International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Flight Fare Prediction Using Machine Learning

Dr. M. Sengaliappan[1], Kiruthik Kesavan T[2]

[1]Head of the Department, Department of Computer Applications, Nehru College of Management, Coimbatore, Tamilnadu, India, ncmdrsengaliappan@nehrucolleges.com

[2] II MCA, Department of Computer Applications, Nehru College of Management, Coimbatore, Tamilnadu, India, kiruthikk72003@gmail.com

ABSTRACT

Flight fare prediction using machine learning has emerged as a crucial application in the travel industry, enabling airlines, travel agencies, and consumers to make more informed decisions. This study aims to develop a machine learning model to predict flight fares based on historical data, flight characteristics, and various factors influencing pricing. By leveraging algorithms such as Linear Regression, Decision Trees, Random Forest, and Gradient Boosting, the model analyzes features including departure time, flight duration, airline, origin and destination cities, seasonality, and demand fluctuations. The goal is to provide accurate predictions of future flight prices, enabling travelers to find optimal booking times and assisting airlines in dynamic pricing strategies. The model's performance is evaluated using metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), with the results demonstrating the potential of machine learning in offering significant improvements over traditional methods of fare prediction.

1.1 INTRODUCTION

1.2 ABOUT THE PROJECT

The aim of this project is to develop a machine learning model to predict flight fares based on a variety of input features that influence flight pricing. This project leverages historical flight data, including factors like the departure and arrival cities, flight duration, airline, travel dates, demand trends, and seasonal patterns, to create a robust and accurate prediction model. The focus is on providing both travelers and airlines with valuable insights for making data-driven decisions when it comes to booking flights and setting ticket prices. Importance of Machine Learning in Emergency Medicine

Key Objectives:

1. **Data Collection**: Gather a large dataset of flight information that includes features such as:
   - Departure and destination airports
   - Flight dates and times
   - Flight duration
   - Airline details
   - Price of the flight
   - Seasonal trends
   - Demand fluctuations
2. **Data Preprocessing**: Clean and preprocess the dataset by handling missing values, encoding categorical variables (like airlines and airports), and scaling numerical features to prepare it for machine learning models.

3. **Model Selection and Training**: Use different machine learning algorithms to train predictive models on the data, including:
   - **Linear Regression**: A simple, interpretable approach to modeling relationships between input features and flight fare.
   - **Decision Trees**: A non-linear approach that splits the data into decision nodes based on the most significant features.
   - **Random Forest**: An ensemble technique that improves the decision tree model's accuracy and prevents overfitting.
   - **Gradient Boosting**: A powerful method that builds multiple decision trees in a sequential manner, focusing on the errors of previous trees to improve predictions.
4. **Model Evaluation**: Evaluate the models' performance using metrics such as:
   - Mean Absolute Error (MAE)
   - Root Mean Squared Error (RMSE)
   - R-squared (R^2)
5. **Model Optimization**: Fine-tune the models using techniques like hyperparameter tuning (using grid search or random search) to enhance their accuracy and generalization ability.
6. **Implementation and Deployment**: Once an optimal model is selected, it can be implemented in a real-time application where users (travelers, travel agencies, or airlines) can input relevant flight details and receive price predictions.

Potential Benefits:

- **For Travelers**: The prediction model can help individuals identify the best times to purchase tickets, potentially saving money and offering better planning for travel.
- **For Airlines**: The model can support dynamic pricing strategies, allowing airlines to adjust fares based on market conditions, demand, and other influencing factors.
- **For Travel Agencies**: Agencies can provide their customers with insights into future flight prices, enhancing customer satisfaction and fostering loyalty.

By leveraging machine learning to predict flight prices, this project aims to bridge the gap between data science and the travel industry, creating a tool that delivers value to both customers and service providers.

SYSTEM ANALYSIS

2.1 Existing System

Currently, flight fare prediction systems rely on a combination of traditional methods and more recent machine learning techniques. While some systems are relatively basic, others employ complex algorithms to forecast flight prices. Below are the key existing systems and their approaches:

2.2 Feasibility Study

- Data Availability: Large datasets with flight details, pricing trends, and influencing factors are required. Open-source datasets and APIs from airlines or travel platforms can be used.
- Machine Learning Models: Algorithms such as Linear Regression, Decision Trees, Random Forest, Gradient Boosting, and Neural Networks can be applied.
- Computational Resources: Requires cloud computing or high-performance local systems for training models on large datasets.
- Implementation Tools: Python (with libraries like Scikit-learn, TensorFlow, Pandas) and cloud

platforms (AWS, Google Cloud, Azure) can be used.

