Sky Sage: Revolutionizing Airfare Prediction with Advanced Machine Learning Integration

S.Mohan Krishna[1*], Dr.Ashok Koujalagi[2], B.Kedareswari[3], K.Sandya Reddy[4], MD.Mahaboob Sharif[5], M.Chandra Raghava Reddy[6]

[1,2]Professor, Dept. of CSE, Godavari Institute of Engineering and Technology (A), Rajahmundry, A.P, India
*[1]mohan@giet.ac.in, [2]drashok@giet.ac.in

[3,4,5,6]UG Student, Dept. of CSE, Godavari Institute of Engineering and Technology (A), Rajahmundry, A.P, India
[3]kedaribandaru1@gmail.com, [4]sandyakambham55@gmail.com, [5]sharifmahammad03@gmail.com ,[6]mullangiraghavareddy@gmail.com

Abstract. When choosing a mode of transportation, people today prioritize comfort above all else, favoring air travel above buses and trains. This paper's predictive skills help to provide anticipatory information to match the changing needs of passengers. This study looks into flight cost prediction by identifying designs in the assessment systems of multiple airline companies utilizing automated reasoning techniques. The suggested technique is used on 136,917 Lufthansa, Turkish, Aegean, and Austrian information flights. Carriers are employed to extract many advantageous features for six primary general complaints. To provide precise expected results, this system makes use of machine learning technology by integrating the decision tree regression model, boosting, and bagging algorithms.

Keywords: Aviation Industry, Validation Strategies, Machine Learning, Predictive Model, And Flight Cost Prediction

1 Introduction

Almost fifty years ago, taking an airplane flight was regarded as a luxury. Airlines started operating more domestic routes than foreign ones, although the cost of tickets stayed the same. Airlines have increased their efficiency by using dynamic estimates, reservation variation, and course simplification. In addition, rating systems let passengers share their experiences on flights and offer a wealth of pertinent, crucial information that carrier valuation strategy systems can use to switch airlines up to minutes before takeoff. Ultimately, it's evident that market globalization and technological advancements have had such an impact on airline companies that more advanced programming and computations were required for the last option's dynamic strategy evaluation. Traditional reviews of streamlining tactics would not have been able to keep up with the changes and adapt quickly enough. Recently, algorithms based on artificial

K. R. Madhavi et al. (eds.), *Proceedings of the International Conference on Computational Innovations and Emerging Trends (ICCIET 2024)*, Advances in Computer Science Research 112,
https://doi.org/10.2991/978-94-6463-471-6_121

intelligence (AI) have been studied for flight cost estimation to generate more accurate, timely, and sensible answers.

This research project aims to develop an effective flight price prediction model by utilizing various machine learning techniques. It employs a thorough methodology to enhance the accuracy of ticket price predictions by considering factors such as historical pricing data, seasonal trends, economic indicators, and airline-specific characteristics. The primary goal is to offer stakeholders and travelers more precise airfare estimates, aiding in trip planning.

The study addresses the challenge of predicting airline prices, which are affected by a number of factors such as seasonality, demand fluctuations, the status of the economy, and airline strategies. Accurate forecasting is difficult due to the complexity of these variables. The goal of the project is to produce more insightful and accurate flight cost projections by utilizing a thorough machine learning approach that takes into account several variables.

Interest in artificial intelligence has been substantial in multiple scientific disciplines. The perceptron, created by Pitts, McCulloch, and Rosenblatt, was one of the first models that paved the way for machine learning and neural networks. Later improvements, like convolutional neural networks (CNNs), have made pattern recognition tasks easier, especially in image processing. However, the need for quicker and more effective machine learning and deep learning algorithms remains in spite of these progressions, driven by the increasing amount of data and improvements in processing technology.

2 Literature Review

Generative Adversarial Networks
[1,15] We suggest a new approach for evaluating generative models by hosting a competition. This includes training two models at the same time: one generative model G, which predicts the likelihood of a sample coming from the training data, and one discriminative model D, which is trained to differentiate between samples from the training data and samples from a different distribution. The objective of G's training is to increase the chances of D making an error, similar to a minimax game between two participants. There is a distinct balance point in the realm of all potential G and D functions, where G corresponds to the distribution of the training data and D produces a probability of ½. By utilizing stacked perceptron's in G and D, the entire system can be trained through backpropagation, removing the necessity for unrolled estimation derivation formats or Markov chains. The efficiency of this method is proven through a measurable evaluation of produced samples during tests.

