Predictive models for organic traffic growth: applying machine learning in digital analytics

Dubinin Kyryl Hennadiiovych[1]

Received: 2025-06-14
Accepted: 2025-07-19
DOI: https://doi.org/10.5281/zenodo.20255912

Abstract. This study builds certified predictive models of organic traffic expansion through incorporation of machine learning techniques to streamline online engagement techniques. The work defines a methodological taxonomy of the process of choosing the right algorithms as well as offers practical guidelines on the implementation of the correct prediction systems depending on the situation of search-driven user acquisition.

In the study, a systematic comparative analysis of twenty recent publications published in 2021-2026 will be performed, representing the current progress in the area. The research methodology integrates the outcomes of research studies assessing the performance of algorithms based on statistical models, supervised learning algorithms, deep neural networks, and hybrid ensemble algorithms considering the analysis of feature engineering and modeling temporal dependencies.

Combination of ensemble algorithms (hybrid ensemble) shows better accuracy in prediction compared to individual algorithm. Attention mechanisms in recurrent neural networks have been shown to achieve good results in modeling temporal dependencies among sequential patterns. Graph neural networks are useful in learning spatial dependencies. The combination of content quality indicators, technical performance measures, and search ranking factors via feature engineering can be seen as the keys to the best accuracy.

First taxonomy for methodology specific to needs of organic traffic forecasting which is different than general network prediction. First evidence-based taxonomy that integrates explainable artifical intelligence with search-driven context and connected advanced machine learning theory with digital marketing application through evidence based algorithm selection frameworks that take into account constraints of enterprise.

Actionable criteria for selecting a deployment architecture for enterprise based on requirements of enterprise. Evidence based guidelines support an iterative model development process from statistical approaches at bottom level to advanced deep learning capabilities match current resource capabilities. Framework provides optimization of resources by accurately forecasting and enhancing content strategy and reducing technical barriers to marketing allocation.

Key words: machine learning; organic traffic prediction; digital analytics; web traffic forecasting; predictive modeling.

[1] SEO Specialist at Academy Sports & Outdoors,
Bachelor's Degree in Marketing,
Simon Kuznets Kharkiv National University of Economics
e-mail: dubinin.kyryl@gmail.com
https://orcid.org/0009-0001-3272-2257

Introduction.

The growth of the digital platform has posed a major analytical problem in relation to studying the online behavior of users. Organic traffic, i.e. visitors who reached the company by using unpaid search results, is an indicator of key performance that is essential to digital companies. When this traffic can be predicted, it enables organizations to make strategic investments of resources and focus on developing contents. Nevertheless, traffic behavior is very complex in the sense of the nonlinear dynamics, seasonal changes, and other externalities and traditional analytical models prove insufficient. Contemporary analytics is challenged with significant challenges of measuring the complex relationships of user engagement measures, content quality measures, and search algorithm adjustments, and macro market trends. The unstable character of online traffic dictates the need of advanced methodologies that are able to process high-dimensional data streams and detect emergent patterns in them. The opportunity of machine learning technologies is great because the technologies learn more advanced relations on the basis of the previous data and apply it to predict the future events.

The contemporary literature on computational intelligence has shown significant advancements within Web Analytics. Ravlinko, et al. [1] illustrated that a key to developing digital competence is by relating employee skill sets to effective use of analytic tools. Petrukha et al. [2], identified how Agile frameworks could be used as an accelerator for implementing machine learning models. The authors (Abd Elminaam, et al. [3]) researched mechanisms of user engagement and found evidence of interactions between the content characteristics and users' behavior. Singh [4] showed potential for using predictive analysis to take data points and transform them into actionable strategies. Gatete [5] expanded our understanding through researching ways to combine machine learning with traditional modeling frameworks showing synergy when both are utilized together. Thakur, et al. [6] developed Hybrid models and evidenced the benefits of utilizing Ensemble Methods over individual models.

