Research and Modelling of the Google AdSense Business Tool in the Context of Selected Website Metrics Using a Machine Learning Approach

Silvia Komara[1] – Marián Čvirik[2]– Martin Kuchta[3]
ORCID iD: 0000-0001-6641-7456[1], 0000-0003-4701-1543[2], 0000-0001-5546-7773[3]
silvia.komara@euba.sk, marian.cvirik@euba.sk, martin.kuchta@euba.sk
[1] Bratislava University of Economics and Business, Faculty of Economic Informatics, Department of Statistics
Bratislava, Slovakia
[2] Bratislava University of Economics and Business, Faculty of Commerce, Research Institute of Trade and Sustainable Business
Bratislava, Slovakia
[3] Bratislava University of Economics and Business, Faculty of Commerce, Department of Marketing
Bratislava, Slovakia

https://doi.org/10.18267/pr.2026.vol.2587.16

Abstract: This paper contributes to the field of digital business marketing by focusing on the factors that lead websites to adopt Google AdSense. The aim is to examine the use of Google AdSense in the context of selected website metrics using a machine learning approach. We focus in particular on how traffic characteristics and user behaviour are reflected in monetisation strategies. We incorporate the SHapley Additive exPlanations method into our analysis, which allows us to go beyond the usual measures of feature importance. We gain a more detailed view of how individual web traffic indicators complement each other and jointly contribute to monetisation results. We can see not only which variables are important to the model but also how they contribute to predicting AdSense adoption in specific cases. This gives us a better understanding of the interactions behind the decision to deploy an advertising platform and allows us to link them to practical website monetisation issues.

Keywords: Google AdSense, business intelligence, machine learning, customer journey, websites metrics

JEL Classification codes: M15, M31, C10

INTRODUCTION

Modern marketing management is built on data, and digital marketing created with the help of data analysis is on the rise (Bajrami et al., 2025; Sak et al., 2025). The use of artificial intelligence, digitalisation and advanced analytics is transforming the marketing paradigm, changing the strategic aspect to a more proactive one with an emphasis on personalisation and customer acceptance (Kita et al., 2024; Kumar et al., 2026).

In the current digital era, effective monetisation of web content represents a key challenge for publishers, marketers and entrepreneurs. Google AdSense is a significant platform in this context with an orientation towards automated advertising mediation. Although this platform has achieved a high level of popularity in terms of both digital marketing and data management, the implementation of this tool is complex and requires deeper knowledge.

Classic linear modelling is losing its relevance in the context of the complexity of the real world, which creates an opportunity for the use of machine learning.

The present conference paper focuses on the research and modelling of the Google AdSense business tool in the context of selected website metrics using a machine learning approach. The aim of the present article is to examine the use of Google AdSense in the context of selected website metrics using a machine learning approach. The article is supported by data from 2000 websites (a systematic sample from 10,000 best websites), for which we obtained detailed traffic statistics from SimilarWeb. The main contribution of the paper can be identified at several levels, linking several areas with an emphasis on digital marketing and data science with an emphasis on machine learning but also helping to better understand consumer behaviour.

From an academic research perspective, the study of AdSense is often part of broader research questions like digital advertising networks, decision-making in the adoption of online advertising, or user behaviour analysis. According to Graham (2017), AdSense functions as the core mechanism of website monetisation, essentially determining which parts of the internet can be profitable. For small and medium-sized websites, publishers, and so-called platform-dependent entrepreneurs (PDEs), AdSense is a fundamental source of income and often the main factor influencing their economic sustainability (Wurfel et al., 2021).

1 LITERATURE REVIEW

In the current era of digitization and the internet, monetizing web traffic has become a key issue for website, blog, and online media owners. Launched in June 2003, Google AdSense is still the most widely used contextual display advertising program in the world. It is a free advertising program from Google that allows website owners to display contextually relevant ads and earn financial rewards based on the number of views or clicks on ads placed directly on their pages (Holland, 2016). The program works by allowing the Google AdSense script to dynamically load ads directly onto the publisher's pages, with Google managing the selection of advertisers and ad content based on the context and traffic of the pages. The model has dramatically lowered the barriers to entry into online advertising for small and medium-sized publishers, who do not have to build their own sales network or negotiate directly with advertisers.