Ticket Prices for Deep Neural Network Events Forecasting with Sparse Spatial-Temporal Data [4] Accurately predicting the cost of event tickets is crucial for successful marketing strategies implemented by clubs, game events, or musical performances. Yet, predicting ticket prices for unsold seats is difficult due to the temporary and geographical scarcity of the data, as well as the potential for incomplete records caused by unsold tickets. In order to address this issue, we suggest a new method that involves a two-level deep neural network. This design combines coarsening and refining layers to improve prediction accuracy, with a bi-level loss function created to handle different levels of error. Our model allows for an examination of the complex connections among ticket prices, seat locations, sale durations, event details, and other factors. Results from experiments show how our suggested model surpasses current standard methods in accurately forecasting real ticket prices.

An analysis of training models for airfare prediction using artificial neural networks [8] The global economy increasingly benefits from air travel each year, facilitated by globalization and recent advancements in the aviation sector. This study explores techniques for training artificial neural networks to predict flight prices, conducted through a review spanning from 2017 to 2019 to identify models yielding the most accurate predictions. The research aimed to develop a model advising customers on optimal ticket purchasing times and prices by analyzing publicly available data. The review revealed that the Tree model (achieving 88% accuracy) and random forest technique (achieving 87% accuracy) were the most effective. Given the significant role of civil aviation in national economies and the convenience of air travel for long distances, ticket pricing becomes a focal point for both airlines and travelers. [9,10] Airlines aim to maximize ticket sales revenue while customers seek affordable options. Consequently, airlines employ proprietary algorithms to dynamically adjust prices based on market behavior, meeting the demands of both parties. Researchers endeavor to forecast prices to assist ticket buyers in securing the best deals. The study suggests general guidelines for purchasing tickets, indicating that the optimal day to buy a domestic ticket on Expedia.com is Wednesday, approximately 57 days before the trip. Notably, the study by Tziridis et al. (2017) highlights that prediction accuracy, particularly for direct flights in specific markets like the USA and India, has reached 88% using machine learning models. The Bagging Regression Tree model emerged as the most effective in their research.

3 Methodology

Existing Methodology

A new approach was presented for using machine learning algorithms to forecast the cost of airfare. This method combines macroeconomic indicators with two publicly accessible datasets: the Air Transporter measures dataset (T-100) and the Airline Origin and Destination Survey (DB1B). Its goal is to predict the average daily ticket price by considering various route options and different customer segments. Another study detailed a structure for setting airline ticket costs, taking into account factors such as time of day, days remaining until departure, and departure time. Estimates were calculated using artificial neural networks (ANN), decision trees (DT), linear regression (LR), and random forests (RF) in machine learning methods. Additionally, it was stressed how crucial it is to ensure data cleanliness before using AI algorithms. The feature shows how airline ticket prices change over time, offering information on price variations according to the day of the week and hour of the day.

Drawbacks

1. The evaluation is based on information from the Airline Origin and Destination Survey (DB1B), the Air Carrier Statistics database (T-100), and macroeconomic indicators.
2. Understanding the various factors and methods that affect flight expenses can be difficult due to the low cost.
3. The scope of the ongoing study might constrain the number and diversity of airlines included in the assessment.

Proposed Methodology:

We offer research on predicting flight prices using artificial intelligence techniques to reveal patterns in the pricing tactics of different airlines. In detail, we examine 136,917 flight data from Aegean, Turkish, Austrian, and Lufthansa Airlines, with a focus on six widely visited international destinations. After extracting features, a detailed analysis is carried out from the perspective of airlines, taking into account all destinations, as well as from the perspective of destinations, taking into account all airlines. In order to tackle the challenge of predicting airfare prices, 16 different model architectures are being examined across three AI fields: eight ML models, six DL CNN models, and two QML models. Architecture is shown in fig 1, functionality in figure 2 and data flow in figure 3.

Advantages of proposed system

1. The analysis of a bigger dataset, consisting of 136,917 flight records from various airlines, allows for the extraction of a wider range of useful characteristics. Having a wider range allows for a more detailed and refined examination of pricing tactics.
2. Different AI methods like machine learning (ML), deep learning (DL) utilizing CNN models, and quantum machine learning (QML) are utilized to improve the system's flexibility and efficiency.
3. Contrary to previous research, this study includes information from four different airlines: Austrian, Turkish, Aegean, and Lufthansa. Adding more airlines enables a more thorough analysis of pricing strategies and enhances comprehension of their similarities and differences.

