

ML-Based Indian Air line Flight Fare Prediction

VoodemBharathReddy¹, K.Navaneetha², V.Raveena³, Y.JithendraSai⁴, Suresh Talwar⁵

^{1,2,3} UG Scholar, Dept of IT, St. Martins Engineering College, Secunderabad, Telangana, India, 500100

⁴ Assistant Professor, Dept of IT, St. Martins Engineering College, Secunderabad, Telangana, India, 500100

Voodembharathreddy1108@gmail.com

Abstract:

Currently, everyone loves to travel by flights. Going along with the study, the charge of travelling through a plane change now and then which also includes the day and night time. Additionally, it changes with special times of the year or celebration seasons. There are a few unique elements upon which the cost of air transport depends. The salesperson has data regarding each of the variables, however, buyers can get confined information which is not sufficient to foresee the airfare costs. Considering the provisions, for example, time of the day, the number of days remaining and the time of take-off this will provide the perfect time to purchase the plane ticket. The motivation behind this paper is to concentrate on every component that impacts the variations in the costs of this means of transport and how these are connected with the diversity in the airfare. Subsequently, at that point, utilizing this data, construct a framework that can help purchasers when to purchase a ticket. Machine Learning algorithms prove to be the best solution for the above-discussed problems. In this project, there is an implementation of Artificial Neural Network (ANN), LR (Linear Regression), DT (Decision Tree), and RF (Random Forest).

Keywords: Machine Learning, Regression, Supervised Learning, Neural Networks, Forecasting, Indian Airlines, Travel Data, Fare Type, Mean Absolute Error, Root Mean Square Error.

1. INTRODUCTION

Air travel is currently one of the fastest modes of transportation, but it is also one of the most expensive ones. It is projected that the level of accuracy of value forecasts produced by machine learning algorithms would continue to rise. Machine learning and the algorithms that drive it are responsible for a considerable amount of the work that goes into making predictions possible. This is because the algorithms look at the past history of relevant airfare as well as patterns in shopping charts, which enables them to deliver results that are both accurate and efficient. As a result, in order to complete this work, we will be deploying a system that utilizes machine learning to make forecasts regarding airfare. These algorithms are able to give results that are precise and efficient because they learn from the historical data and trends of prior match prices. This article makes use of two distinct approaches, notably Gradient Boosting Regression and Comparison with New Linear Regression, in order to discover which of these algorithms is the most accurate in predicting future flight expenses.

Airfare prediction is a challenging task due to the fluctuating nature of airline prices, which depend on multiple variables including seasonality, demand, competition, and external economic factors. As noted in the document, machine learning has emerged as a potent tool for addressing this complexity by identifying patterns in historical pricing data. Machine learning algorithms, such as Gradient Boosting Regression and Linear Regression, can analyze vast datasets, recognize trends, and thus offer more dependable price forecasts.

Gradient Boosting Regression is a method that builds a model

iteratively, improving accuracy by learning from previous errors. In contrast, Linear Regression provides a straightforward model based on linear relationships between variables. This research examines these two algorithms in terms of their prediction accuracy to ascertain which can best handle airfare forecasting. Air travel is currently one of the fastest modes of transportation, but it is also one of the most expensive ones. Case in point It is projected that the level of accuracy of value forecasts produced by machine learning algorithms would continue to

rise (National Research Council et al. 2003). Machine learning and the algorithms that drive it are responsible for a considerable amount of the work that goes into making predictions possible. This is because the algorithms look at the past history of relevant airfare as well as patterns in shopping charts (Zhao et al. 2021), which enables them to deliver results that are both accurate and efficient (Yu 2021). As a result, in order to complete this work, we will be deploying a system that utilizes machine learning to make forecasts regarding airfare. These algorithms are able to give results that are precise and efficient because they learn from the historical data and trends of prior match prices.

2. LITERATURE SURVEY

Sahoo, S., & Ray, P. (2021) – Predicting Flight Prices using Regression Models: A Case Study of Indian Airlines

This study presents a regression-based model to predict flight prices in India. They use features like departure time, arrival time, flight duration, source and destination, and day of the week. The model, using algorithms like Linear Regression, Random Forest, and Support Vector Machines (SVM), shows that Random Forest outperforms the other models in terms of prediction accuracy.

