ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 1, 2024

A HOLISTIC APPROACH ON AIRFARE PRICE PREDICTIONUSING MACHINE LEARNING TECHNIQUES

Mrs. K. Lavanya, Mrs. D. HemaMalini, Mrs. A. Baranishri, Mrs. S. Chandra Priyadharshini

Assistant Professor [1,2,3] Associate Professor [4]

klavanya@actechnology.in, hemamalini.d@actechnology.in, baranishri.a@actechnology.in, chandrapriyadharshini.s@actechnology.in

Department of AI & DS, Arjun College of Technology, Thamaraikulam, Coimbatore-Pollachi Highway, Coimbatore, Tamilnadu-642 120

ABSTRACT

As a result of globalisation, professionals are more competitive on a worldwide scale thanks to new tactics and cost control techniques. There are a lot of aspects that go into determining the optimal pricing policy, and most organisations adjust the initial price based on these policies and algorithms. Due to their generalisability, speed of adaptation, and several potential features, artificial intelligence (AI) models have recently begun to be employed for background labour. With the goal of using smart methods to discover consistency in rates across various organisations, this essay examines the prices of quotations beginning in the air. Aegean, Turkish, Austrian, and Lufthansa flights to six prominent international destinations made up the bulk of the 136,917 flight data used by the system. Afterwards, the extraction method is used to conduct a comprehensive analysis from the standpoint of the end user seeking the most affordable flight ticket. This analysis involves determining the location-based assessment for each aircraft and conducting evaluations of the aircraft with each location. The latter took into account a total of sixteen architectures from three distinct domains—machine learning (ML), eight-state deep learning with eight statheart models—in an effort to resolve the ticket price issue. Considered for the cutting-edge DL CNN model and two QML models. The experimental findings demonstrate that for different locations and aircraft, at least three models from the ML, DL, and QML domains may accomplish an accuracy ranging from 89% to 99% on this regression problem.jargon used in an index Keywords: air ticket pricing, quantum machine learning, cost model, prediction, deep learning, and machine learning.

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 day were developed thanks to deep learning models.

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 1, 2024

e-commerce and bookkeeping platforms that enable optimisation strategies, booking modifications, and dynamic pricing. Airline companies have begun to take revenue management, a new approach to pricing that focusses on knowing, anticipating, and shaping consumer goods for maximum profit, seriously [1]. So, as airlines started offering flights to more foreign locations, passengers started paying more for their choices. Because of this, the airline welcomes any and all prospective passengers, since the competitiveness of airlines is enhanced by dynamic pricing and airline services. Online shopping has also grown popular with regular people in search of the best bargains, quality, and pricing, and it has revolutionised several industries in recent years. Nowadays, you can find a plethora of websites that stress the importance of flying safely and provide a universal flight route that all airlines use to get the best deals. In addition, the review process is a great way for people to share what they know about aviation, which helps airlines and their customers. Cost management uses this information to change pricing right up until a flight takes off. To achieve this goal, it is evident that airlines are not much impacted by changes in the global economy and technology, so no substantial price adjustment is necessary. They willingly adapt to the changes, which allows them to attain the ideal speed. It also makes dynamic pricing policy optimisation more difficult and calls for more sophisticated algorithms and software. Consequently, AI algorithms are being used to decide the pricing of plane tickets.Forecasts yield favourable outcomes and gain clarity more rapidly.The scientific community in several sectors has taken an interest in artificial intelligence. Machine learning (ML) is an area of AI that was first created in 1943 by mathematical models of non-learning neural networks provided by Warren McCulloch and Walter Pitts [2]. In 1950, seven years after that, Frank Rosenblatt put out the perceptron [3] as the first NN that could learn. Many famous machine learning algorithms, including support vector machines (SVMs), k-nearest neighbours (kNNs), and the boosting technique (Boosting), were developed and used because of perceptrons. For statistical reasons, several of these models provide tabular data as their output. Machine learning models aren't useful until they have data to back up the extraction process, although they do learn more, such audio symbols and pictures. Deep learning (DL) handles the latter, which speeds up calculation and decreases execution time. Fukushima invented the convolutional neural network (CNN) in 1980 as a major outcome of DL. Neural networks, motivated by sensors, are used by Fukushima for visual pattern recognition. Yann LeCun's [8] 1990 work, in which he employed the CNN model and iterative learning to identify numerals written in photographs, explicitly motivated this study. A lot of complicated algorithms and apps that people use every .

