International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

Cosmos Impact Factor-5.86

A HOLISTIC APPROACH ON AIRFARE PRICE PREDICTIONUSING MACHINE LEARNING TECHNIQUES

Dr. Adeline Johnsane, Professor, Department Of Data Science, SICET, Hyderabad

B.Kiran Sathya Goud, G.Yashwanth Goud, K.Sriniketh Rao, R.Mallikarjun

UG Student, Department Of Data Science, SICET, Hyderabad

ABSTRACT

Globalization involves new strategies and cost control policies that contribute to global competitiveness for professionals. Most companies change the starting price according to their policies and algorithms, which include many factors to find the right pricing policy. Recently, artificial intelligence (AI) models have started to be used for background work as they have many promising capabilities, fast adaptability, and general features. This article analyzes the prices of quotes starting in the air with the aim of using smart tools to find consistency in prices of different companies. More specifically, the system used was extracted from 136,917 flight records of Aegean Airlines, Turkish Airlines, Austrian Airlines and Lufthansa to six popular international destinations. The extraction process is then used to perform a holistic analysis from the perspective of the end user looking for the cheapest flight ticket, determine locationbased evaluation with each aircraft, and perform aircraft evaluation with each location. For the latter, artificial intelligence models and a total of 16 architectures from three different domains were considered to solve the ticket pricing problem: machine learning (ML), eightstate deep learning with eight statheart models. For the stateoftheart model (DL) CNN model and two models of Quantum Machine Learning (QML). Experimental results show that at least three models in the three domains of ML, DL, and QML can achieve 89% to 99% accuracy on this regression problem for various locations in the world and aircraft.INDEX terminology Air ticket price, artificial intelligence, deep learning, machine learning, prediction model, cost model, regression, quantum machine learning.

I. Introduction

For the last 50 years, flying was considered a luxury. Airlines offer more domestic flight services than international flights, but the price arrangement for flight tickets is the same. To increase profitability, airlines us e management and accounting software to implement optimization methods, booking adjustments and dynamic pricing. One change for airlines is the recognition of revenue management [1], a different pricing concept based on understanding, predicting and influencing users' products to make the most money. As a result, airlines began paying more for passenger choices and flights when adding international destinations. Therefore, the airline is open to all potential customers, as dynamic prices and airline services increase the competitiveness of airlines. In addition, the ability to shop online in recent years has changed many different fields and become a trend for ordinary people looking for quality, most deals and prices. There are now many websites that promote safe flying and list the same flight path on all airlines to get the most competitive flights. Additionally, sharing aviation knowledge through the evaluation process provides valuable information generated by daily airline users and used by cost management to adjust fares even minutes before flight departure. For this purpose, it is clear that the global economy and technological change affect airlines at a level that does not require a significant price adjustment. They reach the desired pace of change by following the changes voluntarily. Second, it increases the need for more complex algorithms and software for dynamic pricing policy optimization. As a result, flight ticket prices are now determined using artificial intelligence (AI) algorithmsPredictions get good results, become clear faster.Artificial intelligence has attracted the attention of the research community in many fields. Machine learning (ML) is a field of artificial intelligence that was first developed in 1943 by Walter Pitts and Warren McCulloch [2], who proposed mathematical models of neural networks that were not capable of learning. Seven years later, in 1950, Frank Rosenblatt proposed the perceptron [3] as the first neural network (NN) capable of learning. Perceptron has inspired researchers to develop and use many wellknown machine learning methods such as SVM [4], kNN [5], and Boosting method [6]. Some of these models produce results in the form of tabular data for statistical purposes. More information is also learned, such as sound symbols and images; but machine learning models cannot be powerful without supporting the extraction process. The latter needs are managed by deep learning (DL), which improves computational efficiency and reduces execution time. The main result of DL is the convolutional neural network (CNN), introduced by Fukushima in 1980; Inspired by the sensor, Fukushima uses neural networks for visual pattern recognition. The explicit motivation for this work was given by Yann LeCun [8] in 1990, who used the CNN model and iterative learning to recognize numbers written in images. Deep learning models have led to the creation of complex algorithms and applications that impact people's daily live s by automating the feature extraction process [9], [10]. But even today there is still a need for faster and more ML and DL algorithms because data is growing and determining that some problems, such as proteins, still cannot be solved even with advances in computing power (GPUs). Simulate the synthesis or production of chemicals during optimization.An attempt to solve the above problems and overcome the limitations faced by ML and DL algorithms is to combine quantum mechanics with ML and DL methods in quantum computing. The field of quantum computing was established in the 1990s, where quantum algorithms were proposed to solve complex problems, such as Shor's numerical factorization algorithm [11] in 1994 or Grover's algorithm [12] in 1997. These algorithms become the rationale for the creation of quantum computers, an area in which IBM is a leader. In the same decade, quantum machine learning (QML) began to develop with the introduction of quantum neural networks (QNN) in 1999 [13]; here quantum circuits and Grover algorithm were used to simulate neural network models. This work inspired many researchers to try QML. Therefore, between 1990 and 2010, many QML algorithms were introduced, including quantum multilayer perceptron (QMLP) [14], quantum support vector machine (QSVM) [15], and others. Although quantum machines are few and far between, quantum machine learning continues to expand even in industry to date, with applications and processes implemented in real quantum devices. The requirements of the QML model of the classical quantum model are also limited. Computer requirements are very high. Additionally, many QML methods are related to classical methods such as QNN training on classical data, where optimizers and losses are calculated based on classical data. The above facts have increased the growth of QML in the business and research field.This study is based on previous work on estimating the initial cost of aircraft [16]. Features of the airline system are extracted and used to highlight their competitive level in air ticket prices for different companies and destinations and to provide good opportunities to solve the problem. Additionally, the scope of applicability and performance of ML, DL and QML models in estimating the initial cost of the aircraft has been comprehensively analyzed. More specifically, two tests were carried out: In the first test, the problem was examined in terms of space for each plane (space-based approach). In particular, the AI models of the three names mentioned above were used in the same place by different aircraft to show the similarity of their operating models. In the second experiment, ML, DL and QML models were applied to all airline data sets (airlinebased approach) regardless of location. It should also be noted here that this work is the first attempt to bring a unified approach to the weather forecasting problem, which has been studied enti rely, including both the destination and the aircraft. company. It is also worth noting that, to the best of our knowledge, QML has never been applied to the ticket price prediction problem.