There are several reasons for examining Google AdSense. First of all, it is worth emphasising that AdSense represents a significant economic resource, which can bring knowledge about the sustainability of content. At the same time, it can also be a source of innovation in the context of other indicators from the point of view of user interaction and the websites themselves. From the point of view of digital marketing, it represents an important metric that can be important in generating budgets. Understanding the multi-channel dynamics of traffic sources can contribute to a higher probability of purchase. In this context, a closer examination of the attribution model and customer journey is necessary, with an emphasis on various channels such as Direct, Organic, Paid, Referral, Display, Social, and Email, which significantly participate in the final income generated by advertisements (Filippou et al., 2024). Direct can be understood as a direct path, i.e., direct traffic, when the visitor directly enters the URL of the given page, which can signal a high level of loyalty and is therefore also an indicator of value (Kakalejčík et al., 2020; O'Connor, 2021). Organic search represents visits coming from unpaid links. The opposite is paid search, which is traffic based on paid links (Ghose & Yang, 2008). Referral traffic is generated through links from external websites, which are usually thematically related, and most often we talk about relevant websites (Stroobant, 2019). Display ads refer to traffic generated by links from visual advertisements (Bayer et al., 2020). Social represents traffic through platforms focused on interactions and sharing. Email

represents links most often from newsletters or other personalised communication sent directly to users (Păvăloaia et al., 2020). It is also important to understand the metrics associated with website evaluation, with traffic being the key (Järvinen & Karjaluoto, 2015). Traffic share is a basic indicator of audience distribution and marketing channel effectiveness (Filippou et al., 2024). Pages per visit expresses the average number of pages viewed during one visit (Plaza, 2011). Visit duration measures the total time spent by users on the site during one visit. Monthly visits capture the total volume of visits per month (exceptionally, other defined periods), which serves as an indicator of trend (Xun, 2015). Yearly change provides essentially the difference between the past and current annual traffic, thus indicating the development trend of the site (Hong, 2007). Desktop Share and Mobile Share express the shares of traffic coming from desktop and mobile devices (Goldstein & Hajaj, 2022). Bounce rate represents the percentage of visits in which a user leaves the site after viewing a single page (Jansen et al., 2022).

2 METHODOLOGY

The use of machine learning (ML) in digital marketing analytics has grown mainly because these methods can reveal patterns in web-traffic data that are too irregular or complex for many traditional statistical techniques. Although advanced ML models often perform well, they are frequently criticized for their limited transparency and the fact that they function as "black boxes" (Ćwiek et al., 2024; Molnar, 2020; Ribeiro et al., 2016; Rudin, 2019). It is therefore difficult to understand why the model arrives at a certain result, which may hinder practical decision-making. To address this limitation, we incorporate the SHAP (SHapley Additive exPlanations) method (Komara et al., 2026). SHAP builds on ideas from cooperative game theory and breaks down a model's prediction into contributions attributable to individual features. Each feature is assigned a value reflecting its average marginal impact across different combinations of predictors. In this way, SHAP helps to clarify the internal mechanics of otherwise opaque models while keeping their predictive power (Li et al., 2024; Lundberg & Lee, 2017; Molnar, 2020).

Our analysis draws on a dataset of 2000 websites (systematic sample from 10000 best website), for which we obtained detailed traffic statistics from SimilarWeb. Systematic selection was aimed at selecting every fifth member, ensuring that it included a diverse range of websites. The variables include standard engagement and volume metrics, such as Traffic Share, Pages per Visit, Visit Duration, Monthly Visits, Yearly Change, Desktop Share, Mobile Share, and Bounce Rate along with information on the main traffic acquisition channels like Direct Search, Organic Search, Paid Search, Referrals, Display Ads, Social, and Email. The outcome of interest is whether a website uses Google AdSense. This variable is binary (1 = adopted, 0 = not adopted), so the task is formulated as a classification problem. To make the features comparable and avoid distortions caused by differences in scale, all variables were standardized using z-score normalization.

To balance interpretability with predictive performance, we selected four machine-learning algorithms. As a straightforward baseline, we included the K-Nearest Neighbours (KNN) classifier due to its simplicity and intuitive decision rule. KNN is a non-parametric approach that assigns a class label to a new observation based on a majority vote among its k nearest neighbours in the feature space. Formally, the predicted class for an instance x is given by:

$$\hat{y}(x) = arg\max_{c \in C} \sum_{i \in N_k(x)} \mathbb{I}(y_i = c) \quad (1)$$

where $N_k(x)$ denotes the set of indices of the k nearest observations, i.e., k nearest neighbours of x, typically computed using Euclidean distance, and $\mathbb{I}(\cdot)$ is the indicator function that equals 1 when the condition holds and 0 otherwise (Cover & Hart, 1967).

