An Integrative Method for Airfare Price Prediction Using AI Techniques

1. RACHAKONDA M D H LAVANYA, Dept. of Computer Science & Engineering, Godavari Institute of Engineering and Technology (Autonomous).

2. Dr. ASHOK KOUJALAGI, Assistant Professor, Dept. of Computer Science & Engineering, Godavari Institute of Engineering and Technology (Autonomous), <u>drashok@giet.ac.in</u>

3. Dr. B. SUJATHA, Professor of CSE, Dept. of Computer Science & Engineering, Godavari Institute of Engineering and Technology (Autonomous), birudusujatha@gmail.com www.giet.ac.in

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 these 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 in use on 136,917 Lufthansa, Turkish, Aegean, and Austrian information flights. Carriers are employed to extract a multitude of advantageous features for six primary general complaints. In order to provide precise expected results, this system makes use of machine learning technology by integrating decision tree regression model, boosting and bagging algorithms.

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

I. 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. By using dynamic estimates, reservation variation, and course simplification, airlines have increased their efficiency. 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 intelligence (AI) are being studied for flight cost estimation in an effort to generate more accurate, timely, and sensible answers.

This research project aims to build a comprehensive and efficient flight price prediction model by utilizing many machine learning techniques. This research employs a comprehensive methodology to enhance the accuracy and reliability of ticket prices by leveraging an array of factors, including past pricing information, seasonal patterns, macroeconomic variables, and airline-specific attributes. The ultimate objective is to provide more precise and current airfare papersions to stakeholders and travelers to assist them in improving trip planning.

This study attempts to solve the problem of airline price's inherent unpredictability and complexity. Seasonality, shifts in demand, the state of the economy, and airline-specific tactics are just a few of the many variables that can affect fluctuations in ticket prices. Accurately forecasting airfare is difficult due to these intricacies. By creating a comprehensive machine learning-based approach that incorporates a number of variables,

this study aims to address this problem by giving travelers and industry participants more precise and insightful flight cost estimates.

Scientists from a variety of fields are very interested in artificial intelligence. Walter Pitts and Warren McCulloch [2] created a mathematical model of a non-learning cell in the actual world in 1943. This served as the model for what is currently referred to as machine learning (ML) or artificial intelligence. Seven years later, in 1950, Frank Rosenblatt created the perceptron [3], the first neural network (NN) with learning capabilities. Many well-known machine learning models, such as SVM [4, 5] and KNN [6, 7], were inspired by the Perceptron.

Without the ability to extract attributes to aid in their development, machine learning models could not advance. The Deep Learning (DL) field achieved this goal, albeit at the expense of increased computer requirements and processing time. Convolutional neural networks, or CNNs, were created in 1980 by Fukushima [7]. CNNs used NNs to recognize patterns in images, which sped up the development of the DL field. In 1990 [8], Yann LeCun gave this attempt a big boost by using CNN models with backpropagation learning to identify the numbers that were manually scribbled on photographs. The feature extraction process can be automated with delayed learning models. Better algorithms and apps that affect people's daily lives can be developed as a result [9, 10]. Nevertheless, faster and smaller ML and DL algorithms are still required due to the rapid expansion of data and advancements in processing technology (GPUs).

LITERATURE SURVEY

Netessine S. and R. The 2002 proposal "Introduction to the theory and practice of yield management" was made by Shumsky, et al. [1].

The 1943 proposal "A logical calculus of the ideas immanent in nervous activity" was made by W. S. McCulloch, W. Pitts, et al. [2].

"The perceptron: A probabilistic model for information storage and organization in the brain" was first proposed by F. Rosenblatt et al. [3] in 1958.

"A training algorithm for optimal margin dassifiers" was proposed by B. E. Boser, I. M. Guyon, V. N. Vapnik, et al. [4] in 1992.

In 1989, E. Fix, J. L. Hodges, et al. [5] introduced "Discriminatory analysis." Consistency properties in nonparametric discrimination ".

