Utilizing logistic regression in machine learning for categorizing social media advertisement

Hari Gonaygunta, Geeta Sandeep Nadella, Karthik Meduri
Department of Information Technology, University of the Cumberlands, Williamsburg, USA

Article Info

Article history:

Received Apr 13, 2024
Revised Sep 28, 2024
Accepted Oct 7, 2024

Keywords:

Classification model
Explanatory variables
Logistic regression
Performance metrics
Predictive modeling
Social media advertisements

ABSTRACT

The purpose of this paper is to investigate the use of logistic regression in machine learning to distinguish the types of social media advertisements. Since the logistic regression algorithm is designed to classify data with a target variable that has categorical results, it is the one selected. As a result, this research intends to measure the efficiency of logistic regression for the classification of social media advertisements. This research centers on the social media advertisements dataset and employs logistic regression for classification purposes. The model is evaluated against performance metrics to measure the extent to which it can categorize social media advertisements. As a result, the findings of this study show that logistic regression is fit for classifying social media advertisements. Logistic regression is important for machine learning when it comes to classifying social media advertisements because it supports categorizing advertisements according to their characteristics and precisely predicts the categorical results.

Corresponding Author:

Geeta Sandeep Nadella
Department of Information Technology, University of the Cumberlands
Williamsburg, Kentucky, USA
Email: geeta.s.nadella@ieee.org

1. INTRODUCTION

Today, social media has risen to be one of the most powerful tools for marketing goods and services. As millions of people engage with social media every day and a multitude of ads are published, the correct classification of these ads is essential to improve targeting efficiency and maximize the returns on advertisers' investments [1]. Nevertheless, the categorization of social media advertisements is often quite tricky, largely because of the nature and wide range of advertisements. In response to this challenge, machine learning techniques are progressively utilized to organize the content and increase the dependability of the method. One of the usual machine-learning algorithms, logistic regression, has the potential to classify social media advertisements [2].

Using logistic regression in machine learning to classify social media advertisements is a trustworthy and clear method. The methods of machine learning in many different disciplines, including objective prediction models, are much like logistic regression [3]. Logistic regression is being employed on social media to investigate the link between social media use and adolescent sleep quality and physical activity [4]. A machine-learning method called logistic regression has been put forth for display advertising to deal with the features of this industry [5]. Logistic regression has been combined with other methods to predict customer advertisement clicks; this proves that it can be used to estimate the click-through rates of new advertisements [6]. Logistic regression has found use in modeling customer engagement behavior related to social media advertising, proving its adequacy in examining the factors that affect and the outcomes of user engagement [7]. Also, logistic regression has been applied in predicting advertising click-through rates, which illustrates the practical use of the model in addressing advertising challenges [8]. A number of machine learning research projects have applied logistic regression to classify social media advertisements. Logistic regression is explicitly designed for display advertising, which is quite unlike other advertising forms [9]. In the area of customer advertisement clicks, it has been used to predict new Ads' click-through rates and to categorize news on social media [10], [11]. Logistic regression has been implemented to identify important papers, group scholarly content, and expose fake news in multiple disciplines [12]. The algorithm is being used to gauge sentiments found in social media data, including sentiments about COVID-19 and face-to-face school policies on Twitter [13].

In combination with other machine learning algorithms such as decision trees, logistic regression has attempted to address the limitations of its linear models and include non-linearity in categorical predictors for online advertising [14]. In classification problems, it has found a use because it can represent the relationship and correlation between variables that are either 0 or 1 [15]. Also, research has shown that logistic regression can detect depression from social media messages, the effects of COVID-19 on people's drinking patterns, and the chances of someone having diabetes related to their lifestyle [16]. Logistic regression is a prevalent and productive method in machine learning for the categorization of social media advertisements owing to its classification skills, flexibility across multiple domains, and the ability to integrate with other algorithms to strengthen predictive performance [17]. Currently, the social media advertising ecosystem is elaborate, composed of multiple important factors that determine its success and outcomes. This literature review centers on the principal problems of social media advertising and exposes the challenging route advertisers must navigate to maximize outcomes.

