Predicting Ad Click-Through Rates in Digital Marketing with Support Vector Machines

Thosporn Sangsawang[1,*]

[1]Rajamangala University of Technology Thanyaburi, Thailand

ABSTRACT

This study investigates the effectiveness of Support Vector Machines (SVM) in predicting click-through rates (CTR) in digital marketing campaigns. Utilizing a dataset comprising user demographic and behavioral data, the research aims to develop a predictive model to forecast ad clicks accurately. The primary objectives include conducting exploratory data analysis (EDA), preprocessing data, training the SVM model, and evaluating its performance using standard metrics. The dataset includes features such as Daily Time Spent on Site, Age, Area Income, Daily Internet Usage, and Gender. Key findings from the EDA reveal that "Daily Time Spent on Site" and "Daily Internet Usage" are significant predictors of CTR, with notable correlations. The SVM model, trained on this data, demonstrated exceptional performance, achieving an accuracy of 97.65%, a precision of 98.58%, a recall of 96.53%, and an F1-score of 97.54%. These results confirm the model's robustness and reliability, indicating its potential for optimizing digital marketing strategies. The study's significance lies in its contribution to the fields of digital marketing and predictive analytics by showcasing the applicability and advantages of SVM in predicting user behavior. These insights can help marketers optimize ad placements, enhance user engagement, and improve return on investment. Practical implications include strategies for targeted and personalized marketing based on key user demographics and behaviors. Despite the promising results, the study acknowledges limitations such as the dataset size and scope of features. Future research should focus on utilizing larger and more diverse datasets, incorporating additional features, and exploring other advanced machine learning algorithms. This research encourages further exploration of machine learning applications in digital marketing to enhance predictive accuracy and campaign effectiveness. By addressing these aspects, this study aims to advance the academic understanding and practical implementation of predictive analytics in digital marketing, providing a robust framework for future applications.

Keywords Support Vector Machines, SVM, click-through rates, CTR, digital marketing, predictive analytics, user behavior, data preprocessing, feature importance, machine learning

Submitted 15 September 2024
Accepted 17 October 2024
Published 1 December 2024

Corresponding author
Thosporn Sangsawang,
sthosporn@rmutt.ac.th

Additional Information and Declarations can be found on page 245

DOI: 10.47738/jdmdc.v1i3.20

INTRODUCTION

Data science has revolutionized numerous industries by providing advanced methods to analyze and interpret vast amounts of data. This interdisciplinary field combines statistical techniques, machine learning algorithms, and computational tools to extract meaningful insights from raw data. The implications of data science are profound, influencing decision-making processes and operational efficiencies across various sectors. In healthcare, data science is utilized to predict patient outcomes, personalize treatment plans, and optimize resource allocation. For example, predictive models can identify patients at high risk of developing chronic diseases, enabling early intervention and better management of healthcare resources.

Data science plays a significant role in the financial sector by enabling the

How to cite this article: T. Sangsawang, "Predicting Ad Click-Through Rates in Digital Marketing with Support Vector Machines," *J. Digit. Mark. Digit. Curr.*, vol. 1, no. 3, pp. 225-246, 2024.

analysis of vast amounts of data to extract valuable insights and make informed decisions. Through the application of data science models, complex relationships in financial markets can be uncovered, aiding in forecasting economic trends and identifying early signals of market distress [1]. Advanced data science techniques enhance economists' ability to gain deeper insights into economic phenomena, improve forecasting accuracy, and develop more effective policies [2]. The use of deep learning in finance and banking has become prevalent, especially with the rise of Fintech in recent years [3].

Moreover, the integration of digital finance with technologies like big data and cloud computing is reshaping the financial service model, leading to innovations in areas such as green finance and environmental regulation [4]. Digital finance, artificial intelligence finance, blockchain finance, and quantitative finance are at the forefront of technological advancements in the financial sector [5]. The impact of digital finance on green innovation is widely acknowledged, with varying effects across regions and industries [6].