Predicting Airline Ticket Prices using Machine Learning Algorithms

[6] This paper discusses airfare. To achieve this goal, many features of a typical flight are selected, with the understanding that these highlights impact the cost of an airline ticket. The cost of a boarding permit varies depending on a number of factors, including the time, location, and length of the journey, as well as unusual events like events or getaways. Therefore, many people will save money and effort when they plan a trip if they have a basic understanding of carrier charges. Following the breakdown of three datasets with experiences into the aircraft fare, the presentation of the seven different Machine Learning (ML) models that are used to measure airline ticket estimating is compared. The objective is to inspect the components that affect flights.

Aerospace Safety-Critical Systems and Explainable Artificial Intelligence (XAI)

UGC Care Group I Journal Volume-14 No.1, 2024

[7] Significant progress has been made in artificial intelligence (AI), particularly in the past ten years. A wide range of aerospace applications, such as air traffic control and the design, operation, construction, and maintenance of aircraft, have made substantial use of various models. In any case, the question is if these AI models incorporate behavior approvals that may further provide assurances that, in a continuous setting, they would continue to function as intended. A unique opportunity to present complex artificial intelligence models to human clients/administrators in a framework that is comprehensible and interpretable is emerging, owing to Explainable AI (XAI), a subfield of AI. This work uses several well-known XAI techniques and classes (the secret parts) to investigate appropriate solutions to such issues that prevent the Aviation simulated intelligence neighborhood from fully elucidating the idea of complex AI models. Therefore, a number of strategies are evaluated to ascertain their degrees of plausibility, including fluffy reasoning, information charts, model skepticism, discovery simulated intelligence, and white-box artificial intelligence. Additionally, the views of designers, insurers, and mediators are explored with regard to the XAI norms for wellness core systems. In order to provide an overview of the XAI skills needed to instill trust in sophisticated AI models, this essay ends by drawing comparisons between different understanding ability levels and the standard elements of the Intelligence Community Directive (ICD).

Generative Adversarial Networks

[8] We propose an interesting methodology for generative model assessment by means of contest. Preparing two models simultaneously: a generative model G that finds out the probability that an example began from the preparation information, and a discriminative model D that learns the probability that an example started from the preparation information as opposed to the information distribution is essential. The objective of G's preparation approach is to improve D's probability of committing an error. This construction is equivalent to that of a minimax game for two players. There is only one arrangement in the arrangement of all potential capabilities G and D, where G gives the preparation information circulation and D is consistently 1/2. When G and D are laid out utilizing layered perceptrons, the whole framework can be prepared by means of backpropagation. You don't require unrolled estimation derivation organizations or Markov chains to produce tests or train your model. The use of the structure is delineated by an immediate and quantitative evaluation of the examples produced in tests.

Ticket Prices for Deep Neural Network Events Forecasting with Sparse Spatial-Temporal Data

[9] Estimating the cost of event tickets is a crucial component of any marketing strategy for a games club, or alternatively, a band. An accurate prediction model could help the marketing team create a more practical and successful promotion strategy. The deal cost is linked to numerous highlights (seat areas, exchange dates, occasion dates, group execution, and so on), and even with all of the verified exchange records, it is difficult to predict the deal cost of the excess seats at any given timestamp due to the dataset's fleeting and spatial sparsity. The cost of a single seat at a game or performance is only apparent during the offer hour. Furthermore, it's possible that some tickets aren't even bought (hence no record open). Information sparsity is in fact a common occurrence in many prediction problems. Here, we suggest using a bi-level ideal deep neural network to overcome the spatio-transient sparsity revile. In order to further improve forecast accuracy, we explicitly combine coarsening and refining layers and build a bi-level misfortune capacity that accounts for several levels of suffering. Our model may be used to investigate the relationships between ticket prices, seat locations, selling periods, event specifics, and other variables. Tests show that the recommended model outperforms the current benchmark comes near in terms of certifiable ticket selling cost prediction.