The main contributions of the proposed project can be summarized as follows:

1) Investigation of the relationship between the price policies of different companies.

2) Examine the impact of features of the weather forecasting problem.

3) QML model was used for weather forecasting in the database for the first time.

4) Comparison of the performance of ML, DL and QML models to estimate startup cost.

This document is organized as follows: Chapter 5 describes the tasks involved in estimating the initial cost of the aircraft. Chapter 5 presents the materials and methods used to accomplish these tasks, as well as the materials and methods used to accomplish these tasks. Section 4 describes the experimental setup, and Section 5 presents and discusses the experimental results. Section 6 presents the results of quantum machine learning and compares them with traditional models. Finally, Chapter 7 concludes the paper and offers suggestions for further research.

II. Related Study

The change in the international market and flight ticket price policy has produced a lot of information on the subject.Thereafter, there was a strong research interest in estimating the cost of operating an aircraft. In terms of artificial intelligence and data analysis, these data are transformed into data with many qualities and quantities that can be called big data, especially in cases where ticket prices and the cost of switching between services are very high. The problem of gambling ticket prices depends on the distribution of customers, time of ticket purchase, need for gambling tickets, etc., as discussed in the analysis of Abdella et al. It can be used in many ways such as. [17] Addresses target application problems and solutions. Overall, the topic of air ticket pricing has gained attention in the last three years. A Scopus search for the term "airline startup cost estimates" returned 24 records from 2003 to present; Most of the work took place in the last three years.