Logistic regression provides an additional level of interpretability (Šoltés et al., 2019). It models the log-odds of adoption as a linear combination of the input features (Agresti, 2013):

$$\log\left(\frac{P(y=1|\mathbf{X})}{1-P(y=1|\mathbf{X})}\right) = \alpha + \boldsymbol{\beta}x \quad (2)$$

where $P(y=1 \mid \mathbf{X})$ is the probability that the outcome variable y takes value 1 (i.e., adoption), $\mathbf{X} = (x_1, x_2, \ldots, x_i)$ is the feature vector, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_i)$ are the model coefficients and α is the intercept. The estimated coefficients β_i make it possible to see how each feature influences the likelihood of adoption. However, because logistic regression assumes a strictly linear relationship between predictors and the log-odds, it may fail to capture more complex nonlinear or interaction effects that can arise in real-world data.

For predicting AdSense adoption, we also relied on Random Forests. This method is particularly useful in settings where relationships between variables are complex or nonlinear, because it can capture such patterns through the hierarchical splits created within individual decision trees. Importantly, these interactions do not need to be specified beforehand. A Random Forest combines the output of many tree-based classifiers into a single prediction. Formally, it consists of a set of classifiers $\{h(x, \Theta_k), k = 1, \ldots.\}$, where $\{\Theta_k\}$ are independent identically distributed random vectors. Every tree contributes one vote for the class it predicts to be the most likely (Breiman, 2001), and the aggregated vote tends to produce a more stable and accurate classification than any single tree alone.

In addition to Random Forests, we employed XGBoost, a scalable gradient-boosting algorithm that constructs trees in a sequential manner. Unlike ensemble methods that build trees independently, XGBoost adds each new tree to correct the errors made by the previous ones, producing a strong predictive model through iterative refinements. Its objective function includes both a loss term and a regularization component:

$$L_{xgb} = \sum_{i=1}^{N} L\big(y_i, F(x_i)\big) + \sum_{m=1}^{M} \Omega\,(h_m) \quad (3)$$

where $L\big(y_i, F(x_i)\big)$ is a differentiable convex loss function, $\Omega(h) = \gamma T + \frac{1}{2}\lambda\|w\|^2$ is the regularization term, T is number of leaves in the tree, γ controls the minimal loss reduction required to split internal nodes, λ controls leaf weight regularization and w are leaf scores (Bentéjac et al., 2021; Friedman, 2001).

We assessed the performance of the models using several commonly applied classification metrics: accuracy, precision, recall, and the F1 score (Powers, 2011). Accuracy reflects the share of correctly classified observations and is defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+F} \quad (4)$$

where TP (true positives) refers to correctly identified AdSense adopters, and TN (true negatives) to correctly identified non-adopters. FP (false positives) and FN (false negatives) capture cases in which the model incorrectly predicts adoption or non-adoption.

Precision indicates how many of the websites that the model classified as adopters were in fact adopters:

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

Recall (or sensitivity) expresses the proportion of actual adopters that the model successfully detected:

$$Recall = \frac{TP}{TP+FN} \tag{6}$$

The F1 score is the harmonic mean of precision and recall and provides a single metric that balances both:

$$F1 = 2\frac{Precision \cdot Recall}{Precision + Recall} \tag{7}$$

To improve the performance of the models, we used Grid Search (implemented through GridSearchCV in Python), which systematically tests combinations of hyperparameters and finds the optimal setting. The procedure evaluates every specified setting—for example, the choice of k in KNN or the maximum tree depth in ensemble models—and selects the configuration that performs best according to the chosen evaluation metric, which in our case was the F1-score.

