

Research Paper

FLIGHT TICKET PRICE PREDICTOR USING MACHINE LEARNING

BONALA JALA

SCHOLAR, DEPARTMENT OF MCA

VAAGESWARI COLLEGE OF ENGINEERING

(S4), KARIMNAGAR

MR,K. NAGENDRA PRASAD

ASST. PROFESSOR DEPT OF MCA

VAAGESWARI COLLEGE OF ENGINEERING

(S4), KARIMNAGAR

DR. P. VENKATESHWARLU

Professor & Head, Department of MCA

Vaageswari College of Engineering(S4), Karimnagar

(Affiliated to JNTUH, Approved by AICTE, New Delhi & Accredited by NAAC with 'A+' Grade)

Karimnagar, Telangana, India – 505 527

Abstract

Flight ticket prices vary significantly depending on factors such as airline, source and destination, departure date, journey duration, number of stops, departure time, and the number of days remaining before the journey. Due to these variations, passengers may find it difficult to determine the appropriate time to purchase flight tickets at a lower price.

The Flight Ticket Price Predictor is a machine-learning-based system designed to predict the approximate price of a flight using historical flight-ticket data. The system collects important flight-related attributes, preprocesses the dataset, performs feature transformation, and trains a machine-learning regression model to learn the relationship between flight characteristics and ticket prices.

The proposed system can use algorithms such as Linear Regression, Decision Tree Regression, Random Forest Regression, Gradient Boosting, or other suitable regression techniques. The trained model receives flight details from the user and predicts the expected ticket price.

The system can provide a simple user interface where users enter the airline, source, destination, journey date, departure time, arrival time, duration, and number of stops. The prediction model processes these features and generates an estimated ticket price. Performance can be evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 score.

The proposed system therefore combines machine learning, data preprocessing, feature engineering, and regression techniques to develop an automated flight ticket price prediction system.

Keywords: Flight Price Prediction, Machine Learning, Regression, Data Preprocessing, Feature Engineering, Random Forest, Ticket Price, Predictive Analytics.

Introduction

Flight ticket prices are dynamic and may change frequently based on several factors. Airline companies may adjust prices according to demand,

availability, travel date, route, season, and other operational factors.

Passengers generally have to search through different flight-booking platforms to compare

prices. It may be difficult for users to understand whether the current ticket price is relatively high or low compared with historical patterns.

Machine learning provides an approach for learning patterns from previously observed flight-ticket data. Historical datasets contain information such as airline, source, destination, departure time, arrival time, duration, number of stops, and ticket price.

The Flight Ticket Price Predictor uses these attributes to train a regression model. After training, the model can accept new flight details and estimate the expected ticket price.

The overall process includes:

Flight Dataset → Data Preprocessing → Feature Engineering → Model Training → Model Evaluation → Price Prediction

Problem Statement

Flight ticket prices depend on multiple factors and can change over time. Users may not have sufficient information to estimate whether a particular ticket price is reasonable.

Traditional flight-booking systems mainly display the current available prices but do not necessarily provide a machine-learning-based prediction of the expected price based on historical patterns.

The problem can therefore be formulated as developing a machine-learning system that can learn from historical flight-price data and predict the approximate ticket price for a given set of flight characteristics.

The proposed system should achieve:

- Accurate flight price prediction.
- Effective preprocessing of flight data.
- Identification of important price-related features.
- Handling of categorical and numerical attributes.
- Reduced prediction error.
- Easy interaction through a user interface.
- Fast generation of predicted prices.

Existing System

Existing flight-booking platforms allow users to search for flights by specifying the source, destination, travel date, and other preferences.

The system displays available flights and their current ticket prices. Users can compare airlines, timings, duration, and number of stops before selecting a flight.

However, conventional booking systems primarily focus on displaying currently available prices. They may not provide a dedicated predictive model that estimates the expected price from historical flight data.

A machine-learning-based prediction system can complement this process by learning historical price patterns and generating an estimated price for a new flight configuration.