Wu et al. [18] implemented an airline price prediction application using two machine learning methods using physical features to identify Vietnam Airlines flights. The planning model is less than the plan, and while one plane is considered, the main point is the customer's use of the application. A different approach is proposed in [19]. A specific Recurrent Neural Network (RNN) for weather forecasting for events such as baseball games was developed and compared with classical ML models. Features describing the basketball game a nd the flight of the plane are combined into a single file to obtain a high estimate. The same approach was adopted by [20]. The authors use customer satisfaction, air ticket availability, distance, etc. to predict air ticket prices using learning models. They proposed a system that collects flight ticket information from various sources such as. Text [21] used ticket price gambling in the economics of the United States and India. The authors used the ML model and reported a prediction rate of 88%. In [22], Joshi et al. A similar approach was achieved by learning new features such as flight time, using fewer ML models, and achieving prediction scores of up to 90%. In [23], feature selection algorithms and hyperparameter methods were used to find the best combination of model parameters and flight descriptions to estimate the initial flight speed. In [24].

In general, all related work is done in a similar way. Requirements vary from: (1) the selection of a particular configuration, (2) the data to be stored, and (3) the goals of the application. Compared to all previous studies in the same field, this study: (4) uses a lot of technology, (5) tries to provide important information about aircraft competition and the consumer, (6) compares different companies.We introduce the algorithm for this problem for the first time, (7) offers two evaluation methods to make a complete analysis of the problem under consideration.

III. Data and methods

This section describes the entire scheme, focusing on the data used and the model chosen. It provides datasets, characterizations and visualizations that show the level of competition and the impact of globalization for air tickets in different airline destinations. Additionally, in this section, the adopted ML, DL and QML models will be introduced and each model will be briefly explained to show the performance and performance differences between them.Figure 1 graphically shows the steps of the proposed method. In the first step of the process, four aircraft and six targets were identified. The extracted features were applied to eight machine learning models and six deep learning models to evaluate the performance of the model. Evaluation is made from two different perspectives. The first test is positionbased testing, where the same set of positions is used for the model regardless of the plane. In the second experiment, an evaluation of the aircraft was carried out based on observations, using data from each aircraft for the model for each location. The two best results of machine learning from the first step of this method will be used in the second step and extended to the quantum domain. More specifically, in the second step of the method, the two bestperforming planes from step 1 and their three bestperforming positions were analyzed to compare the compatibility of the ML mod el and the QML model. The benchmark is determined by the same two assumptions as in step 1.

Data presentation and explanation

This study focuses on predicting air ticket prices for six different destinations of four airline companies. A Airlines are: Aegean Airlines, Austrian Airlines, Lufthansa and Turkish Airlines. Places of interest are as follows:

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)

List Flight information is for one year. To be clear here, flight information is not accurate across a year as some airlines often do not offer the same flights to all destinations throughout the year due to changes in demand. Table I describes the amount of flight data for each destination and each airline.