Zhu [7] identified that using a combination of Deep Learning to model Spatiotemporal Heterogeneity in Neural Networks demonstrated how networks could model dependency. Khatiriolyaee, M., Yang, L., & Pazzi, R. W. [8] reviewed the vast array of AutoML and traditional methodologies and showed evolutionary trends. Hussein et al. [9] established comparative evaluation frameworks for different families of algorithms and evaluated performance. Kablaoui, R., Ahmad, I., Abed, S., & Awad, M. [10] used computer vision techniques to show that time-series data can be represented as an image which demonstrates how visual learning can be applied to identify patterns. Aouedi, O., Le, V. A., Piamrat, K., & Ji, Y. [11] presented an in-depth review of recent developments focusing on innovation with regard to Recurrent Architectures and their superior predictive capability. Lykakis, E., Vardiambasis, I. O., & Kokkinos, E. [12] analyzed 5G-based prediction of traffic volume by taking into account challenges such as those created by heterogeneous streaming sources.

The researchers also have been working on these issues with a variety of studies (Schetakis et al. [14]; Tsalikidis et al. [15], etc.). The researchers used an empirical approach based on real network data for training their models and confirmed previous findings in this area; they showed that models created from real-time traffic data will be different than those created from static simulation data, thus highlighting the need for valid real world testing and evaluation of traffic model-based decision support systems. Schetakis et al. [14] developed and tested some early versions of quantum machine learning algorithms which were shown to be effective for processing large amounts of high dimensional data.

Results.

Machine learning application encompasses diverse methodological approaches. Ravlinko, Z., Petrukha, N., Terebukh, M., Berest, I., & Baran, I. [1] established competency development influences adoption suggesting implementation requires skill initiatives. Petrukha, N., Zhmaiev, A., & Synkevych, M. [2] demonstrated Agile facilitates iterative development enabling adaptation. AbdElminaam, D. S., Mostafa, D., Ibrahim, A. A., Bakry, Y. A., Ibrahim, F. M., & Fekry, N. H. [3] revealed content quality, experience design, and technical performance as determinants informing feature selection. Singh [4] emphasized value transforming observations into decisions. Gatete [5] demonstrated synergistic effects validating ensemble superiority. Thakur, U., Singh, S. K., Kumar, S., Singh, H., Arya, V., Gupta, B. B., Attar, R. W., Alhomoud, A., & Chui, K. T. [6] addressed modelling challenges by proposing architectures leveraging complementary strengths.

Table 1 – Machine Learning Approaches for Organic Traffic Prediction

Approach	Strengths	Limitations	Context
Statistical (ARIMA, SARIMA)	Interpretable, efficient, established	Limited non-linear capacity	Short-term stable patterns
Supervised (RF, XGBoost)	Non-linear handling, robustness	Extensive engineering required	Mid-term rich features
Recurrent (LSTM, GRU)	Temporal dependencies, automatic learning	Computational intensity	Complex sequential patterns
Graph Neural Networks	Spatial modeling, topology effectiveness	Architecture complexity	Multi-source spatial correlations
Hybrid Ensembles	Combined strengths, generalization	Increased complexity	Maximum accuracy requirements

Zhu (7) demonstrated excellent ability to capture heterogeneity. Architectures that incorporate attention mechanisms have shown better accuracy by dynamically assigning a weight of relevance to each observation. Khatiriolyaee, M., Yang, L., & Pazzi, R. W. (8), identified an area of research tracking trajectory and illustrating how the automation of optimization has opened access. The use of auto ML frameworks, reduces barriers to entry and allows non-experts in the field to implement complex systems. Hussein et al. (9), demonstrated that ensemble methods which combine boosting and network methods perform better than individual models across all time frames examined. Kablaoui, R., Ahmad, I., Abed, S., & Awad, M. (10), developed new forms of prediction using data representation to allow for the application of convolutional neural nets. The authors were able to demonstrate that these nets can be used to identify regularities in the data.

Aouedi, O., Le, V. A., Piamrat, K., & Ji, Y. (11), illustrated ways in which innovative ideas are being implemented in industry through their documentation of the identification of recurrent variants in time series that will effectively capture dependency and highlight how transformer architectures can process sequential data efficiently and train them quickly. Lykakis, E., Vardiambasis, I. O., & Kokkinos, E. (12), examined challenges related to dealing with heterogeneous streams of data. Because insights gained from research related to high velocity streaming of data is directly applicable to organic systems that experience volatile patterns of activity. Alekseeva et al. (13), demonstrated through evidence based studies, that models trained on operational data exhibit different characteristics than those trained on simulated or synthetic data and therefore may better represent realistic protocols that account for quality issues experienced during production.

