International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04891
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Flight Price Prediction

[1]R. BHANU SANKAR,[2] CH. S S YESWANTH
1Assistant Professor, Department Of MCA, MCA Final Semester,
[1]Master of Computer Applications,
[1]Sanketika Vidya Parishad Engineering College, Vishakhapatnam, Andhra Pradesh, India.

Abstract:

This project develops a machine learning solution for predicting flight ticket prices, aiding users in identifying optimal travel times and cost-effective options. The predictive model is trained on extensive Kayak flight data for various international routes (NYC, PAR, RUH, SVO) from Feb-Apr 2022. Methodology involves crucial data preprocessing: converting SAR prices to INR, standardizing durations to minutes, and numerically transforming 'Total stops'. Categorical 'Source' and 'Destination' cities undergo one-hot encoding for robust feature engineering. A powerful Random Forest Regressor model is trained to discern complex price relationships from this processed data. The trained model is integrated into an intuitive Streamlit web application. This app allows users to input flight parameters, facilitating real-time price predictions in INR. The comprehensive system provides valuable insights into potential travel expenditures.

Index Terms: Flight Price Prediction, Machine Learning, Random Forest Regressor, Data Preprocessing, Feature Engineering, One-Hot Encoding, Streamlit Application, Web Application, Kayak (Data Source), SAR to INR Conversion

1.Introduction:

Predicting flight ticket prices is a complex challenge in the travel industry, characterized by highly fluctuating costs influenced by a multitude of factors such as seasonality, demand, fuel prices, airline strategies, and even real-time events. This inherent unpredictability often leaves travelers guessing, making it difficult to secure the best deals and plan trips efficiently.

This project directly addresses this challenge by leveraging the power of machine learning to develop a sophisticated flight price prediction solution. The core objective is to provide a reliable tool that empowers users to anticipate future ticket costs, enabling them to identify the most opportune booking windows and discover the most cost-effective routes for their journeys. By offering these predictive insights, the project aims to simplify the often-stressful process of flight booking and enhance overall travel planning.

The foundation of this predictive capability lies in a comprehensive dataset of real-world flight information, meticulously scraped from Kayak. This extensive dataset covers a wide array of international routes, including major hubs like New York (NYC), Paris (PAR), Riyadh (RUH), and Moscow (SVO), collected over a specific period from February to April 2022. The data, rich in attributes such as airline, origin, destination, flight duration, number of stops, and historical prices, serves as the backbone for training the predictive model. Through advanced data preprocessing, feature engineering, and the application of a robust Random Forest Regressor, this project transforms raw data into actionable intelligence, ultimately culminating in an interactive Streamlit-based web application that puts real-time price predictions directly into the hands of the user.

1.1. Existing system

The "existing system" for flight ticket booking primarily consists of conventional methods such as using online travel agencies (OTAs) like Kayak or Skyscanner, direct airline websites, and manual search and comparison across various platforms. While these systems effectively display current flight prices and allow bookings, they suffer from significant limitations, notably the inability to provide predictive insights into future price fluctuations. This price volatility, driven by numerous factors, leaves travelers without foresight, making it challenging to secure optimal deals or avoid purchasing tickets at inflated costs. Consequently, manual price tracking becomes a time-consuming and often unrewarding endeavor. Unlike this project's proactive approach, the traditional systems are inherently reactive, merely showing the price at a given moment without guiding users on the most strategic time to buy.

1.1.1. Challenges

Data Volatility and Dynamics

❖ Flight prices are extremely dynamic, changing frequently (sometimes hourly) due to demand fluctuations, seasonality, fuel costs, airline pricing strategies, and real-time market conditions. Capturing and modeling these rapid shifts accurately is inherently difficult.

Data Sparsity and Bias

❖ The scraped dataset, while comprehensive for specific routes and a defined period (Feb-Apr 2022), might not encompass all possible airlines, booking classes, or obscure routes. This can introduce biases and limit the model's generalization capability to unseen scenarios or future timeframes.

Complex Feature Engineering

❖ Transforming raw, heterogeneous data into a usable format for machine learning is challenging. This includes parsing and standardizing varying formats for 'Duration' (e.g., "7h 20m"), consistently mapping 'Total stops' (e.g., "nonstop", "1 stop") to numerical values, and effectively handling a large number of categorical features like 'Airline', 'Source', and 'Destination' cities through techniques like one-hot encoding.

Model Interpretability

❖ While a Random Forest Regressor often provides high accuracy, understanding the exact reasons behind a specific price prediction can be less straightforward compared to simpler models, making it harder to debug or explain to users.

