EXPLORING PREDICTIVE MODELS FOR AIRFARE FORECASTING: A COMPARATIVE ANALYSIS OF TIME SERIES AND MACHINE LEARNING APPROACHES

Mrs. Sravanthi Boddu\textsuperscript{1}, Mr. K Saketh Reddy\textsuperscript{2}, Gunji Lakshmi Sahitya\textsuperscript{3}, Shaik Afrin\textsuperscript{4}, Kukkamalla Sudha\textsuperscript{5}

\textsuperscript{1}Assistant Professor, Dept Of Artificial Intelligence, KKR & KSR Institute Of Technology And Sciences, Vinjanampadu, Guntur, Andhra Pradesh, India.
\textsuperscript{2}Co-Founder, CODEGNAN, Vijayawada, Andhra Pradesh, India.
\textsuperscript{3,4,5}Student, Dept Of Artificial Intelligence, KKR & KSR Institute Of Technology And Sciences, Vinjanampadu, Guntur, Andhra Pradesh, India.

DOI: https://www.doi.org/10.56726/IRJMETS515210

ABSTRACT

Accurate prediction of airfare prices is essential for both travelers and airlines alike. This paper presents a comprehensive investigation into airfare prediction, focusing on the context of Indian Airways. Leveraging a dataset specific to Indian Airways flight records, spanning both historical and future data, this research explores the efficacy of various predictive modeling techniques. Machine learning algorithms and time series analysis are employed for comparative analysis, evaluating their performance in predicting airfare prices. Through rigorous experimentation and evaluation, this study provides valuable insights into the strengths and limitations of different methodologies. The findings contribute to advancing the understanding of airfare forecasting techniques tailored to the unique characteristics of Indian Airways data. This research not only facilitates informed decision-making for travelers but also aids airlines in strategic pricing.

Keywords: Airfare Prediction, Indian Airways, Predictive Modeling, Machine Learning Algorithms, Time Series Analysis, Comparative Analysis, Historical Data, Future Data, Forecasting Techniques, Travelers, Airlines, Decision-Making, Strategic Pricing, Efficacy Evaluation, Methodologies.

I. INTRODUCTION

Modern airlines employ sophisticated strategies to dynamically set airfare prices, considering various financial, marketing, commercial, and social factors. However, the complexity of these pricing models presents challenges for customers seeking the best deals, as prices often fluctuate unpredictably. Consequently, there has been a rise in the development of techniques aimed at helping buyers time their purchases by accurately predicting airfare prices.

These techniques rely on advanced computational intelligence, notably Machine Learning (ML), to offer timely insights to buyers by accurately forecasting airfare prices. By leveraging ML algorithms, these methods empower customers to make well-informed purchasing decisions. This study delves into airfare prediction, examining how ML techniques assist travelers in navigating airline pricing complexities. Through rigorous analysis, we aim to provide valuable insights into the effectiveness of these prediction models and their practical implications for airline industry consumers.

This research extensively examines predictive models for forecasting fluctuations in airline ticket prices. It employs Linear Regression and Polynomial Regression to predict price decreases, while Logistic Regression and Linear Support Vector Machine (SVM) models are utilized for classification. Additionally, Time Series Analysis techniques are integrated to precisely determine the optimal dates for purchasing tickets. The study sheds light on the effectiveness of these methodologies in assisting travelers to make well-informed decisions about the timing of their ticket purchases.

The proposed paper aims to contribute in the following ways: (1) By predicting airfare prices in Indian airlines using both historical and future data for the first time, (2) Investigating the influence of features on airfare prices, and (3) Conducting a performance analysis of state-of-the-art ML models in airfare prediction.
II. LITERATURE REVIEW

1. In the first paper, authored by Theofanis Kalampokas and Konstantinos Tziridis, this paper delves into airfare price prediction leveraging AI techniques, scrutinizing pricing strategies across diverse airline companies. Utilizing extensive flight data from Aegean, Turkish, Austrian, and Lufthansa Airlines for popular international destinations, the study explores affordability from both destination-based and airline-based perspectives. Employing a range of ML, DL, and QML models, the research attains accuracies between 89% and 99%, unveiling effective strategies for global competitiveness in the airline industry.

2. The second paper, authored by Vishan Lal, Paul Stynes, and Cristina Hava Muntean, this paper examines flight price prediction accuracy in India using machine learning models. Evaluation of basic and advanced regression models, including XG Boost, highlights their superior performance in closely matching actual flight prices. The research aims to optimize flight ticket booking by providing passengers with accurate cost predictions.

