INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 9 ISSUE 6 (2023) PAGES 1-14
RECEIVED:05.11.2023 PUBLISHED:30.12.2023 Dec 30, 2023

A HOLISTIC APPROACH ON AIRFARE PRICE PREDICTION USING MACHINE LEARNING TECHNIQUES

B. SUJAI MATHEW[1], ANITHA M[2]

[1] *Student, Department of CSE, KINGSTON ENGINEERING COLLEGE, VELLORE - 632 059, Tamil Nadu, India*

[2] *Assistant Professor, Department of CSE, KINGSTON ENGINEERING COLLEGE, VELLORE - 632 059, Tamil Nadu, India*

ABSTRACT

Globalization of markets involves new strategies and price policies from professionals that contribute to global competitiveness. Airline companies are changing tickets' prices very often considering a variety of factors based on their proprietary rules and algorithms that are searching for the most suitable price policy. Recently, Artificial Intelligence (AI) models are exploited for the latter task, due to their compactness, fast adaptability, and many potentials in data generalization. This paper represents an analysis of airfare price prediction towards finding similarities in the pricing policies of different Airline companies by using AI Techniques. More specifically, a set of effective features is extracted from 136.917 data flights of Aegean, Turkish, Austrian and Lufthansa Airlines for six popular international destinations. The extracted set of features is then used to conduct a holistic analysis from the perspective of the end user who seeks the most affordable ticket cost, considering a destination-based evaluation including all airlines, and an airline-base devaluation including all destinations. For the latter cause, AI models from three different domains and a total of 16 model architectures are considered to resolve the airfare price prediction problem: Machine Learning(ML) with eight state-of-the-art models, Deep Learning (DL) with six CNN models and Quantum Machine Learning (QML) with two models. Experimental results reveal that at least three models from each domain, ML, DL, and QML, are able to achieve accuracies between 89% and 99% in this regression problem, for different international destinations and airline companies.

INDEX TERMS : Airfare price, artificial intelligence, deep learning, machine learning, prediction model, pricing models, regression, quantum machine learning.

INTRODUCTION

Approximately 50 years ago airline flights were considered a luxury. Airline companies were launching more domestic flights than international while pricing policies for flight tickets were static. To increase profitability, airline companies adopted management and economical software systemsto perform route optimizations, reservation adaptation, and dynamic pricing. An evolution in airline companies was the adoption of yield management, which was a variable pricing strategy based on understanding, anticipating,and influencing consumer behaviour so as to reachthe highest revenues. As a consequence, airline companies started to pay more attention to customers' preferences and experiences during flights, simultaneously increasing the destinationsat an international level. Thus, airline flights becameaccessible to all potential consumers since dynamic pricingand extra flight services increased the competition betweenairline companies. Moreover, in recent years, the ability toshop online revolutionized many different fields and becamea trend among modern people, seeking the most favourableoffers and prices. Currently, there are several websites thatsupport secure flight booking, listing the same flight routesfrom all airline companies towards getting the most competitiveflight deals. Moreover, sharing flight experiences throughrating systems provides a great amount of useful informationproduced daily by airline customers, that are exploited by pricing policy systems to adapt the airfare price, evenminutes before a flight. To this end, it is clear that marketglobalization and technology evolution have affected airlinecompanies at a level where the mainstream price optimizationsystems may not track the changes and reach the adaptationspeed that is required. The latter increased the demand formore sophisticated algorithms and software for dynamic pricepolicy optimization. For this reason, Artificial Intelligence(AI) algorithms are currently considered for airfare price estimation,towards achieving efficient and more realistic resultswith higher speed.Artificial Intelligence attracts high interest from theresearch community in many research fields. Machine Learning(ML) was the first introduced domain of AI by WalterPitts and Warren McCulloch in 1943 where a mathematicalmodel of a biological neuron was proposed with no learningcapabilities. Seven years later, in 1950, Frank Rosenblattproposed the perceptron . . . . . . (()) as the first neural network(NN) with learning abilities. Perceptron was an inspirationfor researchers to design and implement subsequentlymany well-known ML models like SVM, kNN , andBoosting methods. ML models couldn't robustly generalizewithout a supporting feature extraction mechanism.The latter requirement was handled by the Deep Learning(DL) domain,increasing the computational demands andreducing the execution time. The flagship for the rise ofthe DL domain was the introduction of convolutional neuralnetworks (CNN) by Fukushima in 1980 who used a NNfor visual pattern recognition. A distinct boost towards thiseffort came from Yann LeCun in 1990, who used CNNmodels with backpropagation learning in order to recognizehandwritten digits from images. DL models have automatedthe feature extraction process giving the capability to fabricatemore complex algorithms and applicationsthat impact human daily lives. However, even today, dueto the huge data growth rate and despite the evolution ofcomputational hardware (GPUs), there is still a need for fasterand more compact ML and DL algorithms.Based on the above, the main contributions of the proposedwork can be summarized as follows:

1) Investigation of the relation of pricing policies amongdifferent airline companies.