Disadvantages of Existing System

The major limitations of conventional flight-price searching systems include:

1. No Historical Price Prediction

Users may only see current prices without an estimated historical-based price.

2. Price Uncertainty

Passengers may not know whether the displayed price is relatively high or low.

3. Manual Comparison

Users may need to manually compare different airlines, routes, and travel dates.

4. Dynamic Pricing

Flight prices can change based on demand and availability.

5. Complex Price Factors

Many attributes can influence ticket prices, making manual price estimation difficult.

6. Limited Predictive Analysis

Traditional booking interfaces generally focus on flight search and booking rather than machine-learning-based prediction.

Proposed System

The proposed **Flight Ticket Price Predictor** uses machine-learning algorithms to estimate flight ticket prices from historical data.

The system first collects a dataset containing flight information and corresponding ticket prices. The dataset is cleaned and transformed before model training.

Categorical features such as airline, source, destination, and journey information are converted into machine-readable representations. Numerical features such as duration and number of stops are processed appropriately.

After preprocessing, a regression algorithm is trained using the historical data.

During prediction, the user enters flight details such as:

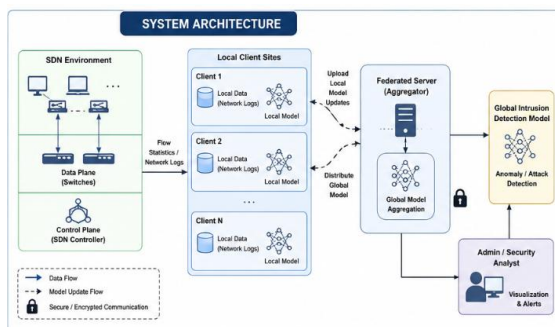
- Airline
- Source
- Destination
- Journey date
- Departure time
- Arrival time
- Duration
- Number of stops

The trained model processes these features and predicts the estimated flight ticket price.

Basic Flow:

User Input → Data Preprocessing → Trained ML Model → Price Prediction → Display Result

System Architecture



Proposed Methodology

The methodology consists of the following sequential stages:

Step 1: Data Collection

Historical flight-ticket data is collected for training the prediction model.

The dataset may contain attributes such as:

- Airline
- Date of Journey
- Source
- Destination
- Route
- Departure Time
- Arrival Time
- Duration
- Total Stops
- Additional Information
- Price

The ticket price is treated as the target variable.

Step 2: Data Preprocessing

The collected dataset may contain missing values, duplicate records, inconsistent formats, or unnecessary attributes.

The preprocessing stage includes:

- Removing duplicate records.
- Handling missing values.
- Removing irrelevant attributes.
- Converting data types.
- Cleaning inconsistent values.

Proper preprocessing improves the quality of the training data.

Step 3: Feature Engineering

Useful features are extracted from the available flight information.

For example, journey date can be divided into:

- Day
- Month
- Year

Departure and arrival times can be converted into meaningful numerical features such as:

- Departure hour
- Departure minute
- Arrival hour
- Arrival minute

Flight duration can also be transformed into numerical values.

This allows the machine-learning model to process the information effectively.

Step 4: Categorical Feature Encoding

Features such as airline, source, and destination are categorical.

Machine-learning algorithms generally require numerical input. Therefore, categorical values can be transformed using suitable encoding techniques.

Examples include:

- Label Encoding
- One-Hot Encoding

The encoded features are then provided to the prediction model.

Step 5: Train-Test Split

The processed dataset is divided into training and testing sets.

For example:

Training Data → Model Learning

Testing Data → Model Evaluation

The training data is used to learn the relationship between flight features and price, while the testing data is used to evaluate prediction performance.

Federated Learning Algorithm

The methodology consists of the following sequential stages:

Step 1: Data Collection

Historical flight-ticket data is collected for training the prediction model.