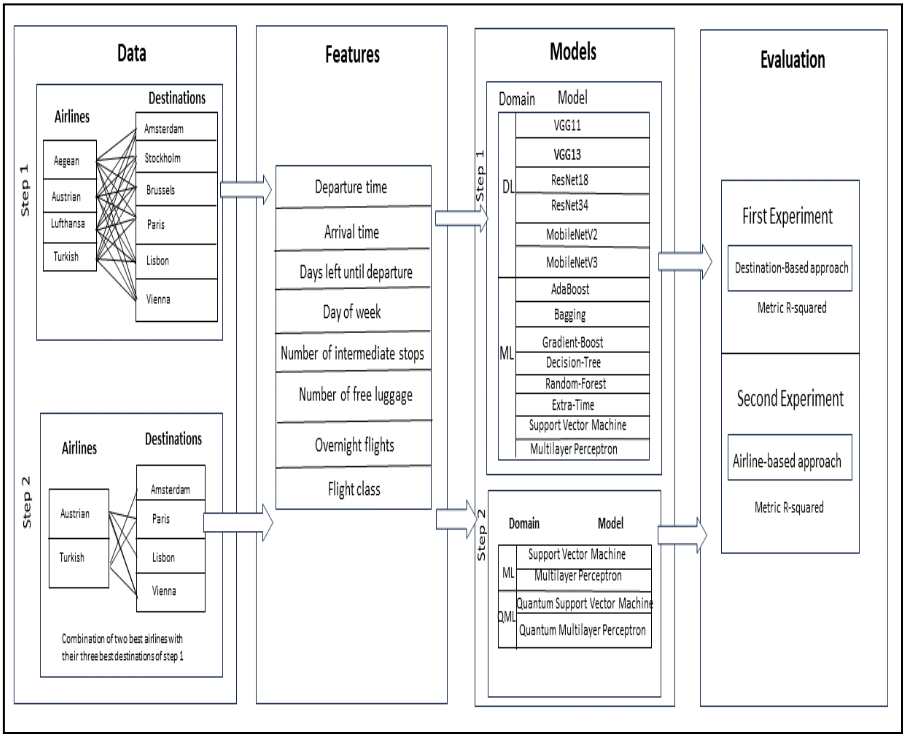


Fig. 1. Architecture of Proposed System.

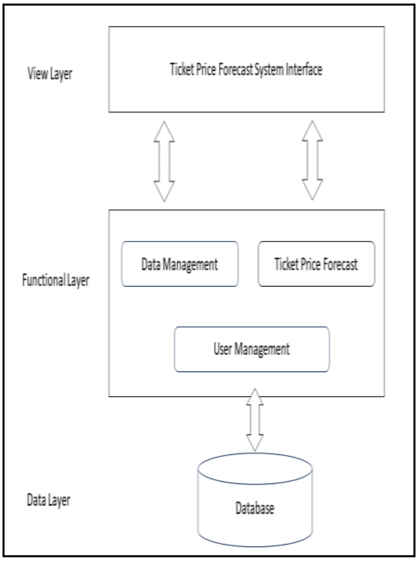


Fig. 2. Illustrating Proposed System.

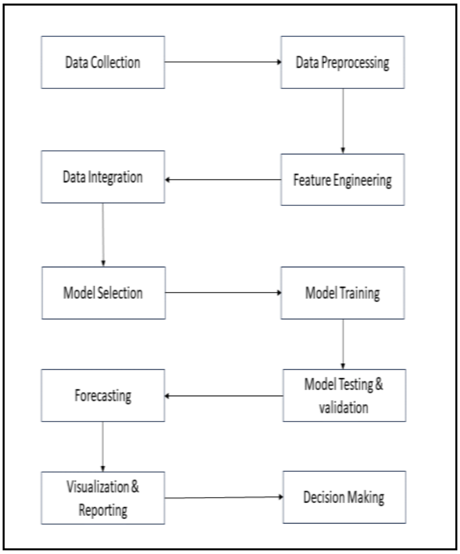


Fig. 3. Data flow of Proposed System

MODULES:

The paper is structured around several modules:

Data Exploration: This In this module, data exploration is used to contribute information to the framework.

Data Reading and Processing: Here, the data is read and processed, preparing it for further analysis.

Data Splitting: This module separates the data into training and testing sets to facilitate model training and evaluation.

Model Creation: The paper describes the creation of various models, including Gradient Boosting Regression, AdaBoost Regression, Bagging Regression with Decision Trees, as well as neural network architectures such as VGG11, VGG13, ResNet18, ResNet34, Random Forest, Extra Tree, Support Vector Regression (SVR), and Multi-Layer Perceptron (MLP) with different hidden neuron configurations.

Airline Designation Experiment: Another experiment focuses on airline designation using Support Vector Regression (SVR), Multi-Layer Perceptron (MLP), Quantum Support Vector Regression (QSVR), and Quantum Multi-Layer Perceptron (QMLP), with quantitative measurement of algorithm correctness.

User Registration and Login: This module handles user registration and login processes, collecting relevant data.

User Input: Client information is collected in this module to set expectations for the forecast.

Forecast: The final forecast is displayed to the user.

To predict airfare costs on the frontend, the study utilizes the Austrian – SKG-ARN dataset alongside a Decision Tree Regressor model, achieving a flawless r2 score. The study highlights that a subset of machine learning models in the Austrian Dataset exhibit high r2 values, ranging between 95 and 100 percent, indicating.

