Vol.15, Issue No 2, 2025

A MACHINE LEARNING-BASED HOLISTIC FRAMEWORK FOR AIRFARE PRICE PREDICTION

[1] Pinjari Moulali, MCA Student, Department Of MCA
[2] Ch Srilakshmi Prasanna, M. Tech,(Ph.D),Assistant Professor, Department Of MCA
[12]Dr KV Subba Reddy Institute of Technology, Dupadu, Kurnool
https://doi.org/10.51470/ijcnwc.2025.v15.i02.pp414-423

ABSTRACT

Professionals have developed new pricing plans and methods as a result of market globalisation, which boosts worldwide competitiveness. In order to determine the best pricing strategy, airline firms often adjust the cost of tickets by taking into account a number of variables based on their own proprietary rules and algorithms. Artificial Intelligence (AI) models have been used recently for the latter job because of its various potentials in data generalisation, compactness, and quick adaptation. This study uses artificial intelligence (AI) techniques to analyse ticket price prediction in order to identify commonalities in the pricing strategies of various airline firms. More precisely, 136.917 data flights of Aegean, Turkish, Austrian, and Lufthansa Airlines for six well-known worldwide locations are used to extract a set of useful attributes. After the characteristics have been retrieved, a comprehensive analysis is carried out from the viewpoint of the customer looking for the cheapest ticket price, taking into account both an airline-based assessment that includes all destinations and a destination-based evaluation that includes all airlines. For the latter, a total of 16 model architectures from three distinct domains—Machine Learning (ML) with eight cutting-edge models, Deep Learning (DL) with six CNN models, and Quantum Machine Learning (QML) with two models—are taken into consideration to solve the airfare price prediction problem. According to experimental data, for various foreign locations and airline businesses, at least three models from each domain—ML, DL, and QML—can achieve accuracies between 89% and 99% in this regression task.

I. INTRODUCTION

Air travel was seen as a luxury around fifty years ago. While flight ticket prices remained unchanged, airline firms were introducing more local routes than foreign ones. Airlines used management and cost-effective software systems to optimise routes, adjust reservations, and implement dynamic pricing in order to boost profitability. The use of yield management [1], a variable pricing approach centred on comprehending, predicting, and influencing customer behaviour in order to maximise revenues, was a development in airline industry. Consequently, airline firms began to focus more on the preferences and experiences of their consumers during flights while also expanding their worldwide locations. Because airline firms grew more competitive due to dynamic pricing and more flight options, airline flights were available to all prospective customers. Furthermore, the ability to purchase online has transformed a number of industries in recent years and has become popular with contemporary people who are looking for the best deals and rates. In order to get the best travel offers, there are now a number of websites that facilitate safe flight booking and show the same flight itineraries from all airline companies. Additionally, the rating systems that airline passengers use to share their flight experiences provide a wealth of valuable information that pricing policy systems utilise to adjust ticket prices even minutes before a departure. In light of this, it is evident that technological advancements and market globalisation have had a significant impact on airline firms, to the point where standard pricing optimisation systems would not be able to keep up with the changes and adjust quickly enough. The latter raised the need for more complex software and algorithms for optimising dynamic pricing policies. In an effort to provide faster, more accurate, and more efficient results, artificial intelligence (AI) algorithms are now being studied for estimating ticket prices.

The scientific community is very interested in artificial intelligence across a wide range of disciplines. When a mathematical model of a biological neurone without learning capabilities was created in 1943, Walter Pitts and Warren McCulloch [2] established machine learning (ML), the first area of artificial intelligence. Frank Rosenblatt introduced the perceptual (()) [3] as the first neural network (NN) with learning capabilities seven years later, in 1950. Researchers used perception as inspiration to create and apply many popular machine learning models, including SVM [4], KNN [5], and boosting techniques [6]. ML models need a supporting feature extraction strategy in order to generalise successfully. The Deep Learning (DL) domain met the latter criteria, which resulted in a reduction in execution time and an increase in computing needs. The invention of convolution neural networks (CNN) [7] by Fukushima in 1980, who used a NN for visual pattern recognition, served as the cornerstone for the emergence of the DL domain. In 1990, Yann LeCun [8] made a significant contribution to this endeavour by using CNN models with back propagation learning to identify handwritten numerals from pictures. By automating the feature extraction process, DL models have made it possible to create more intricate algorithms and applications that have an influence on people's everyday lives [9], [10]. Faster and more compact ML and DL algorithms are still required, nevertheless, because of the enormous pace of data expansion and the advancement of processing technology (GPUs).

