

Interpretable Machine Learning for Profiling and Predicting Marketing Campaign Performance

Raja Venkat Ram V¹, M Raghavender Sharma²

^{1,2}*Department of Statistics, Osmania University, Hyderabad, India*

doi.org/10.64643/IJIRT209699

Abstract—Marketing analysts rarely share campaign data, which makes it difficult to compare analytical methods under controlled conditions. This study uses a synthetic data set of 10,000 marketing campaigns, generated with known channel-by-segment interactions and diminishing returns to budget, to profile campaign performance and to test which modelling choices matter. Rank-based tests showed that channel ($\varepsilon^2 = .763$) and customer segment dominated return on investment (ROI), while budget, duration, and discount added little. Four regression models were compared on a log-transformed return target and on raw ROI. On the log scale, all models performed similarly (test $R^2 = .881$ to $.890$); on raw ROI, a main-effects linear model lagged ($R^2 = .529$) until interaction terms were added ($R^2 = .716$), matching random forest and gradient boosting. Across 20 independent replications, a ridge model with interactions slightly outperformed gradient boosting. Funding only the half of campaigns ranked highest by the model raised portfolio ROI from 1.76 (random selection) to 4.23. The findings suggest that target transformation and interaction terms, rather than model complexity, explain the apparent advantage of ensembles when effects are multiplicative. Because all data are simulated, the results validate a workflow rather than describe real markets.

Index Terms—Marketing analytics, synthetic data, interpretable machine learning, campaign ROI, simulation study, customer segmentation

Marketing teams decide where to spend, whom to target, and which message to run using data about past campaigns (Wedel & Kannan, 2016). Before any prediction is attempted, analysts need a clear statistical profile of that data: how outcomes are distributed, which strategic choices are associated with them, and how much predictive accuracy is realistic. Yet campaign records are seldom available to researchers. They hold customer information, expose competitive strategy, and are restricted by privacy rules, so

published comparisons of analytical methods often rely on a single proprietary data set that others cannot inspect.

Synthetic data offer one way around this barrier (Jordon et al., 2022; Nikolenko, 2021). Learned generators, such as conditional generative adversarial networks, aim to copy the statistical structure of an existing data set (Xu et al., 2019). A researcher can instead write the generating process by hand. This gives up realism but gains a property that real data lack: the true relationships are known, so any analytical method can be judged by whether it recovers them (Morris et al., 2019).

Two questions about modelling remain open for tabular marketing data. First, tree ensembles such as random forests and gradient boosting often outperform linear models on tabular benchmarks (Grinsztajn et al., 2022), yet simple models can remain competitive once the data are suitably prepared (Hand, 2006). It is unclear how much of the ensemble advantage in marketing data comes from the algorithm and how much from the way the outcome is scaled and how interactions are handled. Second, accuracy figures from a single train–test split can mislead; a replicated design shows how stable method rankings are. Interpretability also matters in practice, because managers must justify budget decisions (Doshi-Velez & Kim, 2017).

This study addresses four research questions using a synthetic campaign data set:

RQ1. How is campaign ROI distributed, and which strategic factors are associated with it?

RQ2. How much do the choice of algorithm, the scaling of the outcome, and the inclusion of interaction terms each contribute to predictive accuracy?

RQ3. How stable are the rankings of models across independent replications of the data?

RQ4. Does model-based ranking of campaigns improve portfolio returns compared with funding campaigns at random?

The contribution is a reproducible workflow that joins classical profiling, interpretable and ensemble models, a replication design, and a decision-based evaluation, together with evidence on which modelling choices matter when effects are multiplicative.

I. BACKGROUND

Marketing Analytics and Campaign Return

Marketing analytics connects strategic choices such as channel mix, pricing, and timing to outcomes such as response and profit (Kotler & Keller, 2016). Return on investment, defined here as revenue minus spend divided by spend, is a common summary of campaign success.

Revenue-based measures are typically right-skewed because a few very successful campaigns generate most of the gains, which affects both the choice of statistical test and the scale on which models are fitted (Hastie et al., 2009).

Interpretability and Model Complexity

Interpretable models allow a reader to see how each factor changes the prediction, whereas ensembles are harder to inspect and rely on post hoc tools (Molnar, 2022). Random forests average many decorrelated trees (Breiman, 2001), and gradient boosting builds trees sequentially to correct earlier errors (Friedman, 2001).