The dataset may contain attributes such as:

- Airline
- Date of Journey
- Source
- Destination
- Route

- Departure Time
- Arrival Time
- Duration
- Total Stops
- Additional Information
- Price

The ticket price is treated as the target variable.

Step 2: Data Preprocessing

The collected dataset may contain missing values, duplicate records, inconsistent formats, or unnecessary attributes.

The preprocessing stage includes:

- Removing duplicate records.
- Handling missing values.
- Removing irrelevant attributes.
- Converting data types.
- Cleaning inconsistent values.

Proper preprocessing improves the quality of the training data.

Step 3: Feature Engineering

Useful features are extracted from the available flight information.

For example, journey date can be divided into:

- Day
- Month
- Year

Departure and arrival times can be converted into meaningful numerical features such as:

- Departure hour
- Departure minute
- Arrival hour
- Arrival minute

Flight duration can also be transformed into numerical values.

This allows the machine-learning model to process the information effectively.

Step 4: Categorical Feature Encoding

Features such as airline, source, and destination are categorical.

Machine-learning algorithms generally require numerical input. Therefore, categorical values can be transformed using suitable encoding techniques.

Examples include:

- Label Encoding
- One-Hot Encoding

The encoded features are then provided to the prediction model.

Step 5: Train-Test Split

The processed dataset is divided into training and testing sets.

For example:

Training Data → Model Learning

Testing Data → Model Evaluation

The training data is used to learn the relationship between flight features and price, while the testing data is used to evaluate prediction performance.

System Implementation

The proposed system can be implemented using Python and machine-learning libraries.

Python can be used for:

- Data preprocessing
- Feature engineering
- Model training
- Model evaluation
- Prediction
- Visualization

Libraries such as Pandas and NumPy can be used for data processing, while Scikit-learn can be used for machine-learning algorithms and evaluation metrics.

A web-based interface can be developed to allow users to enter flight details and obtain predictions.

The implementation can contain the following modules:

1. Dataset Upload
2. Data Preprocessing
3. Feature Engineering
4. Feature Encoding
5. Model Training

6. Model Evaluation
7. Price Prediction
8. Prediction History
9. User Interface
10. Result Visualization

Experimental Setup

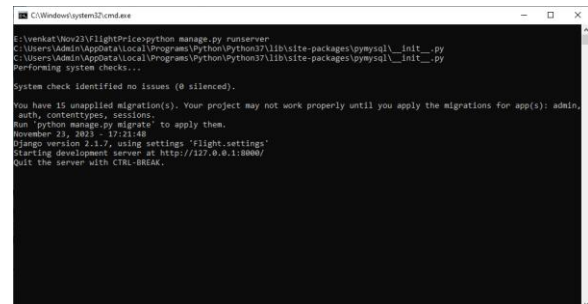
The system can be evaluated using a historical flight-price dataset containing different airlines, routes, journey dates, durations, and ticket prices.

The dataset can be divided into training and testing subsets.

An experimental configuration may use:

Parameter	Configuration
Programming Language	Python
Data Processing	Pandas, NumPy
Machine Learning	Scikit-learn
Dataset	Historical Flight Price Dataset
Problem Type	Regression
Training Data	80%
Testing Data	20%
Evaluation	MAE, MSE, RMSE, R ²
Candidate Models	Linear Regression, Decision Tree, Random Forest, Gradient Boosting

The models can be trained and evaluated using the same testing dataset to compare their prediction errors.