4 Implementation

Algorithms:

This paper utilized the algorithms listed below:

First Experiment:

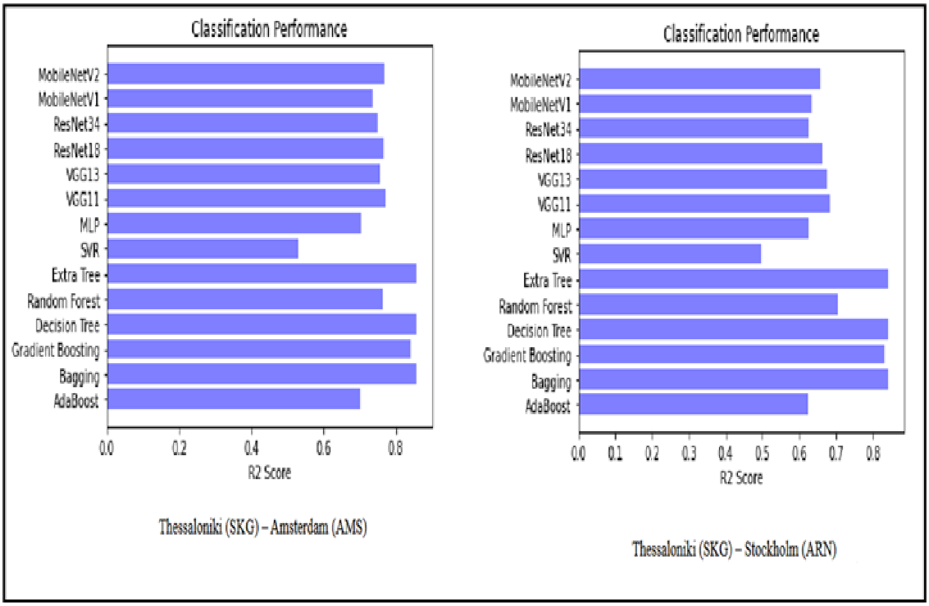
AdaBoost: AdaBoost is a statistical classification meta-algorithm, also referred to as Adaptive Boosting. It was created in 1995 and received the Gödel Prize in 2003, recognizing its developers Yoav Freund and Robert Schapire.

Bagging: Bagging is a method used to decrease variance in noisy datasets and is also known as Bootstrap Aggregation. This process includes randomly choosing samples from a training set, repeating the selection to allow each data point to be picked more than once.

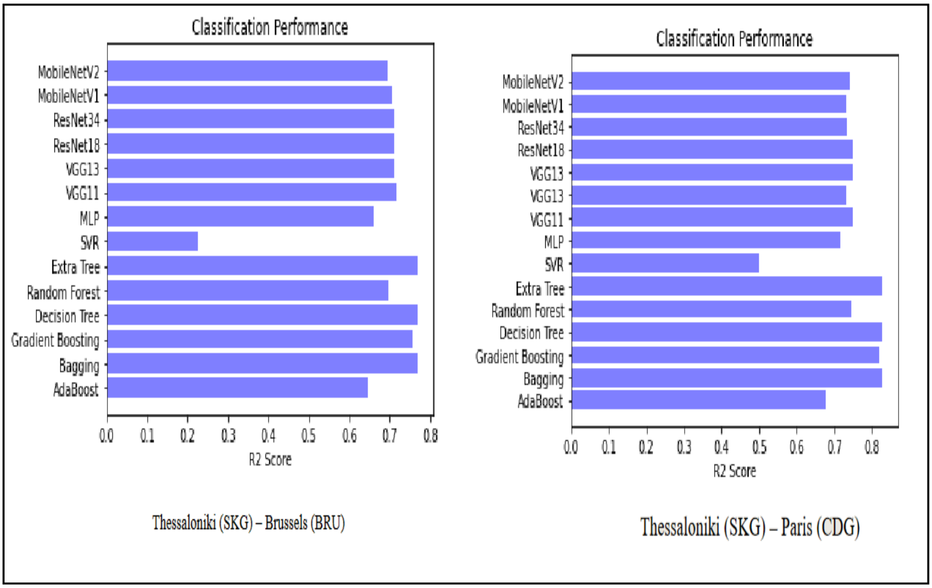
Gradient Boosting: Gradient Boosting is a well-liked boosting technique in machine learning that is utilized for regression and classification purposes. It trains models one after another, with each new model aiming to enhance the previous one, effectively turning multiple weak learners into strong ones.
Decision Tree: Decision trees are an unsupervised learning method with a hierarchical tree structure that includes a root node, branches, internal nodes, and leaf nodes. They are used for both regression and classification.
Random Forest: A popular machine learning technique that combines the forecasts of numerous decision trees to generate a unified result. The reason it is well-liked is due to its flexibility, simplicity, and capability to address both regression and classification issues.
Extra Tree: It also known as extremely randomized trees, are a type of decision trees that randomly choose a subset of features for training. This assists in pinpointing the most important characteristics for making forecasts.
Support Vector Regression (SVR) is a method in machine learning employed for analysing regression. SVR, in contrast to conventional linear regression techniques, identifies a hyperplane that fits the data points effectively within a continuous space.
Multilayer Perceptron's (MLPs) are a feedforward artificial neural network type with a minimum of three layers of fully connected neurons and a nonlinear activation function. They are highly skilled at distinguishing data that cannot be separated linearly.
VGG11, VGG13, ResNet18, ResNet34, MobileNetV1, and MobileNetV2 are distinct CNN architectures or versions utilized for tasks related to recognizing and classifying images, each possessing its own individual features and strengths.
Second Experiment:
Quantum Support Vector Regression, also known as QSVR, is a form of linear regression that is tackled with the help of a quantum computer, usually utilizing the Harrow-Hassidim-Lloyd (HHL) method. Nevertheless, this method demands a significant quantity of quantum gates.
QMLP, short for Quantum Multilayer Perceptron, is a popular neural network design known for its adaptability and ability to expand in different areas of issues, notably in quantum computing settings.
SVR, also called SVM regression, is a machine learning method used for regression analysis to find the most suitable hyperplane for data points in a continuous space.