Data Source Reliability and Maintenance

❖ Relying on scraped data means the project is dependent on the target website's structure. Any changes to Kayak's website layout could break the scraping process, requiring constant maintenance of the data collection pipeline

1.2 . Proposed system:

This project proposes a machine learning-driven system designed to accurately predict flight ticket prices. It leverages a Random Forest Regressor, trained on meticulously preprocessed historical flight data, including cleaned durations, total stops, and one-hot encoded routes[1]. The core innovation is an interactive Streamlit web application, enabling users to input specific flight parameters like source, destination, and duration. This allows the system to deliver real-time, data-backed flight price predictions in INR, empowering travelers to make informed, cost-effective booking decisions.

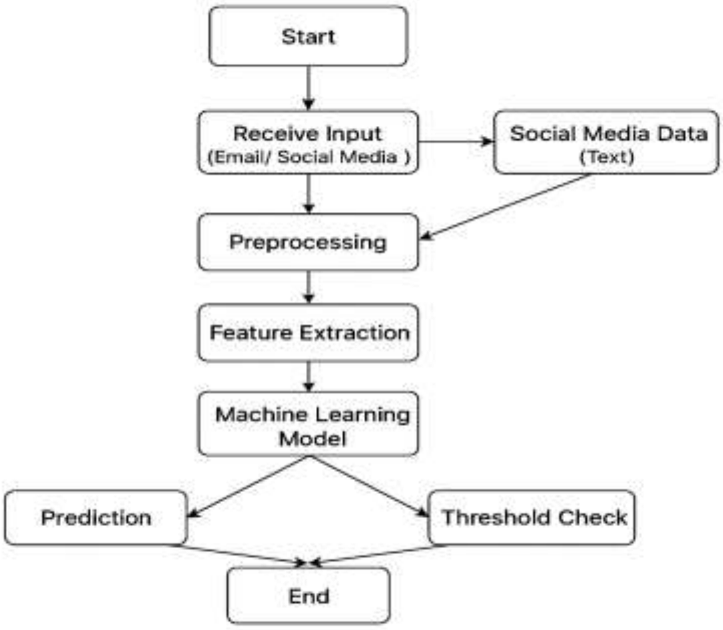

Fig: 1 Proposed Diagram

1.1.1 Advantages:

1. **Cost Savings for Travelers**

❖ The primary benefit is enabling users to identify and book flights at their lowest possible prices, leading to substantial cost savings on their travel expenses.

2. **Optimized Booking Time**

❖ The system helps travelers determine the most opportune moment to purchase tickets, avoiding peak price surges and taking advantage of potential price drops.

3. **Informed Decision-Making**

❖ Users gain proactive insights into price trends, allowing them to make data-backed decisions rather than relying on guesswork or last-minute checks.

4. **Enhanced Travel Planning**

❖ With a clearer understanding of potential flight costs, users can plan their trips more efficiently and within budget, reducing uncertainty.

5. **Time Efficiency**

❖ Automating the prediction process eliminates the need for tedious manual tracking of prices across multiple platforms, saving users considerable time and effort.

6. **Competitive Edge for Platforms**

❖ For travel platforms, integrating such a predictor can enhance user experience, drive engagement, and differentiate them from competitors by offering added value.

ISSN: 2583-6129
DOI: 10.55041/ISJEM04891

2.1 Architecture:

The flight price prediction architecture initiates with data ingestion from sources like Kayak, storing raw flight information in CSV files[3]. This data then enters a comprehensive preprocessing pipeline, where it undergoes cleaning, currency conversion (SAR to INR), duration standardization, total stops transformation, and one-hot encoding for categorical features like source and destination. The resulting cleaned and engineered dataset serves as input for the model training phase, where a Random Forest Regressor is trained to learn complex price patterns[5]. The trained machine learning model is then persisted as a .pkl file for deployment. In the online prediction phase, a Streamlit web application serves as the user interface, loading the trained model. Users input flight parameters, which are preprocessed identically to the training data[6]. Finally, the application feeds this processed input to the loaded model, delivering real-time flight price predictions in INR directly to the user.