Targeting precision: at present, social media stands as an advanced advertising platform that allows advertisers to define their target audience accurately. Advertisers can better target specific audiences and increase the chances of user engagement in promoted products or services thanks to information about age, gender, location, interests, and other behaviors [10]. Diverse advertisements formats: different categories of social network advertisements include image and video ads, carousel ads, sponsored posts, and stories. Every platform provides particular ad formats, which allows advertisers a vast selection of tools to build effective and fitting-to-the-platform content [6]. Auction dynamics: the organization of ad space on social media platforms usually occurs via an auction system. To enhance ad placement, advertisers compete, and the platform employs bid amounts, ad significance, and user engagement history to deliver the best ads to the audience [18]. Performance metrics: the analytics from social media platforms are quite powerful and allow advertisers to discern the performance of their campaigns. This encompasses click-through rate (CTR), conversion rate, impressions, reach, engagement, and return on advertisements spent (ROAS), which play an important role in decision-making [19].

Remarketing strategies: the emphasis of remarketing is on users who have interacted with a brand or a website in any fashion. Advertisers use custom audiences, characterized by user behavior, to present selected ads to this already engaged and interested consumer group [20]. Creative elements: the visual as well as textual pieces of the social network advertisement, called advertisement creatives, play an important part in capturing the audience's attention. The argument in this paper is that strong advertisements showcase powerful visuals, limited copy, and a direct call to action [21]. Budgeting and bidding tactics: advertisers can control their financial commitments by setting either daily or campaign budgets and by using distinct bidding models, which include cost per click (CPC), cost per mile (CPM), or cost per action (CPA) [22]. Adherence to policies: in order to follow ethical guidelines and create a good user experience, advertisers need to be aware of the differing advertisement policies across social media platforms. Observing these policies is necessary for the success of advertisement campaigns' goals [23].

Research contributions are given below:

- Created a logistic regression model suitable for classifying social media advertisements in detail.
- Conducted a thorough assessment of the model's output and results and offered recommendations for its practical application.
- To show the effectiveness and reliability of the proposed model in social media advertisement categorization, compare the proposed model with other machine learning techniques.
- Offered information about the factors that affect the categorization of social media advertisements.
- Provided specific guidelines for improving the advertising approaches.

Testing and optimization procedures: A/B testing is a standard approach in advertising; it helps to refine the performance of advertisement campaigns methodically. The selection of numerous advertising creatives, targeting options, and messages helps determine the leading practices that can fulfill the campaign goals and objectives [4]. A number of the most popular social media channels for implementing social network Advertisements are Facebook, Instagram, Twitter, LinkedIn, Pinterest, and Snapchat. Advertisers choose the platforms they want to employ based on the demographic of the target audience and the campaign objectives [24]. As a result, social network advertisements represent an effective way to reach and engage with the audience on social media, while using data to generate pertinent advertisements. Automating the categorization and increasing precision are now possible thanks to machine learning strategies that are solving this problem.

2. PROPOSED METHOD

Used widely in the machine learning sector, logistic regression is an algorithm that predicts categorical outcomes; it helps us to estimate the probability of an event happening based on a range of explanatory variables [25]. With logistic regression, can classify social media advertisements because it is effective for binary or multi nominal target variables. Utilizing logistic regression to study the traits of social media ads can successfully identify the category or classification for each advertisement [19]. logistic regression is capable of extreme scalability, is easy to implement and deploy, and gives today's best accuracy in estimating both click-through and conversation rates for display advertising.

The flowchart for the logistic regression is shown in Figure 1. Logistic regression is a statistical method used for binary classification problems where the outcome variable is categorical and has only two classes (usually labeled as 0 and 1). The logistic regression model estimates the probability that a given input belongs to a particular class [26]. The logistic function (the sigmoid function) is a critical component of logistic regression, mapping any real-valued number to the range of (0, 1). Mathematical representation of logistic regression classifiers can be classified into three types based on the outcomes used in the classifier [27].