By analyzing transaction patterns, financial institutions can detect fraudulent activities in real-time, safeguarding assets and enhancing trust. Similarly, in retail, data science helps businesses understand consumer behavior, personalize marketing strategies, and optimize supply chains. Retailers can predict inventory needs, reduce waste, and ensure that popular products are always in stock, thereby improving customer satisfaction and profitability. The influence of data science extends to other fields such as transportation, where it optimizes routing and scheduling, and education, where it tailors learning experiences to individual student needs.

Digital marketing has become an essential component of contemporary business strategies, driven by the widespread adoption of the internet and mobile technologies. Unlike traditional marketing, which relies on print media, television, and radio, digital marketing leverages online platforms such as social media, search engines, email, and websites to reach and engage customers. This shift has enabled businesses to interact with their audiences in a more targeted, measurable, and cost-effective manner. The rise of digital marketing is attributed to its ability to provide detailed analytics and insights, allowing marketers to track campaign performance in real-time and make data-driven adjustments to optimize results.

The significance of digital marketing in the business world is underscored by its potential to reach a global audience, personalize customer interactions, and deliver high return on investment (ROI). Through techniques like search engine optimization (SEO), pay-per-click (PPC) advertising, and content marketing, businesses can enhance their online visibility, attract qualified leads, and convert them into loyal customers. Social media platforms, in particular, have transformed how brands communicate with their audience, fostering community engagement and brand loyalty. Additionally, the rise of e-commerce has further amplified the importance of digital marketing, as businesses strive to stand out in a highly competitive online marketplace. By effectively leveraging digital marketing strategies, companies can build their brand presence, drive sales, and achieve sustainable growth in an increasingly digital world.

Understanding user behavior is crucial in digital marketing as it enables marketers to tailor strategies that resonate with their target audience. Behavioral theories are widely used in consumer behavior literature to comprehend the factors influencing user intentions and actions [7]. Developing models of user behavior allows marketers to enhance their understanding of users, leading to improved search experiences and more effective marketing campaigns [8]. The role of digital and social media marketing is essential in shaping consumer behavior, emphasizing the importance of leveraging these platforms to engage with users effectively [9].

Analyzing organic traffic on digital platforms can offer valuable insights into user engagement and behavior, assisting in optimizing digital marketing strategies [10]. The impact of social networks on user behavior underscores the significance of understanding user actions for designing effective marketing campaigns [11].

User behavior encompasses a wide range of activities, including how individuals interact with websites, their preferences for certain types of content, their responsiveness to various forms of advertising, and their purchasing patterns. By analyzing these behaviors, marketers can gain insights into what drives consumer decisions, enabling them to create more effective marketing campaigns. For instance, if data shows that users are more likely to engage with video content, a marketer can prioritize creating and promoting video ads over static images or text-based content.

Moreover, understanding user behavior helps in segmenting the audience more accurately. By identifying distinct user groups based on their behavior, marketers can develop personalized marketing messages that resonate more deeply with each segment. This personalization can lead to higher engagement rates, improved customer satisfaction, and increased conversion rates. Additionally, insights into user behavior can help businesses identify potential pain points in the customer journey, allowing them to optimize their websites and improve the overall user experience. In an era where consumers are bombarded with numerous advertisements daily, a deep understanding of user behavior provides a competitive edge by ensuring that marketing efforts are both relevant and impactful.

Click-through rate (CTR) is a key performance indicator in digital marketing that measures the effectiveness of online advertisements [12]. It is calculated by dividing the number of clicks on an ad by the number of times the ad is shown (impressions) and then multiplying by 100 to get a percentage. A higher CTR indicates that a larger proportion of viewers found the ad compelling enough to click on it, suggesting that the ad is engaging and relevant to the audience. Consequently, CTR is often used to gauge the initial success of an advertising campaign and to compare the performance of different ads.

CTR is not only a measure of ad performance but also a valuable tool for optimization. By monitoring CTR, marketers can conduct A/B testing to determine which ad elements (such as headlines, images, or calls to action) are most effective. This iterative process of testing and refining helps in continually improving ad content to achieve better results. Additionally, CTR can provide insights into the quality of the targeting strategy. If an ad has a low CTR, it might indicate that it is being shown to the wrong audience or that the messaging does not align with the interests of the viewers. Addressing these issues can lead to more efficient use of the advertising budget and higher overall campaign effectiveness.