Combining quantum physics under quantum computing techniques with ML and DL approaches is an effort to get beyond the constraints of these algorithms. In the 1990s, the field of quantum computing emerged, with the proposal of quantum algorithms to address difficult issues such as number factorisation. either Grover's algorithm from 1997 [12] or Shor's algorithm from 1994 [11]. IBM was in the forefront of the development of quantum computers, which were made possible by these algorithms. The advent of quantum neural networks (QNN) [13] in 1999, which combined Grover's algorithm and quantum circuits to simulate neural network processes, marked the beginning of the growth of quantum machine learning (QML) over the same decade. Numerous researchers experimented with the QML domain as a result of that study. Thus, several QML algorithms, such as the quantum multilayer perceptron (QMLP) [14], quantum support vector machine (QSVM) [15], and others, were developed between 1990 and 2010. Despite the restricted amount of accessible quantum hardware and the high computational demands of QML models for a classical computer, quantum machine learning is still growing in the market today with applications and methods that are implemented on genuine quantum hardware. Furthermore, a lot of QML techniques are quite similar to classical techniques. For example, QNN training for classical data computes the optimiser and loss using their classical form. The aforementioned factors accelerated QML's pace of development in both research and the market.

This work is a follow-up to a prior study on predicting airfare prices [16]. In order to show the degree of competition between various airline companies and destinations, a set of features that define a typical flight are extracted and used under the scheme of airfare price prediction. Additionally, the performance and applicability of the various ML, DL, and QML models is examined holistically in airfare price prediction. The first experiment examines the issue from the standpoint of the destination (destination-based approach) for every airline company. The AI models from the three aforementioned domains are then applied to the same set of destinations for various airline companies to show similarities in the models' performance. The second experiment applies the ML, DL, and QML models to datasets for every airline company (airline-based approach), regardless of the destination. Emphasis should be placed on the fact that this study is the first documented effort to tackle the issue of ticket price prediction holistically, examining both techniques from the perspectives of airline businesses and destinations. Furthermore, to the best of our knowledge, QML has never been used to solve the issue of predicting flight prices. In light of the aforementioned, the proposed work's primary contributions may be summed up as follows:

1) Examining the relationship between various airline companies' pricing policies.

2) Examining how attributes affect the challenge of predicting flight costs.

3) This is the first time that QML models have been used in the literature to forecast flight prices.

4) Evaluation of the relative effectiveness of ML, DL, and QML models for forecasting flight prices.

This is how the remainder of the paper is structured: The relevant work on predicting airfare prices is included in Section II. Materials and techniques are presented in Section III, with reference to the data and algorithms used in the execution of this study. The experimental setup is explained in Section IV, and the findings are shown and discussed in Section V. Results from quantum machine learning are shown and contrasted with classical models in Section VI. The work is finally concluded in Section VII, which also offers further possible research areas.

PURPOSE OF THE PROJECT

This project's main goal is to provide a comprehensive strategy for predicting flight prices by using cutting-edge Machine Learning (ML) techniques in conjunction with Deep Learning (DL) and Quantum Machine Learning (QML) approaches. Strong and flexible prediction models are necessary due to the dynamic and intricate nature of airline pricing schemes, which are impacted by several internal and external variables. As end customers' demands for cost-effectiveness and transparency increase, this initiative seeks to:

1. Using historical airfare data, examine price trends and parallels among the main airlines (Aegean, Turkish, Austrian, and Lufthansa).
2. To enable efficient model training and assessment, identify and engineer key characteristics from a dataset of 136,917 flights.
3. Use 16 architectures to apply and evaluate predictive models from three domains—ML, DL, and QML—to the airfare prediction job as a regression issue.
4. Examine ticket costs from both an airline-based (comparing pricing trends across airlines) and destination-based (finding the most economical choices for a certain place) standpoint.
5. Use data-driven insights to help end customers find the most economical travel alternatives, improving the decision-making process.