2) Investigation of features' influence to the airfare pricesprediction problem.

3) Application of QML models in airfare price predictionfor the first time in the literature.

4) Comparative performance analysis of ML, DL andQML models for airfare price prediction.

The rest of this paper is organized as follows: Section IIsummarizes the related work on airfare price prediction.In Section III materials and methods are introduced, refereeingto data and algorithms that have been used for theimplementation of this work. Section IV describes the experimentalsetup, while in Section V the experiment results arepresented and discussed. In Section VI, quantum machinelearning results are presented and compared to classical models.Finally, Section VII concludes the paper and presentsfurther potential research directions.

II. RELATED WORK

Market globalization along with the evolution of airfare pricepolicies resulted in a great amount of relevant informationand, subsequently, high research interest in airfare price prediction.In terms of AI and data analysis, this informationis translated to data with many attributes and in amountsthat could be characterized as big data, especially when thechange rate of air ticket prices and services is such high.The airfare price prediction problem can be exploited undervarious scopes, like customer segmentation, ticket purchasetiming, air tickets demand prediction, and more, as presentedin a review by Abdella et al regarding the target applicationproblem and the solutions. In general, the subject ofairfare price prediction is in the spotlight for three decades;a search on Scopus on the term ''airfare price prediction''returned 24 documents, from 2003 to date, with most of thework being implemented in the last three years. Vu et al. implemented an airfare price prediction application with twoML models, exploiting features around time to describe Vietnamesenational airline company flights. Compared to theproposed approach, fewer models have been presented andonly one airline company has been considered, while themain focus was on consumers' target applications. In ,a different approach was presented. A custom recurrent neuralnetwork (RNN) was constructed and compared to classicalML models in airfare price prediction under events like abasketball match. Features that described basketball matchesand airline flights were combined in one dataset, achievinghigh prediction accuracies. The same approach was followed. The authors proposed a framework that could gatherinformation for air tickets from various sources, such asconsumers' interests, air tickets availability, distance, andmore, to predict airfare prices by using ML models. Inairfare price prediction was implemented in the domesticmarkets of USA and India. The authors exploited ML modelsand reported an 88% score in price prediction. In,Joshi et al. adopted a similar approach with fewer ML models,by investigating new features, like flight duration, andachieved up to 90% prediction score. Infeature selectionalgorithms were applied along with hyperparameter methodsto

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 9 ISSUE 6 (2023) PAGES 1-14
RECEIVED:05.11.2023 PUBLISHED:30.12.2023 Dec 30, 2023

find the optimal model parameters and set of features forflight description in order to predict airfare price prediction.In explain ability for the problem under study has been introduced towards a deeper insight into the models that couldprovide an efficient solution, in order to give robust andexplainable predictions.

III. MATERIALS AND METHODS

In this section, the proposed holistic approach is described,focusing on the used data and the selected methods. Datasets,features description and visualization material are presentedto highlight the level of competition and globalization affectionin airfare tickets between destinations from different airline companies. Moreover, in this section, the ML, DL,and QML models that are employed are presented and a shortdescription for each one is given to underline the differencesin performance and capabilities between them the focus of this work is on the prediction of airfare pricesfor six different destinations for four airline companies.

A. DATA PRESENTATION AND DESCRIPTION

Theairline companies are: Aegean Airlines, Austrian Airlines,Lufthansa Airlines and Turkish Airlines. The destinations of interest are the following:

1) Thessaloniki (SKG) – Amsterdam (AMS), (1907 Km)
2) Thessaloniki (SKG) – Stockholm (ARN), (2157 Km)
3) Thessaloniki (SKG) – Brussels (BRU), (1812 Km)
4) Thessaloniki (SKG) – Paris (CDG), (1863 Km)
5) Thessaloniki (SKG) – Lisbon (LIS), (2747 Km)
6) Thessaloniki (SKG) – Vienna (VIE), (985 Km)

The flight data are collected for the period of one year.1 Itshould be clarified here that flight data are not for exactlyone year, due to the fact that some airline companies didnot provide the same flights for all destinations all over theyear, mainly due to demand variations.

