Maharshi Karve Stree Shikshan Samstha's

Cummins College of Engineering for Women, Pune

(An Autonomous Institute affiliated to Savitribai Phule Pune University)

TY AIML PRACTICAL EXAMINATION-2025

Aritificial Intelligence and Machine Learning Laboratory

23PCCE501L

Airline Fare Dynamics Modeling and Prediction Through Machine Learning

Presented By:

1.Nutan Keshatwar - UCE2023434

2.Pallavi More-UCE2023447

3.Samruddhi Jagtap-UCE2023459

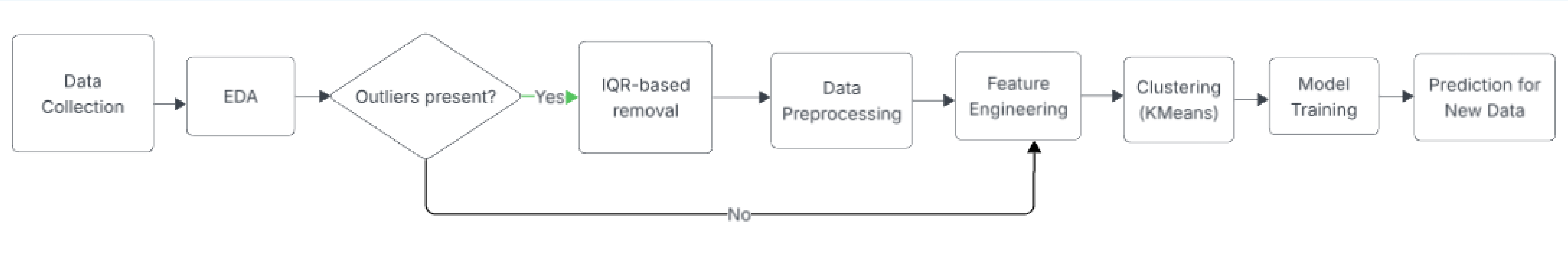
PROBLEM STATEMENT

- Airline ticket prices fluctuate frequently due to demand, timing, and operational factors.
- Prices vary across airlines, routes, and travel duration.
- Relationships between features and ticket prices are often complex and non-linear, making manual prediction difficult.
- The dataset contains both categorical and numerical features that require careful preprocessing.
- Machine learning can provide accurate fare predictions and highlight the key factors influencing ticket prices.

METHODOLOGY

- Follows a structured machine learning pipeline.
- Dataset is cleaned, transformed, and preprocessed for modeling.
- Feature engineering performed to convert categorical and time-based data into numerical representations.
- Models trained and evaluated:
  - Linear Regression – simple and interpretable; captures overall trends but limited for non-linear patterns
  - Random Forest Regressor – captures non-linear relationships.
  - XGBoost Regressor – optimized gradient boosting for higher accuracy.
  - Support Vector Regressor (SVM) – effective for regression with complex patterns.
- Evaluation Metrics: RMSE, R^2 Score, and error analysis for model comparison.

Flow :

