Hyper-Growth Business Impact : Predictive Modeling For Future Revenue From Past Advertising

Introduction:

Can business success be guaranteed? If success in business was guaranteed would you believe it? If only a small number of businesses achieve a hyper-growth business scale then there has to be strong indicators that if followed, if this doesn't guarantee success it will at least give you the highest statistical probability of reaching it. So maybe success cannot be guaranteed in business but what if the framework and strategic approach to conducting business can be guaranteed so that this gives companies the best chances of reaching hyper growth revenue levels. High company revenue figures are not random and one of many evidence-based approaches to massive business growth is applying data science to help drive decisions that have direct high impacts on overall revenue generated (Kraus, 2018). Data science is the study of data to derive insights from and to take actions to reach a predetermined expected outcome.

Within data science and the use of math and economics businesses apply predictive modeling to predict outcomes. In this context, companies use predictive modeling to show what future revenue would look like. This is a standard process companies use so they can adjust business operations accordingly to give them the highest revenue numbers possible currently and in the future (Gajewar, 2016). One of many ways to establish high revenue is from the use and execution of effective advertising to grow business. If advertising is used skillfully along with other revenue generating activities then this helps businesses accomplish a large magnitude of growth.

Additionally, if skillful use of advertising is mastered there is a higher probability business can reach a hyper-growth business phase (Würfel, 2021). In this project, altogether, when businesses utilize previous advertising revenue performance to predict future revenue, assuming their business meets the criteria and has the capacity to enter a hyper-business growth phase, their probability to enter hyper-growth business mode is more likely to occur. The business problem that is investigated or solved is to reach the business status of hyper-growth through advertising from accurately predicting future revenue. If business revenue from past advertising is strategically analyzed then it is possible to predict future revenue and increase the probability of achieving the hyper-growth business status.

Data & Methods:

The dataset used in this project is sourced from a Kaggle.com advertising sales dataset, a platform well-respected for hosting a wide range of high-quality and diverse datasets. Kaggle is a trusted environment among data scientists and business strategists who apply data-driven insights to solve real-world problems, making it an appropriate and credible source for this project. Given the project's aim — helping businesses use advertising to predict and drive toward hyper-growth revenue — it was essential to select data that specifically relates to advertising efforts and corresponding revenue performance. This dataset is a strong match because it reflects the core relationship between advertising activities and revenue generation, two critical components to scale a business aggressively. Just as choosing the right data is crucial, the methodology applied to the data is equally important. For this project, the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework will be used as the guiding approach.

This framework ensures that the process remains structured, strategic, and focused on business objectives. The project will move through the CRISP-DM phases as follows: (1) Confirming a deep business understanding and clearly defining the business problem — in this case, predicting future revenue performance based on past advertising efforts to position companies for hyper-growth. (2) Data understanding by exploring the dataset, identifying key patterns and potential predictors. (3) Preparing the data by cleaning, organizing, and constructing a version suitable for advanced analysis and predictive modeling. (4) Engaging in the modeling phase where predictive algorithms will be developed and tuned to forecast future revenue based on historical advertising performance. (5) Conducting a careful evaluation of the models, assessing how well they meet the business objectives, and refining as needed. (6) Deploying the final predictive model in a way that allows for ongoing adjustments and optimization to maintain alignment with business goals and maximize the probability of achieving hyper-growth success. Through using both a reputable data source and a strategic data science methodology, this project aims to contribute meaningful insights that give businesses their best statistical shot at reaching hyper business growth status.

Data Explore:

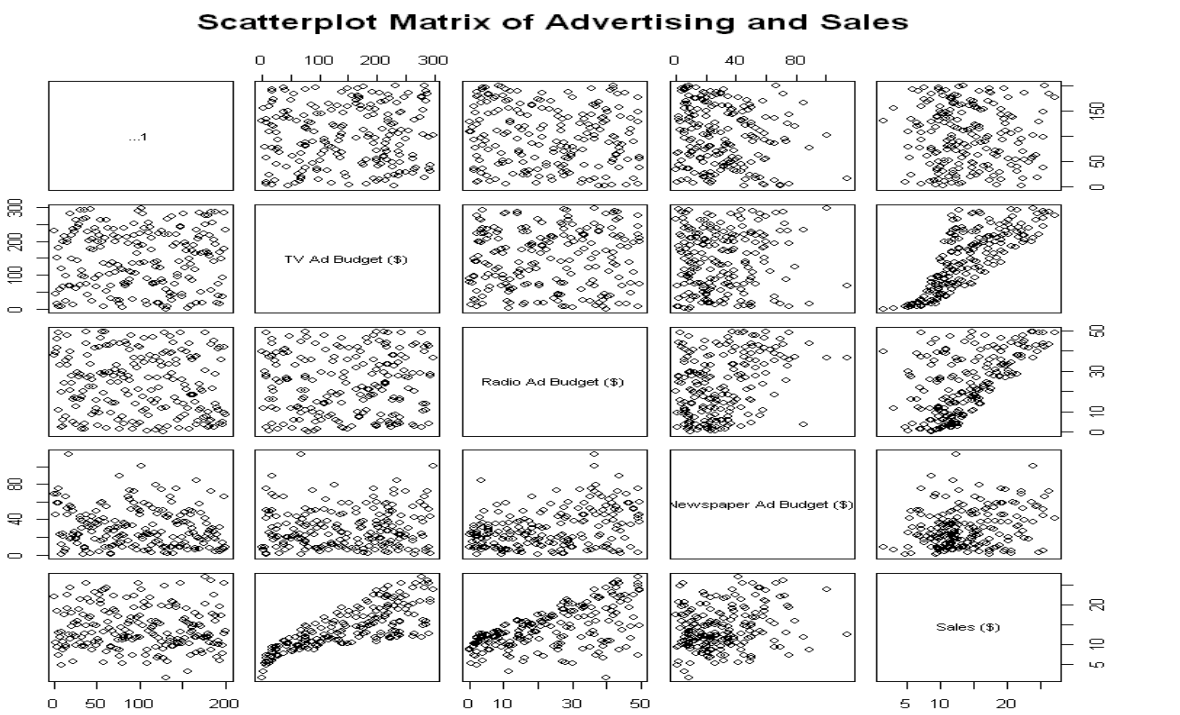
In the data explore phase, I examine the dataset's structure and key variables to understand how advertising performance correlates with revenue generation. The goal was to uncover important patterns that can help forecast future sales based on historical data, offering businesses actionable insights into the impact of their advertising strategies. The Kaggle advertising sales dataset provides detailed information on advertising expenditures across various channels, along with the corresponding revenue outcomes. This makes it a vital resource for analyzing the direct effects of advertising initiatives on sales performance. By exploring the relationships between revenue and advertising spend in different media the dataset only has TV, radio, and newspapers

as advertising channels. The objective was to identify the most high impact factors that drove revenue growth.

This dataset was very small but acts as a resource to be utilized to demonstrate the power of approaching solving a business problem and here it's reaching the hyper growth business phase. As stated earlier, there were only three predictor variables in this dataset. The three data predictors are all key to my analysis, which include TV advertising spend: The amount spent on TV ads. Radio advertising spend: The amount spent on radio ads. Newspaper advertising spend: The amount spent on newspaper ads. The last variable, revenue or sales, will be what I'm looking to predict in the future so revenue generated is the sales or revenue produced as a result of the advertising efforts.

By examining the correlations between advertising spend across these three media and the revenue generated, I can gain a deeper understanding of how each type of advertising impacts sales. This analysis will provide a clearer picture of which advertising channels offer the highest return on investment (ROI) and are most effective at driving revenue. Below is Figure 1, which paints a picture of a matrix. This matrix is a grid of scatter plots where each variable is plotted against every other variable, including the target, which is sales. The advertising spending on the different channels or TV, radio and newspaper are all compared to sales to explore patterns. Upon observing patterns in Figure 1 there is a low positive correlation of TV advertising to go along with sales generated.