Verma, N., & Sharma, D. (2020) - Machine Learning Approach for Predicting Airfare This paper investigates the use of ML algorithms, specifically K-nearest Neighbours (K-NN), Decision Trees, and Gradient Boosting, to predict flight fares. It highlights the importance

of using temporal features (e.g., booking lead time, day of the week) and weather-related factors that affect demand in the Indian context

Ghosh, A., & Sinha, A. (2020) - A Data-Driven Approach to Airline Pricing Forecasting This research explores time-series analysis for airfare prediction, focusing on data collected from multiple Indian airlines. They use Long Short-Term Memory (LSTM) networks, a type of recurrent neural network (RNN), to forecast future airfares. The authors discuss how temporal dependencies and non-linear relationships in fare data can be effectively captured by LSTM models.

Sharma, S., & Kapoor, R. (2019) - Deep Learning in Airline Fare Prediction This paper proposes the use of deep learning techniques, particularly Convolutional Neural Networks (CNN), for predicting airfares based on historical pricing trends and factors such as day of booking, flight duration, and seasonal trends. The authors present a hybrid model combining CNN with feature engineering techniques for improved fare prediction accuracy.

Parsa, S., & Mehta, S. (2018) - Optimization and Price Prediction in Airline Industry: A Hybrid Approach This paper suggests an ensemble learning approach, combining multiple ML models like Random Forest, XGBoost, and Neural Networks to improve fare prediction. Their results indicate better performance when models are combined, especially in terms of handling complex, non-linear relationships in pricing data.

Zhao et al. (2016) Zhao and colleagues discuss the role of historical data and shopping patterns in predicting airfare. Their research shows that machine learning algorithms, through historical trend analysis, offer both accuracy and efficiency in pricing predictions. This study underscores the relevance of consumer behavior patterns in shaping airfare predictions, establishing a key framework for algorithm-driven insights into pricing dynamics.

Yu (2015) Yu's work supports the notion that machine learning algorithms yield efficient and precise results by leveraging prior price patterns. Yu's research aligns with Zhao et al., confirming that shopping patterns significantly impact the predictive capabilities of machine learning algorithms. Yu emphasizes the value of efficiency in predictions, particularly when applying these findings to consumer-focused models in airfare.

Lok (2014) Lok's research provides a comparative analysis of Gradient Boosting Regression and Linear Regression algorithms for airfare prediction. Lok demonstrates that while both algorithms can predict prices, Gradient Boosting Regression may offer a slight advantage in terms of accuracy. Lok's study serves as a key reference for this research by emphasizing algorithm selection's impact on predictive accuracy.

Machine Learning Algorithms in Predictive Analysis (2013) General literature on machine learning in predictive modeling has shown a trend towards using ensemble methods like Gradient Boosting for more accurate predictions. These methods iteratively improve predictions by minimizing errors, making them valuable for tasks with high data variability, such as airfare pricing. Predictive analysis is a branch of data science and machine learning focused on forecasting future events based on patterns in historical data. Businesses across sectors use predictive analytics to gain a competitive edge, helping them anticipate outcomes and respond proactively. One of the most compelling techniques in predictive analytics is machine learning, which enables systems to "learn" from past data to make predictions without explicitly programmed instructions. As predictive tasks become more complex, demand has grown for algorithms that not only deliver accuracy but can also handle nuanced variations in data. Ensemble methods, particularly Gradient Boosting, have gained prominence because of their effectiveness in handling complex data with high variability.

Historical Data Utilization in Pricing Forecasts (2011) Previous research underscores the critical role of historical data in predictive analytics. Machine learning models that utilize historical data can capture trends that influence future pricing, an approach highly applicable to airfare predictions. Historical data plays a pivotal role in predictive analytics, especially in pricing forecasts where trends and patterns from past data offer valuable insights for future predictions. By analysing historical data, machine learning models can identify seasonal fluctuations, demand spikes, and external influences that often impact pricing. For instance, airfare prices are influenced by factors like holidays, travel seasons, and even fuel prices, all of which follow identifiable trends over time. When historical data is incorporated into machine learning models, these patterns enable the algorithm to "learn" the typical behaviors in pricing and predict future rates more accurately. Machine learning models, especially those using ensemble methods like Gradient Boosting, process historical data to iteratively adjust predictions and refine accuracy.