Linear and logistic models with regularization remain standard baselines (James et al., 2021).

Post hoc tools such as permutation importance (Altmann et al., 2010) and partial dependence (Friedman, 2001) make it possible to compare what different models learn.

Simulation Studies

A simulation study specifies a data-generating mechanism, applies methods to the generated data, and judges them against the known truth, ideally across many repetitions (Morris et al., 2019). The present design follows that logic: the generator embeds specific effects, and the analysis asks whether and how well they are recovered.

II. METHOD

Data-Generating Process

Data were generated in Python (McKinney, 2010). The main data set contains 10,000 campaigns (random seed 2026). Each campaign is assigned a channel (six levels), customer segment (four), creative format (four), and quarter, plus a budget, duration, discount level, and audience size (Table 1). Impressions equal budget divided by a channel-specific cost per thousand impressions, multiplied by a diminishing-returns factor of $(\text{budget} / 5,000)^{-0.15}$. Clicks and conversions are binomial draws. Conversion rate depends on channel, creative format, discount, quarter, and a channel-by-segment interaction multiplier; for example, influencer campaigns convert better among younger customers and email converts better among older customers. Revenue is conversions multiplied by a segment-specific order value, reduced by the discount, with random noise. ROI is $(\text{revenue} - \text{budget}) / \text{budget}$, a campaign is labelled profitable when ROI exceeds zero, and the modelling target is log return on ad spend, $\ln(\text{revenue} / \text{budget} + 0.01)$.

Table 1 *Variables and Generating Assumptions*

Variable	Type	Generating assumption
Channel	Categorical	Email, social media, search ads, display, influencer, TV
Customer segment	Categorical	Youth, adults, seniors, premium
Creative format	Categorical	Video, image, text, carousel
Quarter	Categorical	Q1–Q4; Q4 most frequent and strongest
Budget (USD)	Numeric	Log-normal; diminishing returns to reach
Duration (days)	Numeric	Uniform, 7–90
Discount (%)	Numeric	0–30; raises conversion, lowers revenue per sale
Audience size	Numeric	Log-normal
Channel × segment	Interaction	Six multipliers between 0.8 and 1.5 on conversion rate
Outcomes	Derived	Impressions, clicks, conversions, revenue, ROI, profitable

Note. Channel and segment parameters are author-chosen assumptions, not estimates from real firms.

Statistical Profiling

Numeric variables were summarized with means, standard deviations, medians, and skewness. Differences in ROI across channels, segments, quarters, and creative formats were tested with the Kruskal–Wallis test (Kruskal & Wallis, 1952), with epsilon-squared ($\epsilon^2 = H / (n - 1)$) as the effect size. Associations among numeric variables used Spearman correlation, interpreted with the conventional benchmarks of .10, .30, and .50 for small, medium, and large effects (Cohen, 1988).

Predictive Models and Evaluation

Four regression models predicted log return on ad spend: ridge regression with main effects, ridge regression with all pairwise interactions among predictors, random forest, and gradient boosting. Three classifiers (logistic regression, random forest, gradient boosting) predicted profitability. Predictors were channel, segment, creative format, quarter, log budget, duration, discount, and log audience size; outcomes known only after a campaign end were excluded to avoid leakage. The data were split 80% for training and 20% for testing, stratified by profitability. Regression models were scored by five-fold cross-validated R^2 on the training set and by R^2 , root mean squared error (RMSE), and mean absolute error (MAE) on the test set. The regression comparison was repeated on raw ROI to isolate the effect of target scaling. Classifiers were scored by accuracy, F1, area under the ROC curve (AUC), and the Brier score (Brier, 1950), which measures the accuracy of predicted probabilities. All models used scikit-learn (Pedregosa et al., 2011).

Replication Design

To test the stability of model rankings, the entire process, from data generation to test-set scoring, was repeated 20 times with independent seeds and 5,000 campaigns per replication. Differences between models across replications were tested with the Wilcoxon signed-rank test (Wilcoxon, 1945).