3. The third paper, authored by Lee-Yeng Ong, Yee-Fan Tan, and Meng-Chew Leow, delves into time-series forecasting models for dynamic pricing in Digital Signage Advertising (DSA). It aims to optimize pricing decisions considering audience attention and environmental factors. Through reviewing 84 research articles, the study identifies data characteristics and proposes a framework for selecting optimal time-series forecasting models in DSA pricing solutions.

4. The fourth paper, authored by Vansh Sethi, explores the use of Machine Learning Regression techniques for predicting airline fares based on essential parameters. Utilizing two datasets for training and testing, the study evaluates various machine learning approaches to forecast flight ticket prices, addressing the challenge of continual fare fluctuation in air travel.

5. The fifth paper, authored by Tejal Dimble, Nikita Pandey, Harshada Narkhede, and Ruturaj More, focuses on predicting airfare and hotel prices using advanced forecasting models based on machine learning and artificial intelligence. By mining historical data and applying machine learning algorithms, the study aims to assist travelers in understanding price trends and managing their travel costs effectively. The research underscores the importance of leveraging mathematical and scientific tools, along with the power of machine learning, to provide accurate fare predictions and enhance decision-making for travelers.

III. MODELING AND ANALYSIS

Our first approach treats the prediction of air ticket prices as a regression problem, aiming to forecast the exact price. This strategy involves approximating a function that maps data features to airfare prices. Conversely, the second approach transforms the problem into a Time series task, enabling decisions on price ranges or whether to purchase a ticket at a specific price. While the regression approach aims for precise price prediction, the classification approach offers insights into price ranges and purchase decisions.

Initially, the dataset for this study encompasses airfare data from Indian Airlines sourced from Kaggle, spanning from 2019 to 2023. Additionally, recent data collected in March 2024 through web scraping techniques using the Beautiful Soup module supplements the dataset. It includes a diverse range of routes and destinations, including but not limited to the Hong Kong to Vijayawada route.

The current study comprises four distinct phases:

- Identifying the flight features impacting airfare prices
- Gathering sufficient flight data for training and testing ML models
- Choosing regression ML models for comparison and
- Experimentally evaluating the performance of these ML models.
- Performing the Time Series analysis Varying of price

**Phase 1 (Feature Selection for Airfare Dataset)** – This stage involves determining the key features of a flight that significantly influence airfare prices. It holds pivotal importance as it shapes the problem-solving approach.

Each flight is evaluated based on the following features:

- F1: Date of journey
This comprehensive dataset allows for an in-depth analysis of airfare prices across various routes and destinations, facilitating a robust evaluation of predictive models for airfare forecasting.

**Phase 2 (Gathering Sufficient Flight Data)** - The study focuses on predicting a single airfare price without considering return trips. To conduct the experiments, a dataset of flights to various destinations spanning from 2019 to March 2024 is collected. For each flight, eight features (F1:F8) are manually extracted from the web and online resources.

**Phase 3 (ML Models Selection)** - For the current study, state-of-the-art regression ML models were chosen and applied to the flight dataset. The ML models evaluated in this research include:

- Generalized Regression Neural Network
- Polynomial Regression
- Random Forest Regression Tree
- Regression Tree
- Bagging Regression Tree
- Regression SVM (Polynomial and Linear)
- Linear Regression (LR)

**Phase 4 (Evaluation)** - During this stage, the flight data collected in phase 2 underwent cross-validation to train the specified ML models. Comparative analysis of the models was conducted using performance metrics, including prediction accuracy (measured as a percentage - Mean Squared Error between desired and predicted prices) and the training time for each model, expressed in seconds.

In the context of our experiments, a set of simulations was arranged and executed under the MATLA environment on a PC with an Intel Core i5-750 processor and 8GB of memory. The configuration of the machine learning (ML) models was determined through grid search and is summarized in Table 1.

<table>
<thead>
<tr>
<th>ML Model</th>
<th>Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extreme Learning Machine</td>
<td>15 neurons</td>
</tr>
<tr>
<td>Random Forest Regression Tree</td>
<td>200 weak classifiers (decision trees)</td>
</tr>
<tr>
<td>Regression Tree</td>
<td>MinParentSize=15, MinLeafSize=5</td>
</tr>
<tr>
<td></td>
<td>MaxNumSplits=60</td>
</tr>
<tr>
<td>Bagging Regression Tree</td>
<td>600 weak classifiers (decision trees)</td>
</tr>
<tr>
<td>Regression SVM (Polynomial)</td>
<td>order=4</td>
</tr>
<tr>
<td>Regression SVM (Linear)</td>
<td>Switch to mini-batch gradient descent</td>
</tr>
<tr>
<td>Linear Regression</td>
<td>stochastic gradient descent with momentum</td>
</tr>
</tbody>
</table>

The performance of all models for the entire feature set is presented in Table II, indicating the model with the highest accuracy.
Table 2:

<table>
<thead>
<tr>
<th>ML Model</th>
<th>Accuracy (%)</th>
<th>Execution Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalized Regression Neural Network</td>
<td>66.83</td>
<td>0.13</td>
</tr>
<tr>
<td>Extreme Learning Machine</td>
<td>68.68</td>
<td>0.05</td>
</tr>
<tr>
<td>Random Forest Regression Tree</td>
<td>85.91</td>
<td>5.50</td>
</tr>
<tr>
<td>Regression Tree</td>
<td>84.13</td>
<td>0.04</td>
</tr>
<tr>
<td>Bagging Regression Tree</td>
<td>82.42</td>
<td>17.05</td>
</tr>
<tr>
<td>Regression SVM (Polynomial)</td>
<td>77.00</td>
<td>1.23</td>
</tr>
<tr>
<td>Regression SVM (Linear)</td>
<td>49.40</td>
<td>0.34</td>
</tr>
<tr>
<td>Linear Regression</td>
<td>57.25</td>
<td>0.10</td>
</tr>
</tbody>
</table>

After analyzing the models, it was found that the random forest model performed the best. The evaluation of the fitted model and its metrics are depicted in the graph below. The x-axis represents the features, while the y-axis represents the target price.

**Price Variation with Target on Y-Axis**

![Figure 1:](image1)

![Figure 2:](image2)
Phase 5 (Time Series Analysis) - In this study, we conducted a time series analysis of airline ticket prices, focusing on temporal patterns and trends. Utilizing features such as journey dates, airlines, sources, and destinations, we decomposed the time series into its constituent components (trend, seasonal, and residual).

Adding of Time Series Analysis & Price Variation for Each Airline

Our analysis revealed recurring seasonal patterns and long-term trends in ticket prices, providing valuable insights for stakeholders. We further applied modeling techniques for forecasting price variations, enhancing our understanding of price dynamics in the airline industry. This study underscores the significance of time series analysis in uncovering actionable insights for airlines, travelers, and policymakers alike.
The flight price prediction system was implemented by integrating machine learning models with a user-friendly web interface using Flask, a lightweight and flexible web application framework. Flask was chosen for its modular development approach and compatibility with Python-based machine learning models.

The core of the system involves integrating machine learning models trained on historical flight data. These models were serialized and stored using the pickle library for efficient retrieval during runtime. Flask dynamically loaded these serialized models into memory, enabling real-time predictions based on user inputs.

The user interface was designed using HTML templates to ensure a seamless and intuitive experience. Forms were created to capture essential information such as departure time, arrival time, number of stops, airline preference, source, and destination. Flask's templating engine facilitated the dynamic rendering of these forms, resulting in a responsive and interactive interface.

CSS styles were applied to the HTML templates to enhance visual appeal and user experience. Careful design considerations and responsive styling techniques were employed to optimize the web application for various screen sizes and devices, aiming to deliver a visually appealing and user-friendly interface.

Upon completing the development phase, the Flask application was deployed on a chosen server environment. Whether deployed locally or on a cloud platform, configurations were fine-tuned to ensure smooth operation.

Scalability was also a key consideration, with measures taken to ensure that the application could handle multiple user requests concurrently without compromising performance, marking a significant step toward practical and impactful deployment of our model in Airfare settings.

WORKING OF SYSTEM

![Figure 6: Total Flights Data]
Figure 7:

![Diagram of Flight Fare Prediction System]

Figure 8: Error Rate of Model

```python
# Model Error Values
print('MAE:', metrics.mean_absolute_error(y_test, y_pred))
print('MSE:', metrics.mean_squared_error(y_test, y_pred))

# RMSE = sqrt(mean_squared_error(y_test, y_pred))

MAE: 1184.090219202617
MSE: 3473620.600016843
RMSE: 1863.7651676155028
```

Figure 9:

```
# Normalized RMSE
print('Normalized RMSE:', round(np.sqrt(metrics.mean_squared_error(y_test, y_pred)), 2))
print('Max Value: ', max(y), ' Min Value: ', min(y))
```

Normalized RMSE: 0.06
Max Value: 79512.0
Min Value: 1759.0
Our research introduces a groundbreaking approach by combining machine learning models with time series analysis, revolutionizing the way users access real-time flight information and emerging prices. With a focus on accuracy, our system harnesses the power of Random Forest and other cutting-edge algorithms, achieving an impressive accuracy rate of 85%. Furthermore, through our Flask-based application, users benefit from an intuitive interface, empowering them to input their travel details effortlessly and receive precise predictions. This seamless integration of advanced analytics and web technology not only enhances the user experience but also enables travelers to make informed decisions confidently, optimizing their flight booking process like never before.

VI. REFERENCES