s by automating the feature extraction process [9], [10]. But even today there is still a need for faster and m ore 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 productio n of chemicals during optimization.An attempt to solve the above problems and overcome the limitations fa ced by ML and DL algorithms is to combine quantum mechanics with ML and DL methods in quantum co mputing. The field of quantum computing was established in the 1990s, where quantum algorithms were pr oposed 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 wit h the introduction of quantum neural networks (QNN) in 1999 [13]; here quantum circuits and Grover algor ithm were used to simulate neural network models. This work inspired many researchers to try QML. There fore, between 1990 and 2010, many QML algorithms were introduced, including quantum multilayer perce ptron (QMLP) [14], quantum support vector machine (QSVM) [15], and others. Although quantum machin es are few and far between, quantum machine learning continues to expand even in industry to date, with ap plications 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 losse s 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]. Featur es of the airline system are extracted and used to highlight their competitive level in air ticket prices for diff erent 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 airc raft has been comprehensively analyzed. More specifically, two tests were carried out: In the first test, the p roblem 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 si milarity 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 Secti on 5 presents and discusses the experimental results. Section 6 presents the results of quantum machine lear ning and compares them with traditional models. Finally, Chapter 7 concludes the paper and offers suggesti ons for further research.

I. Related Study

The change in the international market and flight ticket price policy has produced a lot of information on th e subject.Thereafter, there was a strong research interest in estimating the cost of operating an aircraft. In te rms 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 betwe en services are very high. The problem of gambling ticket prices depends on the distribution of customers, t ime 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 co st 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 usin g 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 prop osed in [19]. A specific Recurrent Neural Network (RNN) for weather forecasting for events such as baseba ll 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 a dopted by [20]. The authors use customer satisfaction, air ticket availability, distance, etc. to predict air tick et 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 a uthors used the ML model and reported a prediction rate of 88%. In [22], Joshi et al. A similar approach wa s achieved by learning new features such as flight time, using fewer ML models, and achieving prediction s cores 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 particul ar configuration, (2) the data to be stored, and (3) the goals of the application. Compared to all previous stu dies in the same field, this study: (4) uses a lot of technology, (5) tries to provide important information abo ut 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 proble m under consideration.

II. Data and methods

With an emphasis on the selected model and the data utilised, this section details the whole scheme. It shows the degree of competition and the effect of globalisation on airline ticket prices in various locations using data sets, descriptions, and visualisations. Furthermore, this part will present the chosen ML, DL, and QML modules and provide a short explanation of each model to demonstrate their performance and any changes in performance.You may see the stages of the suggested strategy visually in Figure 1. The procedure began with the identification of four planes and six potential targets. In order to assess the model's efficacy, the retrieved characteristics were fed into eight ML models and six DL models. Two viewpoints are used in the evaluation process. In the first kind of test, known as position-based testing, the model is always evaluated using the same set of locations, independent of the plane. In the second trial, data from each plane was input into the model for each site, and the planes were then evaluated based on the results. Step two of this strategy involves taking the two best machine learning results from step one and applying them to the quantum domain. Specifically, the second stage of the procedure included comparing the compatibility of the ML mod with the two best-performing planes from step 1 and their three best-performing locations.

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 fo llows:

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 s ome airlines often do not offer the same flights to all destinations throughout the year due to changes in de mand. Table I describes the amount of flight data for each destination and each airline.