Figure 1



1

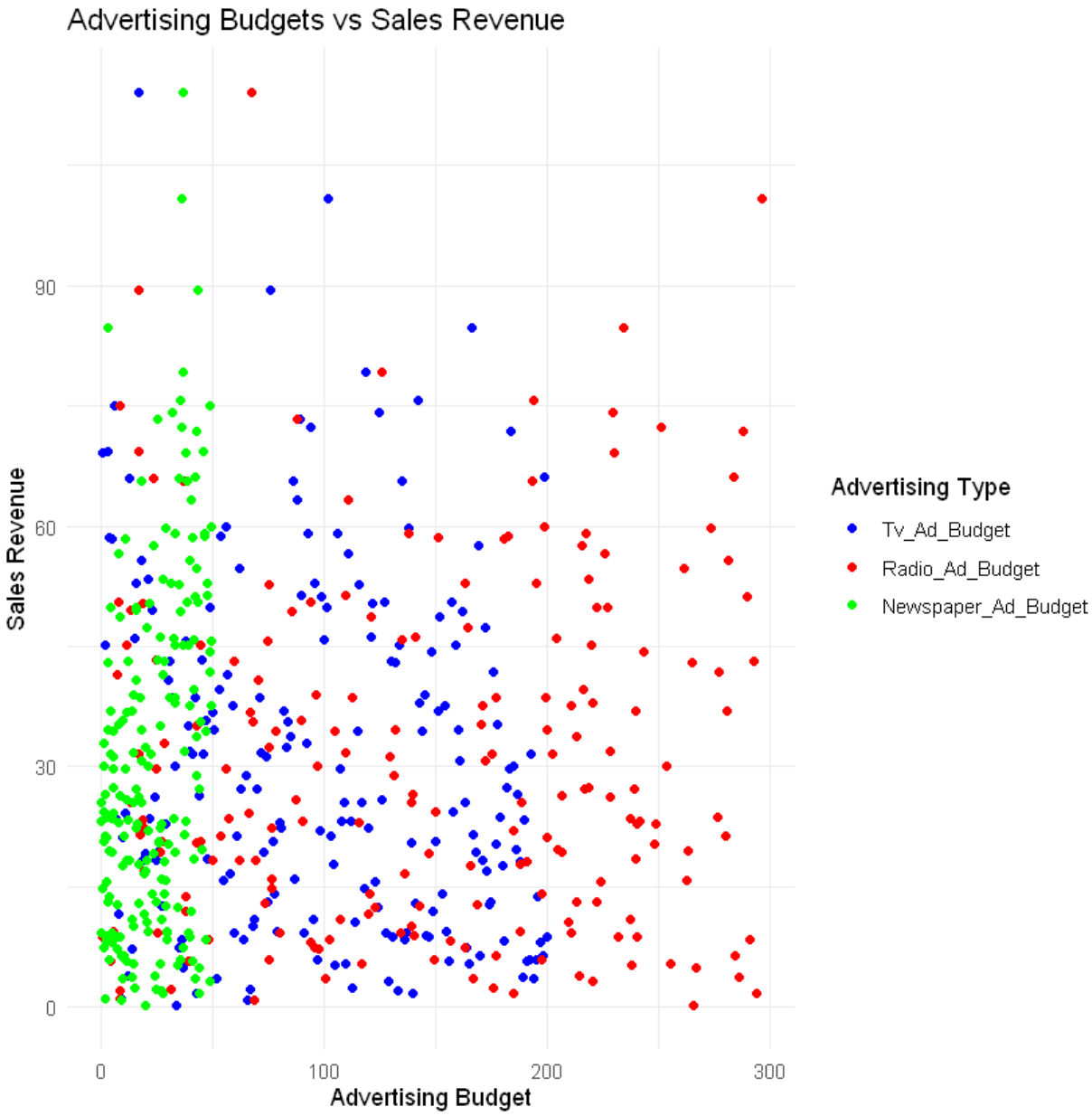
The initial step in this exploration involves calculating the correlation coefficients between these variables. This will help identify any strong linear relationships between advertising expenditures and revenue. If a significant correlation exists, it may suggest that past advertising efforts have a predictable effect on future sales. For instance, if the data shows a high positive correlation between TV advertising spend and revenue, it indicates that increasing TV ad spend could lead to higher sales, which will inform predictive modeling in the next steps of the project.