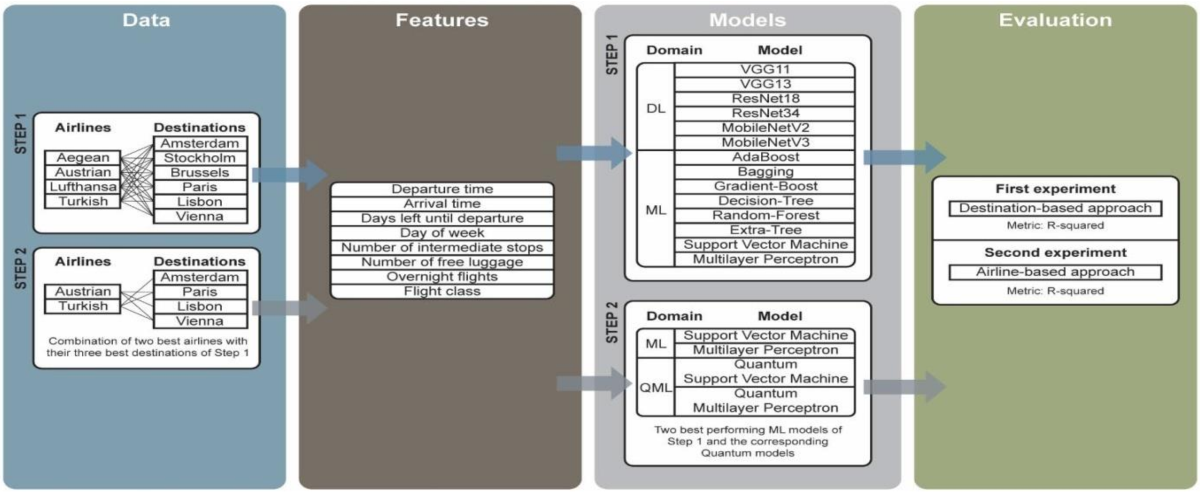


FIGURE 1. The proposed holistic approach to airfare price prediction

In this work, the most descriptive features that affect theairfare price and were publicly available, were selected. Foreach flight data, a set of eight features (0:7) was used. Due tothe difficulty of collecting flight data manually, Data Miningtechniques were applied to acquire as many data as possible.Finally, for each flight the following eight features wereconsidered:

1) Feature 0: Departure time
2) Feature 1: Arrival time
3) Feature 2: Days left until departure (0 - 350+)
4) Feature 3: Day of week (1-7)
5) Feature 4: Number of intermediate stops (0 - 2)
6) Feature 5: Number of free luggage (0 - 2)
7) Feature 6: Overnight flight (yes - 1 or no - 0)
8) Feature 7: Flight class (three-digit number, each digit 0 - 5)

Regarding feature 7, note that flight class is a three-digitinteger number. Each digit independently represents a flightclass, considering up to three correspondences per voyage.For instance, if the third digit of fight class is not zero,it means that the flight had two intermediate stops, thus, thevoyage involved three corresponding flights in total, and eachof the three digits informs about the involved ticket class.

If the third digit is zero, it means that there was no third flight(only two flights) and so on. Every digit's value is rangedfrom 0 to 5, depending on the flight class of each of thecorresponding flights, as follows:

1) Economy class– 1
2) Economy Standard class – 2
3) Economy Premium class – 3
4) Business class – 4
5) First class – 5
6) No flight – 0

Austrian and Turkish airlines, as it can be observed fromFig. 2(b) and Fig. 2(d) have very few flights in the selecteddestinations and, thus, the number of stops (feature 4) hasa low diversity and the correlation coefficients of this valueequal to zero. A first notice is that Aegean displays more lightcolors in its heatmap, translated to less correlations betweenfeatures in destinations of greater distance (SKG_ARN,SKG_LIS) compared to other destinations which seem toave darker colour values, translated to strongercorrelations.Finally, considering Fig. 2and Fig. 3, similarities between airline companies and theirpricepolicies the same observation can be made for Austrian, Turkishand Lufthansa, but only in the destination SKG_ARN. It isalso important to mention that for every airline company anddestination, it seems that flight class (feature 7) and price havea strong correlation despite the differences in the number offlights of each company. Based on this fact, it is easy to

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 9 ISSUE 6 (2023) PAGES 1-14
RECEIVED:05.11.2023 PUBLISHED:30.12.2023

concludethat flight class has a strong impact on the competition Between airline companies. The heatmaps of Pearsoncorrelation coefficient for each airline company are presentedwith the destination as an extra feature (feature 8)

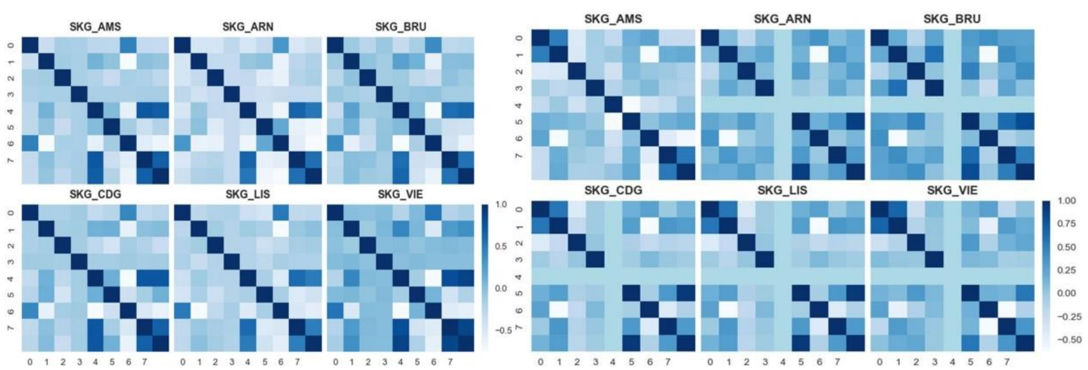
B. MODELS PRESENTATION AND DESCRIPTION

TABLE 2. Selected machine learning (ML) models.