An analysis of training models for airfare prediction using artificial neural networks

UGC Care Group I Journal Volume-14 No.1, 2024

[10] The global economy benefits more from air travel each year. This is now feasible because to globalization and recent developments in the aviation industry. A This study describes techniques for teaching artificial neural networks to predict prices. Through a review process spanning from 2017 to 2019, the models yielding the best accurate paperions were identified. The research's objective was to develop a model that could advise customers on when it would be best to purchase a ticket at the best price. They intended to do this by analyzing data they had personally acquired that was accessible to the public. A review of related literature found that the Tree model (88 percent accurate) and the random forest technique (87 percent accurate) were the best models. Every nation's economy depends on civil aviation. The most practical and effective way to swiftly travel large distances is by air. Airlines give individuals a variety of options for domestic and international travel. The primary point of contention between airlines and travelers is ticket prices. Airlines want to sell more tickets for more money, but customers want to buy more tickets for less money. Consequently, companies generate constantly fluctuating prices using their own proprietary algorithms, and they keep an eye on the market's behavior to react to shifts in customer demand and rivalry. They are able to satisfy the needs of customers and airlines as a result of their efforts. To enable ticket buyers to purchase tickets at the best possible price, researchers are attempting to forecast prices. The results of the study provide us with general guidelines for making purchases. The best day to buy a ticket on expedia.com for a domestic trip is Wednesday, and the best time to do so is 57 days in advance of the trip, according to the report. The research that the authors did to compare machine learning models is summarized in this paper. According to Tziridis et al. (2017), this discipline has only attained 88 percent prediction accuracy and direct flights to one local market (USA, India). The study found that the Bagging Regression Tree model yielded the best results (Tziridis et al., 2017). This learning software can only forecast the amount of free items and the number of days until departure with an accuracy of 88%.

METHODOLOGY

Existing Methodology:

A method was presented for utilizing machine literacy algorithms to compute flying expenses. The system combines macroeconomic statistics with two publicly available datasets, the Air Transporter measures data set (T-100) and the Airline Origin and Destination Survey (DB1B), to predict the daily average ticket price based on potential morning and objective pairings, or request fragmentation. In a different study, they provide a framework for determining airline ticket prices based on several factors like the time of day, the number of days left, and the departure time. They calculated the estimation using machine literacy techniques such artificial neural networks (ANN), decision trees (DT), linear regression (LR), and random forests (RF). Additionally, they have discussed how important it is to have clean data before utilizing AI algorithms. The component illustrates the evolution of airline ticket prices over time, providing figures for price increases and decreases based on day of the week and hour of the day.

Drawbacks

- 1. Airline Origin and Destination Survey (DB1B), Air Carrier Statistics database (T-100), and macroeconomic statistics form the basis of the current assessment. This could make it more challenging to understand the various components and estimating methodologies that affect the cost of flying.
- 2. This broad spectrum of methods isn't specifically mentioned for the framework, which is currently in development, that computes flight costs using machine learning computations.
- 3. The extent of the current study may place limitations on the quantity and variety of carriers that are taken into consideration for assessment.

Proposed Methodology:

We offer a study of flight price prediction using artificial intelligence techniques to identify patterns in the pricing strategies of various airline firms. More precisely, 136.917 data flights of Aegean, Turkish, Austrian, and Lufthansa Airlines are used to extract a set of useful attributes for six well-known worldwide destinations. After the features have been extracted, a comprehensive analysis is carried out from the standpoint of the customer looking for the cheapest ticket price, taking into account both an airline-based assessment that takes into account all destinations and a destination-based evaluation that takes into account all airlines. Due to the latter, 16 model architectures and three distinct areas of AI models are taken into consideration when solving the airfare price prediction problem: There are eight cutting-edge models in Machine Learning (ML), six CNN models in Deep Learning (DL), and two models in Quantum Machine Learning (QML).

Advantages of proposed system

1. By using a bigger dataset of 136,917 flight data from multiple airlines, the current study, on the other hand, extracts a greater variety of advantageous traits. This broader viewpoint makes it possible to analyze pricing strategies with greater accuracy and subtlety.