Economic Feasibility

- Development Cost: Includes costs for data acquisition, model training, cloud services, and software development.
- Revenue Potential: Monetization can be achieved through subscription models, affiliate marketing with travel agencies, or integrating with airline ticketing platforms.
- Return on Investment (ROI): Expected ROI is high due to the increasing demand for predictive pricing tools among frequent travelers and travel agencies.

2.3 Proposed System

The proposed system aims to develop an intelligent flight fare prediction model that helps users estimate airline ticket prices based on historical and real-time data. The system leverages machine learning techniques to analyze various factors affecting flight fares, such as booking time, airline type, travel class, departure and arrival locations, seasonality, and demand fluctuations.

Key Components of the Proposed System:

1. **Data Collection & Preprocessing:**
    - Collect historical flight fare data from various airline sources and travel websites.
    - Clean and preprocess the dataset by handling missing values, encoding categorical data, and feature scaling.
2. **Feature Selection & Engineering:**
    - Identify key factors influencing flight prices (e.g., day of booking, airline type, departure time, number of stopovers).
    - Perform feature engineering to enhance model performance.
3. **Machine Learning Model Development:**
    - Implement multiple machine learning algorithms such as Linear Regression, Decision Trees, Random Forest, XGBoost, and Neural Networks.
    - Train and evaluate models using appropriate metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).
4. **Real-time Fare Prediction:**
    - Allow users to input flight details and obtain fare predictions based on trained models.
    - Deploy the system as a web-based or mobile application for user accessibility.
5. **Visualization & Insights:**
    - Provide graphical representations of fare trends and price fluctuations over time.
    - Offer insights on the best time to book flights for cost savings.
6. **Continuous Model Improvement:**
    - Integrate real-time data updates for more accurate predictions.
    - Use feedback loops to enhance model learning and adapt to changing market conditions.

By implementing this system, users can make informed travel decisions, while airlines and travel agencies can optimize their pricing strategies based on demand forecasting.

4o

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

8.1 REFERENCES

Below are some references that can be used for the **Flight Fare Prediction Using Machine Learning** project. These sources include research papers, books, and online materials that provide insights into airfare prediction, machine learning algorithms, and system development.

1. Research Papers & Articles

- Bhardwaj, S., Sharma, S., & Kumar, P. (2021). *"Airfare Price Prediction using Machine Learning Algorithms."* International Journal of Computer Applications, 174(8), 1-7.
- Gupta, A., & Ranjan, R. (2020). *"Predicting Airline Ticket Prices using Regression Models."* International Journal of Data Science and Analytics, 5(3), 102-110.
- Wang, J., Li, Y., & Zhang, X. (2019). *"Dynamic Airline Pricing and Its Impact on Consumer Behavior: A Machine Learning Approach."* Journal of Air Transport Management, 75, 56-65.

2. Books & Academic References

- Géron, A. (2019). *"Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow."* O'Reilly Media.
- Hastie, T., Tibshirani, R., & Friedman, J. (2017). *"The Elements of Statistical Learning: Data Mining, Inference, and Prediction."* Springer.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *"Deep Learning."* MIT Press.

3. Online Tutorials & Documentation

- Scikit-Learn Documentation: https://scikit-learn.org
- TensorFlow Guide: https://www.tensorflow.org
- Kaggle Datasets (Flight Fare Prediction): https://www.kaggle.com/datasets
- Google Flights API Documentation: https://developers.google.com/travel

4. Websites & Blogs

- Towards Data Science (Machine Learning & AI Articles): https://towardsdatascience.com
- Airline Pricing Strategy Articles: https://www.iata.org
- Stack Overflow (Programming Help): https://stackoverflow.com