For interpreting individual predictions, we relied on SHAP values, a unified framework for feature attribution introduced by Lundberg and Lee (2017). SHAP builds on the concept of Shapley values from cooperative game theory by Lloyd Shapley in 1953 (Peleg & Sudhölter, 2007), providing a principled way to distribute a model's output across the features that contributed to it. In this approach, the difference between the prediction for a specific instance and a baseline value is allocated among all input features. The final prediction $f(x)$ is the sum of the average prediction (ϕ_{base}) and the SHAP values (ϕ_i) for all features:

$$f(x) = \phi_{base} + \sum_{i=1}^{|F|} \phi_i \tag{8}$$

The SHAP value for a feature is computed as the weighted average of its marginal contributions across all possible feature coalitions S:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} \cdot \left[f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - f_S(x_S) \right] \tag{9}$$

where F is the set of all input features, S is a subset of features that does not include feature i. The term $f_{S\cup\{i\}}(x_{S\cup\{i\}}) - f_S(x_S)$ is the marginal contribution of feature i, i.e., how much the model's prediction changes when feature i is added to the subset S.

3 RESULTS AND DISCUSSION

The main objective of the analysis was to examine which factors influence the adoption of Google AdSense and how well different machine-learning models can predict this outcome. The dataset consisting of 2,000 websites was divided into training and testing parts using an 80:20 stratified split to preserve the proportion of adopters and non-adopters. Model parameters were tuned through a five-fold cross-validated grid search carried out on the training set, where we focused mainly on the F1 score because of the class imbalance. The final evaluation was then performed on the unseen test data using accuracy, precision, recall and the F1 score, which together offer a balanced view of model errors and their trade-offs.

During the data preparation, we noticed that the dataset included one extreme outlier — google.com. Its traffic share reached 18.64 %, while the next largest website in the sample had less than 1 %. Such an extreme value can distort both the learning process and performance metrics, so we estimated two sets of models: one with google.com present and one without it. When google.com was kept in the dataset, Random Forest performed best, achieving an accuracy of 0.82 and an F1 score of 0.70, with XGBoost showing comparable results. K-Nearest Neighbours and logistic regression reached lower values (0.73 and 0.71 accuracy, respectively). After removing google.com, all models improved slightly: KNN increased to 0.75 accuracy, logistic regression to 0.72, Random Forest to 0.84 and XGBoost to 0.82. Tab. 1 summarizes the complete set of results across both scenarios.

Tab. 1. Performance Metrics of Machine Learning Models with (W) and without (Wo) *google.com* in the Dataset

Model	Acc (W/WoG)	F1 (W/WoG)	Rec (W/Wo)	Prec (W/WoG)
KNN	0.72/0.75	0.55/0.61	0.52/0.60	0.59/0.63
Logistic Regression	0.70/0.72	0.29/0.32	0.18/0.21	0.69/0.73
Random Forest	0.82/0.84	0.70/0.71	0.63/0.63	0.78/0.83
XGBoost	0.81/0.82	0.70/0.69	0.67/0.63	0.73/0.77

Source: Own processing.

When identifying the best-performing model—in this case, the Random Forest—it was important not only to focus on predictive accuracy but also to understand how individual features shaped the model's decisions. For this reason, we used SHAP, which provides a detailed breakdown of how each feature contributes to each prediction. The SHAP summary plot (Fig. 1) gives an overall view of feature importance and the direction of their effects. The features are ranked according to their influence, with *Pages/Visit* and *Organic_Search_fin* emerging as the strongest predictors of AdSense adoption. High values of Pages/Visit consistently lowered the predicted probability of adoption, which is visible in the plot as red points concentrated on the left side (negative SHAP values). In contrast, higher values of *Organic_Search_fin* increased the likelihood of adoption, reflected by red points grouped on the right side of the axis. Some other variables—such as Social and Mobile Share—show more context-specific behaviour. Their points are more widely spread, indicating that their effect can shift depending on the combination of other features present for a given website. Features that appear lower in the ranking, such as *Monthly Visits*, have SHAP values close to zero and

therefore play a much smaller role in the model's predictions. In several cases, such as *Bounce Rate* and *Mobile Share*, the red points occur on both sides of the SHAP axis. This means that high values of these variables do not always push the prediction in the same direction. Instead, their influence varies depending on the broader feature environment, which suggests that the model is capturing nonlinear relationships or interactions among variables.

Fig. 1 SHAP Summary Dot Plot Showing Feature Importance and Impact on Google AdSense Adoption Predictions

Source: Own processing.