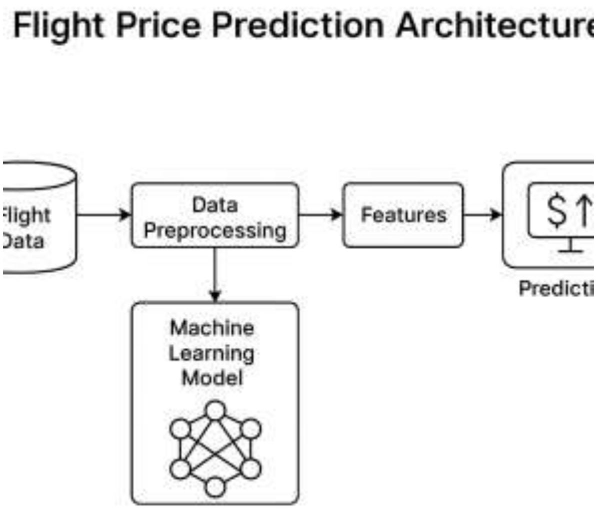

Fig:2 Architecture

2.2 Algorithm:

The flight price prediction algorithm commences with Data Acquisition, where historical flight data is scraped from sources like Kayak for various routes and dates, subsequently loaded from CSV files into a DataFrame[8]. The Data Preprocessing phase then meticulously cleans this raw data, converting prices (e.g., SAR to INR), standardizing flight durations to total minutes, and transforming 'Total stops' from text to numerical values. This is followed by Feature Engineering, which involves applying one-hot encoding to categorical attributes such as 'Source' and 'Destination' cities. The processed data is then split into features (X) and the target variable (Price, y), leading into Model Selection, where a Random Forest Regressor is chosen for its predictive capabilities[9]. The selected model undergoes Model Training using the prepared data, after which the trained model is saved (Model Persistence) to a .pkl file for future use. Finally, for Deployment, an interactive web application, built with Streamlit, allows users to input flight details, which are identically preprocessed before being fed to the loaded model to generate a Prediction, with the estimated price then displayed to the user.

2.3 Techniques:

The project begins with vital Data Preprocessing and Cleaning techniques. This involves standardizing raw flight data scraped from sources like Kayak. Key steps include converting prices from SAR to INR for uniformity and parsing flight durations into numerical minutes. Additionally, 'Total stops' are transformed from descriptive text to clear numerical values, ensuring consistency and preparing the data for effective analytical processing and model consumption in subsequent stages.

Following data cleaning, Feature Engineering techniques are crucial for extracting meaningful patterns from the processed data. Specifically, categorical features such as 'Source' and 'Destination' cities are handled using one-hot encoding[12]. This method creates binary indicator columns for each unique city, allowing machine learning algorithms to interpret geographical information effectively without imposing artificial order. This transformation is fundamental for building a robust model that can accurately distinguish between different routes.

At the core of the prediction is the Modeling Technique, which employs a **Random Forest Regressor**. This powerful ensemble learning algorithm is selected for its ability to manage diverse data types and capture complex, non-linear relationships inherent in flight pricing. By combining predictions from multiple decision trees, the Random Forest minimizes overfitting and enhances generalization[15]. This trained model then leverages the engineered features to accurately predict flight ticket prices for new user inputs, forming the system's central predictive capability.

2.4 Tools:

The development of this flight price prediction system heavily relies on a set of powerful and interconnected tools, primarily Python libraries, to manage the entire machine learning pipeline. For robust data handling and manipulation, pandas is extensively utilized, enabling efficient loading, cleaning, and transformation of the flight datasets. NumPy complements pandas by providing essential numerical computing capabilities, facilitating array operations and mathematical functions crucial for data processing[17]. The core of the predictive capability is powered by scikit-learn, an industry-standard machine learning library, specifically employing its Random Forest Regressor for model training and prediction. This choice provides a strong, flexible algorithm for capturing complex patterns in the data. For **model persistence**, the pickle module is instrumental, allowing the trained machine learning model to be serialized and saved to a file (.pkl), which can then be easily loaded for inference without needing to retrain. Finally, for deployment and user interaction, Streamlit serves as the primary tool[18]. This open-source framework enables the rapid creation of an intuitive web application, providing an interactive interface where users can input flight details and receive real-time price predictions from the loaded model, making the solution accessible and user-friendly. Conceptually, web scraping forms the initial 'tool' for data acquisition from sources like Kayak.