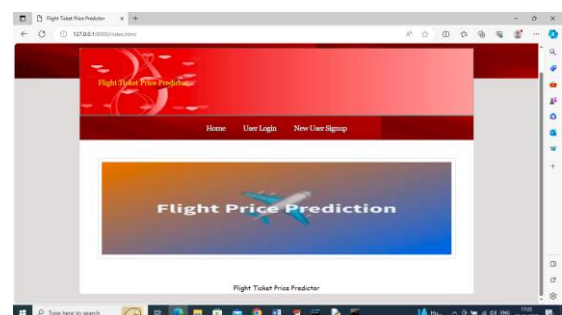


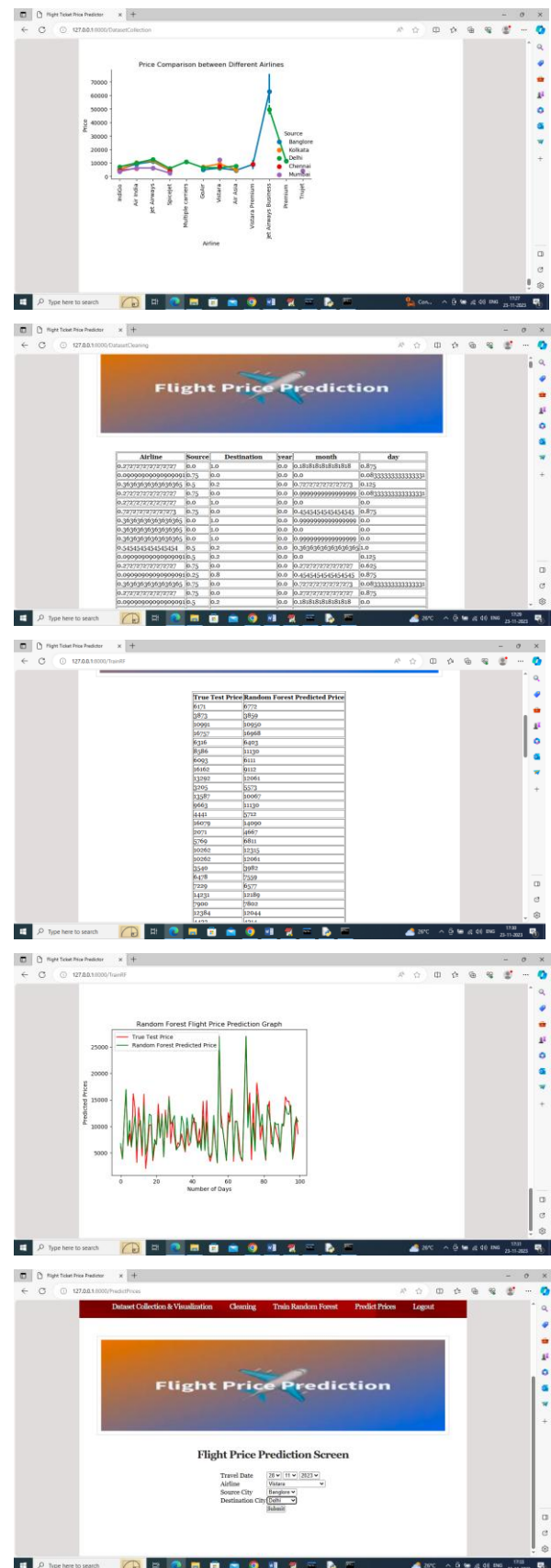
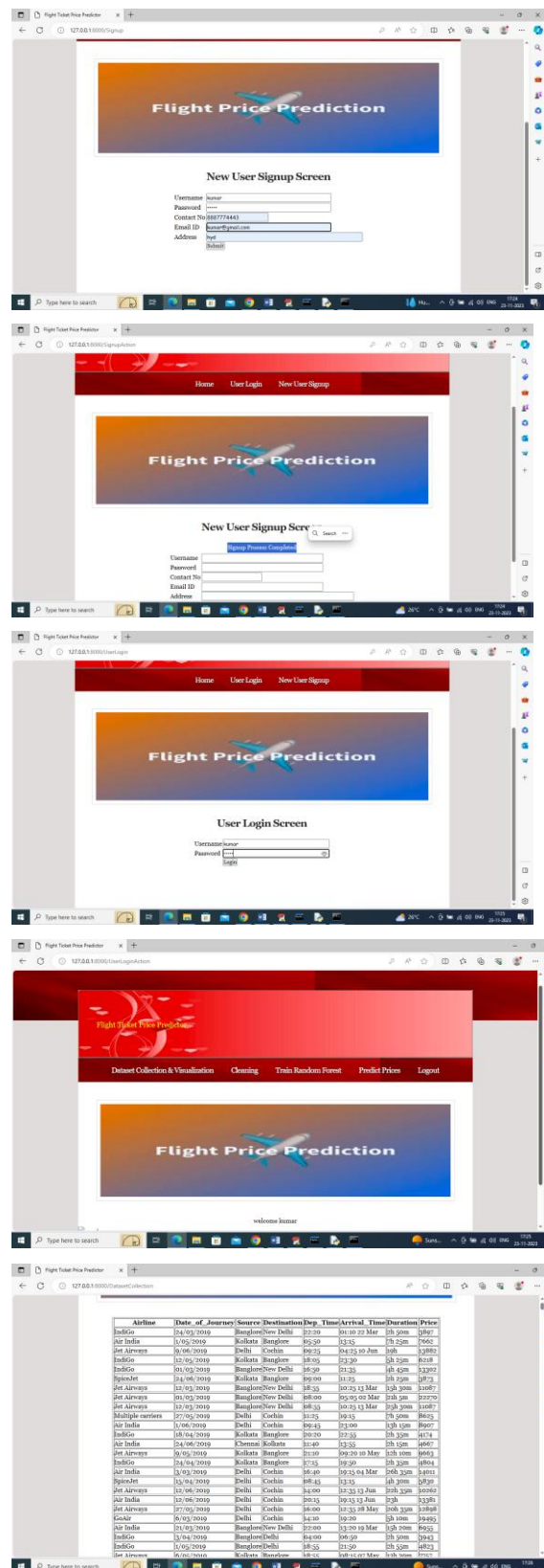
```

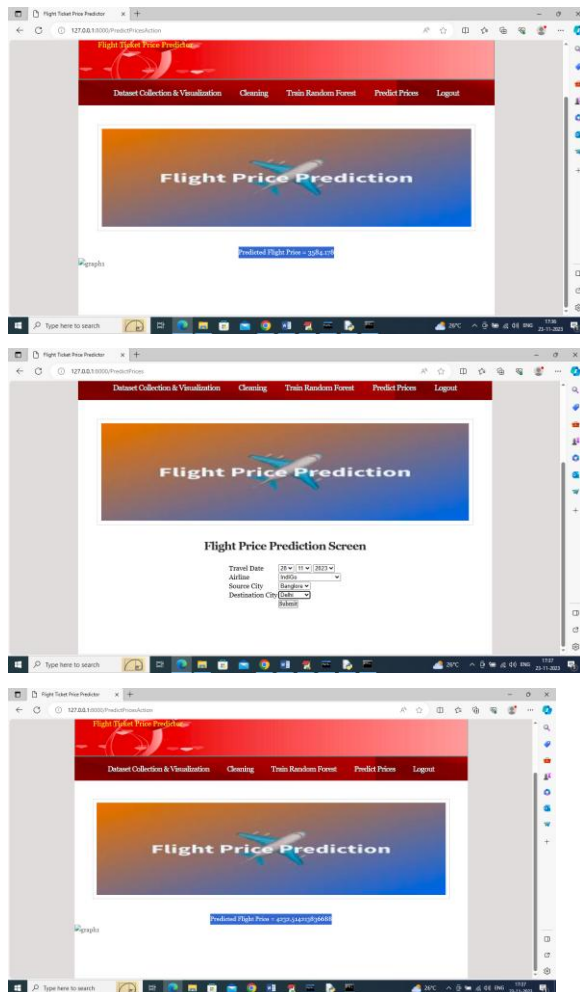
C:\Windows\system32\cmd.exe
E:\venkat\Wov23\FlightPrice>python manage.py runserver
C:\Users\Admin\AppData\Local\Programs\Python\Python37\lib\site-packages\pymysql\__init__.py
C:\Users\Admin\AppData\Local\Programs\Python\Python37\lib\site-packages\pymysql\__init__.py
Performing system checks...

System check identified no issues (0 silenced).

You have 15 unapplied migration(s). Your project may not work properly until you apply the migrations for app(s): admin,
auth, contenttypes, sessions.
Run 'python manage.py migrate' to apply them.
November 23, 2021, 17:21:48
Django version 2.1.7, using settings 'flight.settings'
Starting development server at http://127.0.0.1:8000/
Quit the server with CTRL-BREAK.
  