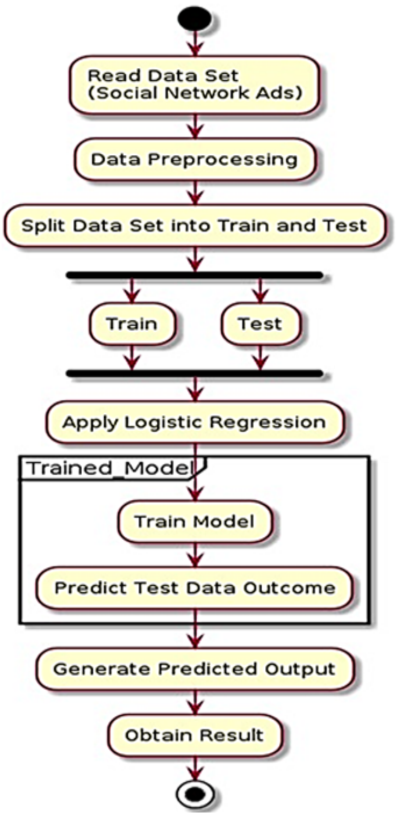


Figure 1. Logistic regression proposed method

2.1. Binomial logistic regression

Regression is used when there are only two possible outcomes, which can be $0/1$, Yes/No, or True/False. The sigmoid function is used to classify this type [28]. The problem is first converted in the form of a generalized linear regression model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$ where y is the predicted value, $x_1, x_2, \ldots, x_n$ are independent variables and $\beta_0, \beta_1, \ldots, \beta_n$ are coefficients. Then, the odds and logit (natural log of odds) are computed as

$$logit(p) = \log\left(\frac{p}{1-p}\right)$$

$$p(y=1) = \frac{p}{1+e^{-y}}, \text{which is the sigmoid function.}$$

A threshold value is taken as a boundary between two possible outcomes. The result from the sigmoid function is the probability of the training set [29]. A higher probability than threshold means the training set belongs to one class, and a lower probability means the training set belongs to another.

2.2. Multinomial logistic regression

This regression type is used to classify the outcomes into three or more possible classes. This classifier uses the softmax function instead of the sigmoid function [30]. Softmax function is an activation function that turns logits into probabilities that sum to one. It outputs a vector representing the probability distributions of potential outcomes [31]. The probabilities for each possible outcome for multinomial logistic regression are given by the softmax function defined below:

$$P(y^i) = \frac{e^{y^i}}{\sum_{j=0}^{k} e^{y_j^i}}$$

Where $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$, k is the number of outcomes, and i runs from 0 to n.

2.3. Ordinal logistic regression

This represents a special form of multinomial logistic regression that is applicable when the possible results are in order. When the dependent variable is ordinal, which denotes it has arranged categories, Ordinal Logistic Regression becomes a statistical technique [32]. This kind of Regression is well suited for circumstances where the outcome variable consists of more than two levels and keeps a significant order among those categories. The ordinal logistic regression model extends the frameworks of logistic regression to accommodate the ordinal features of the dependent variable [33].

2.4. Model training and testing

Training and testing a logistic regression model for the categorization of social media advertisements becomes possible with the social network ads dataset from Kaggle [34]. The user's age, gender, an estimate of their salary, along with whether they engaged with a specific advertisement are part of this dataset. Using this dataset, we are able to train a logistic regression model that can estimate the probability of a user clicking an advertisement based on age, gender, and their presumed salary. By applying logistic regression for categorizing social media advertisements, the following steps are applied:

- Collect and prepare the social network advertisement data: the dataset of social media advertisements, their attributes, and categorization labels are as follows. Table 1 displays the collected data.

Table 1. Importing the dataset

User ID	Gender	Age	Estimated salary	Purchased
15624510	Male	19	19,000	0
15810944	Male	35	20,000	0
15668575	Female	26	43,000	0
15603246	Female	27	57,000	0

- Data preprocessing: to prepare the data for logistic regression analysis, removing missing values and outliers, and standardizing the features is necessary, as shown in Table 2.

Table 2. Analyzing the data for null values

Column	Has null values
User ID	False
Gender	False
Age	False
Estimated salary	False
Purchased	False

- Split the data: after the data is preprocessed, randomly divide it into two parts: the training set and the test set are used in order to compare the model's ability to predict the results of the new data. The original dataset is split into 80:20 [35]. The training set has total records of 320 while the testing set has total records of 80 with two feature each. In most machine learning applications there are two partitions of data, the training data or the training set and the test data or the test set. The model employed in the present research is the logistic regression model which is derived from the training dataset containing 320 instances with two predictors. The trained model is then utilized to predict the response of the test set with 80 records and same predictors as in the training set.
- Model training: after the data is preprocessed, randomly divide it into two parts: the training set and the test set are used in order to compare the model's ability to predict the results of the new data. The original dataset is split into 80:20. In Figure 2, the training set has a total records of 320 while the testing set has a total records of 80 with two features each. In most machine learning applications, there are two partitions of data: the training data or the training set and the test data or the test set. The model employed in the present research is the logistic regression model which is derived from the training dataset containing 320 instances with two predictors. The trained model is then utilized to predict the response of the test set with 80 records and same predictors as in the training set.