Interpretation and Decision-Value Evaluation

Variable importance was measured by permutation importance on the test set with 20 repeats (Altmann et

al., 2010), and the effect of budget was examined with a partial dependence curve (Friedman, 2001). For decision value, the model with the best test R^2 ranked the 2,000 test campaigns by predicted return, and the top half was treated as funded. Portfolio ROI of this selection was compared with funding all campaigns, a random half (mean of 500 draws), and an oracle that ranks by true return. This comparison takes spend as fixed and does not model reallocation of the unfunded budget.

Segmentation

K-means clustering (Lloyd, 1982) was applied to standardized log budget, click-through rate (CTR), conversion rate, log revenue, and capped ROI, with 2 to 7 clusters compared by the silhouette coefficient (Rousseeuw, 1987).

III. RESULTS

Profile of Campaign Performance (RQ1)

Table 2 summarizes the data. ROI was strongly right-skewed (skewness = 4.4); the mean of 2.16 was far above the median of 0.20, and 55% of campaigns were profitable. Medians and rank-based tests were therefore used for group comparisons.

Table 2 *Descriptive Statistics (N = 10,000)*

Variable	M	SD	Mdn	Skew
Budget (USD)	6090	5704	4421	3.33
Impressions (000)	387	430	256	4.08
CTR (%)	1.63	1.01	1.47	0.75
Conversion rate (%)	2.61	1.49	2.26	1.20
Revenue (USD)	17281	37149	5471	7.05
ROI	2.16	5.33	0.20	4.43

Note. CTR = click-through rate; ROI = (revenue – budget) / budget.

Channel was strongly related to ROI, $H(5) = 7630.84$, $p < .001$, $\epsilon^2 = .76$ (Table 3). Email ($Mdn = 6.15$) and search ads ($Mdn = 1.86$) were the only channels with clearly positive returns; television ($Mdn = -0.90$) and display ($Mdn = -0.65$) lost most of their spend.

Customer segment was also associated with ROI, $H(3) = 930.67$, $p < .001$, $\epsilon^2 = .09$, with premium customers highest ($Mdn = 1.85$) and youth lowest

($Mdn = -0.14$). Quarter, $H(3) = 69.97$, $p < .001$, $\varepsilon^2 = .01$, and creative format, $H(3) = 100.17$, $p < .001$, $\varepsilon^2 =$

.01, were statistically significant but with small effect sizes, so they matter less than their p values suggest.

Table 3 Campaign Performance by Channel

Channel	n	M ROI	Mdn ROI	CTR (%)	Conv. rate (%)	Profitable (%)
Email	1,733	8.82	6.15	2.17	3.64	100
Search Ads	2,177	3.34	1.86	2.93	4.07	95
Influencer	1,175	0.19	-0.09	1.73	2.69	45
Social Media	2,424	0.17	-0.13	1.19	1.83	42
Display	1,414	-0.51	-0.65	0.54	1.09	11
TV	1,077	-0.86	-0.90	0.43	1.65	0

Note. Channels are sorted by median ROI. Conv. rate = conversion rate.

The channel-by-segment pattern (Figure 1) shows the built-in interactions. Premium customers raised ROI in every channel, but by different amounts: email reached a median ROI of 17.7 for premium customers against 3.4 for youth. Television stayed unprofitable for every segment.

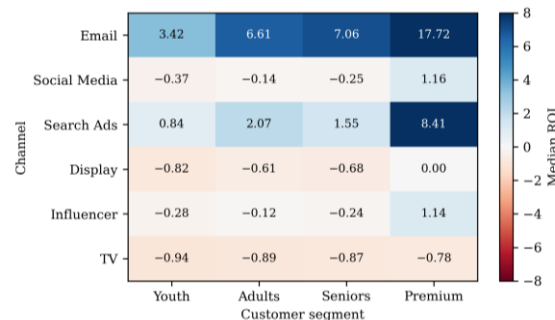


Figure 1 Median ROI by Channel and Customer Segment

Note. Cell values are median ROI of simulated campaigns ($N = 10,000$).

Among numeric variables, ROI was strongly associated with CTR ($\rho = .80$) and conversion rate ($\rho = .73$). Discount raised conversion rate ($\rho = .22$, a

small-to-medium effect) but was essentially unrelated to ROI ($\rho = .02$), because deeper discounts also reduce revenue per sale.

