, integrating real-time data and deploying the system as a web or mobile application can further enhance its functionality and reach. In conclusion, the project successfully demonstrates how machine learning can be combined with user interfaces to create intelligent and practical applications. It provides a strong foundation for further research and development in predictive analytics within the aviation domain.

REFERENCES

1. Meepganithage, A. et al., “Airfare Forecasting: A Deep Learning Approach,” 2024.
2. Korkmaz, H., “Prediction of Airline Ticket Price Using ML,” 2024.
3. Lal, V. et al., “Predicting Flight Fares in India,” 2023.
4. Degife, W. A., “GRU Model for Flight Fare Forecasting,” 2023.
5. Sruneethi, C., “Flight Price Prediction Using ML,” 2025.
6. Bollack, J. et al., “Using ML Algorithms for Flight Prices,” 2023.
7. Panda, B. et al., “Flight Price Prediction Techniques,” 2022.
8. Sarao, P., “Flight Fare Prediction Using ML,” 2022.

9. Gupta, S., “Flight Fare Prediction Analysis,” 2024.

10. Wong, P. et al., “Predictive Analysis Using Spark ML,” 2023.

11. Dai, M., “Flight Delay Prediction with ML,” 2024.

12. Chaleshtori, A., “Elastic Net for Price Prediction,” 2024.

13. Jha, R. K., “Hybrid ML for Flight Prediction,” 2024.

14. IJRCSEIT Journal, “Flight Price Prediction Research,” 2025.

15. MDPI Applied Sciences, “AI in Airfare Forecasting,” 2023.