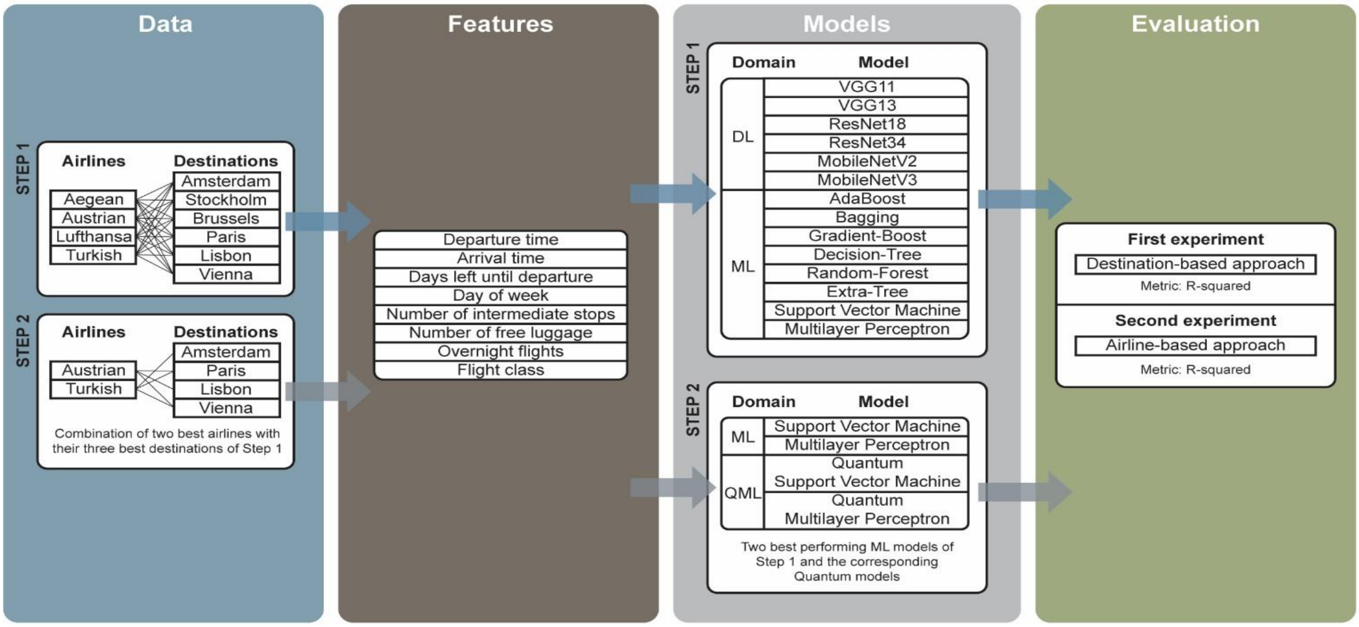

FIGURE 1. The proposed holistic approach to airfare price prediction.

As can be seen from Table 1, the Aegean has the most airlines; This is because Aegean is a Greek company and therefore has its headquarters at SKG Airport in Thessaloniki, Greece. It is also seen that the number of flights reveals the consistency of the airlines. In particular, Aegean Airlines, Turkish Airlines and Austrian Airlines have similar flights to some destinations (e.g. Turkish Airlines, SKG to ARN and BRU, etc.). Amsterdam (AMS) and Paris (CDG) are the most popular destinations for Aegean, Turkish and Austrian airlines. According to Table 1, as the distance to the destination increases, the number of flights also increases. One of the reasons for this may be that the ticket (including service) can be differentiated, making it more profitable for airlines.

Unfortunately, the QML field suffers from many limitations, such as the availability of quantum materials and the high demands of quantum simulations of classical systems. Therefore, hybrid algorithms are intended to work on classical and quantum data, not quantum data at all. To date, we do not know how to express everything according to the principles of quantum mechanics. In other words, depending on the problem of the application, the QML model can vary between full and hybrid form. When using the QML model to solve quantum problems, since the input data and the main goal are the quantum state, quantum mechanics and computational rules can be taught to every part of the process. In contrast, when the QML model is used for classical data, the target value must be encoded in the quantum state and the output quantum state must be determined in order to be obtained. The optimization algorithm used for learning in this strategy is still classical. This fact is often seen in regression problems where the output is numerical rather than in distributions where the output is mapped to qubit states defined by the probability of the distribution. QML models utilize the principles of quantum mechanics; because qubits are the main source of information, because the 2Ω classical state can be found for N qubits.Qubits. Compared to the former, the latter provides massive data encoding capabilities. It can also be seen whether qubits based on quantum mechanics are being discussed. This causes data to be stored in particle form and processed as symbols. Another advantage is entanglement, where all states of the qubit can exist and be similar at the same time. The disadvantage is that the entanglement state is lost during qubit measurement and therefore a new circuit must be created and run. Additionally, entanglement can create noise that disturbs neighboring qubits and destabilizes the quantum state.For this purpose, this article refers to QSVM [15], which describes the basic function from the same num ber of entangled qubits as the configuration. Entanglement is used by the revolving door and the visual weight value is estimated as the plane that best separates the data. Considering that qubits can go into many more states than classical objects, the characteristic space dimensions of quantum nuclei may be higher than classical ones. Therefore, split hyperplane data can be approximated better and faster on quantum hardware. Especially in classification problems, QSVM has been shown to be superior to classical SVM in many well known data for related applications such as cancer and fraud. Unfortunately, practical applications of QSVM are very limited because the number of features increases, hence the number of qubits, and hence the requirements of the game of Classical mechanics satisfy the limits of quantum materials. all available. Finally, all qubits are measured. The measured value represents the weight of the network given the input data, and the classical linear model is used to estimate the predicted value based on the output. In general, there are no predetermined rules for quantum circuit design and quantum gate selection for each problem. This proves that this technology is very new and can provide solutions to current and future problems of ML and DL models. In the study, classical optimization algorithms and classical loss energy are needed to adjust the quantum phase parameters.