2. A range of AI techniques are used in this work, such as machine learning (ML), deep learning (DL) with CNN models, and quantum machine learning (QML).

3. In contrast, the study in this paper makes use of data from four different airlines: Austrian, Turkish, Aegean, and Lufthansa. More airlines mean that pricing plans can be examined in greater detail and the similarities and variations between them may be better understood.



Fig .1 Architecture of Proposed System

MODULES

The Paper is developed by the following modules

- Data exploration: Information will be added to the framework using this module.
- We'll read and process the data.
- This module will separate the data into train and test sets.
- Model creation: Model building The initial trial, which focused on airlines in the following states: Turkish, Aegean, Austrain, and Lufthansa Gradient Boosting Regression, AdaBoost Regression, and Bagging Regression - Regression with Decision Trees - VGG11 (128 Hidden Neurons) - VGG13 (64 Hidden Neurons) - ResNet18 (64 Hidden Neurons) - ResNet34 (516 Hidden Neurons) - Random Forest -Extra Tree - SVR - MLP 32 Hidden Neurons in MObileNEtV1 and 1024 Hidden Neurons in MobileNetV2.
- Airline Designation Based on Second Experiment: SVR, MLP, QSVR, and QMLP Algorithm correctness is measured quantitatively.
- User registration and login: This module gathers registration and login data.
- User input: In order to set expectations, this module collects client information.
- Forecast: The final forecast is displayed.

In order to predict the cost of an AirFare on the frontend, this study used the austrain - SKG-ARN dataset in conjunction with a Decision Tree Regressor model that achieved a flawless r2 score. A small number of machine learning models in the Austrian Dataset have r2 values between 95 and 100 percent, which is high compared to comparable datasets.

II. IMPLEMENTATION

Algorithms:

Here in this project we are used the following algorithms

First experiment:

AdaBoost: Also known as Adaptive Boosting, AdaBoost is a statistical classification meta-algorithm that was developed in 1995 and for which Yoav Freund and Robert Schapire were awarded the Gödel Prize in 2003.

Bagging: Another name for this ensemble learning technique that's frequently used to lower variance in a noisy dataset is bootstrap aggregation, or bagging. In bagging, a training set's random sample of data is picked with replacement, allowing each data point to be selected more than once.

Gradient Boosting: Applied to regression and classification applications, gradient boosting is a well-liked boosting approach in machine learning. One type of ensemble learning technique is called "boosting," in which the model is trained successively, with each new model attempting to improve upon the one before it. It turns a number of ineffective learners into effective ones.

DT: A non-parametric managed learning approach that is utilized for both relapse and characterization applications is the decision tree. With a root hub, branches, inward nodes, and leaf nodes, it has a hierarchical tree structure.

RF: Leo Breiman and Adele Cutler are the brand name holders of the broadly utilized AI method known as "random forest," which totals the result of a few decision trees to deliver a solitary outcome. Its adaptability and convenience, joined with its capacity to deal with both relapse and grouping issues, have driven its prominence.

UGC Care Group I Journal Volume-14 No.1, 2024

Extra Tree: An approach for classification and regression tasks is called an extra tree. A Decision Tree is trained on a subset of features that are chosen at random. After that, the tree is only left with the characteristics that are most crucial for making predictions.

SVR: Support Vector Regression, often known as SVM regression, is a machine learning technique that is applied to regression analysis. As opposed to fitting a line to the data of interest, which is to be expected with normal direct relapse techniques, it decides a hyperplane that best matches the data of interest in a constant space.

Multilayer perceptrons, or MLPs: for short, are a misnomer for a type of contemporary feedforward artificial neural network. They are comprised of no less than three layers of completely associated neurons with a nonlinear enactment capability, and are recognized by their ability to isolate information that isn't directly distinguishable. It is a misnomer because, unlike modern networks, the original perceptron employed a nonlinear type of activation function—the Heaviside step function.