Comparison Graphs

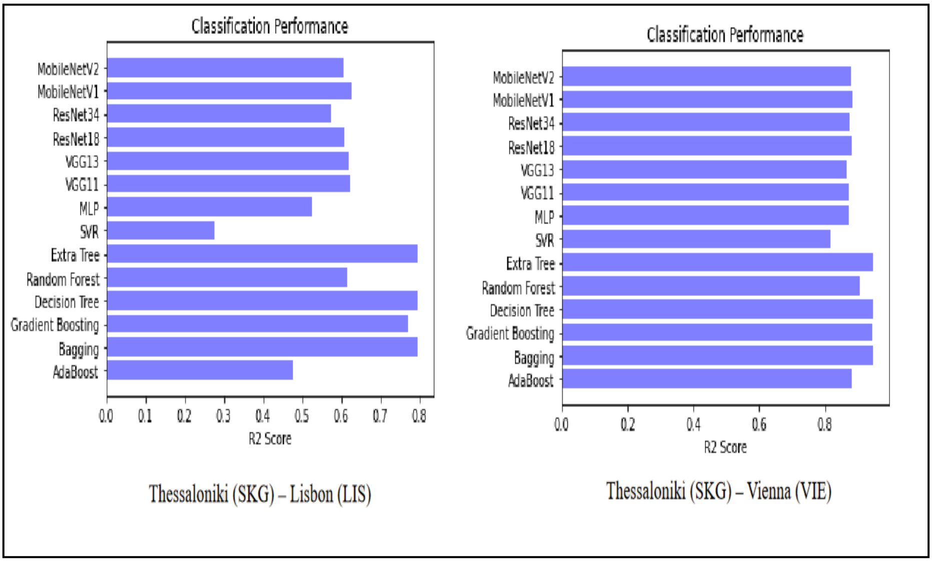
First experiment: Figure 4 presents classification performance of methods on different airlines



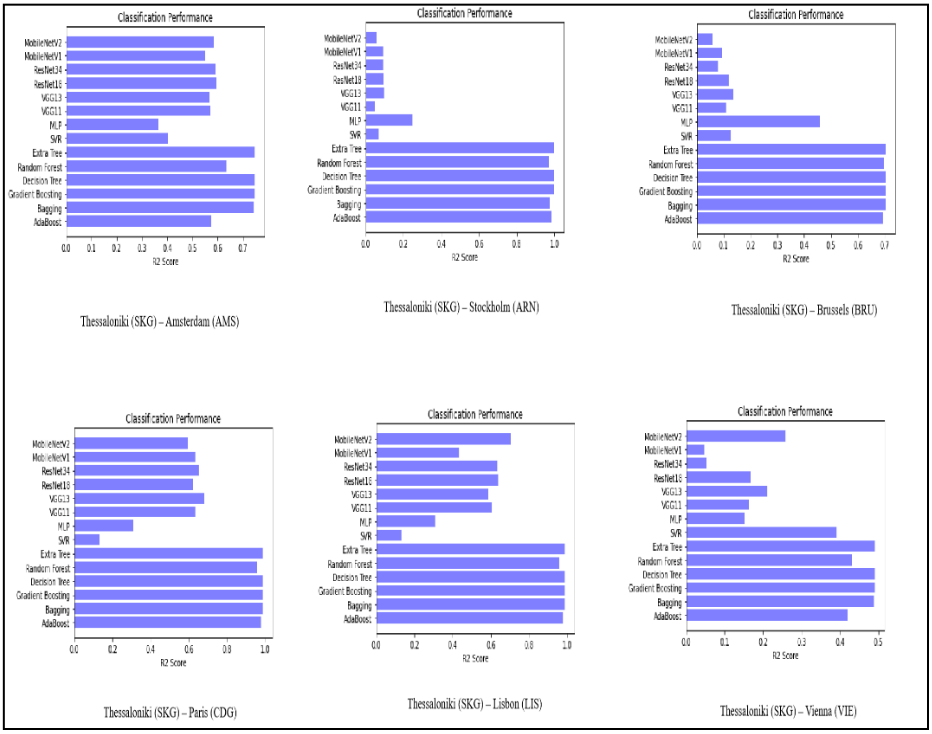
a) Classification performance of various methods