Circuit based on reducing the loss between the predicted and actual target. The second is the most commonly used quantum neural network model, similar to the classical model but with some additions. Quantum circuits or quantum layers have a more compact structure because qubits can encode large amounts of information, so complex features can be extracted even from small architectures. Using entanglement in quantum mechanics, QMLP can achieve the speed of light by processing all the entanglements of a given object simultaneously. Unfortunately, this speed cannot be seen yet because the transfer of information from the classical state to the quantum state will take a huge amount of time in the process. Another problem is that there is general use of quantum functions based on linear models and nonlinear models; therefore the QMLP architecture uses classical functions (usually sigmoid) to assign input and quantum weights to the output.
According to the above, the scope of the entire proposal is to use all the above 16 models in the fields ML, DL and QML for flight pricing and comparison of benefits. In the following section, the experimental setup followed in this study is explained in detail.

IV. Test setup

In this study, two tests were conducted to detect the problem of the application. The first experiment used the locationbased approach, which selects models from the ML, DL, and QML domains to find the best option for each location for each aircraft. Through this experiment, a set of good models can be derived that describe the same destination for different airlines and have similar ticket price prediction. To achieve this, all data is distributed everywhere for every plane. Specifically, 24 records were created for four aircraft and six locations. In the second test, the general approach was followed, the same strategy as in the first test; this time the goal was to find the best possible model that simultaneously describes all six addresses for each plane. Therefore, the data set was divided into four parts according to the four selected planes. Then, a new number describing the position from 0 to 6 is added to the four files to clarify the data of the second test in the ML, DL and QML models. For deep learning models, dataset feature values are normalized and converted into images to be used as input to the CNN model.QML models are excluded at this stage because it would take a long time to process 28 different tests considering the flights in Table I.

CONCLUSION

This exception is also confirmed by the time units in the training model for each manager; here for ML and in DL the training process takes few hours and hence QML model takes few days to train. There is only one goal. To verify this calculation, consider that a 64dimensional vector is needed to simulate the dimensional qubit state in a classical computer, since 26 = 64. For a classical computer, this calculation is difficult.Because objects can only be in one state at a time. According to behavioral data, 8 qubits are used in the first experiment and 9 qubits are used in the second experiment. So the length of each profile is 256 and 512 respectively. Additionally, the branches included in QMLP double the weight of features, so ultimately 65,536 and 262,144 dimensions are needed respectively, requiring millions of floating point operations for a classical machine.

REFERENCES

Allen, I. E., & Seaman, J. (2008). ***Staying the course: Online education in the united states, 2008.*** ERIC.

Bambrick-Santoyo, P. (2010). ***Driven by data: A practical guide to improve instruction***. John Wiley & Sons.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2023). ***lme4: Linear mixed-effects models using eigen and S4***. https://github.com/lme4/lme4/

Betebenner, D. W. (2021). ***randomNames: Generate random given and surnames***. https://CenterForAssessment.github.io/randomNames

Bransford, J. D., Brown, A. L., Cocking, R. R., et al. (2000). ***How people learn*** (Vol. 11). Washington, DC: National academy press.

Breiman, L. (2001). Random forests. ***Machine Learning***, ***45***(1), 5–32.

Bryan, J. (2017). ***Project-oriented workflow***. https://www.tidyverse.org/blog/2017/12/workflow-vs-script/

Bryan, J. (2019). ***Reproducible examples and the 'reprex' package***. https://community.rstudio.com/t/video-reproducible-examples-and-the-reprex-package/14732

Bryan, J. (2020). ***Happy git with r***. https://happygitwithr.com/

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). ***Learning to improve: How america's schools can get better at getting better***. Harvard Education Press.

Campaign, D. Q. (2018). ***Teachers see the power of data - but don't have the time to use it***. https://dataqualitycampaign.org/wp-content/uploads/2018/09/DQC_DataEmpowers-Infographic.pdf

Conway, D. (2010). The data science venn diagram. ***Drew Conway***, ***10***. http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

Datnow, A., & Hubbard, L. (2015). Teachers' use of assessment data to inform instruction: Lessons from the past and prospects for the future. ***Teachers College Record***, ***117***(4), n4.