Model	Algorithm type
AdaBoost Regressor	Boosting family
Bagging Regressor	Boosting family
Gradient Boost Regressor	Boosting family
Decision Tree Regressor	Tree based
Random Forest Regressor	Tree based
Extra Tree Regressor	Tree based
Support Vector Regressor (SVM)	Kernel function
Multi-Layer Perceptron (MLP)	Neural Network

Starting from theMLdomain, eight state-of-the-art modelswere selected and presented in Table 2. *AdaBoost regressor*comes from the 'Boosting' familyof algorithms, forming a strong learner froma compositionof weak learners. Very often these learners are DecisionTreeswhere iteratively AdaBoost adapts their errorsand combines them sequentially to create a strong ensemblemodel that will decrease bias and variance in the training

data. A disadvantage of this algorithm is its sensitivity tonoise and overfitting with the increase of dataset features andsize. *Bagging regressor* adopts a variation of the sameapproach as AdaBoost.Weak learners in Bagging are createdin parallel and, thus, independently of each other, while inAdaBoost they are created sequentially. In addition, Baggingdecreases the variance more than the bias and it is proposedto resolve overfitting issues. A reported disadvantage is itssensitivity to noise data and the construction of ideal globalsolutions in a large number of features and data. Finallyfrom Boosting family, *Gradient Boost algorithm*is alsoselected. Gradient Boost can produce new models (oftenDecision Trees) to be maximally correlated with the negativegradients of a loss function (often Mean Squared Error) tominimize it with minimum iterations.



INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 9 ISSUE 6 (2023) PAGES 1-14
RECEIVED:05.11.2023 PUBLISHED:30.12.2023 Dec 30, 2023

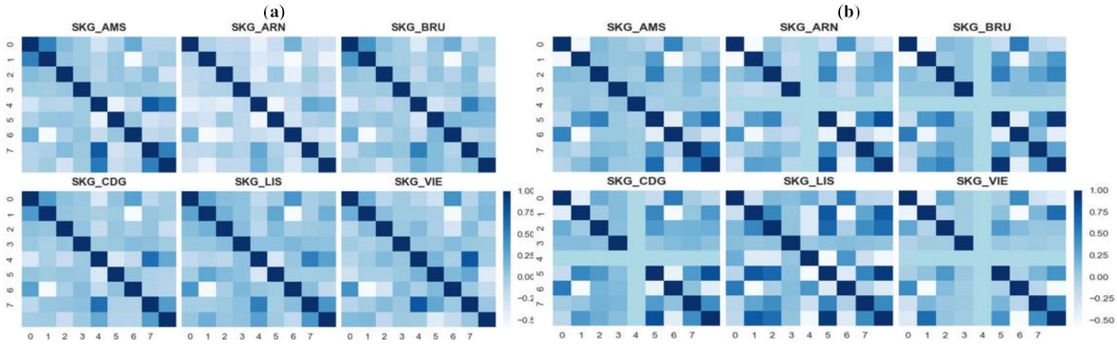


FIGURE2. Pearson correlation coefficients heat maps for each destination: (a) Aegean airlines correlation coefficients; (b) Austrian airlines correlation coefficients; (c) Lufthansa airlines correlation coefficients; (d) Turkish airlines correlation coefficients.

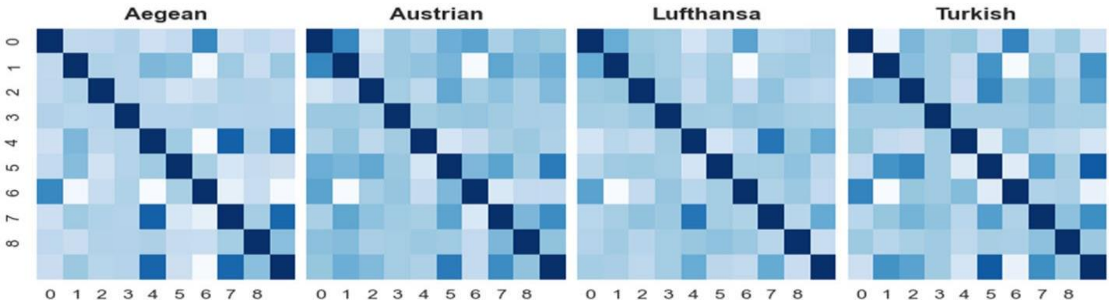


TABLE 3. Selected deep learning (DL) models.

Model	Training parameters (Millions)
VGG11	133
VGG13	133
Resnet18	11.4
Resnet34	21.5
MobileNetV2	3.4
MobileNetV3	2.4

TABLE 4. Selected quantum machine learning (QML) models.