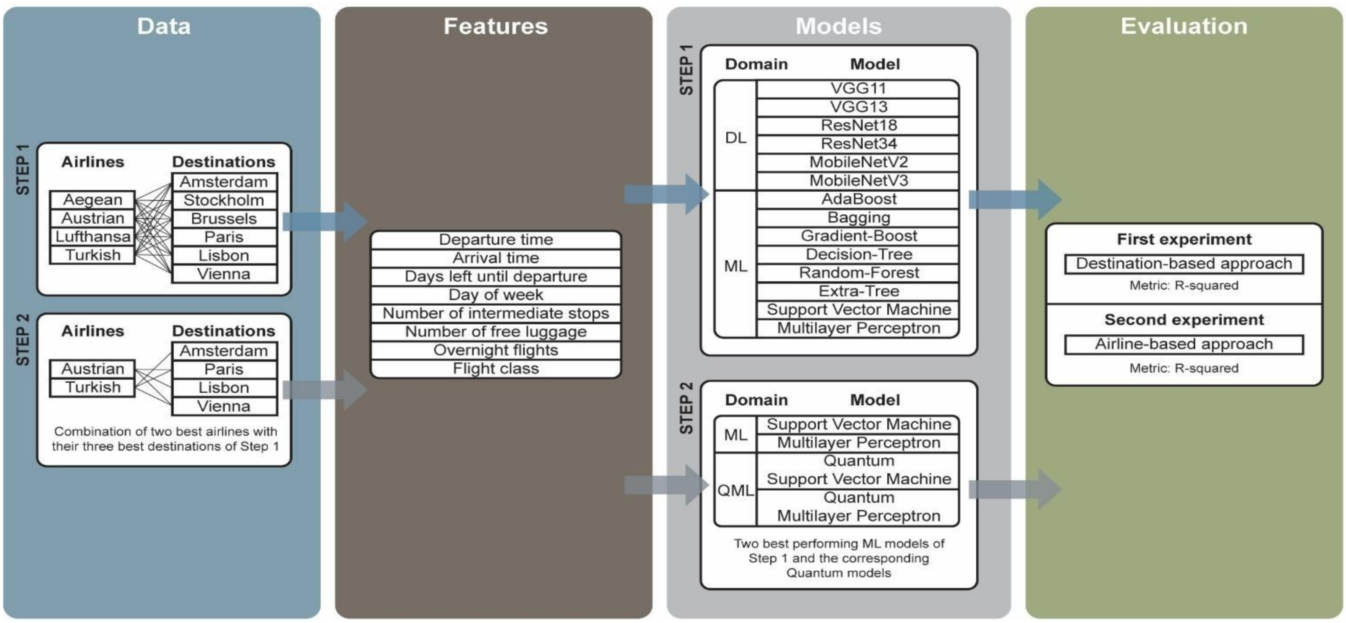

FIGURE 1. The proposed holistic approach to airfare price prediction.

Because it is a Greek firm with its headquarters at SKG Airport in Thessaloniki, Greece, Aegean has the most airlines, as seen in Table 1. Additionally, the frequency of flights provides insight on the reliability of the carriers. For example, SKG to ARN and BRU on Turkish Airlines, etc., are flights that are comparable to those of Austrian Airlines, Aegean Airlines, and Turkish Airlines. The most common destinations for Aegean, Turkish, and Austrian flights are Amsterdam (AMS) and Paris (CDG). The number of flights rises as the distance to the destination increases, according to Table 1. Possible explanations include the fact that airlines may increase their profits by differentiating the ticket (including service). A lot of problems plague the QML area, including the fact that quantum materials aren't readily available and that quantum simulations of classical systems are quite demanding. Hence, hybrid algorithms aren't meant to handle pure quantum data, but rather a combination of classical and quantum metrics. According to the rules of quantum physics, we still don't know how to express everything. That is to say, the QML model may take either a complete or hybrid form, depending on the application's problem. Quantum mechanics and computational principles may be taught to every phase of the process while utilising QML mode l to solve quantum issues, because the input data and major aim are the quantum state. On the other hand, while working with classical data and the QML model, you have to encode the goal value in the quantum state and then figure out the output quantum state to retrieve it. This tactic still makes use of a standard optimisation algorithm for its learning processes. This is more common in numerical output issues like regressions than in distributions, where the output is transformed into qubit states determined by the distribution's probability. Since qubits are the primary data source and the 2Ω classical state may be identified for N qubits, QML models employ the concepts of quantum mechanics.The quantum bits. The latter offers far more data encoding possibilities than the former. Whether or not qubits grounded in quantum physics are under discussion is also visible. This results in the data being processed as symbols and stored as particles. Entanglement is another perk; it allows all possible qubit states to exist and be comparable simultaneously. The loss of the entangled state during qubit measurement necessitates the creation and execution of a new circuit, which is a drawback. Noise produced by entangled qubits may destabilise the quantum state and disrupt nearby qubits.Quasi Support Vector Machines [15] explains the fundamental function using the same number of examples as this article.