The SHAP dot plot highlighted Pages/Visit as the feature with the strongest influence on AdSense adoption. To better understand this relationship, we also examined the point-biserial correlation between the two variables. Although the correlation was relatively weak (r_pb = –0.19) it was highly significant with *p-value < 0.0001*, supporting the general pattern that websites with more pages viewed per visit tend to be less likely to use AdSense. Because the SHAP results suggested that the effect might not be purely linear, we explored whether a threshold or interaction pattern could explain the relationship more clearly. We divided the sample into four groups based on the medians of the two most influential predictors—*Pages/Visit* and *Organic_Search_fin*. This created four combinations: low versus high *Pages/Visit*, and low versus high organic search traffic. The differences between groups were substantial. Websites with low *Pages/Visit* and high organic search traffic showed the highest adoption rate (59%). At the opposite end, sites with high *Pages/Visit* and low organic traffic had the lowest adoption rate (16%). The remaining two groups fell somewhere in between, with adoption rates of 31% and 24%.

These patterns show that the negative effect of *Pages/Visit* becomes much more pronounced once websites pass the median threshold, and that its impact depends heavily on how much

organic search traffic the site receives. This interaction—captured by SHAP and confirmed by the grouped adoption rates—explains why *Pages/Visit* emerges as the top feature in the model despite only a modest linear correlation with the target variable. Similar interactions may exist for other variables as well, and different threshold definitions could be used depending on the focus of the analysis.

To illustrate how the model behaves for a specific website, we used a SHAP force plot, which breaks down a single prediction into the contributions of individual features. As an example, we examined the website *weather.com*, a site that is known to use AdSense. The force plot shows how the prediction for this instance is built up, starting from the model's base value—the average predicted probability across all training observations. For *weather.com*, the base value is 0.3317 (see Fig. 2).

Fig. 2 SHAP Force Plot Visualizing Feature Contributions to the Predicted Probability of Google AdSense Adoption for the Website weather.com.

Source: Own processing.

In the force plot (Fig. 2), a series of red bars appears to the right of this baseline. These red segments represent features such as *Mobile Share*, *Bounce Rate*, *Social*, *Pages/Visit*, and *Organic_Search_fin*, all of which push the prediction upward. The red colouring indicates a positive contribution toward a higher probability of AdSense adoption. Notably, there are no blue bars in this case, meaning that none of the features substantially reduce the predicted probability. When the individual contributions are summed, the model arrives at a final predicted probability of 0.89 for *weather.com*, which aligns well with the fact that the website actually uses AdSense. Overall, the force plot shows that the high predicted probability is the result of several features consistently pushing the prediction in the same direction. This type of local interpretability is valuable because it highlights not only which features matter on average but also how they combine for particular, real-world cases.

Implementing advanced machine learning models allows marketing managers to predict the needs for generating market entities' strategies and optimise data-driven decision-making processes (Sun, 2025). From a digital marketing perspective, the analysis results prove that maximising the share of organic search is a key determinant for successful integration of advertising networks. The key is setting up the business model (Kita et al., 2024), while the identified negative correlation between the number of pages viewed per visit and advertising adoption indicates that platforms with a high level of deep user interaction require different business models.

3.1 Limits and future research

One limitation of our analysis is that we did not classify the websites according to their main purpose or content type. Although the dataset contains the URL for each site, we did not assign them to categories such as news, e-commerce, blogs, or other functional groups. Because of this, our results reflect broader average patterns drawn from a mixed and diverse sample, rather than capturing the distinct behaviours that specific categories might exhibit. This heterogeneity may blur or reshape the underlying relationships between variables. For instance, the factors that drive AdSense adoption on news platforms may differ quite

substantially from those influencing adoption on e-commerce websites. Future work could address this issue by categorizing websites based on their domain characteristics or content, allowing for a more refined analysis of the determinants of AdSense adoption within particular website segments.

CONCLUSION

The aim of this paper was to identify factors associated with the adoption of Google AdSense at the website level and to show how predictive performance can be balanced with model interpretability. We compared four classification approaches (KNN, Logistic Regression, Random Forest, and XGBoost) and evaluated their performance with an emphasis on F1 score metric due to class imbalance. The results consistently showed that ensemble methods provided the best performance. Removing the outlier google.com led to a slight improvement in performance across all models. From a methodological point of view, this confirms the importance of working with outliers in web traffic data, where the dominance of large platforms can distort both model training and evaluation.