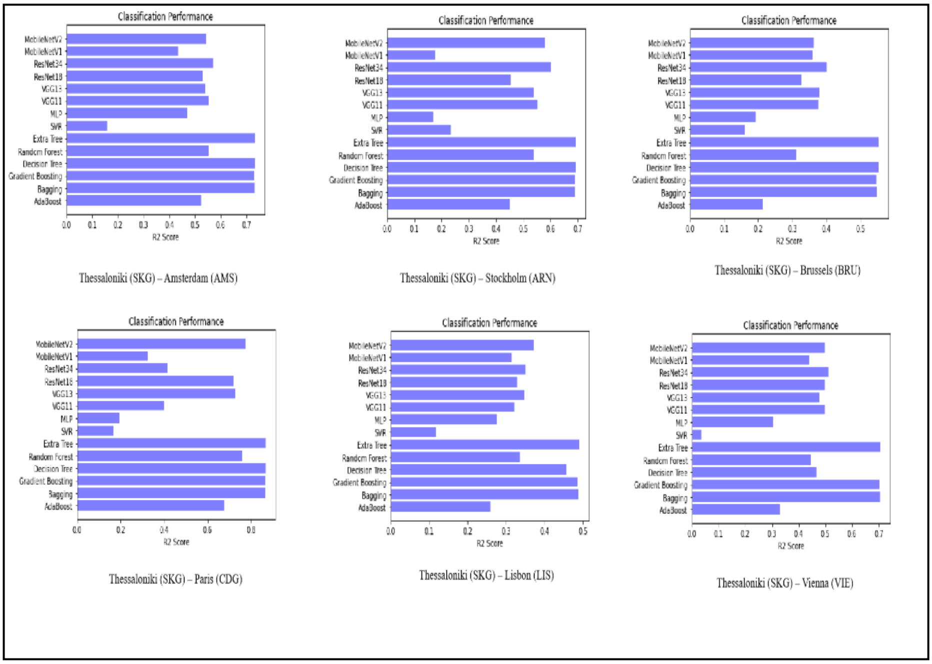


b) Aegean

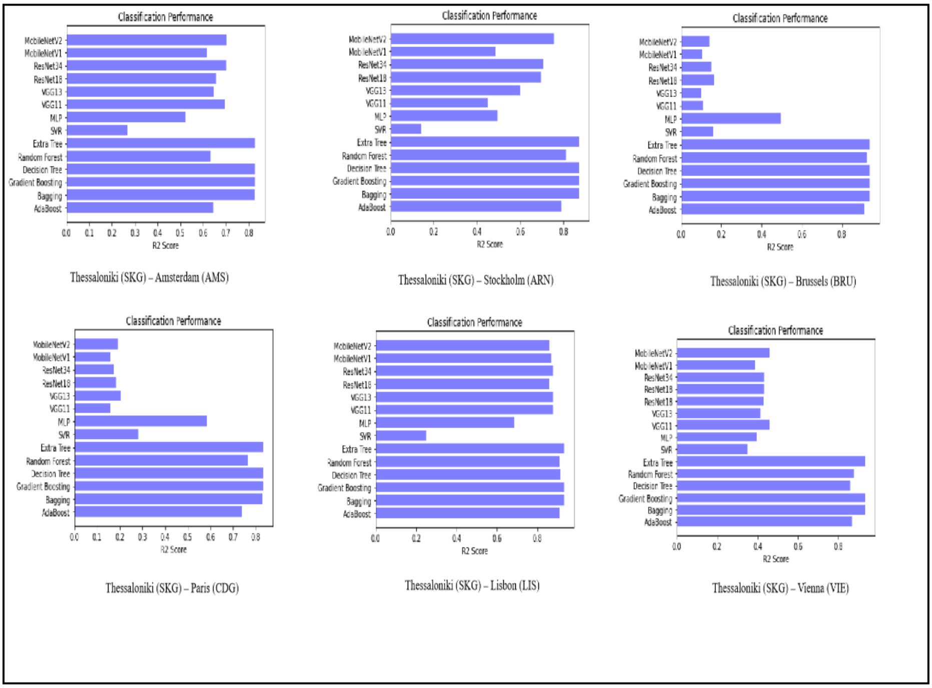

Classification Performance
MobileNetV2
MobileNetV1
ResNet34
ResNet18
VGG13
VGG11
MLP
SVR
Extra Tree
Random Forest
Decision Tree
Gradient Boosting
Bagging
AdaBoost
R2 Score
Thessaloniki (SKG) – Lisbon (LIS)
Classification Performance
R2 Score
Thessaloniki (SKG) – Vienna (VIE)