of qubits that are entangled as the setup. The revolving door makes advantage of entanglement, and the ideal plane to separate the data is evaluated as the visual weight value. It is possible that quantum nuclei have greater characteristic space dimensions than classical ones due to the fact that qubits may enter more states than classical items. Thus, quantum technology allows for quicker and more accurate approximations of split hyperplane data. Several well-known datasets from relevant applications, such cancer and fraud, have shown that QSVM is better than standard SVM, particularly in classification difficulties. Regrettably, the practical uses of QSV M are severely constrained due to the fact that when the number of features increases—and hence the number of qubits—the demands of classical mechanics' game meet the limitations of quantum materials. they are all at your fingertips. At last, every qubit is measured. The measured value stands for the network's weight in relation to the input data, and the output is utilised to estimate the predicted value using the classical linear model. As a general rule, there are no hard and fast guidelines for designing quantum circuits or selecting quantum gates for any given situation. The fact that this technique can solve both the present and future issues with ML and DL models demonstrates how novel it is. The research requires the use of classical loss energy and classical optimisation techniques in order to fine-tune the quantum phase parameters. Circuit designed to minimise the deviation between the intended and actual outcome. Second, there is the most popular model of quantum neural networks, which is essentially the same as the classical model with a few tweaks here and there. Because qubits may encode huge quantities of information, quantum circuits or quantum layers can have a more compact structure. This allows for the extraction of sophisticated characteristics from even the smallest of structures. By simultaneously analysing all of an object's entanglements, QMLP is able to use quantum mechanics' entanglement to reach the speed of light. It will take a very long time for information to get from the classical to the quantum state, thus we haven't witnessed this speed yet. The QMLP architecture also has the issue of using quantum functions that are based on linear and nonlinear models, which means that inputs are assigned using classical functions, often sigmoid, and outputs are given quantum weights. Based on the information provided, the proposal aims to use all sixteen models in the domains of ML, DL, and QML to determine flight prices and compare advantages. The experimental setup used in this work is described in detail in the section that follows.

III. Test setup

Two tests were carried out in this research to identify the application's difficulty. To determine the optimal optimisation for each aircraft at each site, the first experiment made use of the location-based technique, which draws models from the ML, DL, and QML domains. By doing this experiment, we may construct a set of excellent models that can estimate ticket prices similarly and identify the same destination for various airlines. All data is dispersed over all planes to accomplish this. Four planes and six places were the subjects of the twenty-four entries. Following the same basic approach and technique as the first test, the second one aimed to identify the optimal model that could explain all six addresses for each plan concurrently. So, the four chosen planes served as a dividing line in the data set. The second test data for the ML, DL, and QML models is then clarified by adding a new number representing the location from 0 to 6 to the four files. Before being fed into a convolutional neural network (CNN) model, dataset feature values undergo normalisation and image conversion for deep learning models.Because processing 28 separate tests while taking into account the flights in Table I would take too long, QML models are not being considered at this level.

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.Beca use objects can only be in one state at a time. According to behavioral data, 8 qubits are used in the first exp eriment and 9 qubits are used in the second experiment. So the length of each profile is 256 and 512 respect ively. 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***.