An important contribution of this paper is the use of SHAP method to interpret the decisions of the so-called machine learning black-box models. SHAP analysis showed that the strongest determinants of AdSense adoption are Pages/Visit and the share of organic traffic (Organic Search). A higher number of pages per visit was associated with a lower probability of adoption, while higher organic traffic increased it. At the same time, it was found that these relationships are not strictly linear: the effect of Pages/Visit is accentuated after exceeding a threshold level and depends significantly on the level of organic traffic, which was evident in the SHAP decomposition as well as in the group adoption rates. Furthermore, a local interpretation using the example of the website weather.com demonstrated the practical value of SHAP in explaining a specific model decision (why the model evaluated the website as a high-probability adopter). From a digital marketing perspective, the results suggest that traffic volume metrics alone may not be key. User behaviour indicators and acquisition channel structure appear to be more important, with their effect potentially influenced by a combination of factors. This is relevant for marketers and website owners who want to better understand the profiles of sites typically associated with monetization through AdSense, as well as for analysts who need models that are not only accurate but also explainable.

ACKNOWLEDGEMENT

Paper is output from the project A-25-102/3020-15 "Enhancement of retail business models based on sensory, digital, and geographic marketing management through modern approaches in market research and statistical data analysis" and VEGA 1/0002/27 "Progressive socially-oriented food retail models generated based on data science and their role in reflecting consumer trends and values for at-risk customer groups with an emphasis on well-being".

REFERENCES

Agresti, A. (2013). *Categorical data analysis* (3rd edition). Wiley-Interscience, a John Wiley & sons, inc., publication.

Bajrami, R., Gashi, A., Bajraktari, K., & Namligjiu, E. (2025). The impact of big data analytics on digital marketing decision-making: A comprehensive analysis. *Innovative Marketing, 21*(3), 171–182. https://doi.org/10.21511/im.21(3).2025.13

Bayer, E., Srinivasan, S., Riedl, E. J., & Skiera, B. (2020). The impact of online display advertising and paid search advertising relative to offline advertising on firm performance and firm value. *International Journal of Research in Marketing, 37*(4), 789–804. https://doi.org/10.1016/j.ijresmar.2020.02.002

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review, 54(*3), 1937–1967. https://doi.org/10.1007/s10462-020-09896-5

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory, 13*(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964

Ćwiek, M., Ulman, P., & Sadko, M. (2024). Evaluation of housing conditions in Europe using the TOPSIS method. *Ekonomista*, 1–24. https://doi.org/10.52335/ekon/189393

Filippou, G., Georgiadis, A. G., & Jha, A. K. (2024). Establishing the link: Does web traffic from various marketing channels influence direct traffic source purchases? *Marketing Letters, 35*(1), 59–71. https://doi.org/10.1007/s11002-023-09700-8

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics, 29*(5), 1189–1232.

Ghose, A., & Yang, S. (2008). Comparing performance metrics in organic search with sponsored search advertising. *Proceedings of the 2nd International Workshop on Data Mining and Audience Intelligence for Advertisin*g, 18–26. https://doi.org/10.1145/1517472.1517475

Goldstein, A., & Hajaj, C. (2022). The hidden conversion funnel of mobile vs. Desktop consumers. Electronic *Commerce Research and Applications, 53*, 101135. https://doi.org/10.1016/j.elerap.2022.101135

Graham, R. (2017). Google and advertising: Digital capitalism in the context of Post-Fordism, the reification of language, and the rise of fake news. *Palgrave Communications, 3*(1), 45. https://doi.org/10.1057/s41599-017-0021-4

Holland, J. (Ed.). (2016). *Wearable Technology and Mobile Innovations for Next-Generation Education: IGI Global*. https://doi.org/10.4018/978-1-5225-0069-8

Hong, I. B. (2007). A survey of web site success metrics used by Internet-dependent organizations in Korea. *Internet Research, 17*(3), 272–290. https://doi.org/10.1108/10662240710758920

Jansen, B. J., Jung, S., & Salminen, J. (2022). Measuring user interactions with websites: A comparison of two industry standard analytics approaches using data of 86 websites. *PLOS ONE, 17*(5), e0268212. https://doi.org/10.1371/journal.pone.0268212

Järvinen, J., & Karjaluoto, H. (2015). The use of Web analytics for digital marketing performance measurement. *Industrial Marketing Management, 50*, 117–127. https://doi.org/10.1016/j.indmarman.2015.04.009

Kakalejčík, L., Bucko, J., & Danko, J. (2020). Impact of direct traffic effect on online sales. *Journal of Research in Interactive Marketing, 14*(1), 17–32. https://doi.org/10.1108/JRIM-01-2019-0012