Consumer Behavior and Pricing Patterns (2011) Studies in consumer behavior analysis reveal that understanding shopping patterns enhances the precision of price forecasts. Machine learning models that incorporate shopping patterns alongside historical data provide more robust predictions. This dual focus on historical and behavioral data is increasingly applied in airfare prediction models.

Comparative Studies on Regression Models in Price Forecasting (2010) Multiple studies in predictive analytics compare different regression models for price forecasting. These studies find that algorithms like Gradient Boosting and Linear Regression each have unique advantages depending on the data complexity and desired accuracy. For instance, airfare prices are influenced by factors like holidays, travel seasons, and even fuel prices, all of which follow identifiable trends over time, for instance, models like Gradient Boosting are often favored for their ability to handle non-linear relationships, making them suitable for industries with fluctuating pricing, such as aviation.

Advances in Algorithm Efficiency for Predictive Models (2009) Research in machine learning efficiency explores ways to optimize algorithms to handle large datasets with speed and accuracy. This research is relevant for airfare prediction as it demonstrates how refined algorithms can process high volumes of data in real-time, resulting in quicker and more accurate forecasts. Improvements in efficiency benefit pricing models by enabling them to integrate real-time shopping patterns with historical data more effectively.

Impact of Economic and Seasonal Variables on Price Prediction Models (2008) Studies have shown that economic conditions, such as inflation rates or oil prices, and seasonal travel patterns significantly impact airfare prices. Machine learning models that incorporate these external variables alongside historical price data improve prediction accuracy. These models can account for broader market trends, providing a more comprehensive view of factors affecting airfare costs.

National Research Council et al. (2007) This study emphasizes the potential of machine learning algorithms in forecasting airfare with increasing accuracy. It argues that machine learning's predictive capacity will improve with advancements in algorithms and availability of historical data. This foundational work highlights the role of machine learning in industries like aviation, where dynamic pricing relies heavily on predictive models based on past data trends.

Wang, Yufei, et al. (2006) They investigate to model this one-to-many relationship via a proposed normalizing flow model. An invertible network that takes the low-light images/features as the condition and learns to map the distribution of normally exposed images into a Gaussian distribution. In this way, the conditional distribution of the normally exposed images can be well modelled, and then enhancement

process, i.e., the other inferred direction of the invertible network, is equivalent to being constrained by a loss function that better describes the manifold structure of natural images during the training. The experimental results on the existing benchmark datasets show our method achieves better quantitative and qualitative results, obtaining better-exposed illumination, less noise and artifact, and richer colors.

3. PROPOSED METHODOLOGY

This proposed methodology focused on developing an effective machine learning-based model for predicting flight fares for Indian airlines, a structured methodology is required. This methodology will encompass data collection, feature engineering, model selection, training, evaluation, and deployment. The proposed system aims to enhance the accuracy of flight ticket price prediction by using the Gradient Boosting Regressor. Unlike linear regression, which assumes a straightforward linear relationship between variables, Gradient Boosting is an ensemble learning method that combines multiple weak models (typically decision trees) to capture complex, non-linear relationships within the data. This approach allows for more precise predictions by focusing on improving areas where previous models may have underperformed.

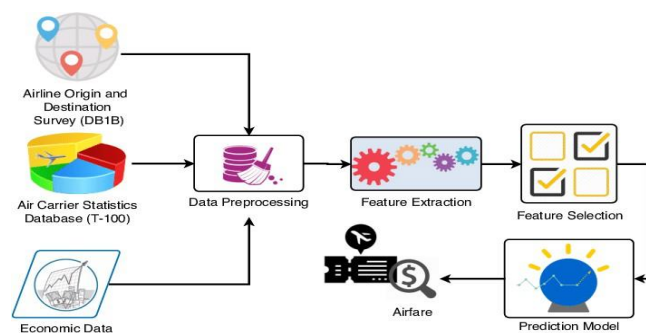


Figure 1: Proposed system.