EXISTING SYSTEM

The growth of airline pricing policies and market globalisation produced a wealth of pertinent data, which in turn led to a high level of academic interest in airfare price prediction. This information is converted into data with several properties and in quantities that may be referred to as big data in terms of artificial intelligence and data analysis, particularly given the rapid changes in airline ticket rates and services. According to a study by Abdella et al. [17] on the target application issue and the solutions, the airfare price prediction problem may be used in a variety of contexts, such as customer segmentation, ticket purchase timing, air ticket demand forecast, and more. Generally speaking, the topic of airline price prediction has been in the news for thirty years; a search for the keyword "airfare price prediction" on Scopus yielded twenty-four papers from 2003 to the present, the majority of which were completed during the previous three years.

Using two machine learning models, Vu et al. [18] developed an application for predicting ticket prices by using time-related data to characterise flights operated by the Vietnamese national airline. While the primary emphasis was on the target applications of customers, fewer models were offered and just one airline firm was taken into consideration in comparison to the suggested strategy. A alternative strategy was introduced in [19]. In order to anticipate flight prices for events such as basketball games, a bespoke recurrent neural network (RNN) was built and contrasted with traditional machine learning models. High prediction accuracies were obtained by combining features that characterised airline trips and basketball games into a single dataset. In [20], the similar strategy was used. The authors suggested a system that could collect data on airline tickets from several sources, including distance, customer interests, and ticket availability, in order to use machine learning models to forecast flight costs. Predicting flight prices was introduced in the domestic markets of India and the United States in [21].

By using machine learning models, the authors were able to anticipate prices with an 88% accuracy rate. A similar strategy using fewer machine learning models was used by Joshi et al. in [22], who investigated novel characteristics such as flight time and were able to attain prediction scores of up to 90%. In [23], feature selection algorithms and hyperparameter techniques were used to determine the best model parameters and feature set for flight description in order to forecast ticket prices. In order to deliver reliable and comprehensible forecasts, [24] provided explainability for the issue being studied in order to get a greater understanding of the models that may offer an effective solution.

Disadvantages

• Data complexity: To identify airfare price prediction, the majority of machine learning models now in use need to be able to effectively comprehend huge and complicated information.

• Data availability: In order to provide precise predictions, the majority of machine learning models need a lot of data. The accuracy of the model may degrade if data is not accessible in large enough amounts.

• Inaccurate labelling: The accuracy of the machine learning models that are now in use depends on how well the input dataset was used for training. Inaccurate labelling of the data prevents the model from producing reliable predictions.

PROPOSED SYSTEM

The chosen feature sets, the data collecting sources, and the application's aim are the three main areas where the planned effort varies. The current work, in comparison to all prior research in the same field,: (4) makes use of more technologies; (5) aims to extract information that can be used to analyse the competition of airline companies and the behaviour of consumers; (6) compares the performance of various algorithms that are introduced to the problem for the first time; and (7) offers two evaluation perspective approaches to a comprehensive investigation of the problem under study. Given the multiplicity of the suggested methodology, it is clear that a direct comparison with other approaches is pointless since there is no shared point of reference, and no conclusions can be drawn. Therefore, in this study, the chosen ML, DL, and QML models for the issue of flight price prediction are compared using common points of reference (dataset, features, and application objective).

With an emphasis on the chosen techniques and the data employed, the suggested holistic approach is explained. To illustrate the degree of competition and globalisation in airfare tickets between locations from various airline companies, datasets, feature descriptions, and visualisation materials are provided. Additionally, this part presents the ML, DL, and QML models that are used and provides a brief overview of each to highlight the variations in their capabilities and performance.

Advantages

1) Examining the relationship between various airline companies' pricing policies.
2) Examining how attributes affect the challenge of predicting flight costs.
3) This is the first time that QML models have been used in the literature to forecast flight prices.
4) Evaluation of the relative effectiveness of ML, DL, and QML models for forecasting flight prices.