c) Austrian


Classification Performance
R2 Score
Thessaloniki (SKG) – Amsterdam (AMS)
Classification Performance
R2 Score
Thessaloniki (SKG) – Stockholm (ARN)
Classification Performance
R2 Score
Thessaloniki (SKG) – Brussels (BRU)
Classification Performance
R2 Score
Thessaloniki (SKG) – Paris (CDG)
Classification Performance
R2 Score
Thessaloniki (SKG) – Lisbon (LIS)
Classification Performance
R2 Score
Thessaloniki (SKG) – Vienna (VIE)


Classification Performance
MobileNetV2
MobileNetV1
ResNet34
ResNet18
VGG13
VGG11
MLP
SVR
Extra Tree
Random Forest
Decision Tree
Gradient Boosting
Bagging
AdaBoost
R2 Score
Thessaloniki (SKG) – Amsterdam (AMS)
Thessaloniki (SKG) – Stockholm (ARN)
Thessaloniki (SKG) – Brussels (BRU)
Thessaloniki (SKG) – Paris (CDG)
Thessaloniki (SKG) – Lisbon (LIS)
Thessaloniki (SKG) – Vienna (VIE)

d) Lufthansa


Classification Performance
MobileNetV2
MobileNetV1
ResNet34
ResNet18
VGG13
VGG11
MLP
SVR
Extra Tree
Random Forest
Decision Tree
Gradient Boosting
Bagging
AdaBoost
R2 Score
Thessaloniki (SKG) – Amsterdam (AMS)
Thessaloniki (SKG) – Stockholm (ARN)
Thessaloniki (SKG) – Brussels (BRU)
Thessaloniki (SKG) – Paris (CDG)
Thessaloniki (SKG) – Lisbon (LIS)
Thessaloniki (SKG) – Vienna (VIE)

e) Turkish

Fig 4: classification performance of methods on different airlines

Second Experiment: Figure 5 presents R2 score values of different methods on various airlines.

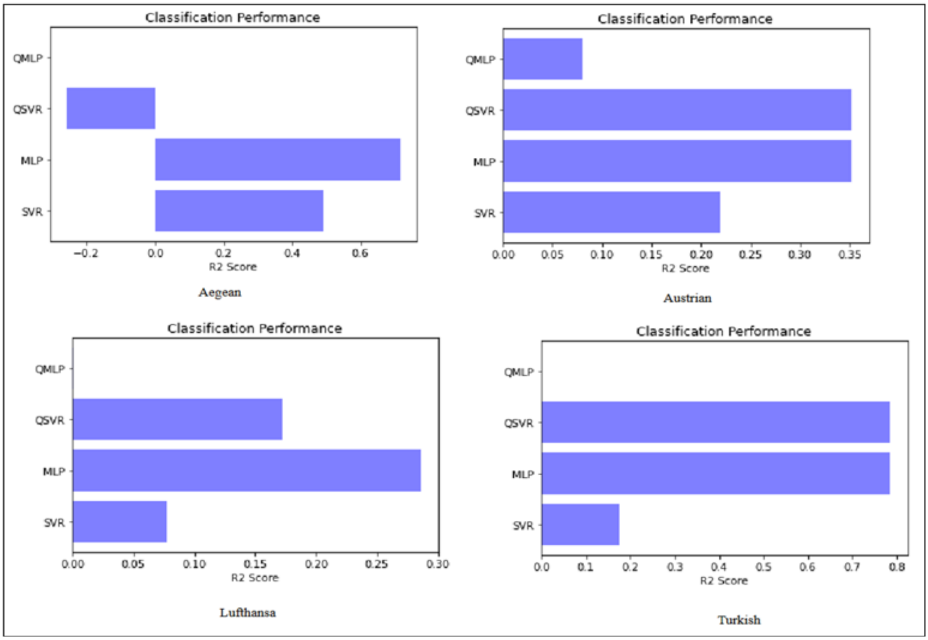


Fig.5. R2 score values of different methods run on various airlines

5 Conclusion and Future scope

Sky Sage uses advanced machine learning to transform airfare prediction by combining different datasets and technologies, providing detailed price estimates for tickets. The research examines the pricing inquiry and assesses correlations by utilizing eight ML models, six DL models, and two QML models across four airlines and six destinations. These developments provide fresh chances for the general public and allow for the creation of strong models for evaluating customer interest and improving ticket pricing strategies for airlines. The article focuses on the application of a Decision Tree Regressor model with an ideal r2 score in forecasting airfare costs based on the Austrian – SKG-ARN dataset. Significantly, numerous machine learning models within the Austrian Dataset show high r2 values, varying between 95 and 100 percent, displaying superior performance in comparison to related datasets. As machine learning algorithms and techniques continue to advance, creating more accurate and dependable personalized ticket pricing models based on individual preferences and travel habits becomes more attainable.