Additionally, I will visualize these relationships through scatter plots and correlation heatmaps, allowing for a deep understanding of how each advertising medium influences sales performance. The visual representation will further highlight any outliers or patterns that could inform more broader insights for future advertising campaigns.

This exploration will also help uncover potential areas where certain advertising channels may not contribute as significantly to revenue, enabling businesses to reallocate resources more effectively. Ultimately, the aim is to develop a comprehensive understanding of past advertising effectiveness, laying the foundation for predictive models that can project future sales with greater accuracy. By thoroughly exploring the dataset, this phase will undergo data preparation and modeling efforts, ensuring that the right features are used to predict future sales outcomes.

[1] Figure one illustrates a scatterplot matrix to express the relationships between three different advertising channels compared to sales generated. A matrix is a grid of scatter plots where each variable is plotted against every other variable, including the target, which is sales

Figure 2



[2]

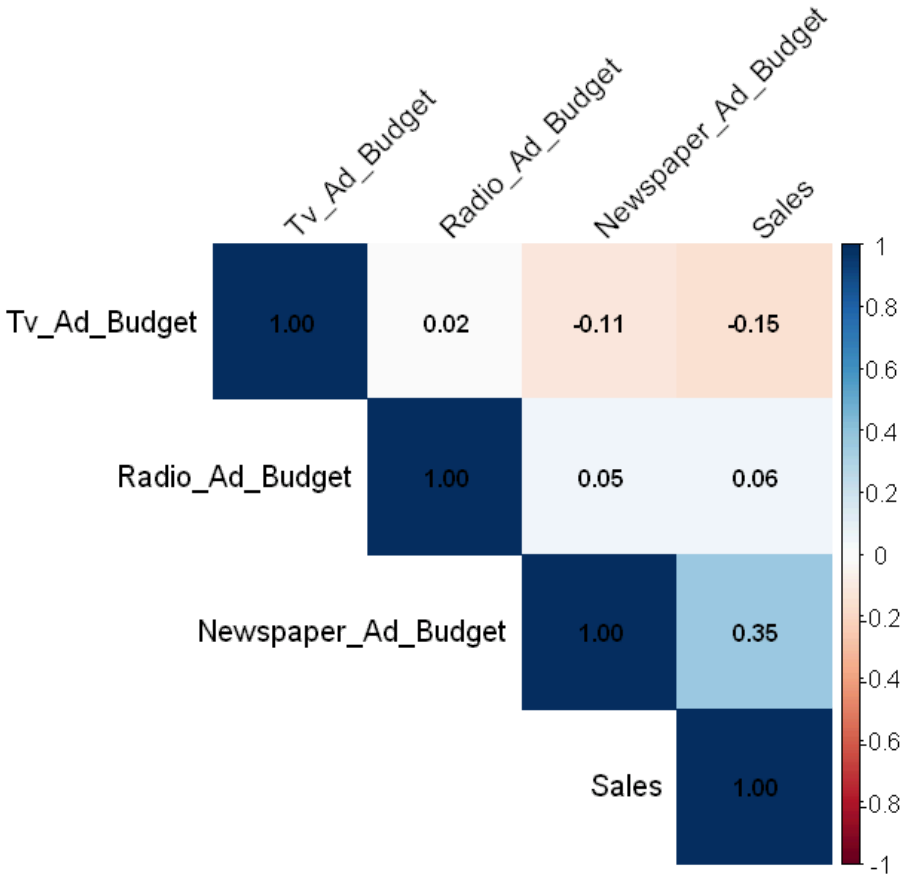
Figure 2 above are scatterplots that display the relationships between each advertising channel and sales, allowing each variable to be compared independently of each other. We can see that sales are being generated from advertising but there are no real strong correlations. As a default businesses will likely grow sales from advertising but the magnitude of getting results that the

[2] Figure 2 is 2D scatterplots that display the relationships between each advertising channel and sales, allowing each variable to be compared independently of each other

sales are higher than the cost of advertising is not always accomplished. Overall, from the patterns above there is no strong correlation of these advertising marketing channels compared to sales created. Its likely newspapers have the worst impact on sales while TV and radio have greater impacts generally then newspaper advertising.