II. REQUIREMENTS & ANALYSIS

2.1. LITERATURE SURVEY

"An Introduction to the Theory and Practice of Yield Management," by S. Netessine and R. Shumsky In September 2002, INFORMS Trans. Educ., vol. 3, no. 1, pp. 34–44.

In addition to outlining the similarities between yield management and the newsvendor framework, a crucial inventory management model, this paper aims to present the core ideas and trade-offs of yield management. This note specifically addresses how a manager may distribute perishable goods among various client categories. We will focus the most of our time on an application that incorporates two different kinds of clients in order to get some insight into the issue.

"A logical calculus of the ideas immanent in nervous activity," by W. S. McCulloch and W. Pitts Biophys. Bull. Math.

Because neurological activity is "all-or-none," propositional logic may be used to describe neural events and the relationships between them. It is discovered that every net's behaviour can be explained in these terms, with the inclusion of more intricate logical methods for nets with circles; moreover, any logical statement that meets certain requirements may be used to locate a net that behaves in the manner it describes. It has been shown that a large number of specific options among potential neurophysiological assumptions are similar, meaning that for every net acting under one assumption, there is another net acting under the other and producing the same outcomes, although maybe not at the same moment. There is discussion of several calculus applications.

F. Rosenblatt, "The perceptron: A probabilistic model for the brain's organisation and storage of information"

A theory is constructed for a hypothetical neural system called a perceptron in order to address the following questions: how is information about the physical world detected, what form is information recalled, and how does information maintained in memory affect recognition and behaviour? The hypothesis acts as a link between psychology and biophysics. Neurological characteristics may be used to predict learning curves, and vice versa. The organisation of cognitive systems may be better understood via the use of quantitative statistical methods.

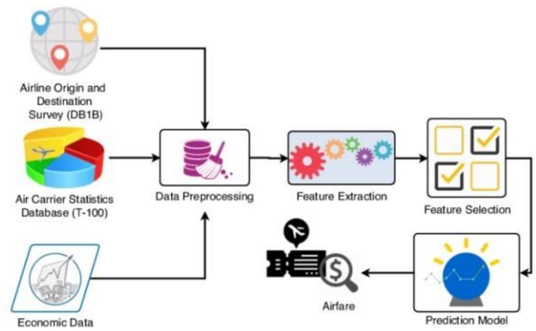
An optimum margin classifier training method by B. E. Boser, I. M. Guyon, and V. N. Vapnik

We provide a training method that maximises the gap between the decision border and the training patterns. The method works with a broad range of classification functions, such as polynomials, radial basis functions, and perceptrons. The effective number of parameters is automatically modified to reflect the problem's complexity. Supporting patterns are combined linearly to represent the answer. The training pattern subset that is closest to the decision boundary is this one. There are limits on the generalisation performance dependent on the VC-dimension and the leave-one-out approach. When compared to other learning methods, experimental findings on optical character recognition challenges show the excellent generalisation achieved.

J. L. Hodges and E. Fix, "Discriminatory analysis." Discrimination that is not parametric: Properties of consistency

In supervised learning, algorithms may categorise one instance from the minority class as belonging to the majority class because to the unequal number of occurrences across the classes in a dataset. The KNN algorithm serves as a foundation for other balancing techniques in an effort to address this issue. In this paper, these balancing techniques are reviewed, and a novel and straightforward KNN undersampling technique is suggested. The KNN undersampling approach fared better than other sampling techniques, according to the trials. The suggested approach also performed better than the findings of previous research, suggesting that the ease of use of KNN may serve as a foundation for effective machine learning and knowledge discovery algorithms.

SYSTEM ARCHITECTURE



2.2. MODULES DESCRIPTION

Modules

Service Provider

The Service Provider must use a working user name and password to log in to this module. Following a successful login, he may do several tasks including training and testing datasets, See the Bar Chart for Trained and Tested Accuracy. View the results of trained and tested accuracy, view the kind of prediction for airfare prices, view the ratio of the type of prediction, and download the predicted data sets. View All Remote Users and the Results of the Airfare Price Prediction Type Ratio.