This is a study of "frontier" research by schetakis et al. [14]. The researchers have used their theoretically based approach to demonstrate that it has potential for using re- upload techniques in demonstrating promise. In terms of approaches for developing quantum; as it develops we will be able to develop ways that are capable of providing a significant breakthrough capability for dealing with large amounts of data on-line in real time. Tsalikidis et al. [15] showed how an additional source of information (external) can add-value. A multi-step methodology was developed that shows that environment influences behavior, especially for lbs. It also highlights that taking into consideration domain specific external variable(s) can improve accuracy significantly if they are integrated.

Guo, X., Zhang, Q., Jiang, J., Peng, M., Zhu, M., & Yang, H. F. [16], dealt with the challenge of interpretability. An application of a language model was developed to generate explanations illustrating the ability to close the gap between mathematical models and practical knowledge. Explainability provides great value in many situations where team members need understandable justifications to inform strategic decisions and resource allocations. Zhao et al. [17], expanded upon graph methodologies. They proposed multi-granularity representations of dynamic systems that capture changes occurring over different levels of resolution. An architecture was developed that demonstrates that by combining spatial dependency relationships between sections of a system and its temporal evolution one can achieve higher predictive accuracy. This suggests that prediction should take into account both the interconnectivity of the content of a website and the navigational patterns users follow.

Kumar et al. [18], demonstrated that there is a wide range of applicability in terms of assessing current market trends which highlights the versatility of this method. The fundamental aspects of the engineering and validation processes are similar for all areas (networks or organic), however the choice of methods and the interpretation of results are highly dependent on domain specific knowledge. This reinforces the fact that the methodologies developed in network contexts can be transferred towards organic applications.

Ananthakrishnan et al. (2019) have highlighted the importance of determining how much reliance can be placed on the confidence level associated with a point prediction. According to the authors, a forecasting system that provides both a point forecast and confidence interval will allow users to consider the uncertainty related to the plan they are creating as part of the decision making process. This is particularly important for decision-makers when dealing with volatile markets which could lead to considerable variance in forecast accuracy.

The suggested architectural design is based on the best practice developed as the framework focuses on an iterative mechanism in which continuous monitoring and retraining models keep predictions relevant with time. The framework is specifically designed to overcome typical interpretability issues by explicitly integrating the elements of explainability, and it can promote learning and continuous improvement in the organization with the help of feedback mechanisms. Overall, the literature review has been synthesized to show a number of consistent patterns. To begin with, hybrid approaches involving a combination of several algorithms are always superior to the ones based on a single method, which means that the ensemble technique should be the new default in the production setting. Second, the role of domain-specific feature engineering can hardly be underestimated - generic models should be improved with content quality indicators, technical performance measures, and features associated with search engines to be as accurate as possible. Third, interpretability and explainability have become very important factors due to the need of stakeholder confidence and practical business understanding of business predictions where the reasoning behind them is transparent.