Model	Algorithm type
Quantum Support Vector Regressor (QSVM)	Quantum kernel
Quantum Multiplayer Perceptron (QMLP)	Quantum neural network with 48 parameters

The selected six models in this work from the DL domainare included in Table 3.First, the most important layer is the Convolutional Layer, which consists of convolutional filters (or kernels). Each filteris convolved with the input 2D data to produce feature maps. Kernels are randomly initialized, and they slide in the input data where the dot product is calculated in each slide. Kernel values, namely weights, adjust during training. Second, thepooling layers are applied to sub-sample the feature mapsto produce smaller maps, maintaining most of the dominantfeatures. The pooling process is applied with variousmethods like average, min-max, or custom methods. Then, activation functions take place to map input data with targetvalues through the weighted summation of convolutionallayers neurons weights. Thus, it is determined if neurons arecontributing to thecorresponding target value of a given inputdata or not.In the next year,*ResNet*was proposed to overcome the gradient vanish problem. However, while the depth of the CNN network increased the dimension of the features also increased and in contradiction, the loss was optimized to local minima. In that case, a part of the network usually at the start had a low contribution to the prediction. This phenomenon was notice din VGG and attempted to be resolved by ResNet where multiple residual blocks were used to shorten the connections between layers and, thus, the network could take more layerswith stable performance and simpler architecture. It was also proposed under various architectures with 85 millionparameters in Resnet50. Finally, *MobileNetV2* was also selected in this work, as a CNN architecture that focuses onthe balance between performance and speed. It consists of3 convolutional layers with a filter size of 1×1 in orderto reduce computation time. In addition, the latest versionMobileNetV3 was proposed for mobile processing unitshaving less than 2 million parameters.A disadvantage of the DL technology is that operatesnon-optimally and is based on statistical methods, consideringthat CNN treats neuron weights as a whole, even thoughsome weights might not have a high contribution to thepredictions of an input datum. This fact justifies the longtraining times that are required. Based on the above, it isclear that the CNN models' design needs improvement, andthus an effort was given by the research community duringthe last years to produce sophisticated mechanisms that willmake CNN architectures more robust and exclusive to theproblem through attention mechanisms, custom losses, andlayers or even model design under new domains on whichthese models will be structured in a more compact way andwith more generalization capabilities.

IV. EXPERIMENTAL SETUP

In this work, two experiments are conducted inorder tocover the proposed holistic approach for the target applicationproblem. In the firstexperiment, namely the destination-basedapproach, the selected models from ML, DL, and QMLdomains are used to find the best choice for each destinationper Airline Company. With thisexperiment, it can beconcluded the optimal set of models that describe the samedestinations for separate airline companies, having similarairfare price prediction accuracies. To accomplish that, the entire dataset was split for each destination for each airlinecompany. From PyTorch all thepresented CNN [7] models wereapplied along with the learning process on a GPU unit. FromScikit-Learn all the ML models were used and fitted on aCPU unit. From PennyLane QMLP network was formedand executed on a simulator that benefited CPU unit. Underthe same principles, QSVM was applied from the Qiskitframework. The hardware specifications where all theabove experiments have been conducted arepresented below:

• CPU: AMD Ryzen™Threadripper™2920X, 12 cores
(24 threads), 3.5GHz base clock.
• RAM: 32 GB DDR 4.
• GPU: NVIDIA GeForce RTX 2060 SUPER 8 GB VRAM.
• STORAGE: Viper M.2 vpn100 3450 MB/s-read, 3000 MB/s-write.

V. EXPERIMENTAL RESULTS OF STEP 1: ML VS DL

In this section the results for both experimental approaches, for each ML and DL model are presented, by using the data of four airlines and six destinations.