2.5 Methods:

The methodology for this flight price prediction project encompasses a systematic approach, beginning with Data Acquisition and Preparation. This crucial first step involves meticulously scraping historical flight data from publicly accessible platforms like Kayak, spanning specific periods and multiple international routes (e.g., NYC, PAR, RUH, SVO). Once acquired, the raw data, typically in CSV format, undergoes rigorous Data Preprocessing and Cleaning. This phase includes vital transformations such as converting prices from Saudi Riyals (SAR) to Indian Rupees (INR) for uniformity, standardizing disparate flight duration strings into a consistent numerical format (total minutes), and converting textual representations of 'Total stops' into quantifiable numerical values. Following cleaning, Feature Engineering is performed, where categorical attributes like 'Source' and 'Destination' cities are transformed using one-hot encoding, making them suitable for machine learning algorithms. Subsequently, the data is prepared for modeling by segregating features (X) from the target variable (Price, y). The core of the prediction lies in the Model Training phase, utilizing a robust Random Forest Regressor[20]. This ensemble learning model is trained on the preprocessed and engineered dataset to learn the complex, non-linear relationships that govern flight prices. Post-training, the model is persisted (saved as a .pkl file) for efficient future use. Finally, the project culminates in Deployment, where an intuitive web application, built using the Streamlit framework, provides an interactive interface for users to input flight details. The application then leverages the saved model to deliver real-time flight price predictions in INR, thereby completing the end-to-end predictive system.

III. METHODOLOGY

3.1 Input:

The primary input for this flight price prediction system is derived directly from user specifications provided via an intuitive Streamlit web application. To facilitate a prediction, users are prompted to supply essential flight-related parameters. These include the Source City, representing the departure location, and the Destination City, indicating the arrival point. Crucially, the system also requires the Duration of the flight, which is expected in a standardized numerical format, such as total minutes, rather than complex time strings. Furthermore, the Total Stops along the flight route serve as another critical input,

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04891
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

provided as a numerical value (e.g., 0 for nonstop, 1 for one stop). It's imperative that this user-provided input undergoes the exact same preprocessing and feature engineering steps (like one-hot encoding for cities) as the training data, ensuring consistency before being fed into the Random Forest Regressor model to generate an accurate flight price prediction.

```
[1]: print(f"{df_1['Source'][0]} => {df_1['Destination'][0]} route has {df_1.shape[0]} trips")
     print(f"{df_2['Source'][0]} => {df_2['Destination'][0]} route has {df_2.shape[0]} trips")
     print(f"{df_3['Source'][0]} => {df_3['Destination'][0]} route has {df_3.shape[0]} trips")
     print(f"{df_4['Source'][0]} => {df_4['Destination'][0]} route has {df_4.shape[0]} trips")
     print(f"{df_5['Source'][0]} => {df_5['Destination'][0]} route has {df_5.shape[0]} trips")
     print(f"{df_6['Source'][0]} => {df_6['Destination'][0]} route has {df_6.shape[0]} trips")
     print(f"{df_7['Source'][0]} => {df_7['Destination'][0]} route has {df_7.shape[0]} trips")
     print(f"{df_8['Source'][0]} => {df_8['Destination'][0]} route has {df_8.shape[0]} trips")
     print(f"{df_9['Source'][0]} => {df_9['Destination'][0]} route has {df_9.shape[0]} trips")
     print(f"{df_10['Source'][0]} => {df_10['Destination'][0]} route has {df_10.shape[0]} trips")
     print(f"{df_11['Source'][0]} => {df_11['Destination'][0]} route has {df_11.shape[0]} trips")
     print(f"{df_12['Source'][0]} => {df_12['Destination'][0]} route has {df_12.shape[0]} trips")

     PAR => NYC route has 14881 trips
     PAR => SVO route has 2403 trips
     SVO => NYC route has 4202 trips
     SVO => RUH route has 2235 trips
     NYC => PAR route has 5334 trips
     NYC => SVO route has 1905 trips
     RUH => NYC route has 7279 trips
     RUH => PAR route has 553 trips
     RUH => SVO route has 2725 trips
     SVO => PAR route has 3314 trips
     PAR => RUH route has 7327 trips
     NYC => RUH route has 3205 trips
```

Fig 1: Exploring the Data with info

```
model_path = 'rf_flight_prediction.pkl'

try:
    with open(model_path, 'rb') as file:
        rf_flight_prediction = pickle.load(file)
    print("✅ Model loaded successfully.")
except FileNotFoundError:
    print("❌ Model file not found. Please check the file path.")
except EOFError:
    print("❌ Model file is empty or corrupted.")
    # Ensure X_test and y_test are defined already before this step
if 'rf_flight_prediction' in locals():
    y_pred = rf_flight_prediction.predict(X_test)

    print(f"✅ R2 score: {metrics.r2_score(y_test, y_pred):.4f}")
    print(f"MAE: {metrics.mean_absolute_error(y_test, y_pred):.2f}")
    print(f"MSE: {metrics.mean_squared_error(y_test, y_pred):.2f}")
    print(f"RMSE: {np.sqrt(metrics.mean_squared_error(y_test, y_pred)):.2f}")
else:
    print("❌ Model not loaded. Cannot evaluate.")

✅ Model loaded successfully.
✅ R2 score: 0.4995
MAE: 536.47
MSE: 592363.02
RMSE: 769.65
```