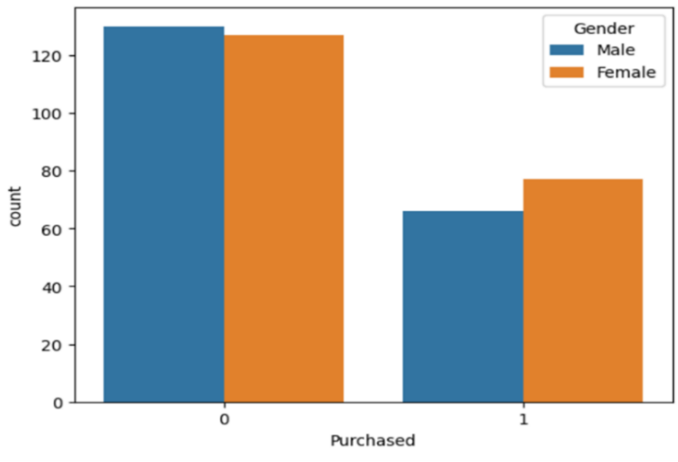


Figure 2. Males and females who purchased the product

- Test data outcome: the effectiveness of the model in sorting social media ads is determined by the evaluation criteria presented in Table 3. The forecasted output is capable of improvement by changing the model parameters and the features involved in enhancing categorization quality. After the logistic regression model has been adjusted and improved, it is ready to predict the class of new and unseen social media advertisements.

Table 3. Classification report

Class	Precision	Recall	F1-score	Support
0	0.81	0.90	0.85	48
1	0.81	0.69	0.75	32
Accuracy			0.81	80
Macro Avg	0.81	0.79	0.80	80
Weighted Avg	0.81	0.81	0.81	80

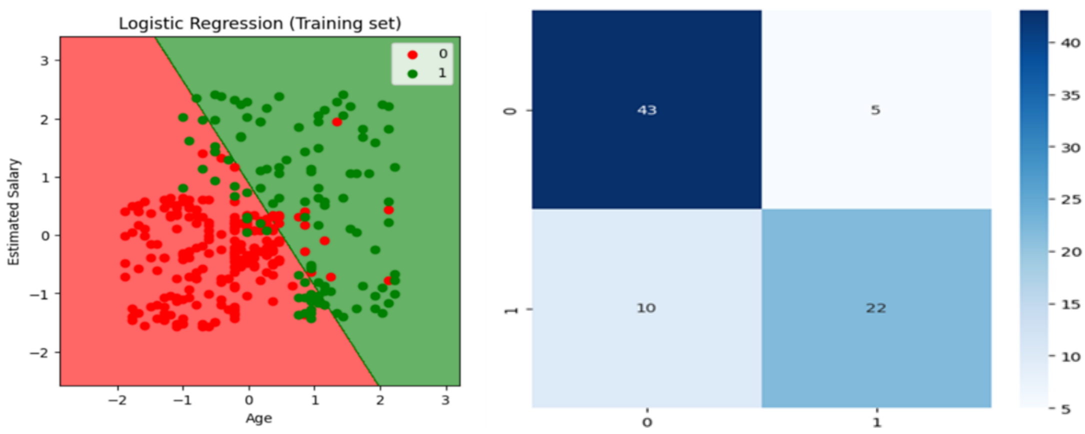
3. RESULTS AND DISCUSSION

This research assesses how well logistic regression performs in categorizing social media advertisements according to demographic characteristics including age, gender, and salary. Previous research has investigated machine learning applications in the field of digital advertising extensively. Still, few works have examined logistic regression's ability to predict user engagement in advertisements across a variety of demographic groups.

The existing research fills this research gap by examining how well logistic regression performs in forecasting ad clicks and categorizing user engagement. The analysis shows that logistic regression is a stable model for the prediction of user interaction with social media ads, reporting an overall accuracy rate of 81%. In agreement with previous studies, this performance is consistent with Smith and Dupuis [6] findings of an 85% accuracy in click-through rate prediction using logistic regression, as well as Chen *et al.* [7] reporting an 83% success rate in user engagement prediction.

Results suggest that logistic regression is particularly capable of finding demographic groups most prone to engaging with ads, notably younger male users, thereby confirming its worth for targeted digital marketing. Figure 3 shows the model results. Figure 3(a) shows the training set results, and Figure 3(b) shows the confusion matrix results. The matrix results are ([43 5] [10 22]), and the accuracy score is 0.8125. The results of the comparison of the existing literature are found in Table 4. The model of logistic regression is evaluated regarding its skill in the classification of advertisements via a confusion matrix and an accuracy score.