A. FIRST EXPERIMENT OF STEP 1: THEDESTINATION-BASED APPROACH

Tables 6 to 9 include the experimental results for each airline company and destination for the first experiment. The best scores for each destination are marked in bold in the Tables. An observation that can be derived from the following tables is regarding the model with the best score for each destination, as for all destinations by considering the Meanperformance (last column of each Table). Therefore, informationabout airfare price policies and competition levelsbetween airline companies can be extracted.From Table 6 it is obvious that the best models for eachdestination are the neural networks from the DL domain. Bagging,Multilayer Perceptron, Random Forest, and Extra-Treefrom the ML domain are following in performanceit can be concluded that for the Aegean airline,AMS and VIE are the most important destinations comparedto the rest of the destinations, since at these destinationsalmost all models achieve their highest scores, greater than86%. Based on Table 1, AMS is the destination with thehighest number of flights for Aegean and based on Fig. 2(a) itseems that for AMS there are many available flights despitethe variety of ticket classes, so the distribution of prices isnormal. The same fact involves the VIE destination.Additionally,in Fig. 2(a), VIE has darker colour compared to therest destinations, and since it is the closest destination toSKG it can be concluded that there are many flights to VIEwith similar prices. Based on the above it can be assumedthat Aegean ticket price strategy aims to attract a varietyof consumer groups for AMS destination, rather than VIE,where ticket classes and a variety of services are limited. The best model for the Austrian airline isExtra-Tree-Regressor with 99% in VIE destination. It canbe observed that ML and DL achieve the highestscoreswith less difference between them,compared to the previousairline companyperformances. This is justified consideringthe number of flights from Table 1, as Austrian airlines have atleast 50% fewer flights than Aegean. Even with fewer flightsfor each destination compared to Aegean in CDG, LIS andVIE, all models achieve high performance scores. In addition,according to Fig. 2(b), Austrian airline has stronger correlationsfor these destinations. Thus, it seems that Austrianairline attempts a competitive policy with many flights thathave similar ticket classes and number of stops along withthe amount of luggage. On the contrary, for the destinationsAMS and BRU, price and ticket classes have high variationwith small number of flights, which justifies the results of the DL models. It seems that AMS and BRU are not among thedestination that Austrian tries to compete with.For the case of Lufthansa,the results were very poor compared to the other airlines ingeneral. The best model is the MLP in CDG destination fromthe ML domain. The highest and similarresults of the models are in AMS, ARN and CDG, which canbe justified by the number of flights in Table 1. It seems thatLufthansa tries to be more competitive with Aegean Airlinerather than Austrian since VIE is not in the scope of concernfor its price policy. Finally, Lufthansa seems to differ from allairline companies in its general price strategy, since for fourout of six destinations ML and DL models have the highestdifference in performance compared to the rest of the airlines.Finally, for the Turkish airline, it can be observedthat the best scoring destinationsinclude LIS, AMS, CDGand VIE. Morespecifically, for destination AMS, the bestmodels are AdaBoost and Random Forest with a score of93%. For LIS the best models are from both ML and DLdomains with scores of up to 97%. For Turkish airline, almostin all destinations, the models bring similar results, exceptfor ARN destination, revealing that it is not so preferabledue to its price strategy. In general, based on the results ofTable 9, Turkish airline attempts to be competitive throughsimilar ticket classes and prices, considering its number offlights. A final notice is that Turkish and Austrian have moresimilar price strategies since ML and DL models for four outof six destinations share common performances.The three best scoring destinations for eachairline company are presented. The destination AMS is the best for Aegean, VIE forAustrian, CDG for Lufthansa and LIS for Turkish airlines.Another fact is that destination CDG is among the best performingfor all airlinecompanies. In general, it seems that allairline companies are being competitive withAegean airline,which has the most flights. The latter can be observed especiallyin VIEdestination, which is the nearest to SKG and,thus, the ticket prices for each airline company are similar butwith different number of flights and services. Another noticethat justifies this fact is, that even Lufthansa is the second that even Lufthansa is the second.

B. SECOND EXPERIMENT OF STEP 1: THE AIRLINE-BASEDAPPROACH

Turkish Airlines reporta higher performance, reaching 97% with MobilNetV3.ML models did not perform so well in the experiment withTurkish airline, compared to the first experiment. Ingeneral,the superiority of DL models in a higher amount of data isclear compared to the ML domain.A reason for this poor performance could be that Turkishairline has 8391 data flights for six different destinations,having a distribution of many low-price tickets and a smallnumber of more expensive tickets, which mainly affects theensemble models. In contrast, DL models prove that theycan adjust weights along with the features and targets in amore flexible way, and by using supplementary methods likepooling, they can achieve a higher score under more complexdata.

VI. EXPERIMENTAL RESULTS OF STEP 2: ML VS QML

A. FIRST EXPERIMENT OF STEP 2: THEDESTINATION-BASED APPROACH

QMLP clearly holds the first placecompared to classical MLP and SVM, achieving enhancedperformances by 3% and 8%, respectively, based on theMean performance for all three destinations. Since QMLPhas a more compact structure based on Table 4 with a similarfeature enhancement capacity to a CNN, its generalizationcapability is very high. Regarding the pair of modelsSVM and QSVM, quantum kernels have proven better,comparedto the classical since they can construct larger featuredimensions that might lead to linear data separation evenunder complex data structures or sparse patterns. In addition,it should be noted that QML models are examined not at theirfull potential, since not all the capabilities of their learningprocess are feasible to be explored due to huge time andresource requirements. In Fig. 4, a comparative illustrationis presented, including all domain models for the Austrianairline and the three selected destinations.This performance similaritybetween QML and DL models is justified from the dimensionsof features space underquantum principles that areloser to CNN rather to ML models, but with a simpler structureconsidering that 16 qubits represent 8 flight featuresand their corresponding neurons. The dimension of aqubit is 2N in classical machines, where N is the number ofqubits, therefore, for the proposed problem QMLP constructsa 65.356-dimensional feature map. Moreover, the classical gradient-based optimization algorithm requires the constructionand evaluation of several quantum circuits in its gradientiteration which is computation costly. All above justify higher hardware resource demands of QML in a classical machine,compared to the other two domains.