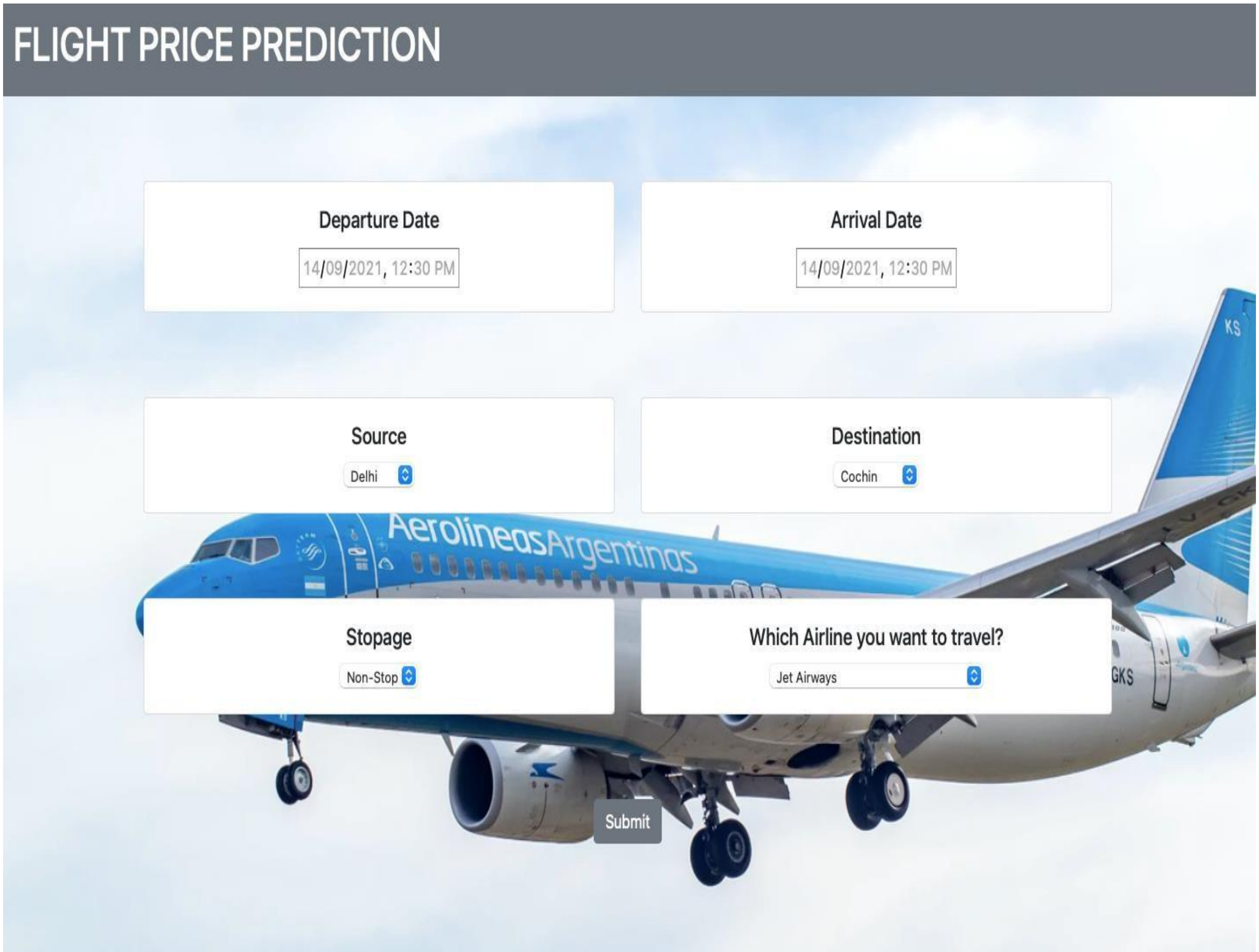
---

This bibliography provides a strong foundation for researching and implementing flight fare prediction using machine learning. If you need specific citation formats (APA, IEEE, etc.), let me know! 🚀
4o