Figure 3

Heat Map Correlations (Advertising Spend & Revenue Variables)



[3] Figure 3 expresses a correlation heatmap of the strength and direction of relationships between multiple numerical variables using a color-coded matrix. The number variables are the advertising channels and revenue

Figure 3 above is a heatmap correlation of all my variables. In my analysis, the correlation heatmap is highly likely to be misleading data because the dataset is small, 200 x 5 is the data set size. There are limited observations, correlation coefficients here are unstable and overly sensitive to outliers or random irrelevant fluctuations, which causes exaggerations or hides the actual relationships between variables. For example, the heatmap displays a weak general positive correlation between newspaper advertising and sales (0.35), illustrating a stronger relationship than what is seen by other visualizations.

The heatmap is saying television and radio advertising are weaker or have negative correlations, despite other visuals like previous scatter plots showing a more clear and consistent positive association with sales, even though not a big one. This inaccuracy put the spotlight on key limitations of correlation analysis on small datasets. With small datasets, heatmap correlations only discover linear relationships and don't consider non-linear patterns or the effects of multicollinearity between predictors. As a result, greater weight should be given to visual patterns observed in scatter plots and to results from multiple regression models, which give more reliable determinants of the true influence of each advertising channel on sales.

Model & Evaluation:

During my modeling phase I applied a combination of mathematical or computational infrastructure that captures relationships within the data. After careful review and reading other publications within this same core subject, I went with a 70/30 split for predicting future revenue from past advertising performance because having a larger test set allows for an effective evaluation of how well the model will perform on new, unseen data, such as actual future revenue. It’s critical to believe in the model’s predictions in real-world situations, so by me reserving 30% of the data for testing allows the evaluation to be more accurate. Additionally, using the 70% for training this lets the model keep a good amount of data to learn and highlight trends. Overall, when I considered the dataset and the project goals this split was a great balance and makes sense for this kind of forecasting objective.

For this project, the objective is to predict future revenue based on historical advertising performance. By analyzing past advertising data, including TV, radio, and digital ad budgets, I sought to forecast the revenue generated from advertising expenditures. The modeling process helps in understanding how different advertising investments impact future revenue results so this allows business leaders to optimize future advertising effectively. To achieve high potential for maximum future revenue, I constructed different machine learning models to test so I could determine the most accurate method for predicting future revenue. My models were designed to discover complex, non-linear relationships between advertising spending across multiple channels and the revenue as an outcome that was generated. The Root Mean Squared Error

(RMSE) metric was used to evaluate the models, as it expresses a measure of how far off the predicted revenue values are from the actual values.

The first model I did the test with was a Multiple Linear Regression, which assumes a linear relationship between advertising budgets (TV, radio, and digital) and the revenue generated. This model acted as a general guide to understand how simple linear assumptions perform in this environment. The model provided insights into the relationship between advertising spend and revenue but it did not highlight the more complex patterns in the data, leading to a bad RMSE score.

Next, I applied the Random Forest model. Random Forest builds multiple decision trees and aggregates their predictions to improve accuracy. By identifying non-linear interactions and reducing overfitting, Random Forest is ideal for more complicated data relationships. In this case, its RMSE evaluation performance was bad and is under the threshold of an acceptable or passable model. The third model I tested was XGBoost, a gradient boosting method, which usually has a strong predictive power. The XGBoost corrects errors from previous models and optimizes for accuracy. This approach allowed XGBoost to appear to look like a good model but after evaluation the RMSE showed its accuracy was off by a wide margin but was better then other models tried.

The fourth model, Linear Model using only TV & Radio Advertising spending, was a simple version of the linear regression model that considered only the TV and radio advertising efforts as predictors. This model helped isolate the effects of these two channels, it performed poorly in comparison to the more complex models. By not including newspaper advertising spend, it was likely that I would have a better predictive model since newspapers contributed the least when growing revenue compared to TV and radio. Regardless, the model did not meet standard performance metrics to be considered applicable for business use.

Finally, the fifth model was the TV Ad Budget Only. Considering only the TV advertising efforts as the predictor for revenue, this model further simplified the problem by ignoring other advertising channels. After evaluation, it performed horribly, the RMSE indicated its predictions are largely off from the actual real revenue generated by advertising. The single predictor approach failed to account for the contributions of radio and newspaper ads, which are possibly critical factors or elements of success in generating revenue.