```







Future Work

Future improvements can enhance the Flight Ticket Price Predictor in several directions.

1. Real-Time Flight Data

The system can be integrated with live flight-price APIs to obtain current market information.

2. Price Trend Prediction

The system can be extended to analyze price trends over time.

3. Deep Learning

Advanced neural-network models can be investigated for complex price patterns.

4. Recommendation System

The system can provide alternative flight combinations based on the user's budget and preferences.

5. Multi-City Prediction

The system can be extended to support multiple destinations and complex travel routes.

6. Mobile Application

A mobile application can be developed so users can access predictions through smartphones.

7. Explainable Predictions

Explainable AI techniques can be incorporated to show which factors contributed significantly to the predicted ticket price.

8. Real-Time Price Monitoring

The system can be extended to monitor price changes and provide updated predictions when new flight-price information becomes available.

Conclusion

The Flight Ticket Price Predictor provides a machine-learning-based approach for estimating flight ticket prices using historical flight information. The proposed system preprocesses flight data, performs feature engineering and encoding, trains regression models, and generates an estimated price for new flight configurations.

By considering factors such as airline, source, destination, journey date, departure time, arrival time, duration, and number of stops, the system can learn patterns associated with flight-ticket prices.

The proposed approach combines data preprocessing, machine learning, regression analysis, and predictive analytics to create an automated flight price prediction system. The final performance can be measured using standard regression metrics such as MAE, MSE, RMSE, and R^2 score.

REFERENCES

1. Lu, J. (2017). *Machine Learning Modeling for Time Series Problem: Predicting Flight Ticket Prices*. arXiv.
2. Tziridis, K., Kalogeropoulos, A., Papakostas, G. A., and others. (2019). *Airline Ticket Price and Demand Prediction: A Survey*. Journal of

- King Saud University – Computer and Information Sciences.
3. Bhargavi, K., Sai, A. L., & Rao, K. T. (2022). *Airlines Ticket Price Prediction Using Machine Learning Approach*. International Journal of Computer Sciences and Engineering, 10(2). DOI: 10.26438/ijcse/v10i2.4953.
 4. Wong, P., Thant, P., Yadav, P., Antaliya, R., & Woo, J. (2023). *Using Spark Machine Learning Models to Perform Predictive Analysis on Flight Ticket Pricing Data*. arXiv.
 5. Vaishnavi, K. D. V. N., Bindu, L. H., Satwika, M., Udaya Lakshmi, K., Harini, M., & Ashok, N. (2023). *Flight Fare Prediction Using Machine Learning*. EPRA International Journal of Research & Development, 8(10).
 6. Nagagopiraju, V., Umamaheswari, S., Sailakshmi, M., Hussain, M. Y., & Ahmad, S. N. (2025). *Machine Learning-Based Airline Ticket Price Prediction*. International Journal of Communication Networks and Information Security, 17(3), 176–181.
 7. Ertürk, M., Emeç, M., Atılğan Sarıdoğan, A., & Küçükgergerli, N. (2025). *Large-Scale Airline Ticket Price Prediction Using Ensemble Machine Learning Algorithms*. Nişantaşı University Social Sciences Journal, 13(1).
 8. *Flight Fare Prediction Using Machine Learning*. (2025). 2025 2nd International Conference on Advanced Computing and Emerging Technologies (ACET), IEEE. DOI: 10.1109/ACET67282.2025.11430271.