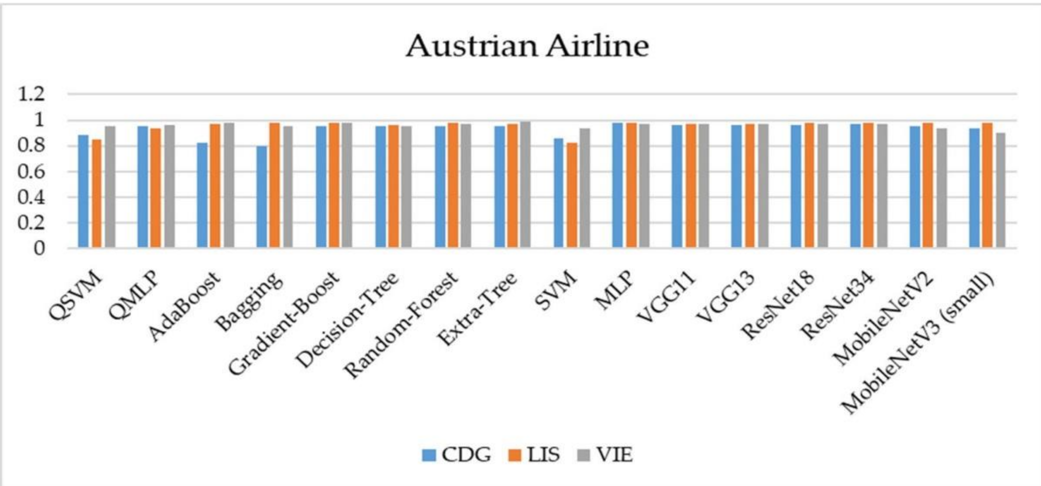


FIGURE 4. Bar plot for all domain models in three destinations of Austrian airline.

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 9 ISSUE 6 (2023) PAGES 1-14
RECEIVED:05.11.2023 PUBLISHED:30.12.2023 Dec 30, 2023

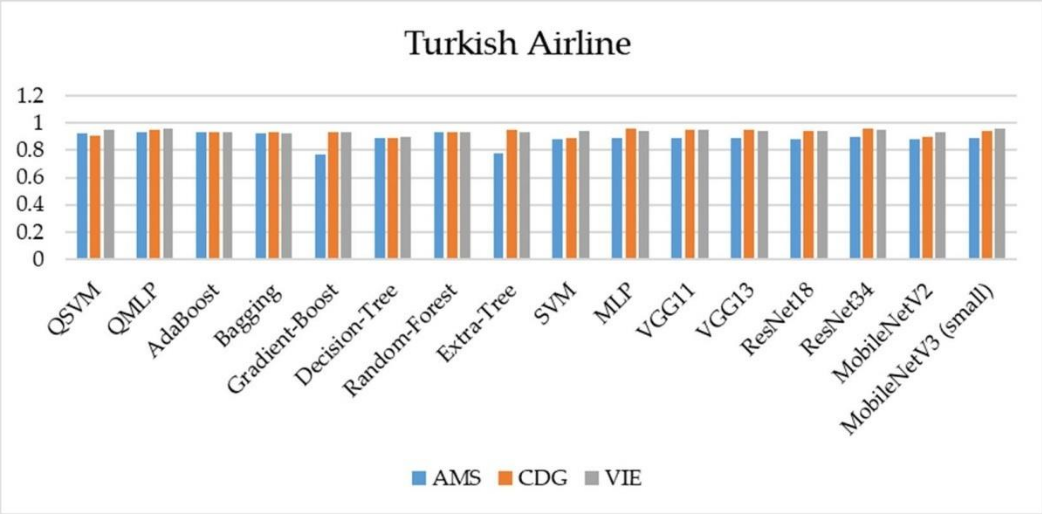


FIGURE 5. Bar plot for all domain models in three destinations of Turkish airline