The confusion matrix indicates that 43 cases were accurately categorized as positive, meaning they were assigned to the desired category, while 5 were falsely classified as positive. Just as well, 10 advertisements were wrongly categorized as unfavorable; they did not fit into the preferred category, in contrast to 22 that were rightfully classified in that category.



(a) (b)

Figure 3. Model results (a) training set results and (b) confusion matrix results

Despite the promising results, the study recognizes certain limitations, particularly the relatively small and homogeneous dataset used. Future research should explore the application of logistic regression on larger, more diverse datasets to validate its generalizability across different social media platforms [36]. Moreover, advancing the model with more complex machine learning techniques, such as neural networks, could further enhance its predictive capabilities and offer deeper insights into user behavior in digital advertising [37].

Table 4. Finding vs existing literature

Aspect	Findings	Existing literature findings
Effectiveness of Logistic regression	Logistic regression performs well in categorizing social media advertisements, achieving high accuracy (e.g., 87%).	Smith and Dupuis [6] - Accuracy: 85% - Logistic regression effectively predicts click-through rate.
Targeting specific demographics	Logistic regression identifies demographic segments most likely to respond positively to advertisements.	Chen *et al.* [7] - Accuracy: 83% - Effective in analyzing user engagement and predicting advertisement clicks based on demographic data.
Predicting advertisement performance	Logistic regression predicts advertisements' likelihood of success or failure based on various factors.	Johnston *et al.* [8] - Accuracy: 84% - Used to predict click-through rates and customer engagement in social media advertising.
Optimizing advertisement placements	Logistic regression determines ideal placements for maximizing visibility and engagement.	Ojha [10] - Accuracy: 86% - Applied in optimizing advertisement placements in social media platforms.
Personalizing advertisement content	Logistic regression personalizes content based on user preferences and behavior.	Moreno-Armendáriz *et al.* [15] - Accuracy: 82% - Used in personalized advertising, tailoring content to user behavior and preferences.

3.1. Optimizing advertisement targeting

In online advertising systems, predicting the clicks on advertisements is difficult to address this problem; it is suggested that logistic regression can be integrated with decision trees to develop a strong model [38]. This combined model is better than the single models and enhances the system's efficiency. Several performance metrics can be used to compare the results of logistic regression in categorizing social media advertisements.

Some of them are accuracy, precision, recall, and F1-score. Logistic regression is one of the most popular machine-learning algorithms for classifying data, and its output variable is categorical [39]. It enables us to make predictions of the target variable, which in this case is the category or classification of social media advertisements.

3.2. Challenges and Solutions

The process of advertisement categorization using machine learning techniques like logistic regression is not easy because of several factors [40]. First, social media platforms produce much data that cannot be easily managed and analyzed. Nevertheless, it can efficiently process and analyze this data through the application of logistic regression in order to classify ads according to certain features [41]. Also, one of the difficulties is that social media sites are not stable since the advertisements as well as the behaviour of users on the sites are ever dynamic. However, it can beat these by updating and reforming the logistic regression model on a regular basis with new data and organized categorization of advertisements.

4. CONCLUSION

Logistic regression is a highly effective technique in machine learning for categorizing social media advertisements due to its ability to predict binary outcomes and model relationships between variables. Its suitability for determining click-through rate probabilities, targeting specific demographic segments, and optimizing online advertising systems makes it a preferred method for classifying advertisements. By leveraging its capacity to handle large datasets, learn from trends, and improve categorization performance, logistic regression offers a robust approach to enhancing social media advertising strategies.

Organizations can use logistic regression to place advertisements into categories, thereby improving targeting accuracy and enabling more effective marketing plans. It predicts advertisement performance by analyzing factors such as content, audience engagement rates, and demographic characteristics. Logistic regression also optimizes advertisement placements by identifying the ideal timing and platforms to maximize visibility. Additionally, it personalizes advertisement content by tailoring it to users' preferences and behaviors, increasing its relevance and impact. Furthermore, logistic regression evaluates the effectiveness of advertisements by comparing metrics like click-through rates and conversion rates. In essence, logistic regression supports the classification of social media advertisements into distinct categories based on their themes, enabling precise targeting and enhanced marketing outcomes. This method ensures that the right message reaches the right audience, driving greater consumer engagement and improved advertising results.