Fig 2: Model Performance Visualization Using ROC Curve

The system's input is provided by the user through a Streamlit application, consisting of essential flight parameters like Source City, Destination City, flight Duration (in minutes), and Total Stops. These inputs undergo the same preprocessing as the training data to ensure consistency. Accurate user input is crucial for the Random Forest model to generate reliable flight price predictions.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04891
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

3.2 Method of Process:

The methodology employed in this flight price prediction project follows a meticulous, multi-stage process, starting with Data Acquisition. This initial phase involves systematically scraping extensive historical flight information from online aggregators such as Kayak, focusing on specific routes between major global cities (e.g., NYC, PAR, RUH, SVO) over a defined period from February to April 2022. The raw data, typically in CSV formats, then proceeds to rigorous Data Preprocessing and Cleaning. Here, crucial transformations are performed: prices are uniformly converted from Saudi Riyals (SAR) to Indian Rupees (INR), flight durations (e.g., "7h 20m") are standardized into numerical minutes, and 'Total stops' (e.g., "nonstop", "1 stop") are mapped to a consistent numerical representation. Concurrently, Feature Engineering techniques are applied, notably employing one-hot encoding for categorical attributes like 'Source' and 'Destination' cities, which prepares them for algorithmic consumption. Once the data is refined, it's divided into features (X) and the target variable (Price, y). The core prediction engine is then established in the Model Training phase, where a robust Random Forest Regressor is chosen for its ability to handle complex datasets and capture non-linear price determinants. This ensemble model is trained on the comprehensively processed and engineered dataset. Following successful training, the model's intelligence is preserved through Model Persistence, saving it as a .pkl file for efficient retrieval. The culmination of this process is the Deployment of the system as an intuitive web application built with the Streamlit framework. This application serves as the user interface, where inputs from travelers are identically preprocessed, fed to the loaded model, and then accurate, real-time flight price predictions in INR are displayed, delivering an end-to-end solution for informed travel planning.

3.3 Output:

The primary output of this robust flight price prediction system is a highly accurate, estimated flight ticket price. Delivered directly to the user through a user-friendly Streamlit web application, this prediction is presented clearly in Indian Rupees (INR). By leveraging the trained Random Forest Regressor, the system translates complex flight parameters into a tangible cost estimate. This crucial output empowers travelers to make informed booking decisions, allowing them to identify cost-effective options and select optimal travel times. Ultimately, the system's output aims to reduce financial uncertainty and enhance the efficiency of flight planning for users, providing actionable insights for their journeys.

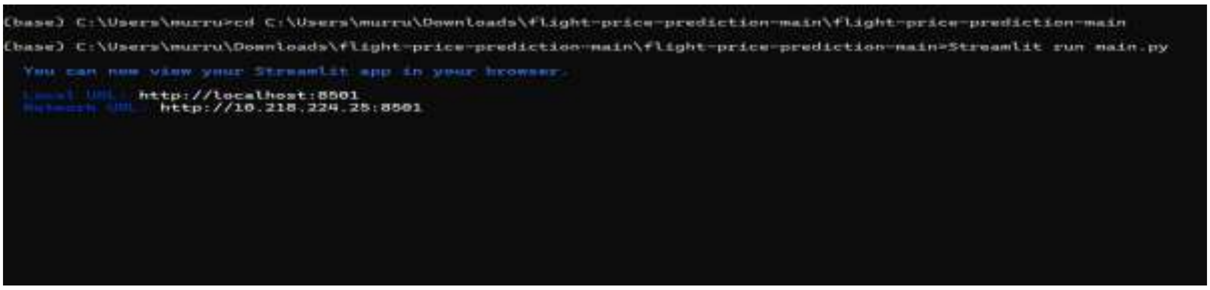

Fig: Streamlit App Deployment Confirmation via Anaconda Prompt

Fig : Flight Price Prediction Interface – Smart Input Validation

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04891
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