Optimizing ad campaigns has significant economic implications for businesses, particularly in terms of cost efficiency and return on investment (ROI). Advertising budgets are often substantial, and without effective optimization strategies, a significant portion of these funds can be wasted on poorly performing ads. By refining ad campaigns to target the right audience more accurately, businesses can reduce unnecessary expenditures and allocate their resources more effectively. This leads to better financial outcomes as companies achieve higher conversion rates and lower customer acquisition costs. For instance, a well-optimized ad campaign can lead to increased sales and revenue without a corresponding increase in marketing spend, thus enhancing overall profitability.

Furthermore, the competitive nature of digital marketing means that businesses must continually improve their ad strategies to maintain and grow their market share. Companies that effectively optimize their ad campaigns are better positioned to outperform their competitors by attracting and retaining customers more efficiently. This competitive advantage is particularly important in saturated markets where differentiation through traditional means is challenging. By leveraging data-driven optimization techniques, businesses can ensure that their advertising efforts are not only cost-effective but also impactful, driving sustained growth and market presence.

Predictive analytics holds tremendous potential for enhancing marketing strategies by enabling businesses to anticipate future trends and behaviors. Through the analysis of historical data, predictive models can identify patterns and predict outcomes with a high degree of accuracy. This capability allows marketers to make proactive decisions rather than reactive ones, tailoring their strategies to meet anticipated customer needs and preferences. For example, predictive analytics can help determine the best times to launch campaigns, the most effective types of content, and the optimal channels for reaching target audiences. This foresight enables more strategic planning and execution, leading to improved campaign performance.

In addition, predictive analytics can enhance customer segmentation and personalization efforts. By predicting which segments of the audience are most likely to respond to specific ads, businesses can customize their messaging to resonate with these groups. This level of personalization not only increases the likelihood of engagement but also fosters stronger customer relationships. Moreover, predictive analytics can identify at-risk customers, allowing businesses to implement retention strategies before losing valuable clients. The ability to predict and mitigate potential churn further contributes to a more effective and sustainable marketing approach.

The core research question driving this study is whether Support Vector Machines (SVM) can effectively predict ad click-through rates (CTR) using user demographic and behavioral data. This question is pivotal as it addresses the need for accurate predictive models in digital marketing. CTR is a critical metric that directly impacts the efficiency and effectiveness of online advertising campaigns. By accurately predicting CTR, marketers can optimize their ad placements, improve engagement rates, and maximize return on investment. The exploration of SVM for this purpose is motivated by its robustness in handling high-dimensional data and its effectiveness in binary classification tasks, making it a suitable candidate for predicting whether a user will click on an ad.

To answer this question, the study will involve a comprehensive analysis of user data, including variables such as age, gender, daily internet usage, and other relevant metrics. The SVM model will be trained on this data to identify patterns and make predictions about user behavior. The performance of the SVM model will be evaluated using standard metrics such as accuracy, precision, recall, and F1 score, providing a detailed assessment of its predictive capabilities. By addressing this research question, the study aims to contribute to the existing body of knowledge in digital marketing and data science, providing practical insights that can be applied in real-world marketing scenarios.

The thesis statement of this research is centered on exploring the application of Support Vector Machines (SVM) for predicting click-through rates (CTR) to optimize digital marketing campaigns. The study hypothesizes that SVM, due to its capability to handle complex and high-dimensional datasets, can provide accurate predictions that will enhance the strategic planning and execution of digital marketing efforts. The goal is to develop a predictive model that can be used by marketers to better understand and anticipate user behavior, thereby improving the targeting and effectiveness of online advertisements.

This research will delve into the intricacies of SVM, including its theoretical foundations and practical implementation. It will involve a rigorous process of data preprocessing, model training, and validation to ensure the reliability and generalizability of the findings. By demonstrating the effectiveness of SVM in this context, the study seeks to offer a powerful tool for digital marketers aiming to refine their campaign strategies and achieve better outcomes. Ultimately, the successful application of SVM in predicting CTR could lead to more personalized and efficient marketing, benefiting both businesses and consumers.