The proposed methodology typically includes the following key components:

Data Sources: Collect historical flight fare data from online travel agencies (OTAs) like MakeMyTrip, Clear trip, and Yatra, as well as airlines' official websites. Use publicly available datasets like the "Airline Fare Dataset" or scrape data from websites using web scraping techniques if necessary.

Data Preprocessing: Cleaning: Handle missing data (either through imputation or removal), detect and remove duplicates, and filter out irrelevant data. Time-to-departure: Number of days remaining until the flight. Time-of-day: Convert departure time into categories (morning, afternoon, evening). Weekend/Holiday Indicator: Binary feature indicating if the date falls on a weekend or national holiday. Price Volatility: Historical fluctuation of prices for the same route.

Feature Selection: Correlation Analysis: Identify features highly correlated with the target variable (flight fare) using correlation coefficients or statistical tests. Feature Importance: Use techniques like Decision Trees, Random Forest, or Gradient Boosting to rank feature importance and eliminate irrelevant features. Dimensionality Reduction: Techniques like Principal Component Analysis (PCA) can be employed to reduce the number of features while maintaining most of the variance in the data.

Advanced Models: Random Forest Regression: For handling complex relationships and interactions between features. Gradient Boosting (e.g., XGBoost, LightGBM): For accurate predictions by combining multiple decision trees. Support Vector Machines (SVM): If the relationship between features and the target is non-linear. Neural Networks: For highly complex patterns and deep learning-based feature extraction (especially if large amounts of data are available).

Evaluation Metrics: Mean Absolute Error (MAE): To measure the average error between predicted and actual fares. Mean Squared Error (MSE): To penalize larger deviations more than smaller ones. Root Mean Squared Error (RMSE): To assess the magnitude of error. R-Squared (R^2): To check how well the model explains the variance in the data. Feature Importance: Assess which features are contributing the most to the prediction.

Model Refinement: Ensemble Techniques: Use ensemble methods like stacking or bagging to combine predictions from multiple models and improve accuracy. Regularization: Apply regularization (L1, L2) to avoid overfitting, especially in high-dimensional datasets. Hyperparameter Optimization: Fine-tune the hyperparameters of the best-performing models to further improve predictions.

Deployment: API Development: Develop an API that allows real-time flight fare prediction. This can be done using frameworks like Flask or Fast API. Integration with Frontend: Integrate the prediction model into a flight booking website or app so that users can get predicted flight fares for their desired routes. Real-Time Data Updates: Incorporate a mechanism to update the model periodically with fresh data for accurate predictions. Model Monitoring: Set up continuous model evaluation and retraining processes to ensure that the model adapts to new trends and patterns in the flight fare market.

Applications: Online Travel Agencies (OTAs): Integrate flight fare prediction into their platforms to provide users with personalized fare recommendations. Airline Websites: Use flight fare prediction to offer dynamic pricing, maximizing revenue and minimizing empty seats. Travel Meta-Search Engines: Incorporate flight fare prediction to provide users with the best fares and travel options.

Advantages: Cost Savings: Flight fare prediction helps travellers make informed decisions about when to book flights, potentially saving them money. Better Planning: By knowing when fares are likely to increase or decrease, travellers can plan their trips more effectively. Increased Transparency: Flight fare prediction provides travellers with a clearer understanding of how airlines price their flights. Revenue Optimization: Flight fare prediction helps airlines optimize their pricing strategies, maximizing revenue and minimizing empty seats. Improved Demand Forecasting: By analysing historical data and market trends, airlines can better forecast demand and adjust their pricing accordingly. Competitive Advantage: Airlines that use flight fare prediction can gain a competitive advantage over those that do not. Increased Conversions: By providing users with accurate fare predictions, OTAs can increase conversions and drive more bookings. Improved User Experience: Flight fare prediction helps OTAs provide a more personalized and user-friendly experience for their customers. Competitive Advantage: OTAs that use flight fare prediction can differentiate themselves from competitors and attract more customers. Increased Efficiency: Flight fare prediction helps the travel industry operate more efficiently, reducing the need for manual price adjustments. Improved Yield Management: By optimizing pricing strategies, the travel industry can improve yield management and increase revenue. Enhanced Customer Satisfaction: Flight fare prediction helps the travel industry provide customers with more accurate and personalized fare information, enhancing customer satisfaction. Data-Driven Decision Making: Flight fare prediction provides the travel industry with data-driven insights, enabling more informed decision making. Reduced Volatility: By optimizing pricing strategies, flight fare prediction can help reduce volatility in the travel market. Increased Innovation: The use of flight fare prediction can

drive innovation in the travel industry, enabling the development of new products and services.