Budget showed a weak negative association with ROI ($\rho = -.08$). Median ROI by budget quintile fell from 0.44 in the lowest quintile to 0.03 in the highest, consistent with the diminishing returns built into the generator.

Predictive Accuracy and the Role of Scaling (RQ2)

Table 4 compares the four regression models on both outcome scales. On the log return scale, all models performed almost identically (test R^2 from .881 to .890), and the highest test R^2 belonged to ridge regression with interactions (.890), not to an ensemble. On raw ROI the picture changed. Ridge regression with main effects explained only .529 of the variances, whereas adding interaction terms raised this to .716, which was in line with random forest (.714) and gradient boosting (.720). The pattern suggests that the multiplicative structure of returns, rather than any special strength of tree ensembles, produces the gap on the raw scale. A log transformation turns multiplicative effects into additive ones, so a main-effects linear model suffices, and when the outcome is left unscaled, interaction terms recover the same structure.

Table 4 Test-Set Performance of Regression Models on Two Outcome Scales

Model	CV R^2 (log)	Test R^2 (log)	RMSE (log)	Test R^2 (raw)	RMSE (raw)
Ridge (main effects)	.882 $\pm$.005	.885	0.50	.529	3.83
Ridge (with interactions)	.885 $\pm$.005	.890	0.48	.716	2.98
Random forest	.879 $\pm$.006	.881	0.50	.714	2.99
Gradient boosting	.883 $\pm$.005	.885	0.50	.720	2.96

Note. Log = log return on ad spends, ln (revenue / budget + 0.01); raw = ROI. CV = five-fold cross-validation on the training set ($M \pm SD$). Test metrics use the 20% hold-out set.

Stability Across Replications (RQ3)

Across 20 independent replications (Table 5), ridge regression with interactions had the highest mean test R^2 ($M = .885$), followed by ridge with main effects ($M = .884$), random forest ($M = .878$), and gradient boosting ($M = .875$). Gradient boosting scored significantly below both the main-effects ridge model, $W = 1, p < .001$, and random forest, $W = 28, p = .003$. The gain from adding interactions to ridge was small and not significant at the .05 level, $W = 54, p = .058$. The absolute differences were all under .01 in R^2 , so the ranking is consistent but practically negligible. A single split, as in Table 4, would have been enough to reach the same substantive conclusion here, but it could not have shown that.

Table 5 Test R^2 on the Log Return Scale Across 20 Replications

Model	M	SD	Min	Max
Ridge (main effects)	.884	.008	.859	.898
Ridge (with interactions)	.885	.007	.862	.895
Random forest	.878	.007	.860	.890
Gradient boosting	.875	.008	.853	.886

Note. Each replication generated 5,000 new campaigns with an independent seed.

Classification of Profitable Campaigns

The three classifiers were nearly indistinguishable (Table 6): accuracy was about .885, AUC about .962, and Brier scores between .078 and .080. Logistic regression had the lowest Brier score, so its predicted probabilities were at least as well calibrated as those of the ensembles.

Table 6 Classification of Profitable Campaigns on the Hold-Out Set

Model	Accuracy	F1	AUC	Brier score
Logistic regression	.89	.89	.96	.08
Random forest	.88	.89	.96	.08
Gradient boosting	.89	.89	.96	.08

Note. Lower Brier scores indicate better-calibrated probabilities. The profitable class is the positive class.

Interpretation

Permutation importance for the best-scoring model (Figure 2) ranked channel first (R^2 drop = 1.48) and customer segment second (0.23). Quarter (.02), creative format (.01), and budget (.01) each contributed little, and duration, discount, and audience size contributed almost nothing.

The partial dependence curve for budget (Figure 3) declines steadily as budget grows, the signature of diminishing returns: other things equal, larger campaigns return less per dollar.

This agrees with the generator, which supports the use of these tools for recovering known effects. Even so, the small importance of budget on the whole does not mean budget is unimportant for any single campaign; the effect is gradual and swamped by channel differences.