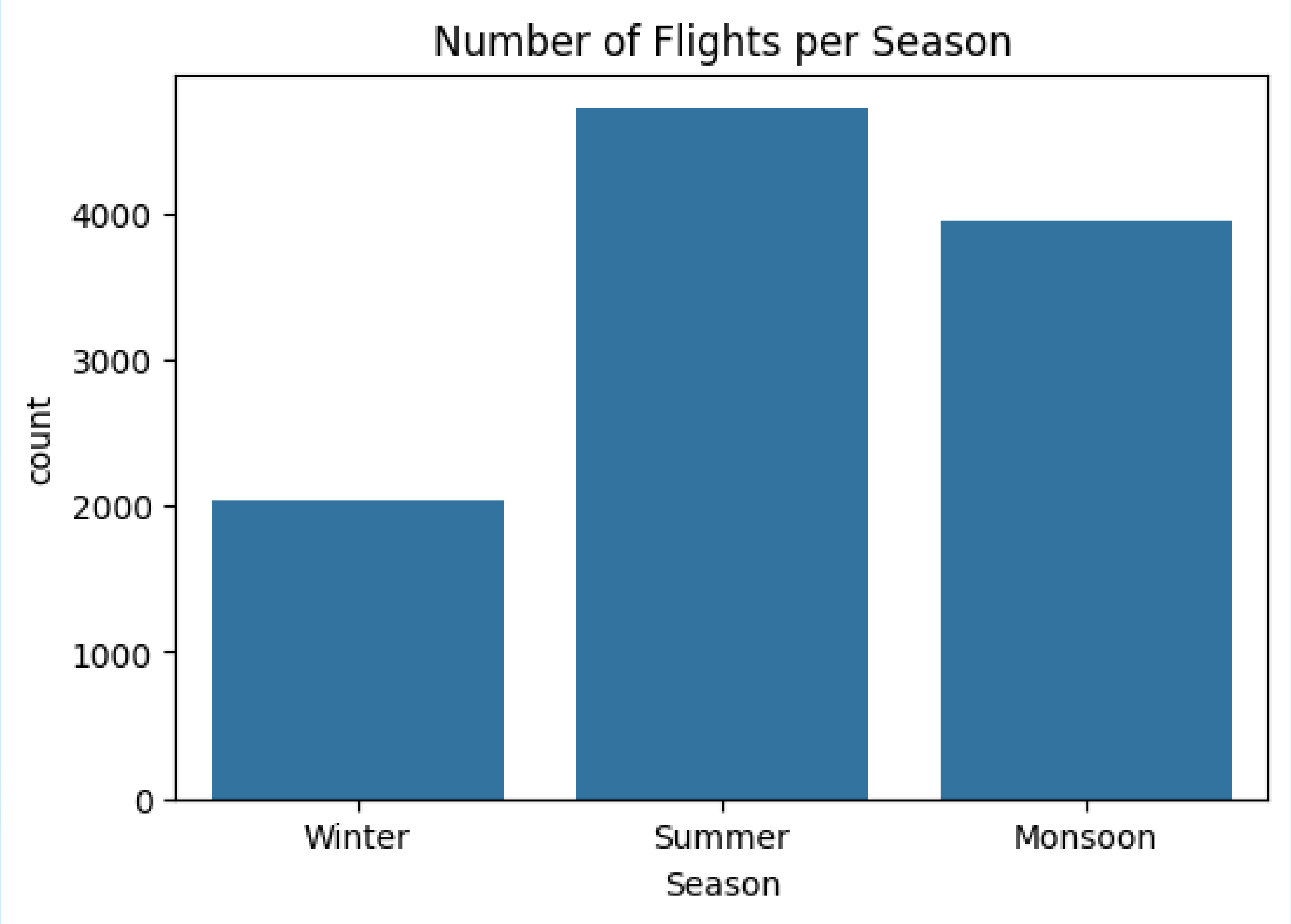


Data Collection
EDA
Outliers present?
Yes
No
IQR-based removal
Data Preprocessing
Feature Engineering
Clustering (KMeans)
Model Training
Prediction for New Data

FEATURE ENGINEERING

<u>Extracted features:</u>

- Airline
- Source City
- Destination City
- Departure Time
- Duration
- Class



Number of flights per Season

MODEL DEVELOPMENT

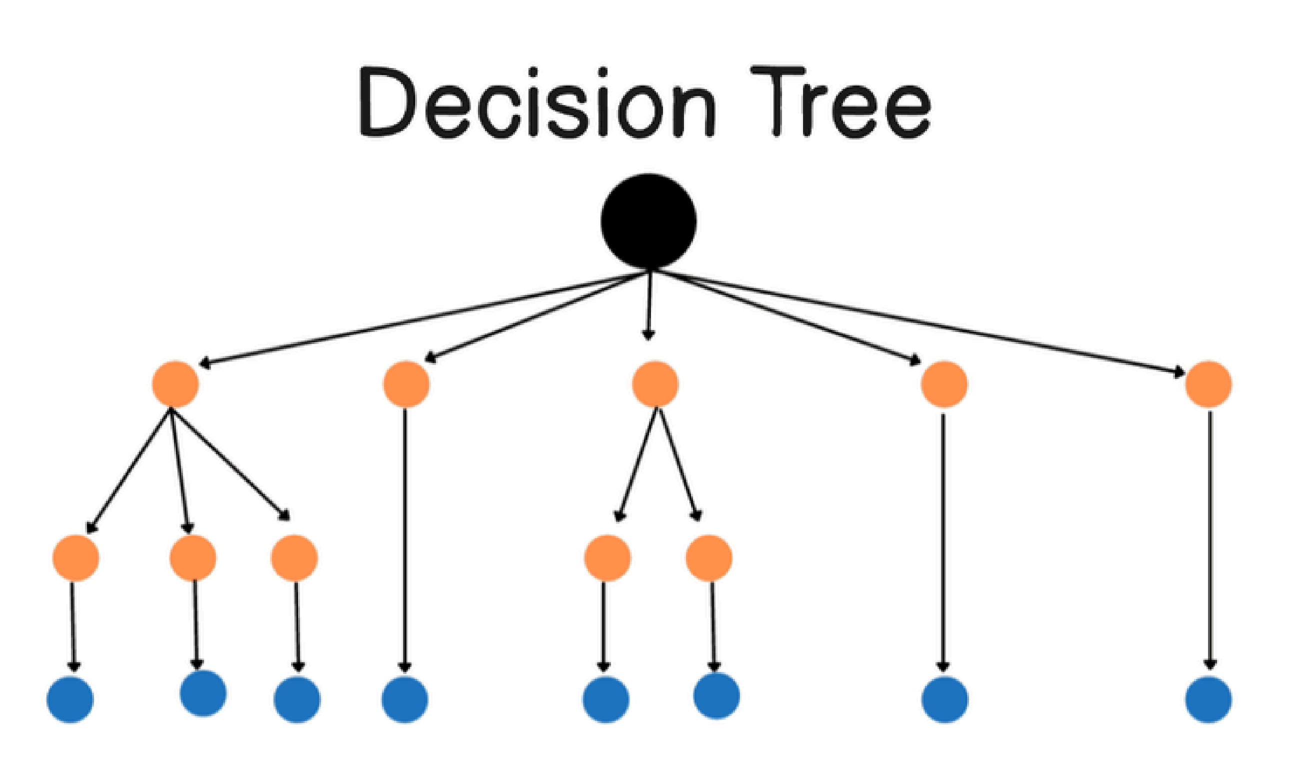
- **Linear Regression:**
  - Used as the baseline model for performance comparison.
  - Simple and interpretable but limited in capturing non-linear pricing patterns.
  - $\mathbf{Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n}$