Dirksen, J. (2015). ***Design for how people learn***. New Riders.

Dweck, C. (2015). Carol dweck revisits the growth mindset. ***Education Week***, ***35***(5), 20–24.

Education Statistics U.S. Department of Education, N. C. for. (2019). Concentration of public school students eligible for free or reduced-price lunch. ***The Condition of Education 2019***. https://nces.ed.gov/fastfacts/display.asp?id=898

Elbers, B. (2020). ***Tidylog: Logging for dplyr and tidyr functions***. https://github.com/elbersb/tidylog/

Emdin, C. (2016). ***For white folks who teach in the hood... And the rest of y'all too: Reality pedagogy and urban education***. Beacon Press.

Estrellado, R. A., Bovee, E. A., Motsipak, J., Rosenberg, J. M., & Vel’asquez, I. C. (2019). ***Taylor and francis book proposal for data science in education***. https://github.com/data-edu/DSIEUR_support_files/blob/master/planning/T%26F%20Book%20Proposal%20for%20Data%20Science%20in%20Education.docx

Estrellado, R., Bovee, E., Mostipak, J., Rosenberg, J., & Vel’asquez, I. (2024). ***Dataedu: Package for data science in education using r***. https://github.com/data-edu/dataedu

Firke, S. (2023). ***Janitor: Simple tools for examining and cleaning dirty data***. https://github.com/sfirke/janitor

for Education Statistics, N. C. (2018). ***Public elementary/secondary school universe survey***. https://nces.ed.gov/programs/digest/d17/tables/dt17_204.10.asp?current=yes

Gelman, A., & Hill, J. (2006). ***Data analysis using regression and multilevel/hierarchical models***. Cambridge university press.

Grimm, K. J., Ram, N., & Estabrook, R. (2016). ***Growth modeling: Structural equation and multilevel modeling approaches***. Guilford Publications.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). ***The elements of statistical learning: Data mining, inference, and prediction***. Springer Science & Business Media.

Hattie, J. (2012). ***Visible learning for teachers: Maximizing impact on learning***. Routledge.

Healy, K. (2019). ***Data visualization: A practical introduction***. Princeton University Press.

Hill, A. (2017). ***Up and running with blogdown***. https://alison.rbind.io/post/2017-06-12-up-and-running-with-blogdown/

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. ***Science***, ***349***(6245), 261–266.

Ismay, C., & Kim, A. Y. (2019). ***Statistical inference via data science***. CRC Press.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). ***An introduction to statistical learning*** (Vol. 112). Springer.

Jarvis, C. (2019). ***Creating calling***. HarperCollins.

Jordan, R. (2015). ***High-poverty schools undermine education for children of color***. https://www.urban.org/urban-wire/high-poverty-schools-undermine-education-children-color

Kahneman, D. (2011). ***Thinking fast and slow***.

Kearney, Michael W. (2016). Rtweet: Collecting twitter data. ***Comprehensive R Archive Network. Available at: Https://Cran. R-Project. Org/Package= Rtweet***.

Kearney, Michael W., Revilla Sancho, L., & Wickham, H. (2023). ***Rtweet: Collecting twitter data***. https://docs.ropensci.org/rtweet/

Kleon, A. (2012). ***Steal like an artist: 10 things nobody told you about being creative***. Workman Publishing.

Kozol, J. (2012). ***Savage inequalities: Children in america's schools***. Broadway Books.

Krist, C., Schwarz, C. V., & Reiser, B. J. (2019). Identifying essential epistemic heuristics for guiding mechanistic reasoni ng in science learning. ***Journal of the Learning Sciences***, ***28***(2), 160–205.

Kuhn, M. et al. (2008). Building predictive models in r using the caret package. ***Journal of Statistical Software***, ***28***(5), 1–26.

Kuhn, M. (2023). ***Caret: Classification and regression training***. https://github.com/topepo/caret/

Kurz, S. (2019). ***Statistical rethinking with brms, ggplot2, and the tidyverse***.