Kita, P., Čvirik, M., & Maciejewski, G. (2024). What is the meaning and importance of customer orientation as part of the business models? The case of retail chains in the food market. *Polish Journal of Management Studies, 30*(1), 169–186. https://doi.org/10.17512/pjms.2024.30.1.10

Komara, S., Cvirik, M., Kosikova, M., & Páleš, M. (2026). Utilizing Machine Learning and Explainable AI for Assessing Income Poverty Risk: Evidence from EU-SILC 2023 in Slovakia. *Engineering, Technology & Applied Science Research, 16(*2), 34219–34225. https://doi.org/10.48084/etasr.16958

Kumar, V., Kotler, P., & Kumar, A. (2026). Transformative marketing strategies in the era of new-age technologies: Principles, plan, purpose, and practice. *Journal of the Academy of Marketing Science, 54*(1), 1–27. https://doi.org/10.1007/s11747-025-01120-6

Li, M., Sun, H., Huang, Y., & Chen, H. (2024). Shapley value: From cooperative game to explainable artificial intelligence. *Autonomous Intelligent Systems, 4*(1), 2. https://doi.org/10.1007/s43684-023-00060-8

Lundberg, S. M., & Lee, S.-I. (2017, May 22). A Unified Approach to Interpreting Model Predictions. *Neural Information Processing Systems*. https://www.semanticscholar.org/paper/A-Unified-Approach-to-Interpreting-Model-Lundberg-Lee/442e10a3c6640ded9408622005e3c2a8906ce4c2

Molnar, C. (2020). *Interpretable machine learning*. Lulu.com.

O'Connor, P. (2021). Loyalty Programs and Direct Website Performance: An Empirical Analysis of Global Hotel Brands. In W. Wörndl, C. Koo, & J. L. Stienmetz (Eds.), Information and Communication Technologies in Tourism 2021 (pp. 150–161). *Springer International Publishing*. https://doi.org/10.1007/978-3-030-65785-7_13

Păvăloaia, V.-D., Anastasiei, I.-D., & Fotache, D. (2020). Social Media and E-mail Marketing Campaigns: Symmetry versus Convergence. *Symmetry, 12*(12), 1940. https://doi.org/10.3390/sym12121940

Peleg, B., & Sudhölter, P. (Eds.). (2007). *The Shapley Value. In Introduction to the Theory of Cooperative Games* (pp. 151–179). Springer. https://doi.org/10.1007/978-3-540-72945-7_8

Plaza, B. (2011). Google Analytics for measuring website performance. *Tourism Management, 32*(3), 477–481. https://doi.org/10.1016/j.tourman.2010.03.015

Powers, D. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies, 2*(1), 37–63.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 1135–1144. https://doi.org/10.1145/2939672.2939778

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Sak, T., Mylko, I., Ivashko, O., Chulipa, I., Wołowiec, T., & Sokołowska-Woźniak, J. (2025). How the Digital Economy Is Revolutionizing Marketing Management and Driving Sustainable Development. *Sustainability, 17*(3), 1130. https://doi.org/10.3390/su17031130

Šoltés, E., Zelinová, S., & Bilíková, M. (2019). General Linear Model: An Effective Tool For Analysis Of Claim Severity In Motor Third Party Liability Insurance. *Statistics in Transition New Series, 20*(4), 13–31. https://doi.org/10.21307/stattrans-2019-032

Stroobant, J. (2019). Finding the news and mapping the links: A case study of hypertextuality in Dutch-language health news websites. *Information, Communication & Society, 22*(14), 2138–2155. https://doi.org/10.1080/1369118X.2018.1477971

Sun, B. (2025). Data-driven personalized marketing strategy optimization based on user behavior modeling and predictive analytics: Sustainable market segmentation and targeting. *PLOS One, 20*(7), e0328151. https://doi.org/10.1371/journal.pone.0328151

Wurfel, M., Han, Q., & Kaiser, M. (2021). Online Advertising Revenue Forecasting: An Interpretable Deep Learning Approach. *2021 IEEE International Conference on Big Data (Big Data)*, 1980–1989. https://doi.org/10.1109/BigData52589.2021.9672010

Xun, J. (2015). Return on web site visit duration: Applying web analytics data. *Journal of Direct, Data and Digital Marketing Practice, 17*(1), 54–70. https://doi.org/10.1057/dddmp.2015.33