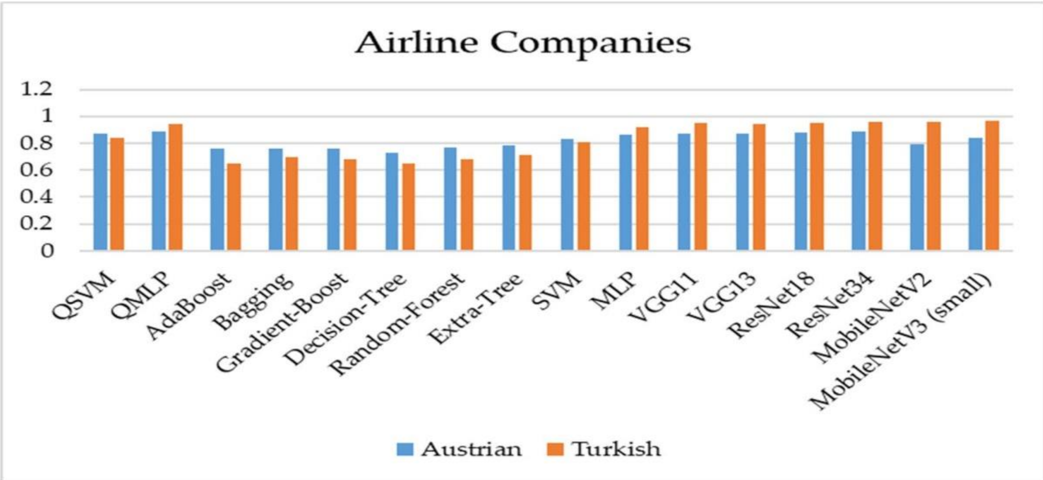


FIGURE 6. Bar plot for all domain models in three destinations of Turkish and Austrian airlines.

B. SECOND EXPERIMENT OF STEP 2: THE AIRLINE-BASEDAPPROACH

Based on the experimental results, QML modelsreveal the optimal performance for both airlines. DespiteAustrian airline's imbalanced distribution of airfare pricegroups and a few number of flights, still results are rankinghigh for all models. However, for QSVM, the above factseems to have a smaller impact compared to SVM, reportinga score difference of 4%. For QMLP and MLP the samenotice can be made. The Turkish airline shares commonstrategies with Austrian, but with a more normal distributionin airfare prices and services groups. Same with the previousairline company, QML models come first compared to MLmodels for Turkish airline flights. Another similar conclusionto the previous experiment's QML models achieveperformance scores closer to DL models rather than to MLmodels, as it can be observed from the bar plots illustrated inFig. 6. Despite their similar performance to DL models, QMLmodels are competitive and among the best models withalmost similar performances for both airline companies. performances for both airline companies.

VII. DISCUSSION AND CONCLUSION

In this work, the focus is on the airfare price prediction holisticapproach, considering different datasets and technologiesthat could be applied. To this end, four airlines and six destinationswere considered. To resolve the problem under study, eight ML models, six DL models, and two QML modelshave been employed and comparatively evaluated. Experimentalresults reveal that at least three models from eachdomain ML, DL, and QML are able to achieve accuraciesbetween 89% and 99% in this regression problem, for differentinternational destinations and airline companies. Resultsreveal that by using AI models and flight features that areavailable to customers before purchase, the airline companyticket price policy can be efficiently analysed. More featuresare publicly available and by using the abovetechnologies,robust simulations for flight tickets' price optimization andcustomer demand could be approximated, towards providingrich information to airline companies to build their optimalprice strategy. However, even under a small set of features,all model domains are able to extract patterns from the givenflight data and can find similarities between them. In thiswork, two different approaches have been investigated and analyzed: one based on the destinations (for all airlines) andone based on the airline companies (for all destinations).Future work from the perspective of the airline-based targetapplication, could include the same airline companies and destinations studied from different airports to examine ifthe information could be efficiently extracted. Moreover, thesame problem could be studied as a classification problemthrough customer segmentation, based on the flight features set.From a technological point of view, QML models havebeen studied under a regression application, which is limitedin the literature, since the advantage of QML models inclassical data is controversial, considering the limitations and the available quantum resources along with the computationaldemands in classical machines. Despite limitations like thenumber of qubits and noise levels in quantum machines,the availability of quantum hardware must be increased and become friendlier in order to pave the way for QML solutionsto be applied to more real-world applications.In this work, QML models for airfare price predictionachieved higher results in most cases compared to ML andDL models, despite the reported disadvantages and confronteddifficulties. It could be therefore concluded that futureapproaches to airfare price prediction based on the QMLdomain could provide efficient solutions, especially with theexpectation that the amount, complexity, and diversity ofdata will grow. Future work around QML methods in airfareprice prediction, includes the investigation of various differentmethods for data encoding in quantum states, and more quantum models like quantum Boltzmann machines, whichwill be able to generate flight data based on given air tickets feature sets and distributions. The resulted QML-based applicationcould be used as an airfare price policy generator.

REFERENCES

[1] S. Netessine and R. Shumsky, ‘‘Introduction to the theory and practiceof 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 immanentin nervous activity,’’ *Bull. Math. Biophys.*, vol. 5, no. 4, pp. 115–133,Dec. 1943.

[3] F. Rosenblatt, ‘‘The perceptron: A probabilistic model for informationstorage and organization in the brain,’’ *Psychol. Rev.*, vol. 65, no. 6pp. 386–408, 1958.

[4] B. E. Boser, I. M. Guyon, and V. N. Vapnik, ‘‘A training algorithm foroptimal 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 modelfor 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 learningapplied 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,I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, andD. Hassabis, ‘‘Mastering the game of go with deep neural networks andtree 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 logarithmsand 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.