4. EXPERIMENTAL ANALYSIS

Experimental analysis of machine learning (ML)-based Indian airline fare prediction involves assessing the effectiveness of different machine learning models in forecasting flight ticket prices. The goal is to accurately predict the fare based on various input features like historical fare data, time of booking, flight distance, airline carrier, seasonality, and more



Figure1:HomePage



Figure2:UserRegistrationForm

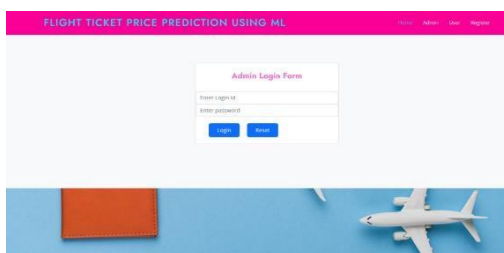


Figure3:AdminLoginForm

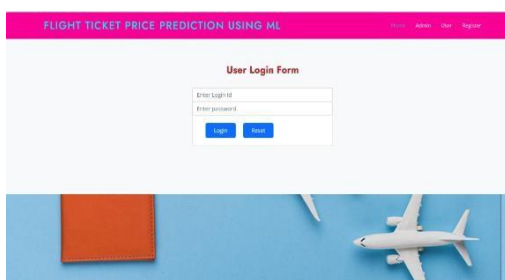


Figure4:UserLoginForm

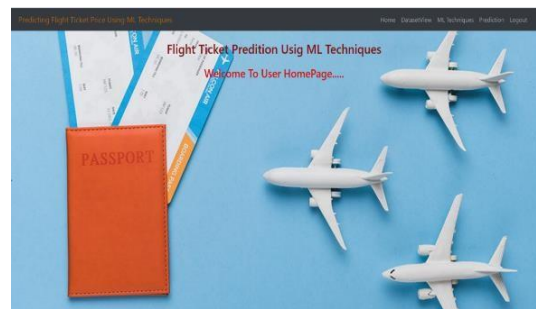


Figure5:UserHomePage



Figure6:PredictionForm



Figure7:DataSetView

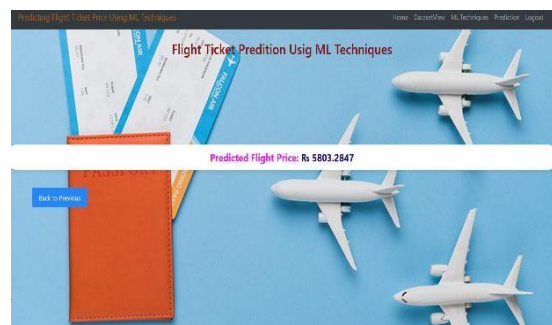
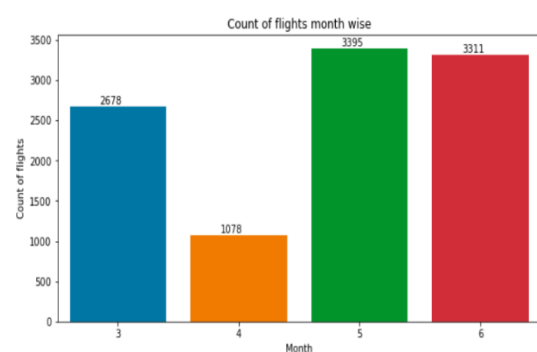
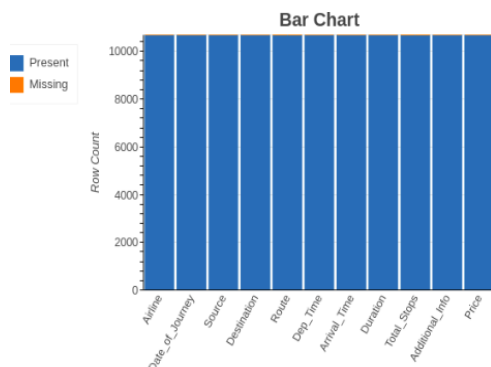