    Where, y is the predicted price,

    $x_1, x_2, \ldots, x_n$ represent feature values,

    $\beta_0, \beta_1, \beta_2, \ldots, \beta n$ are model coefficients.

- **Random Forest Regression:**
  - Steps followed
    1. Create many decision trees on random subsets
    2. Each tree predicts price
    3. Average all predictions
    4. Final value = Predicted ticket price



MODEL DEVELOPMENT

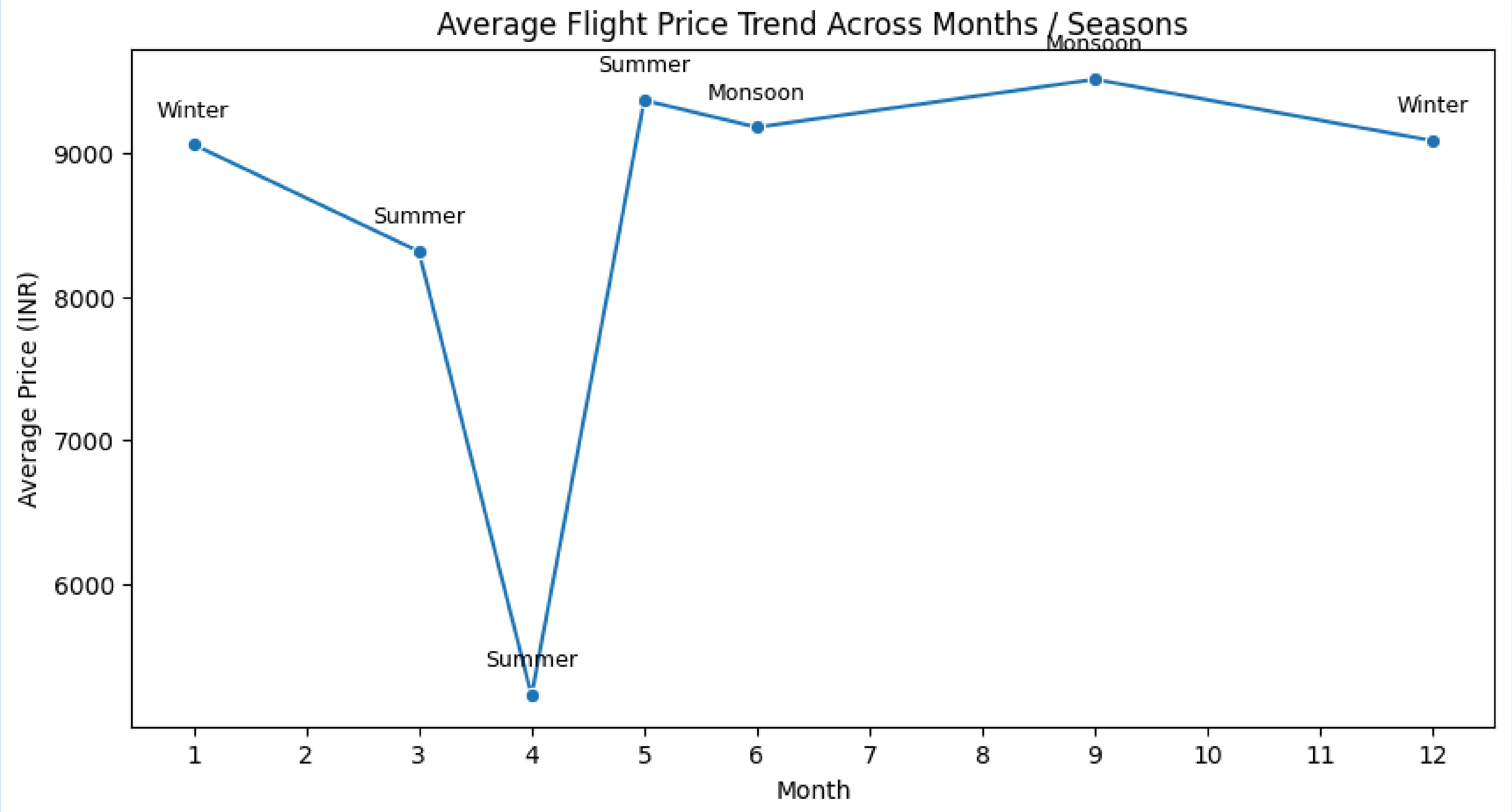
- **XGBoost Regression:**
  - Steps followed:
    1. Build trees sequentially, each tree correcting the errors of the previous ones.
    2. Apply gradient boosting to minimize prediction error.
    3. Combine all tree predictions with weighted sum.
    4. Final value = Predicted ticket price
    5. Efficient and handles complex non-linear relationships well