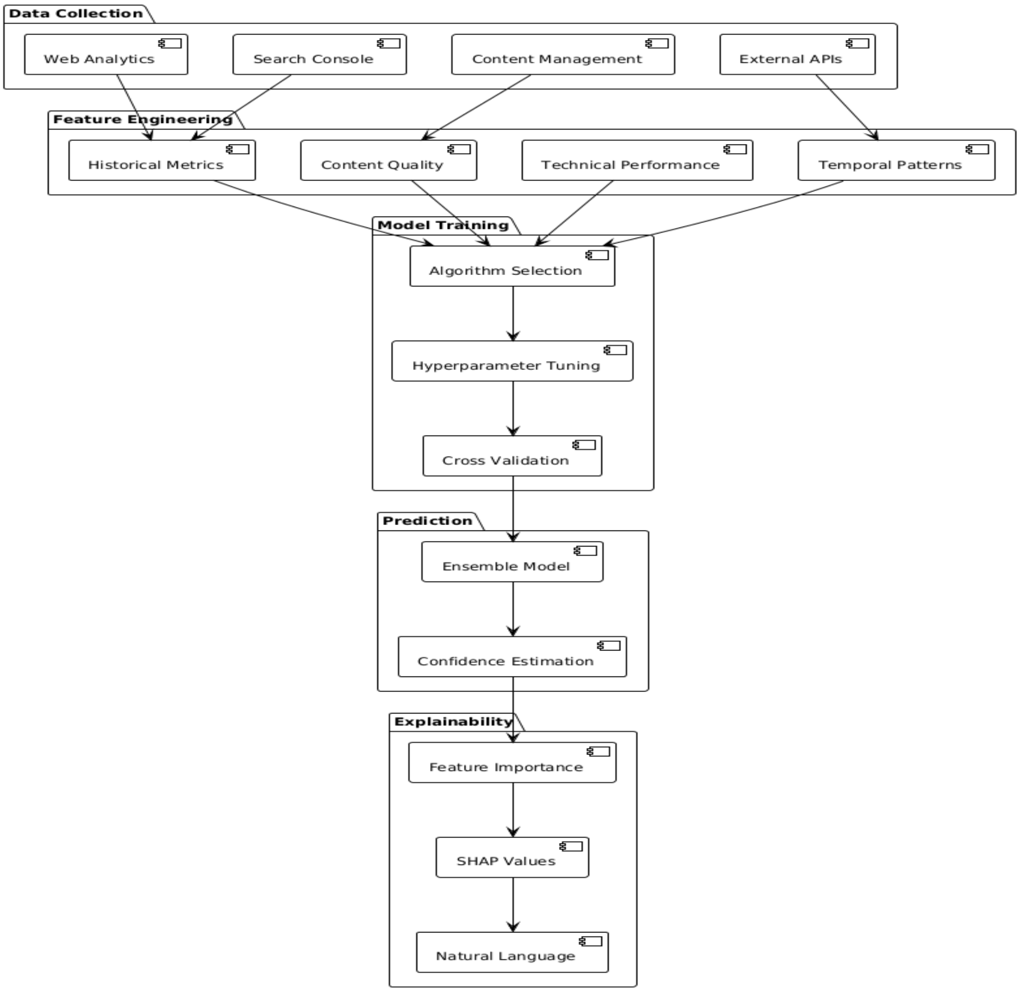

Figure 1 – Integrated ML Architecture for Organic Traffic Prediction (PlantUML)

Discussion of the results. The systematic examination of the current research shows significant advances in the area as well as the unresolved issues. The exhibited hybrid superiority justifies the original hopes that numerous algorithms in a hybrid model will promote the reduction of the biases of individual models and enhance the generalization ability. Nevertheless, empirical experience with implementation points toward ensemble complexity introducing issues of operational complexity, such as greater computational demands, greater maintenance cost, and issues in the interpretability of models. The study determines deep learning structures as innovative techniques that present superior performance rates. But these sophisticated methods demand large volumes of training data, very technical expertise, and big quantities of infrastructure - which are problematic when applied in small firms with small resources. As it has been observed, democratization of accessing such technologies via AutoML platforms can become one of the determiners of usage of technologies by organizations of all sizes.

Explainable mechanisms (how interpretable do explainable mechanism models allow you to understand why they made their prediction) will be critical for moving past simply being able to measure technical performance. Stakeholders (business leaders, customers) need to trust the predictions; and have the ability to take action based upon the insights derived from data driven decision making; and make better decisions by understanding what is driving the predictions. Large language models in the near future may create human readable explanation of their models, thereby closing the gap between the complexity of mathematical models and the utilization of such models within business. In the next several years I predict that explainability will become an essential component of each and every predictive model, rather than an optional feature. As well, there has been increasing importance placed on feature engineering with respect to specific domains, therefore it follows that generic modeling approaches will require significant advancements in the area of three key elements; Content Quality Features, Technical Performance Features, and Search Engine Optimization Features that increase prediction precision. Such advancements will likely result in truly collaborative cross-functional work between Data Scientists, Digital Marketing Specialists, and Content Strategy Specialists.

Conclusions. The studies developed an overall guideline using ML for predicting organic traffic growth. Hybrid Ensemble Methods will require structured evaluation and have been found to be the best method. There is evidence that deep Neural Networks such as LSTM networks, GRU Variants, Graph Networks show exceptional ability to capture Temporal Interactions (Multi-Aspect) and Spatial Heterogeneities. The research identified the Critical Success Factors: the appropriate Algorithm Selection depending upon the Horizon, the Full Feature Engineering, which includes Domain-Based Indicators; a well-established Data Infrastructure to enable Continuous Training Process and Explainability Mechanisms to provide confidence to the Stakeholder.