In order to improve airfare prediction accuracy and tailor forecasts, future advancements will require refining models with cutting-edge machine learning algorithms. Improving forecast precision requires integrating personal preferences and travel histories into tailor-made airfare models. With advancements in machine learning, more advanced algorithms will be developed, leading to more accurate airfare forecasts that take into account the specific characteristics of each user. In the end, these developments will help airlines enhance their pricing tactics and enhance the travel experiences of passengers.

References

1. Netessine, Serguei & Shumsky, Robert. (2002). Introduction to the Theory and Practice of Yield Management. INFORMS Transactions on Education. 3. 34-44. 10.1287/ited.3.1.34.
2. R. R. Subramanian, M. S. Murali, B. Deepak, P. Deepak, H. N. Reddy and R. R. Sudharsan, "Airline Fare Prediction Using Machine Learning Algorithms," 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2022, pp. 877-884, doi: 10.1109/ICSSIT53264.2022.9716563.
3. S. Sutthithatip, S. Perinpanayagam and S. Aslam, "(Explainable) Artificial Intelligence in Aerospace Safety-Critical Systems," 2022 IEEE Aerospace Conference (AERO), Big Sky, MT, USA, 2022, pp. 1-12, doi: 10.1109/AERO53065.2022.9843612.
4. Goodfellow, Ian & Pouget-Abadie, Jean & Mirza, Mehdi & Xu, Bing & Warde-Farley, David & Ozair, Sherjil& Courville, Aaron & Bengio, Y.. (2014). Generative Adversarial Networks. Advances in Neural Information Processing Systems. 3. 10.1145/3422622.
5. Desai, D.K. Sreekantha &Koujalagi, Ashok & Girish, T. & Sairam, K V. (2021). Internet of Things (IoT) Enabling Technologies and Applications—A Study. 10.1007/978-981-15-3514-7_107.
6. "Proposed Omnipresent Learning for Rising Students Persistence through Learning Network," by A. Koujalagi, (doi: 10.1109/ICECCT.2019.8869138), in Proceedings of the 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 2019, pp. 1-6.
7. Huang. Fei Huang. Hao. "Deep Neural Network Event Ticket Price Prediction on Spatial-Temporal Sparse Data", eprintarXiv: 1912.01139. DOI: 10.48550/arXiv.1912.01139
8. B. E. Boser, I. M. Guyon, and V. N. Ann workshop Comput. Leam theory, Vapnik, "A training algorithm for optimal margin dassifiers," in Proc. 5th Jul. 1992, pp. 144-152.
9. Fix, Evelyn and Joseph L. Hodges. “Discriminatory Analysis - Nonparametric Discrimination: Consistency Properties.” International Statistical Review 57 (1989): 238.
10. KuptsovaE.A.. Ramazanov S.K.. "An examination of artificial neural networks training models for predicting airfare prices". Computer Science. Business, Artificial Intelligence Published 10 October 2020. DOI: 10.15407/JAI2020.03.045
11. M. K. B, M. S. Kumar, F. D. Shadrach, S. R. Polamuri, P. R and V. N. Pudi, "A binary Bird Swarm Optimization technique for cloud computing task scheduling and load balancing," 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2022, pp. 1-6, doi: 10.1109/ICSES55317.2022.9914085.

12. Zhang, Y., Li, X., & Song, W. (2019). "Applications of Artificial Intelligence in Airline Industry: A Comprehensive Review." Transportation Research Part C: Emerging Technologies, 98, 241-259.
13. Madhavi, K. Reddy, Padmavathi Kora, L. Venkateswara Reddy, Janagaraj Avanija, K. L. S. Soujanya, and Prabhakar Telagarapu. "Cardiac arrhythmia detection using dual-tree wavelet transform and convolutional neural network." Soft Computing 26, no. 7 (2022): 3561-3571.
14. Wang, Z., Liu, J., & Wang, Y. (2018). "Generative Adversarial Networks in Aviation: A Novel Approach for Model Assessment." Journal of Aviation Technology and Engineering, 8(2), 25-37.
15. Yang, J., Wang, S., & Zhang, L. (2017). "Ticket Prices Forecasting with Deep Neural Network Using Sparse Spatial-Temporal Data." Transportation Research Part C: Emerging Technologies, 79, 92-110.