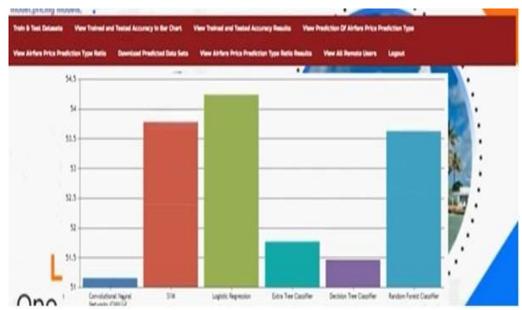
View and Authorize Users

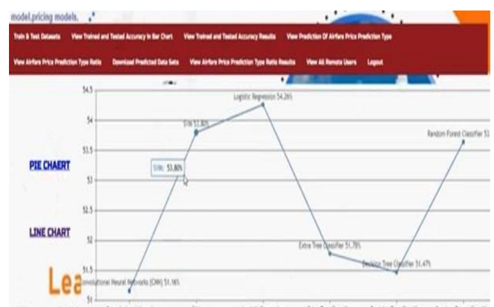
The administrator may see a list of all registered users in this module. Here, the administrator may see the user's information, like name, email, and address, and they can also grant the user permissions.

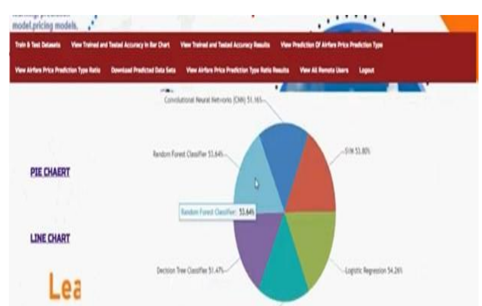
Remote User

A total of n users are present in this module. Before beginning any actions, the user needs register. Following registration, the user's information will be entered into the database. Following a successful registration, he must use his password and authorised user name to log in. Following a successful login, the user may do tasks including registering and logging in, predicting the kind of airfare, and seeing their profile.

III. SCREEN SHOTS







IV. CONCLUSION

The holistic approach to ticket price prediction, taking into account various datasets and potentially useful technologies, is the main emphasis of this study. Six locations and four airlines were taken into consideration in order to achieve this. Eight ML models, six DL models, and two QML models have been used and compared in order to answer the issue under investigation. According to experimental data, for various foreign locations and airline firms, at least three models from each of the domains of machine learning, deep learning, and QML may obtain accuracies ranging from 89% to 99% in this regression issue. The results show that the airline company's ticket pricing policy may be effectively examined by applying AI models and flight details that consumers can see before to purchase. Additional features are openly accessible, and with the help of the aforementioned technologies, reliable simulations for estimating client demand and flight ticket pricing optimisation might be produced, giving airline firms extensive data to help them choose their best pricing strategy. All model domains, however, are able to identify similarities and extract patterns from the provided flight data, even when just a limited number of characteristics are present. Two distinct methods—one based on the destinations (for all airlines) and one based on the airline firms (for all destinations)—have been examined and analysed in this study. In order to determine if the information could be effectively retrieved, future study from the standpoint of the airline-based target application might use the same airline firms and destinations evaluated from other airports. Furthermore, using customer segmentation based on the flight feature set, the same topic might be examined as a classification problem.

From a technological perspective, QML models have been examined in the context of regression applications, which are scarce in the literature because the benefits of QML models in classical data are debatable given their limitations, the quantum resources that are available, and the computational requirements of classical machines. To enable the use of QML solutions in more practical applications, the availability of quantum hardware has to be expanded and made more accessible, notwithstanding constraints such as the quantity of qubits and noise levels in quantum machines.

Despite the documented drawbacks and challenges, QML models for flight price prediction in this study often produced better results than ML and DL models. Therefore, it can be said that future methods of predicting flight prices based on the QML domain may provide effective solutions, particularly given the anticipated growth in data volume, complexity, and variety. The investigation of various techniques for data encoding in quantum states and additional quantum models, such as quantum Boltzmann machines, which will be able to generate flight data based on given air ticket feature sets and distributions, are examples of future work surrounding QML methods in airfare price prediction. The resulting QML-based program might be used to generate policies for flight prices.

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 classifiers,’’ in *Proc. 5th Annu. 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 Internationale deStatistique*, vol. 57, no. 3, p. 238, Dec. 1989.