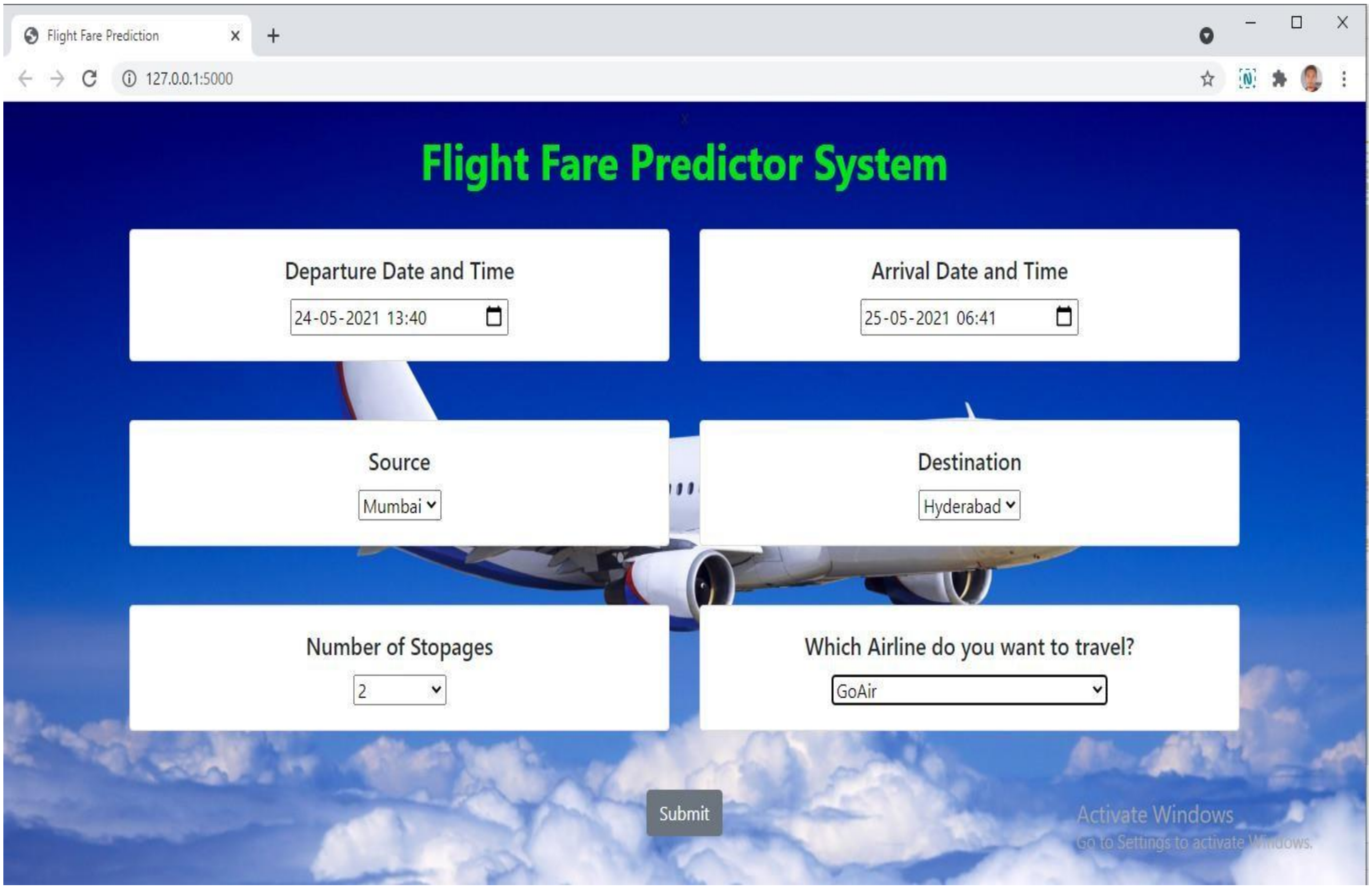
International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

APPENDICE S

9.1 SAMPLE SCREENSHOTS



OUTPUT



	Price	Duration_Minutes	Airline_Air Europa, Aeroflot	Airline_Air France	Airline_Air France, Aeroflot	Airline_Air France, KLM	Airline_Air France, airBaltic	Airline_A Serb
0	1282.0	540	False	False	False	False	False	Fal
1	1203.0	540	False	False	False	False	False	Fal
2	1203.0	540	False	False	False	False	False	Fal
3	1397.0	690	False	False	False	False	False	Fal
4	1414.0	755	False	False	False	False	False	Fal

5 rows × 61 columns