Practical recommendations may include to apply iterative development techniques beginning with simple statistical models that get more complex in methods as data and technical capacity may permit. The companies should aim at investing in the data infrastructure and improving the skills of the personnel besides developing the algorithms. In the research, the idea of AutoML platforms is discussed as the possible environment that will allow democratizing the access with minimum technical requirements and maintaining the quality of the predictions. The directions of future researches must deal with the problems of concept drift management in the swiftly changing online world, come up with the efficient training tactics to execute in deep architecture, the improve the procedure of explainability to execute in the complex ensembles, and the create the standard evaluation procedure to execute on an organic prediction case. Convergence in technology such as quantum machine learning and large language models are on their way and there might be opportunities of realizing improvements in precision and explanation in regard to modifying how companies will think about predicting and optimizing organically as regards to growth of traffic.

References

1. Ravlinko, Z., Petrukha, N., Terebukh, M., Berest, I., & Baran, I. (2022). Formation of digital competencies of personnel in the context of security aspects of the digital economy. International Journal of Computer Science and Network Security, 22(11), 628–634. DOI: https://doi.org/10.22937/IJCSNS.2022.22.11.89 URL: http://paper.ijcsns.org/07_book/202211/20221189.pdf
2. Petrukha, N., Zhmaiev, A., & Synkevych, M. (2024). Innovative approaches to it project management using agile project and management methods. Science and Technology Today, 8(36), 824–839. DOI: https://doi.org/10.52058/2786-6025-2024-8(36)-824-839
3. AbdElminaam, D. S., Mostafa, D., Ibrahim, A. A., Bakry, Y. A., Ibrahim, F. M., & Fekry, N. H. (2024). Unlocking the secrets of online engagement: A deep dive into website traffic. In 2024 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC) (pp. 235–242). IEEE. DOI: https://doi.org/10.1109/miucc62295.2024.10783519 URL: https://ieeexplore.ieee.org/document/10783519
4. Sumit Kumar Singh. (2025). Predictive analytics: Transforming historical data into strategic future insights. World Journal of Advanced Engineering Technology and Sciences, 15(3), 1774–1781. DOI: https://doi.org/10.30574/wjaets.2025.15.3.1119 URL: https://wjaets.com/node/1133
5. Gatete, O. (2025). Advancing predictive analytics: Integrating machine learning and data modelling for enhanced decision-making. International Journal of Latest