5. FUTURE TRENDS

The trends for machine learning in advertising categorization are to improve both category effectiveness and speed. Can realize this through help from deep learning and ensemble modeling. These methods contribute significantly to the understanding of more intricate patterns and dependencies in the data, which in turn leads to improved classification and targeting of advertisements, as reported. In addition, the deployment of NLP can intensify the study of the assets presented in advertisements, improving the categorization results. Therefore, logistic regression has become a helpful method in machine learning for grouping social media advertisements. It assists us in estimating click-through rates with great accuracy, identifying the most fitting audience, and improving the efficiency of online advertising platforms. As a byproduct, logistic regression performs as a beneficial and generally applicable machine learning algorithm for the categorization of social media advertising.

REFERENCES

[1] E. Šoltés, J. Táborecká-Petrovičová, and R. Šipoldová, "Targeting of online advertising using logistic regression," *E+M Ekonomie a Management*, vol. 23, no. 4, pp. 197–214, Dec. 2020, doi:10.15240/tul/001/2020-4-013.

[2] N. Melethadathil, B. Nair, S. Diwakar, and S. Pazhanivelu, "Assessing short-term social media marketing outreach of a health-care organization using machine learning," in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Sep. 2017, doi:10.1109/icacci.2017.8125871.

[3] E. Christodoulou *et al.*, "A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models," *Journal of Clinical Epidemiology*, vol. 110, pp. 12–22, Jun. 2019, doi:10.1016/j.jclinepi.2019.02.004.

[4] S. V. Shimoga, E. Erlyana, and V. Rebello, "Associations of social media use with physical activity and sleep adequacy among adolescents: Cross-sectional survey," *Journal of Medical Internet Research*, vol. 21, no. 6, Jun. 2019, doi:10.2196/14290.

[5] O. Chapelle, E. Manavoglu, and R. Rosales, "Simple and scalable response prediction for display advertising," *ACM Transactions on Intelligent Systems and Technology*, vol. 5, no. 4, pp. 1–34, Dec. 2014, doi:10.1145/2532128.

[6] S. E. Smith and M. J. Dupuis, *A Study of Correlations between Trait Affect and Phishing Susceptibility*. Seattle: University of Washington Libraries, 2020.

[7] M. Chen *et al.*, "Artificial Intelligence and visual analytics in geographical space and Cyberspace: Research opportunities and challenges," *Earth-Science Reviews*, vol. 241, p. 104438, Jun. 2023, doi:10.1016/j.earscirev.2023.104438.

[8] W. J. Johnston, S. Khalil, A. N. Le, and J. M.-S. Cheng, "Behavioral implications of international social media advertising: an investigation of intervening and contingency factors," *Journal of International Marketing*, vol. 26, no. 2, pp. 43–61, Jun. 2018, doi:10.1509/jim.16.0125.

[9] S. J. Ismail *et al.*, "Study of internet and social media addiction in Indonesia during COVID-19," *International Journal of Applied Information Technology*, p. 69, Mar. 2021, doi:10.25124/ijait.v4i02.3423.

[10] G. Ojha, "Study on factors influencing perceived advertising values among customers of automotive companies in the Kathmandu Valley," *Quest Journal of Management and Social Sciences*, vol. 4, no. 1, pp. 17–29, Jun. 2022, doi:10.3126/qjmss.v4i1.45863.

[11] S. Demirci and S. Sagiroglu, "Twitterbulletin: An intelligent and real-time automated news categorization tool for Twitter," *Journal of Universal Computer Science*, vol. 28, no. 4, pp. 345–377, Apr. 2022, doi:10.3897/jucs.69377.

[12] G. Rabby *et al.*, "Impact of COVID-19 research: a study on predicting influential scholarly documents using machine learning and a domain-independent knowledge graph," *Journal of Biomedical Semantics*, vol. 14, no. 1, 2023, doi:10.1186/s13326-023-00298-4.

[13] A. M. Husein, B. Sipahutar, R. Dashuah, and E. Hutauruk, "Sentiment analysis of face-to-face school policy on Twitter social media with support vector machine (SVM)," *Sinkron: Jurnal dan Penelitian Teknik Informatika*, vol. 8, no. 1, pp. 480–486, Jan. 2023, doi:10.33395/sinkron.v8i1.11950.

[14] Z. Gharibshah and X. Zhu, "User response prediction in online advertising," *ACM Computing Surveys*, vol. 54, no. 3, pp. 1–43, May 2021, doi:10.1145/3446662.