[6] R. E. Schapire, ‘‘The strength of weak learnability,’’ *Mach. Learn.*, vol. 5, no. 2, pp. 197–227, Jun. 1990.

[7] K. Fukushima, ‘‘Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,’’ *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, Apr. 1980.

[8] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, ‘‘Gradient-based learning applied to document recognition,’’ *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[9] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, . Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and

D. Hassabis, ‘‘Mastering the game of go with deep neural networks and tree search,’’ *Nature*, vol. 529, no. 7587, pp. 484–489, Jan. 2016.

[10] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, ‘‘Generative adversarial networks,’’ *Commun. ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020.

[11] P. W. Shor, ‘‘Algorithms for quantum computation: Discrete logarithms and factoring,’’ in *Proc. 35th Annu. Symp. Found. Comput. Sci.*, 1994, pp. 124–134.

[12] L. K. Grover, ‘‘A framework for fast quantum mechanical algorithms,’’ in *Proc. 30th Annu. ACM Symp. Theory Comput. (STOC)*, 1998, pp. 53–62.

[13] M. Andrecut and M. K. Ali, ‘‘Quantum associative memory,’’ *Int. J. Mod. Phys. B*, vol. 17, no. 12, pp. 2447–2472, May 2003. [14] M. V. Altaisky, N. E. Kaputkina, and V. A. Krylov, ‘‘Quantum neural networks: Current status and prospects for development,’’ *Phys. Particles Nuclei*, vol. 45, no. 6, pp. 1013–1032, Nov. 2014.

[15] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, ‘‘Supervised learning with quantum enhanced feature spaces,’’ *Nature*, vol. 567, no. 7747, pp. 209–212, Mar. 2019.

[16] K. Tziridis, T. Kalampokas, G. A. Papakostas, and K. I. Diamantaras, ‘‘Airfare prices prediction using machine learning techniques,’’ in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2017, pp. 1036–1039.

[17] J. A. Abdella, N. Zaki, and K. Shuaib, ‘‘Automatic detection of airline ticket price and demand: A review,’’ in *Proc. Int. Conf. Innov. Inf. Technol. (IIT)*, Nov. 2018, pp. 169–174.

[18] V. H. Vu, Q. T. Minh, and P. H. Phung, ‘‘An airfare prediction model for developing markets,’’ in *Proc. Int. Conf. Inf. Netw. (ICOIN)*, Jan. 2018, pp. 765–770.

[19] F. Huang and H. Huang, ‘‘Event ticket price prediction with deep neural network on spatial–temporal sparse data,’’ in *Proc. 35th Annu. ACM Symp. Appl. Comput.*, Mar. 2020, pp. 1013–1020.

[20] T. Wang, S. Pouyanfar, H. Tian, Y. Tao, M. Alonso, S. Luis, and S. Chen, ‘‘A framework for airfare price prediction: A machine learning approach,’’ in *Proc. IEEE 20th Int. Conf. Inf.*

Reuse Integr. Data Sci. (IRI), Jul. 2019, pp. 200–207.

[21] E. A. Kuptsova and S. K. Ramazanov, ‘‘Analysis of artificial neural networks training models for airfare price prediction,’’ *Artif. Intell.*, vol. 25, no. 3, pp. 45–50, Oct. 2020.

[22] N. Joshi, G. Singh, S. Kumar, R. Jain, and P. Nagrath, ‘‘Airline prices analysis and prediction using decision tree regressor,’’ in *Proc. Int. Conf. Recent Develop. Sci., Eng. Technol.*, in Communications in Computer and Information Science, 2020, pp. 170–186.

[23] 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.

[24] S. Sutthithatip, S. Perinpanayagam, and S. Aslam, ‘‘(Explainable) artificial ntelligence in aerospace safety-critical systems,’’ in *Proc. IEEE Aerosp. Conf. (AERO)*, Mar. 2022, pp. 1–12.

[25] K. Pearson, ‘‘VII. Note on regression and inheritance in the case of two parents,’’ *Proc. Roy. Soc. London*, vol. 58, nos. 347–352, pp. 240–242, Dec. 1895.