Technology in Engineering Management & Applied Science, 14(4), 169–189. DOI: https://doi.org/10.51583/ijltemas.2025.140400020 URL: https://www.ijltemas.in/submission/index.php/online/article/view/1857
6. Thakur, U., Singh, S. K., Kumar, S., Singh, H., Arya, V., Gupta, B. B., Attar, R. W., Alhomoud, A., & Chui, K. T. (2025). Advanced Web Traffic Modelling and Forecasting with a Hybrid Predictive Approach. Journal of Web Engineering. DOI: https://doi.org/10.13052/jwe1540-9589.2434 URL: https://ieeexplore.ieee.org/document/11037626
7. Zhu, S. (2025). Deep Learning Approaches for Traffic Forecasting: Spatiotemporal Heterogeneity and Model Advances. ITM Web of Conferences, 80, 01041. DOI: https://doi.org/10.1051/itmconf/20258001041 URL: https://www.itm-conferences.org/articles/itmconf/abs/2025/11/itmconf_acaai2025_01041/itmconf_acaai2025_01041.html
8. Khatiriolyaee, M., Yang, L., & Pazzi, R. W. (2026). A comprehensive review of traffic prediction: From traditional machine learning to AutoML. Neurocomputing, 666, 132148. DOI: https://doi.org/10.1016/j.neucom.2025.132148 URL: https://www.sciencedirect.com/science/article/pii/S0925231225028206
9. Hussein, A., Gan Goh, G. G., Anbar, M. A., Nassr, M., Semenova, M. I., & Aksenova, T. V. (2025). Comparative Evaluation of Machine Learning Models for Time Series Forecasting in Network Traffic Analysis. 2025 7th International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE), 1–6. DOI: https://doi.org/10.1109/reepe63962.2025.10971155 URL: https://ieeexplore.ieee.org/document/10971155
10. Kablaoui, R., Ahmad, I., Abed, S., & Awad, M. (2024). Network traffic prediction by learning time series as images. Engineering Science and Technology, an International Journal, 55, 101754. DOI: https://doi.org/10.1016/j.jestch.2024.101754 URL: https://www.sciencedirect.com/science/article/pii/S221509862400140X?dgcid=rss_sd_all&
11. Aouedi, O., Le, V. A., Piamrat, K., & Ji, Y. (2025). Deep Learning on Network Traffic Prediction: Recent Advances, Analysis, and Future Directions. ACM Computing Surveys, 57(6), 1–37. DOI: https://doi.org/10.1145/3703447 URL: https://dl.acm.org/doi/10.1145/3703447
12. Lykakis, E., Vardiambasis, I. O., & Kokkinos, E. (2025). Data Traffic Prediction for 5G and Beyond: Emerging Trends, Challenges, and Future Directions: A Scoping Review. Electronics, 14(23), 4611. DOI: https://doi.org/10.3390/electronics14234611 URL: https://www.mdpi.com/2079-9292/14/23/4611
13. Alekseeva, D., Stepanov, N., Veprev, A., Sharapova, A., Lohan, E. S., & Ometov, A. (2021). Comparison of Machine Learning Techniques Applied to Traffic Prediction of Real Wireless Network. IEEE Access, 9, 159495–159514. DOI: https://doi.org/10.1109/access.2021.3129850 URL: https://ieeexplore.ieee.org/document/9623523
14. Schetakis, N., Bonfini, P., Alisoltani, N., Blazakis, K., Tsintzos, S. I., Askitopoulos, A., Aghamalyan, D., Fafoutellis, P., & Vlahogianni, E. I. (2025). Data re-uploading in Quantum Machine Learning for time series: application to traffic forecasting (Version 1). arXiv. DOI: https://doi.org/10.48550/ARXIV.2501.12776 URL: https://arxiv.org/abs/2501.12776v1
15. Tsalikidis, N., Mystakidis, A., Koukaras, P., Ivaškevičius, M., Morkūnaitė, L., Ioannidis, D., Fokaides, P. A., Tjortjis, C., & Tzovaras, D. (2024). Urban Traffic Congestion Prediction: A Multi-Step Approach Utilizing Sensor Data and Weather Information. Smart Cities, 7(1), 233–253. DOI: https://doi.org/10.3390/smartcities7010010 URL: https://www.mdpi.com/2624-6511/7/1/10

16. Guo, X., Zhang, Q., Jiang, J., Peng, M., Zhu, M., & Yang, H. F. (2024). Towards explainable traffic flow prediction with large language models. Communications in Transportation Research, 4, 100150. DOI: https://doi.org/10.1016/j.commtr.2024.100150 URL: https://www.sciencedirect.com/science/article/pii/S2772424724000337?dgcid=rss_sd_all&
17. Zhao, X., Yu, W., & Zhang, Y. (2024). MST-GNN: graph neural network with multi-granularity in space and time for traffic prediction. GeoInformatica, 29(3), 377–401. DOI: https://doi.org/10.1007/s10707-024-00532-w URL: https://link.springer.com/article/10.1007/s10707-024-00532-w
18. Kumar M, V. T. R. P., Satish Kumar, S., Arun Kamal Sai, G., Tejaswini, B., Bhargavi, E., & Poojitha, T. (2025). Analyzing and Predicting Organic Pesticide Market Trends with Machine Learning. 2025 International Conference on Inventive Computation Technologies (ICICT), 1204–1208. DOI: https://doi.org/10.1109/icict64420.2025.11005234 URL: https://ieeexplore.ieee.org/document/11005234
19. Ananthakrishnan, A. V., Maya, D. K., & Krishnan, M. S. (2025). Predictive analytics: Unveiling road accident risks with machine learning algorithms. AIP Conference Proceedings, 3281, 060043. DOI: https://doi.org/10.1063/5.0247104 URL: https://pubs.aip.org/aip/acp/article-abstract/3237/1/060043/3337730/Predictive-analytics-Unveiling-road-accident-risks?redirectedFrom=fulltext
20. Subramanian, V. (2025). Driving value in digital marketing: Navigating from applied machine learning to applied artificial intelligence. Applied Marketing Analytics: The Peer-Reviewed Journal, 11(2), 109. DOI: https://doi.org/10.69554/ndmy9137 URL: https://hstalks.com/article/10202/driving-value-in-digital-marketing-navigating-from/?business