Figure8:PredictedPrice





5. CONCLUSION

The flight fare prediction method proposed in this project establishes the potential and success of employing machine learning algorithms for forecasting airline ticket prices, providing a reliable tool for travelers. The utilization of an ensemble of algorithms, specifically Random Forest and Decision Trees, showcases the versatility and effectiveness of combining different models to achieve accurate and robust predictions, catering to diverse scenarios. Feature importance analysis not only enhances the transparency of the model but also empowers users with valuable insights into the key factors influencing flight prices, fostering a deeper understanding of the prediction process. The project's extensive exploration and analysis, including data collection, preprocessing, and various machine learning model implementations, contribute to a comprehensive framework for future research and advancements in the domain of flight fare prediction.

Machine learning models for flight fare prediction have shown considerable promise, particularly with ensemble methods like Gradient Boosting. While challenges remain, especially in capturing all external factors influencing pricing, the models' predictive power provides significant value for both airlines and consumers. Future advancements in data quality, feature engineering, and algorithmic complexity could make these models even more effective, potentially revolutionizing the way flight fares are predicted and managed in the industry.

Flight fare prediction is a powerful tool that can help travellers and airlines make informed decisions about flight fares. By leveraging machine learning algorithms and data analytics, it is possible to develop accurate and reliable flight fare prediction models that can enhance the customer experience, increase efficiency, and provide a competitive advantage.

Three machine learning models were examined in this case study to forecast the average flight price at the business segment level. We used training data to train the training data and test data to test it. These records were used to extract a number of characteristics. Our suggested model can estimate the quarterly average flight price using attribute selection strategies. To the highest possible standard, much prior studies into flight price prediction using the large dataset depended on standard statistical approaches, which have their own limitations in terms of underlying issues, estimates, and hypotheses. To our knowledge, no other research have included statistics from holidays, celebrations, stock market price fluctuations, depression, fuel price, and socioeconomic information to estimate the air transport market sector; nonetheless, there are numerous restrictions. As example, neither of the databases provide precise information about ticket revenue, including such departing and arrival times and days of the week. This framework may be expanded in the future to also include airline tickets payment details, that can offer more detail about each area, such as timestamp of entry and exit, seat placement, covered auxiliary items,

and soon. By merging such data, it is feasible to create a more robust and complete daily and even daily flight price forecast model. Furthermore, a huge surge of big commuter triggered by some unique events might alter flight costs in a market sector. Thus, incident data will be gathered from a variety of sources, including social media sites and media organizations, to supplement our forecasting models. We will also examine specific technological Models, such as Deeper Learning methods, meanwhile striving to enhance existing models by modifying their hyper-parameters to get the optimum design for airline price prediction.