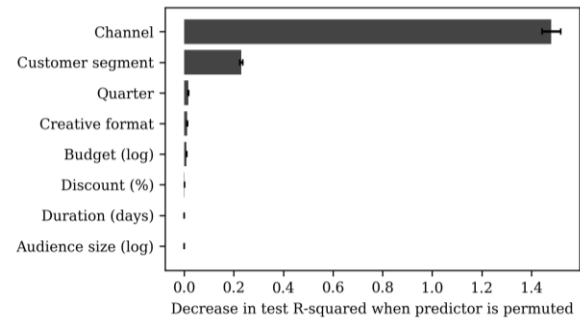


Figure 2 Permutation Importance of Predictors

Note. Bars show mean decrease in test R^2 over 20 permutations; whiskers show standard deviations. Model: ridge (with interactions).

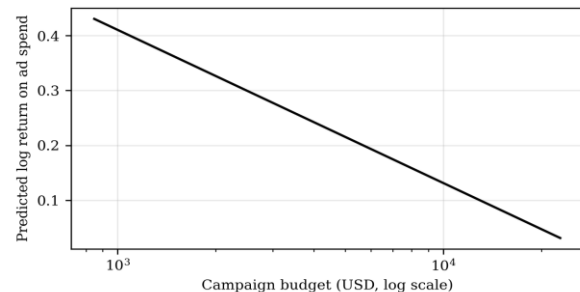


Figure 3 Partial Dependence of Predicted Return on Budget

Note. Average predicted log return on ad spend over the test set as budget varies from the 2nd to the 98th percentile.

Decision Value (RQ4)

Ranking test campaigns with the model and funding the top half produced a portfolio ROI of 4.23, compared with 1.76 for random selection and for funding all campaigns (Table 7). The model-based selection recovered 98% of the net profit of the oracle selection (USD 24.1 million against 24.5 million). Net profit also exceeded that of funding every campaign, because the excluded half contained unprofitable campaigns. These comparisons hold total spend per campaign fixed and ignore any reallocation of the saved budget, so they indicate the value of ranking rather than an achievable total return.

Table 7 *Portfolio Outcomes of Campaign Selection Strategies (Test Set)*

Strategy	Campaigns funded	Net profit (USD million)	Portfolio ROI
Fund all campaigns	2,000	21.3	1.76
Fund random half	1,000	10.7	1.76
Fund top half by model	1,000	24.1	4.23
Fund top half (oracle)	1,000	24.5	4.35

Note. Portfolio ROI = (total revenue – total budget) / total budget for the funded set. Random-half values are means of 500 draws.

Campaign Segments

The silhouette coefficient was highest for two clusters (.39) and lower for every larger number, so two groups were retained.

The high-performing cluster (3,832 campaigns, 38%) had a mean CTR of 2.60%, a mean conversion rate of 3.98%, a median ROI of 3.73, and 99% profitable campaigns, most often in Search Ads. The low-performing cluster (6,168 campaigns) had a mean CTR of 1.02%, a mean conversion rate of 1.76%, a median ROI of –0.40, and 28% profitable campaigns, most often in Social Media. A silhouette value near .39 indicates moderate separation, so the clusters are best read as a broad split between engaging and non-engaging campaigns.

IV. DISCUSSION

This study profiled a synthetic marketing data set and examined which modelling choices drive predictive accuracy. Three findings stand out. First, channel and customer segment, and the interaction between them, accounted for nearly all of the explainable variation in ROI, whereas budget, duration, and discount mattered little. Second, apparent differences between algorithms depended on how the outcome was scaled. With a log return target, a ridge regression matched or beat random forest and gradient boosting in both the main split and the replications. With raw ROI, a main-effects linear model fell behind, and adding interaction terms closed the gap. Third, ranking campaigns by predicted return captured almost all of the profit available to a perfect ranking.

These results qualify the common expectation that tree ensembles are the default choice for tabular data (Grinsztajn et al., 2022). In this setting, simple and interpretable models were competitive, which echoes the argument that complex classifiers often add less than expected (Hand, 2006). For practitioners, the implication is a sequence of checks: start with a transformation suited to skewed return data, include interactions among the strongest drivers, and use ensembles as a benchmark rather than as the first choice. Interpretable models also let managers read off the contribution of each factor directly, which supports accountability for budget decisions (Doshi-Velez & Kim, 2017).