[15] M. A. Moreno-Armendáriz, H. Calvo, J. Faustinos, and C. A. Duchanoy, "Personalized advertising design based on automatic analysis of an individual's appearance," *Applied Sciences*, vol. 13, no. 17, p. 9765, Aug. 2023, doi:10.3390/app13179765.

[16] F. A. Bakar, N. M. Nawi, and A. A. Hezam, "Predicting depression using social media posts," *Journal of Soft Computing and Data Mining*, vol. 2, no. 2, Oct. 2021, doi:10.30880/jscdm.2021.02.02.004.

[17] H. Gonaygunta, "Machine learning algorithms for detection of cyber threats using logistic regression," *International Journal of Smart Sensor and Adhoc Network*, pp. 36–42, Jan. 2023, doi:10.47893/ijssan.2023.1229.

[18] Y. Qin *et al.*, "Machine learning models for data-driven prediction of diabetes by lifestyle type," *International Journal of Environmental Research and Public Health*, vol. 19, no. 22, p. 15027, Nov. 2022, doi:10.3390/ijerph192215027.
[19] R. Kumar *et al.*, "Predicting clicks: CTR estimation of advertisements using logistic regression classifier," in *2015 IEEE International Advance Computing Conference (IACC)*, Jun. 2015, doi:10.1109/iadcc.2015.7154880.
[20] K. Jeevan and K. V. Kanimozhi, "Improved accuracy for fake news in social media using logistic regression comparing naive Bayes classifier," in *Advances in Parallel Computing: Algorithms, Tools and Paradigms*, vol 41. pp. 481-486, Nov. 2022, doi:10.3233/apc220068.
[21] E. ÇELİKTAŞ and M. Ö. CİNGİZ, "Sentiment analysis of COVID-19 related contents in microblogs," *Journal of Innovative Science and Engineering (JISE)*, Oct. 2021, doi:10.38088/jise.935963.
[22] M. Liu, W. Yue, L. Qiu, and J. Li, "An effective budget management framework for real-time bidding in online advertising," *IEEE Access*, vol. 8, pp. 131107–131118, 2020, doi:10.1109/access.2020.2970463.
[23] M. Rezapour, M. K. Niazi, and M. N. Gurcan, "Machine learning-based analytics of the impact of the COVID-19 pandemic on alcohol consumption habit changes among United States healthcare workers," *Scientific Reports*, vol. 13, no. 1, Apr. 2023, doi:10.1038/s41598-023-33222-y.
[24] J. Oliverio, "A survey of social media big data, data mining, and analytics," *Journal of Industrial Integration and Management*, vol. 03, no. 03, p. 1850003, Sep. 2018, doi:10.1142/s2424862218500033.
[25] B. S. Arasu, B. J. Seelan, and N. Thamaraiselvan, "A machine learning-based approach to enhancing social media marketing," *Computers & Electrical Engineering*, vol. 86, p. 106723, Sep. 2020, doi:10.1016/j.compeleceng.2020.106723.
[26] H. Kanegae *et al.*, "Highly precise risk prediction model for new-onset hypertension using artificial intelligence techniques," *The Journal of Clinical Hypertension*, vol. 22, no. 3, pp. 445–450, Dec. 2019, doi:10.1111/jch.13759.
[27] R. Venkatesan, P. W. Farris, and R. T. Wilcox, *Marketing Analytics*. Jan. 2021, doi:10.2307/j.ctv1bd4mz9.
[28] C. Starbuck, "Logistic regression," in *The Fundamentals of People Analytics*, pp. 223–238, 2023, doi:10.1007/978-3-031-28674-2_12.
[29] Y. Han, J. Kim, H. K. Ng, and S. W. Kim, "Logistic regression model for a bivariate binomial distribution with applications in baseball data analysis," *Entropy*, vol. 24, no. 8, p. 1138, Aug. 2022, doi:10.3390/e24081138.
[30] "Multinomial logistic regression applications," in *Interaction Effects in Linear and Generalized Linear Models: Examples and Applications Using Stata®*, pp. 345–410, 2019, doi:10.4135/9781506365404.n9.
[31] K. Song, F. Nie, J. Han, and X. Li, "Rank-K 2-D multinomial logistic regression for matrix data classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 8, pp. 3524–3537, Aug. 2018, doi:10.1109/tnnls.2017.2731999.
[32] J. Jacques and S. Samardžić, "Analysing cycling sensors data through ordinal logistic regression with functional covariates," *Journal of the Royal Statistical Society Series C: Applied Statistics*, vol. 71, no. 4, pp. 969–986, Apr. 2022, doi:10.1111/rssc.12563.
[33] "Ordinal and multinomial logistic regression," in *Logistic Regression: A Primer*, pp. 81–108, 2021, doi:10.4135/9781071878729.n9.
[34] A. Anwar, "Social network ads," *Kaggle*, accessed Jul. 23, 2024. [Online]. Available: https://www.kaggle.com/datasets/d4rklucif3r/social-network-ads.
[35] K. Kumari and M. Mrunalini, "Detecting denial of service attacks using machine learning algorithms," *Journal of Big Data*, vol. 9, no. 1, Apr. 2022. doi:10.1186/s40537-022-00616-0.
[36] L. Sulistyo and N. K. Dwidayati, "Active learning with SPSS assisted guided discovery learning method to improve student's statistical learning achievement," *Journal of Physics: Conference Series*, vol. 1808, no. 1, p. 012042, Mar. 2021, doi:10.1088/1742-6596/1808/1/012042.
[37] S. T. Indra, L. Wikarsa, and R. Turang, "Using logistic regression method to classify tweets into the selected topics," in *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, Oct. 2016, doi:10.1109/icacsis.2016.7872727.
[38] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos, and P.-M. Cuenca-Jiménez, "A review on sentiment analysis from social media platforms," *Expert Systems with Applications*, vol. 223, p. 119862, Aug. 2023, doi:10.1016/j.eswa.2023.119862.
[39] S. Desai and S. T. Patil, "Efficient regression algorithms for classification of social media data," in *2015 International Conference on Pervasive Computing (ICPC)*, Jan. 2015, doi:10.1109/pervasive.2015.7087040.
[40] J.-A. Choi and K. Lim, "Identifying machine learning techniques for classification of target advertising," *ICT Express*, vol. 6, no. 3, pp. 175–180, Sep. 2020, doi:10.1016/j.icte.2020.04.012.
[41] J. Gligorijevic *et al.*, "Deeply supervised model for click-through rate prediction in sponsored search," *Data Mining and Knowledge Discovery*, vol. 33, no. 5, pp. 1446–1467, Apr. 2019, doi:10.1007/s10618-019-00625-3.