VGG11: There are three completely linked nodes in VGG-11. The diagrammatic representation of the 4096 channels in the upper two fully-connected have been supplied in the image below. There are 1000 channels in the third fully connected tier. this is so that each fits into a single class.

VGG13: "Very Deep Convolutional Networks For Large-Scale Image Recognition" is the configuration "B" of the VGG 13-layer model. The model requires a minimum input size of 32 by 32. pretrained (bool): Returns a model that has been pre-trained on ImageNet if True.

ResNet18: ResNet-18 is an eighteen-layer convolutional brain organization. A pretrained rendition of the organization, prepared on more than 1,000,000 photographs from the ImageNet assortment, is accessible for download. Pictures of 1000 different item classifications, including a console, mouse, pencil, and various creatures, can be ordered by the pretrained network. Consequently, a vast array of image rich feature representations have been trained by the network. The network can handle 224×224 picture input sizes.

ResNet34: Pre-trained on the ImageNet dataset, ResNet 34 is an image categorization model. The architecture outlined in the paper "Deep Residual Learning for Image Recognition" in the TorchVision package (see here) is the basis for this PyTorch* implementation. A blob made up of a single image with the RGB values 1, 3, 224, and 224 is the model's input.

MobileNetV1: A review is conducted on Google's MobileNetV1. To minimize the size and complexity of the model, Depthwise Separable Convolution is employed. Applications involving mobile and embedded vision can benefit most from it.

MobileNetV2: MobileNet-v2 is a 53-layer profound convolutional brain organization. A pretrained variant of the organization, prepared on more than 1,000,000 photographs from the ImageNet assortment, is accessible for download [1]. Pictures of 1000 different item classifications, including a console, mouse, pencil, and various creatures, can be grouped by the pretrained network.

Second Experiment:

QSVR: Using a quantum computer, a quantum support vector machine solves this linear problem. Typically, the Harrow-Hassidim-Lloyd (HHL) technique is used to solve linear equations. This approach, however, needs a large number of quantum gates.

QMLP: Because of its versatility and capacity to scale for a wide range of problem solutions, the multilayer perceptron is a well-liked neural network architecture.

Page | 14

UGC Care Group I Journal Volume-14 No.1, 2024

SVR: Support Vector Regression, often known as SVM regression, is a machine learning technique that is applied to regression analysis. As opposed to fitting a line to the data of interest, which is to be expected with commonplace straight relapse strategies, it decides a hyperplane that best matches the data of interest in a ceaseless space.

Multilayer perceptrons, or MLPs: for short, are a misnomer for a type of contemporary feedforward artificial neural network. They are comprised of no less than three layers of completely associated neurons with a nonlinear initiation capability, and are recognized by their ability to isolate information that isn't straightly divisible. It is a misnomer in light of the fact that, dissimilar to present day organizations, the first perceptron utilized a nonlinear kind of enactment capability — the Heaviside step capability.



Fig.2 Block diagram of Proposed System

Fig.3 Data flow of Proposed System

III. EXPECTED RESULTS

	Welcome
WELCOME TO DASHBOARD	2 Americanian
A Holistic Approach	Anaros .
On Airfare Price	
Prediction Using	Advector Resident

Fig 2 Home Page

Fig 3 Signup Page



A LA LA LA
OUTCOME: AIRFARE PRICE IS \$ 2070.58 BASED ON INPUT DATA:
ц.

Fig 8 Upload another input values



Comparison graphs

Second experiment:



Classification Performance QMLP OSVR MLP SVR. 0.10 0.15 0.20 0.00 0.05 0.25 0.30 R2 Score Lufthansa

Classification Performance QMLP QSVR MLP SVR 0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 R2 Score

Austrian



Turkish

First experiment:



Thessaloniki (SKG) - Amsterdam (AMS)



Thessaloniki (SKG) - Stockholm (ARN)

UGC Care Group I Journal Volume-14 No.1, 2024







UGC Care Group I Journal Volume-14 No.1, 2024



Thessaloniki (SKG) - Paris (CDG)



Austrian:

26 27

9.5



Thessaloniki (SKG) - Stockholm (ARN)