V. LIMITATIONS

The central limitation is that all effects were written by the author. The generator is multiplicative, so a log transformation makes it additive by construction, which favours the log-linear model. Real campaigns may include thresholds, saturation, seasonality, and delayed conversions that favour more flexible learners; the conclusion about ensembles should not be assumed to transfer. The strong effects of channel and segment, the near-zero effect of duration, and the discount trade-off all reflect assumed parameters. The decision-value analysis treats campaigns as independent and ignores reallocation of budget and interactions between concurrent campaigns. Finally, the data contain no missing values, measurement error, or causal confounding.

VI. FUTURE RESEARCH

Future work should repeat the design with data from learned generators and with anonymized real campaign records, include generating mechanisms that are not multiplicative so that the conditions favouring ensembles can be identified, and extend the decision analysis to budget reallocation with uncertainty intervals.

VII. CONCLUSION

A replicated simulation on 10,000 synthetic campaigns showed that channel and customer segment drive marketing returns, that model complexity contributes little once the outcome is scaled suitably or interactions are included, and that model-guided selection of campaigns substantially raises portfolio ROI. The workflow is transparent and reproducible, and it can be rerun on real data as it becomes available. Because the evidence comes from simulation, it supports methodological conclusions and not claims about specific markets.

DECLARATIONS

Conflict of interest. The author declares no conflict of interest. [Edit as needed.]

Funding. [State funding source or "This research received no external funding."]

Data availability. The synthetic data set and the Python code used to generate it and to reproduce all tables and figures are available from the author on request (random seeds are reported in the text).

REFERENCES

- [1] Altmann, L. Toloşi, O. Sander, and T. Lengauer, "Permutation importance: A corrected feature importance measure," *Bioinformatics*, vol. 26, no. 10, pp. 1340–1347, 2010, doi: 10.1093/bioinformatics/btq134.
- [2] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [3] G. W. Brier, "Verification of forecasts expressed in terms of probability," *Monthly Weather Review*, vol. 78, no. 1, pp. 1–3, 1950.
- [4] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Lawrence Erlbaum Associates, 1988.
- [5] F. Doshi-Velez and B. Kim, *Towards a Rigorous Science of Interpretable Machine Learning*, arXiv, 2017, doi: 10.48550/arXiv.1702.08608.
- [6] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001, doi: 10.1214/aos/1013203451.
- [7] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on typical tabular data?" in *Advances in Neural Information Processing Systems*, vol. 35. Curran Associates, 2022.
- [8] D. J. Hand, "Classifier technology and the illusion of progress," *Statistical Science*, vol. 21, no. 1, pp. 1–14, 2006, doi: 10.1214/088342306000000060.
- [9] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, 2009.
- [10] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R*, 2nd ed. Springer, 2021.
- [11] J. Jordon, L. Szpruch, F. Houssiau, M. Bottarelli, G. Cherubin, C. Maple, S. N. Cohen, and A. Weller, *Synthetic Data—What, Why and How?* arXiv, 2022, doi: 10.48550/arXiv.2205.03257.
- [12] P. Kotler and K. L. Keller, *Marketing Management*, 15th ed. Pearson, 2016.
- [13] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 583–621, 1952, doi: 10.1080/01621459.1952.10483441.
- [14] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982, doi: 10.1109/TIT.1982.1056489.
- [15] W. McKinney, "Data structures for statistical computing in Python," in *Proc. 9th Python in Science Conf.*, S. van der Walt and J. Millman, Eds., 2010, pp. 56–61, doi: 10.25080/Majora-92bf1922-00a.
- [16] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, 2nd ed., 2022. [Online]. Available: <https://christophm.github.io/interpretable-ml-book>

- [17] T. P. Morris, I. R. White, and M. J. Crowther, "Using simulation studies to evaluate statistical methods," *Statistics in Medicine*, vol. 38, no. 11, pp. 2074–2102, 2019, doi: 10.1002/sim.8086.
- [18] S. I. Nikolenko, *Synthetic Data for Deep Learning*. Springer, 2021.
- [19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [20] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.
- [21] M. Wedel and P. K. Kannan, "Marketing analytics for data-rich environments," *Journal of Marketing*, vol. 80, no. 6, pp. 97–121, 2016, doi: 10.1509/jm.15.0413.
- [22] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945, doi: 10.2307/3001968.
- [23] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni, "Modeling tabular data using conditional GAN," in *Advances in Neural Information Processing Systems*, vol. 32. Curran Associates, 2019.