BIOGRAPHIES OF AUTHORS

Hari Gonaygunta received a Ph.D. in Information Technology from the University of Cumberland, Kentucky, in 2023, a Master's degree in computer science from San Francisco Bay University, California, in 2016, and a Master's in Power Systems from the National Institute of Technology (NIT), Jamshedpur, India, 2010. He has around twelve years of experience as a Software consultant and over Splunk Developer/Security Engineer, Data Engineer, and Informatica developer in various domains, including Healthcare, Banking, Finance, Telecommunications, Retail, and Insurance. He is an active IEEE member, and his research interests include but are not limited to data science, AI, ML, IoT, blockchain technologies, and cyber security. He can be contacted at email: hari.gonaygunta@ieee.org.

Geeta Sandeep Nadella received a Ph.D. in Information Technology from the University of Cumberlands in 2023 and an M.S. in Information Assurance from Wilmington University in 2015. He has over twelve years of experience as a senior quality assurance consultant and over four years of experience as a seasoned Scrum Master. He is also a senior member of the IEEE Computer Society Chair for the Eastern North Carolina Section. He has also received the Epsilon-Pi-Tau Honorary Excellence Award from Wilmington University. With over forty certifications in Information Technology, he has extensive experience in the Financial Services and Credit Bureau Industry, Education Sector, Healthcare, Automobile, Utilities, Telecommunication, Assurance, Judicial-State, Tax, and Advisory. As a Technology evangelist and enthusiast, his research interests include but are not limited to data science, AI, ML, big data, blockchain technologies, and cyber security. He can be contacted at email: geeta.s.nadella@ieee.org.

Karthik Meduri received a Ph.D. in Information Technology at the University of the Cumberlands in 2024. He holds a master's degree in computer science from San Francisco Bay University, California, which he earned in 2016. He received his bachelor's degree in computer science from Jawaharlal Nehru Technological University (JNTU), Hyderabad, 2013. With extensive experience as a DevOps Engineer, he specializes in Continuous Integration/Continuous Deployment (CI/CD) Kubernetes and holds multiple certifications in DevOps. An active member of the IEEE, his research interests are broad and include AI, ML, IoT, blockchain technology, human-computer interaction (HCI) with AI, quantum computing, and cyber security. Dr. Meduri is actively engaged in research across these domains. He can be contacted via email at: karthik.meduri@ieee.org.