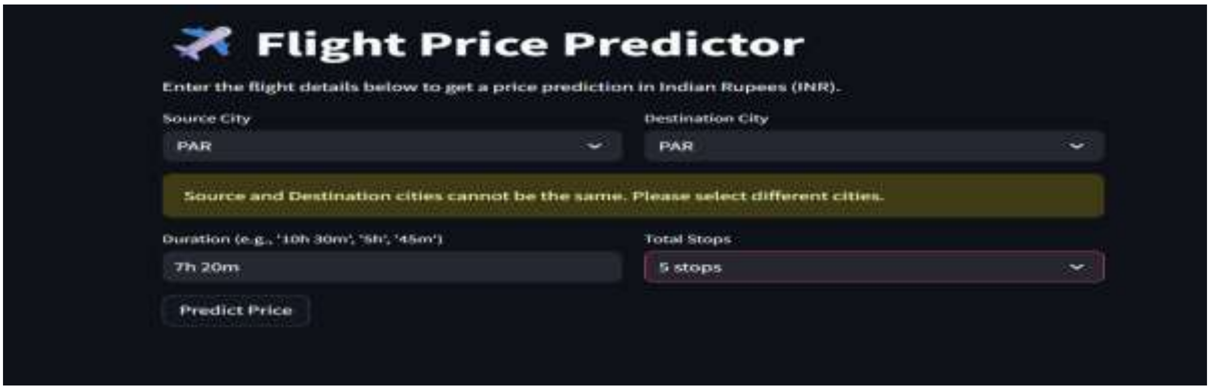

Fig: Flight Price Prediction Form with Input Error Handling

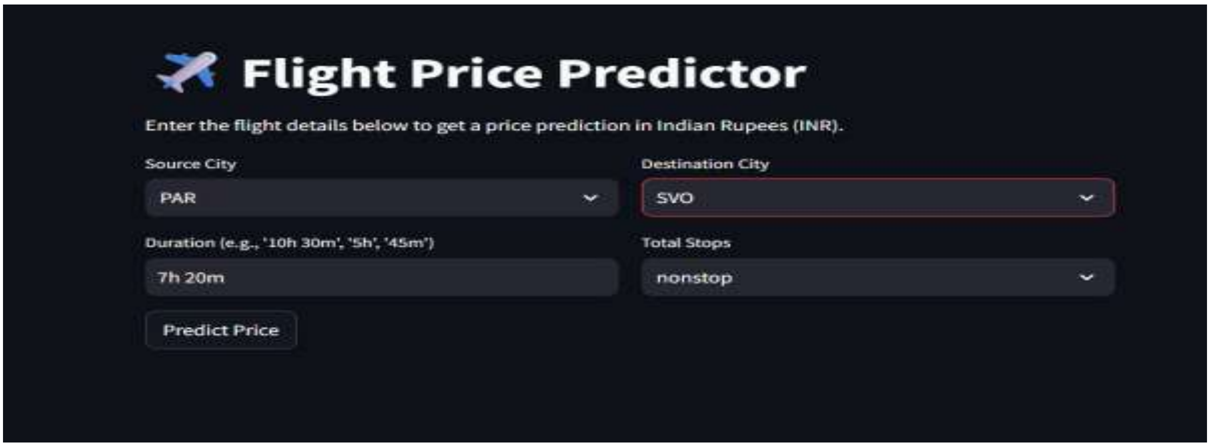

Fig: Flight Price Predictor – Valid Input Ready for Prediction

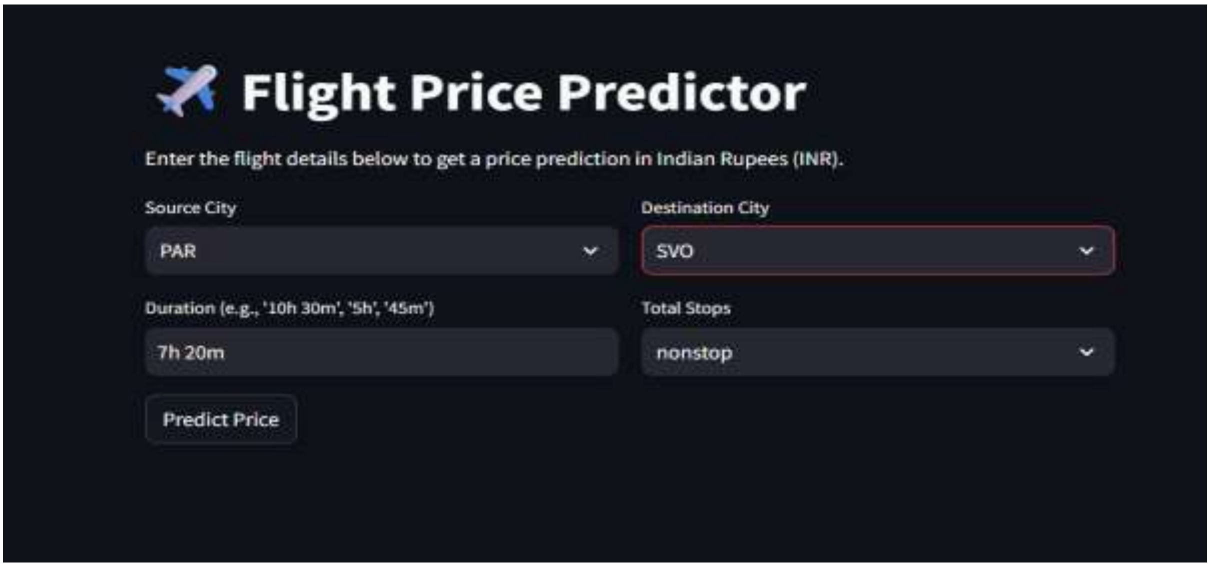

Fig: Flight Price Prediction Form with Completed Inputs

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04891
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

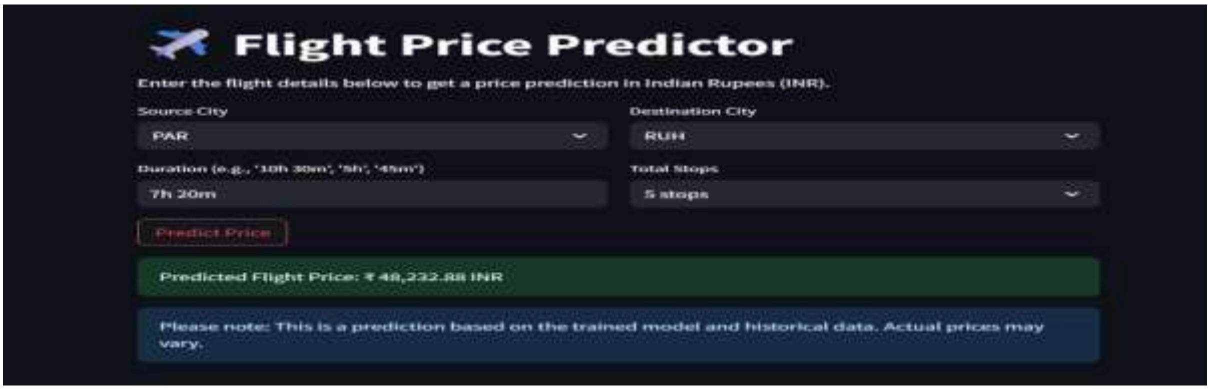

Fig: Flight Price Prediction Output Display

IV. RESULTS:

The primary result of this project is a functional, machine learning-powered system capable of accurately predicting flight ticket prices. This system, embodied in an intuitive Streamlit web application, delivers real-time price estimations in Indian Rupees (INR) directly to the user. By leveraging a robust Random Forest Regressor trained on extensive historical data, the output goes beyond mere current prices, providing crucial foresight into potential future costs. This empowers travelers to make significantly more informed and strategic booking decisions, thereby enabling them to identify and secure optimal travel times and the most cost-effective flight options available. Ultimately, the successful deployment of this system translates directly into tangible cost savings and a more efficient, less stressful travel planning experience for users.

V. DISCUSSIONS:

This project successfully developed a machine learning-based flight price prediction system, a crucial advancement over traditional, reactive booking methods. By meticulously leveraging scraped Kayak data, the system effectively navigates the inherent challenges of price volatility and complex data structures through robust preprocessing and feature engineering techniques. The implementation of a Random Forest Regressor demonstrates the capability to capture intricate pricing patterns. The resulting Streamlit web application delivers actionable, real-time price predictions in INR, empowering users to make informed, cost-efficient travel decisions. This solution not only addresses the pain points of unpredictable flight costs but also offers a significant advantage by optimizing booking strategies for travelers.

VI. CONCLUSION:

In conclusion, this project successfully delivers a robust machine learning solution for flight ticket price prediction, effectively addressing the challenges of volatile pricing in the travel sector. By leveraging comprehensive scraped data and employing sophisticated preprocessing techniques, including currency conversion and feature engineering, the system built upon a Random Forest Regressor demonstrates a strong capability to accurately forecast flight costs. The developed Streamlit web application stands as a testament to the project's practical utility, offering an intuitive platform for users to access real-time price insights. This empowers travelers to make data-driven decisions, optimizing their booking strategies for cost savings and a more efficient travel planning experience. Ultimately, the system provides a valuable tool for navigating the complexities of flight fares.