Each of these models was evaluated using RMSE to quantify their predictive performance. Simple models such as Multiple Linear Regression and Linear Model using TV & Radio advertising efforts only provided useful insights, they were unable to highlight the complexity of the relationships between advertising spend and revenue. More advanced models like Random Forest and XGBoost were able to account for these complexities but still weren't able to make

strong and accurate predictions, resulting in RMSE scores that expressed the models as highly inaccurate.

Results, Discussion & Conclusion

The business problem to solve was to increase business revenue from creating a business model that predicts future revenue from past performance of advertising spending. From the predictions of future revenue along with other business strategies, assuming business has the capacity to grow exponentially, or economies of scale, a potential to have a greater probability of reaching the hyper growth business phases could be an added benefit from predicting future revenue. The results of all of the models do not meet the guidelines from standard model performance evaluations. This finding doesn't mean the project was a failure but simply shows the evidence-based approach that uses data science to predict future revenue and add other elements of business growth that are outside of the context of this project. In conclusion, The project was a success because it aimed at showing one of many approaches to increase the realistic possibility of achieving the legendary status of a hyper growth business organization.

Although all the predictive models failed the performance evaluation there are reasons why. There were various limitations and a couple are that the dataset only contained three advertising channels. Other constraints are that there were no time stamps nor a data sample large enough for a valid usable predictive model. The dataset does not show how sales and advertising spend change over time. Future improvement would be to add more data and have the project be able to see the impacts of advertising over a span of time. To predict the future, you usually need data that shows what happened month by month, week by week, or day by day — so you can see trends, patterns, and delays. Lastly, regardless of the findings, this research can still be used because it's a universal approach. Professionals and businesses can derive value to then customize it to fit into their own pursuits to reach a result or final outcome that satisfies them.

References:

Agu, C., et al. "Utilizing Advanced Data Analytics to Boost Revenue Growth and Operational Efficiency in Technology Firms." ResearchGate, 2024. https://www.researchgate.net/publication/386276269

Bose, Amitabh. "Supporting Revenue Growth with Predictive Analytics in Marketing." WNS Global Services, 2024. https://www.wns.com/perspectives/articles/articledetail/889

Dekimpe, Marnik G., and Dominique M. Hanssens. "Persistence Modeling in Marketing: Descriptive, Predictive, and Normative Uses." Journal of Marketing Research, 2024. https://journals.sagepub.com/doi/10.1177/14413582231222311

Gajewar, Amita, and Gagan Bansal. "Revenue Forecasting for Enterprise Products." arXiv preprint arXiv:1701.06624, 2016. https://arxiv.org/abs/1701.06624

"How Predictive Analytics Can Boost Revenue Growth." FasterCapital, 2024. https://fastercapital.com/articles/How-Predictive-Analytics-Can-Boost-Revenue-Growth.html

Jacks, T., et al. "Data-Driven Strategies for Business Expansion: Utilizing Predictive Analytics for Enhanced Profitability and Opportunity Identification." International Journal of Frontiers in Engineering and Technology Research, vol. 6, no. 2, 2024, pp. 71–81. https://www.researchgate.net/publication/381000569

General Use Purposes: OpenAi

Kraus, Mathias, Stefan Feuerriegel, and Asil Oztekin. "Deep Learning in Business Analytics and Operations Research: Models, Applications and Managerial Implications." arXiv preprint arXiv:1806.10897, 2018. https://arxiv.org/abs/1806.10897

Lu, Wei, et al. "Show Me the Money: Dynamic Recommendations for Revenue Maximization." arXiv preprint arXiv:1409.0080, 2014. https://arxiv.org/abs/1409.0080

"Predictive Modeling in Business Analytics: Leveraging AI & Machine Learning." International Journal of Science and Research, vol. 13, no. 4, 2024, pp. 875–880. https://www.researchgate.net/publication/380029952

Würfel, Max, Qiwei Han, and Maximilian Kaiser. "Online Advertising Revenue Forecasting: An Interpretable Deep Learning Approach." arXiv preprint arXiv:2111.08840, 2021. https://arxiv.org/abs/2111.08840