- **Support Vector Regression (SVM):**
  - Steps followed:
    1. Find a function that deviates from actual prices by a value no greater than a specified margin (ε).
    2. Use kernel functions (e.g., RBF) to capture non-linear patterns.
    3. Fit the model to minimize error while maximizing generalization.
    4. Final value = Predicted ticket price
    5. Effective for small-to-medium datasets and complex patterns

RESULT

- Linear Regression: $R^2 \approx 0.576$, RMSE ≈ 2672
  - Captures general trends but struggles with non-linear fare patterns.
- SVM: $R^2 \approx 0.584$, RMSE ≈ 2645
  - Slight improvement over Linear Regression; handles some non-linear relationships.
- Random Forest: $R^2 \approx 0.710$, RMSE ≈ 2210
  - Effectively learns complex interactions between features.
  - Provides robust and reliable fare predictions.
- XGBoost: $R^2 \approx 0.721$, RMSE ≈ 2168
  - Best performing model.
  - Accurately captures non-linear relationships and yields the lowest prediction error.

Average Flight Price trend



Comparison of ActualPrices vs Predicted Prices

	Actual_Price	Predicted_Price
9684	4823.0	4762.679688
4794	12898.0	14325.129883
574	3850.0	4367.193848
4534	3573.0	3915.262207
5643	9968.0	9998.362305
2180	12284.0	12556.928711
940	13376.0	12381.162109
2602	4878.0	6033.486328
8968	3597.0	3723.037598
4286	11188.0	8638.856445

ENHANCEMENTS

- Real-Time Data Scraping – Collect live fares from APIs/websites.
- Add More Features – Demand level, holidays, fuel cost, aircraft type, seat availability.
- Advanced ML Models – XGBoost, LightGBM, CatBoost for higher accuracy.
- Time-Series Forecasting – ARIMA, LSTM, Prophet for daily price trend modeling.
- Web App Deployment – Users enter route & date to get instant predictions.
- Price Alert System – Notify users when fares drop or are likely to change.

LIMITATIONS

- Dataset is limited and missing key factors like demand, seasonality, weather, holidays, and remaining seats.
- Data is static; no real-time or continuously updated pricing.
- No time-series component, so rapid price fluctuations aren’t captured.
- Only two models used (Linear Regression, Random Forest); advanced models not explored.

FUTURE SCOPE

- Live Airline API Integration – Real-time price prediction.
- Travel Recommendation System – Cheapest day, best airline, alternative routes.
- Price Trend Forecasting – Predict price hikes & best booking time.
- Mobile App Deployment – Quick predictions and fare alerts.
- Hybrid Model – Combine ML, time-series, and rule-based logic.
- Commercial Use – Flight recommendations, price monitoring, sales optimization.

CONCLUSION

- Machine learning can effectively predict airline ticket prices using only dataset features.
- Random Forest captures complex patterns better than simple linear models.
- The approach is scalable and reproducible for structured flight data.
- Insights can guide pricing strategies and decision-making.
- Future enhancements: integrate trends, seasonal factors, and interpretability techniques.

REFERENCES

- P. Sharma and A. Gupta, “Flight Ticket Price Prediction Using Regression Models,” International Journal for Research in Applied Science and Engineering Technology (IJRASET), vol. 10, no. 4, Apr. 2022.
- Scikit-learn, “Linear Models — User Guide,” scikit-learn.org, 2024. [Online]. Available: https://scikit-learn.org/stable/modules/linear_model.html
- A. Sharma, “Ensemble Learning: Bagging, Boosting, and Stacking,” TowardsAI, 2022.
- Scikit-learn, “Preprocessing Data — User Guide,” scikit-learn.org, 2024. [Online]. Available: https://scikit-learn.org/stable/modules/preprocessing.html

THANK YOU