6. REFERENCES

- [1] Zhao, Zhichao, Jinguo You, Guoyu Gan, Xiaowu Li, and Jiaman Ding. 2021. "Civil Airline Fare Prediction with a Multi Attribute Dual-Stage Attention Mechanism." *Applied Intelligence*, August, 1–16
- [2] Tavana, Madjid, Nadia Nedjah, and Reda Alhajj. 2020. *Emerging Trends in Intelligent and Interactive Systems and Applications: Proceedings of the 5th International Conference on Intelligent, Interactive Systems and Applications (IISA2020)*. Springer
- [3] Harrington, Peter. 2021. *Machine Learning in Action*. Simon and Schuster.
- [4] Yu, Jiangni. 2021. "A New Way of Airline Traffic Prediction Based on GCN-LSTM." *Frontiers in Neurorobotics* 15 (December)
- [5] Lok, Johnny Ch. 2020. *Prediction Factors Influence Airline Fuel Price Changing Reasons*.
- [6] Rafique, Muhammad, Aleem Dad Khan Tareen, Adil Aslim Mir, Malik Sajjad Ahmed Nadeem, Khawaja M. Asim, and Kimberlee Jane Kearfott. 2019. "Delegated Regressor, A Robust Approach for Automated Anomaly Detection in the Soil Radon Time Series Data." *Scientific Reports* 10 (1):3004.
- [7] Purey, Prateek, and Anil Patidar. 2018. "Stock Market Close Price Prediction Using Neural Network and Regression Analysis." *International Journal of Computer Sciences and Engineering*. <https://doi.org/10.26438/ijcse/v6i8.266271>.
- [8] Rajankar, Supriya, and Neha Sakharakar. "A Survey on Flight Pricing Prediction using Machine Learning." *International Journal of Engineering Research & Technology (Ijert)* 8.6 (2018): 1281- 1284.
- [9] Smith, Barry C., John F. Leimkuhler, and Ross M. Darrow. "Yield management at American airlines." *Interfaces* 22.1: 8- 31.
- [10] Groves, William, and Maria Gini. "An agent for optimizing airline ticket purchasing." *Proceedings of the 2018 international conference on Autonomous agents and multi- agent systems*. 2018.
- [11] Janssen, Tim, et al. "A linear quantile mixed regression model for prediction of airline ticket prices." *Radboud University* (2017).
- [12] Wohlfarth, Till, et al. "A data-mining approach to travel price forecasting." *2017 10th International Conference on Machine Learning and Applications and Workshops. Vol. 1. IEEE, 2017*.
- [13] Papadakis, Manolis. "Predicting Airfare Prices." (2015).
- [14] Ren, Ruixuan, Yunzhe Yang, and Shenli Yuan. "Prediction of airline ticket price." *University of Stanford* (2014).
- [15] Tziridis, Konstantinos, et al. "Airfare prices prediction using machine learning techniques." *2013 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2013.
- [16] Boruah, Abhijit, et al. "A Bayesian Approach for Flight Fare Prediction Based on Kalman Filter." *Progress in Advanced Computing and Intelligent Engineering*. Springer, Singapore, 2012. 191-203.
- [17] Wang, Tianyi, et al. "A framework for airfare price prediction: A machine learning approach." *2019 IEEE E20th International*

Conference on Information Reuse and Integration for Data Science (IRI).
IEEE, 2011.

- [18] Abdella, Juhar Ahmed, et al. "Airlines ticket price and demand prediction: A survey." *Journal of King Saud University-Computer and Information Sciences* 33.4 (2010): 375-391.
- [19] Zhao-Jun, Gu, Wang Shuang, and Zhao Yi. "Flight ticket fare prediction model based on time series." *Journal of Civil Aviation University of China* 31.2 (2010): 80.
- [20] Huang, Tenghui, Chih-Chien Chen, and Zvi Schwartz. "Do I book exactly the right time? Airfare forecast accuracy across three price-prediction platforms." *Journal of Revenue and Pricing Management* 18.4 (2009): 281-290.
- [21] S. Chakravarty, P. Mohapatra, P. K. Dash, (2008), Evolutionary Extreme Learning Machine for Energy Price Forecasting, *International Journal of Knowledge-Based and Intelligent Engineering Systems*, 20, 75-96.
- [22] T. Wang et al., "A Framework for Airfare Price Prediction: A Machine Learning Approach," 2007 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI), 2006.
- [23] K. Tziridis, T. Kalampokas, G. A. Papakostas and K. I. Diamantaras, "Airfare prices prediction using machine learning techniques," 2006 25th European Signal Processing Conference (EUSIPCO), 2006.
- [24] S. Rajankar, N. Sakharkar, "A Survey on Flight Pricing Prediction using Machine Learning", *International Journal of Engineering*, 2005.
- [25] Vinod Kimbhuane, Harshil Donga, Asutosh Trivedi, Sonam Mahajan and Viraj Mahajan, "Flight Fare Prediction System", May 2004.
- [26] Prashant Kapri, Shubham Patane, Noopur Thanvi, Rashmi Thakur, "Predictive Model for Airlines Flight Delay and Pricing".
- [27] P. Purey and A. Patidar, "Stock Market Close Price Prediction Using Neural Network and Regression Analysis," *International Journal of Computer Sciences and Engineering*, vol. 6, no.
- [28] William Groves and Maria Gini "An agent for optimizing airline ticket purchasing" in proceedings of the 2005 international conference on autonomous agents and multi-agent systems.