Thessaloniki (SKG) - Brussels (BRU)



0.4 RJ Scott

Thessaloniki (SKG) - Amsterdam (AMS)

6.3

25 97

Thessaloniki (SKG) – Paris (CDG)



Thessaloniki (SKG) - Lisbon (LIS)



Thessaloniki (SKG) - Vienna (VIE)

Lufthansa:



Thesnaloniki (SKG) - Amsterdam (AMS)



Thessaloniki (SKG) - Paris (CDG)



Thessaloniki (SKG) - Stockholm (ARN)





Thessaloniki (SKG) – Brussels (BRU)



Thessaloniki (SKG) - Vienna (VIE)

Turkish:



IV. CONCLUSION AND FUTURE SCOPE

A variety of datasets and technologies are combined in SkySage: Revolutionizing Airfare Prediction with Advanced Machine Learning to provide comprehensive ticket price estimation. Four carriers and six locations were taken into account. Eight ML models, six DL models, and two QML models were used to address the question, and their correlations were examined. Additional capabilities are now accessible to the general public, and the aforementioned developments can be leveraged to establish robust models for the assessment of customer interest and the valuation enhancement of carrier tickets, giving aircraft companies extensive data to support their ideal evaluation scheme.

This article used a Decision Tree Regressor model with a perfect r2 score to predict the front-end pricing of an AirFare using the austrain - SKG-ARN dataset. A small number of machine learning models in the Austrian Dataset have r2 values between 95 and 100 percent, which is high compared to comparable datasets.

With the advancement of machine learning algorithms and techniques, more precise and reliable customized ticket models based on personal preferences and travel patterns are now possible.

FUTURE SCOPE:

In order to deliver even more precise and customized forecasts, future advancements in airfare prediction will require fine-tuning models through the use of cutting-edge machine learning algorithms. Prediction accuracy will increase when customized airfare models include individual preferences and trip history. As machine learning techniques continue to progress, complex algorithms will emerge, making airfare predictions more accurate and

UGC Care Group I Journal Volume-14 No.1, 2024

representative of individual user profiles. This will ultimately help airlines optimize their pricing policies and enhance passengers' travel experiences.

REFERENCES

1. S. Netessine and R. Shumsky, "Introduction to the theory and practice of yield management," INFORMS Trans Educ, vol. 3, no. 1, pp. 34-44, Sep. 2002.

2. W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," Bull. Math Biophys, vol. 5, no. 4, pp. 115-133, Dec. 1943.

3. F. Rosenblatt, "The perceptron A probabilistic model for information storage and organization in the brain" Psychol. Rev., vol. 65, no. 6, pp. 386-408, 1958.

4. B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin dassifiers," in Proc. 5th Ann workshop Comput. Learn theory, Jul. 1992, pp. 144-152.

5. E. Fix and J. L. Hodges, "Discriminatory analysis. Nonparametric discrimination Consistency properties," Int Stat. Rev/Revue International de Statuesque, vol. 57, no. 3, p. 238, Dec. 1989.

6. R. R. Subramanian, M. S. Murali, B. Deepak, P. Deepak, H. N. Reddy, and R. R. Sudharsan, "Airline fare prediction using machine learning algorithms" in Proc. 4th Int. Conf. Smart Syst Inventive Technol. (ICSSIT), Jan. 2022, pp. 877-884.

7. S. Sumthithutip, S. Perinpanayagam, and S. Adam, "Explainable) artificial inteligence in aerospace safety critical systems," in Proc. IEEE Aroup Corf (AERO), Mar. 2022, pp. 1-12.

8. Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, "Generative Adversarial Networks" published in Statistics, Machine Learning, DOI: https://doi.org/10.48550/arXiv.1406.2661

9. Huang, Fei ; Huang, Hao, "Deep Neural Network Event Ticket Price Prediction on Spatial-Temporal Sparse Data", eprint arXiv:1912.01139, DOI: 10.48550/arXiv.1912.01139

10. Kuptsova E.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