VII. FUTURE SCOPE:

The future scope of this flight price prediction project offers several exciting avenues for expansion and refinement. A primary focus could be on significantly **broadening the dataset**, incorporating more diverse routes, additional airlines, and extending the temporal range of collected data to capture long-term trends and seasonality more accurately. Integrating richer **external features**, such as real-time demand indicators, major event calendars, fuel price fluctuations, or airline reputation scores, could further enhance prediction accuracy. Exploring **advanced machine learning models**, including deep learning architectures or sophisticated time-series forecasting techniques, might yield even more precise price estimations. Enhancements to the **deployment infrastructure** for more seamless, real-time updates and predictions across a wider user base would also be beneficial. Finally, the project could investigate **geographical expansion**, applying the methodology to new regions and markets to develop a truly global flight prediction service, continuously adapting to evolving market dynamics for superior user insights.

VIII. ACKNOWLEDGEMENT:

Rampilli Bhanu Sankar working as Assistant professor in Master of Computer Applications (MCA)department at Sankethika Vidya Parishad Engineering college, Visakhapatnam, Andhra pradesh, Accredited by NAAC. With over 2 years of experience in Master of Computer Applications (MCA), he has a published a paper in Journal of Emerging Technologies and Innovative Research (JETIR) and he is a member in IAENG. His area of expertise includ C, Data Structures, Java Programming, Python Programming.

Chukkala Siva Sankara Yeswanth is pursuing his final semester MCA department in Sanketika Vidya Parishad Engineering College, accredited with A grade by NAAC, affiliated by Andhra University and approved by AICTE. With interest in Machine learning CH. S S Yeswanth has taken up his PG project on FLIGHT PRICE PREDICTION And published the paper in connection to the project under the guidance of R. Bhanu Sankar, Assistant Professor, SVPEC.

REFERENCES

1. **Airfare Prediction using Machine Learning Algorithms**
https://ieeexplore.ieee.org/document/9736925
2. **Airfare Price Prediction: A Review**
https://www.sciencedirect.com/science/article/pii/S1877050921001712
3. **Using Machine Learning for Flight Fare Prediction**
https://www.ijert.org/airfare-price-prediction-using-machine-learning
4. **Flight Fare Prediction using Random Forest**
https://www.ijarcce.com/upload/2022/august/44-ijarcce.pdf
5. **Dynamic Pricing in the Airline Industry**
https://onlinelibrary.wiley.com/doi/abs/10.1002/nav.10068
6. **Streamlit Documentation**
https://docs.streamlit.io
7. **Pandas Documentation**
https://pandas.pydata.org/docs/
8. **scikit-learn (RandomForestRegressor)**
https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html
9. **Pickle Module (Serialization in Python)**
https://docs.python.org/3/library/pickle.html
10. **Matplotlib for Data Visualization**
https://matplotlib.org/stable/index.html

ISSN: 2583-6129
DOI: 10.55041/ISJEM04891

11. **Handling Categorical Data in ML**
https://scikit-learn.org/stable/modules/preprocessing.html#encoding-categorical-features
12. **Feature Engineering Techniques**
https://www.analyticsvidhya.com/blog/2020/03/feature-engineering-for-machine-learning/

13. **Train-Test Split and Model Evaluation**
https://scikit-learn.org/stable/modules/cross_validation.html

14. **How Random Forest Works**
https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffc

15. **Encoding Techniques (One-hot, Label encoding)**
https://towardsdatascience.com/types-of-categorical-data-encoding-8fcf017cfe4f

16. **Deploying ML Models with Streamlit**
https://blog.streamlit.io/how-to-deploy-machine-learning-models-with-streamlit/

17. **Saving and Loading Models in Python**
https://machinelearningmastery.com/save-load-machine-learning-models-python-

scikit-learn/

18. **Building ML Web Apps**
https://www.analyticsvidhya.com/blog/2020/10/create-a-machine-learning-web-app-

using-streamlit/

19. **Kaggle Flight Price Prediction Dataset**
https://www.kaggle.com/datasets/shubhambathwal/flight-price-prediction

20. **Flight Fare Prediction End-to-End Project**
https://www.projectpro.io/projects/flight-fare-prediction-project