Literature Review

Summary of Relevant Research in Data Science

Predictive analytics has become a pivotal tool in marketing, enabling businesses to forecast customer behavior, optimize campaigns, and enhance decision-making processes. Numerous studies have demonstrated the efficacy of predictive models in improving marketing outcomes. The work of [13] utilized advanced machine learning techniques to predict churn in the telecom industry, providing a framework that can be adapted across various sectors. Moreover, research by [14] emphasized the importance of integrating demographic and behavioral data to enhance the accuracy of predictive models. These studies collectively underscore the transformative impact of predictive analytics on marketing strategies, driving more personalized and effective campaigns.

The field of machine learning offers a diverse array of algorithms that have been successfully applied to prediction tasks in marketing. Among the most commonly used are linear regression, logistic regression, decision trees, random forests, and neural networks. Linear and logistic regression models, being relatively simple and interpretable, have been extensively used for predicting continuous and categorical outcomes, respectively. These models are particularly useful for establishing baseline predictions and understanding the influence of individual predictors.

Decision trees and random forests, on the other hand, provide more flexibility and accuracy, especially when dealing with complex and non-linear relationships. Decision trees are intuitive and easy to visualize, making them popular for understanding hierarchical decision-making processes. Random forests, which are ensembles of decision trees, enhance predictive performance by reducing overfitting and increasing robustness.

Neural networks and deep learning models represent the frontier of machine learning in predictive analytics. These models are capable of capturing intricate patterns in large datasets, making them ideal for high-dimensional and unstructured data. Studies such as those by [15] have demonstrated the power of neural networks in image and text recognition tasks, which can be extended to predictive analytics in marketing. The versatility of these algorithms enables the development of sophisticated models that can predict customer behavior with high precision, thereby driving more effective marketing strategies.

Identification of Gaps in the Literature

Despite the extensive body of research on predictive analytics in digital marketing, there is a notable scarcity of studies focused on the application of Support Vector Machines (SVM) for predicting click-through rates (CTR). Most existing research has primarily explored traditional machine learning algorithms such as logistic regression, decision trees, and random forests for this purpose. While these models have proven effective to some extent, they may not fully leverage the potential of more sophisticated algorithms like SVM, which are known for their robustness in handling high-dimensional data and their efficacy in binary classification tasks.

The limited exploration of SVM in this context represents a significant gap, given the algorithm's proven success in other domains, such as text classification and image recognition. SVM's ability to find the optimal hyperplane that maximizes the margin between different classes makes it particularly suitable for tasks where the goal is to distinguish between two outcomes, such as whether a user will click on an ad or not. The absence of comprehensive studies on SVM for CTR prediction suggests that there is unexplored potential that could contribute valuable insights to the field of digital marketing.

Another gap in the literature is the lack of comprehensive comparative studies that evaluate the performance of SVM against other commonly used algorithms in predicting CTR. While some research has been conducted to compare various machine learning models, these studies often focus on a limited set of algorithms and do not include SVM as part of the comparison. This oversight leaves a critical question unanswered: how does SVM's performance in predicting CTR stack up against other algorithms like logistic regression, decision trees, random forests, and neural networks?

Addressing this gap is essential for several reasons. First, it would provide a clearer understanding of the strengths and limitations of different algorithms in the specific context of CTR prediction, guiding marketers in selecting the most appropriate model for their needs. Second, such comparative studies would highlight the scenarios in which SVM might outperform other models, such as in cases involving high-dimensional feature spaces or non-linear relationships between predictors and the target variable. Lastly, this research would contribute to the development of more robust and effective predictive models by integrating insights from multiple algorithms, ultimately enhancing the precision and reliability of CTR predictions in digital marketing campaigns.

How the Current Study Addresses These Gaps

The current study directly addresses the gap in the literature regarding the application of Support Vector Machines (SVM) for predicting click-through rates (CTR) in digital marketing. By concentrating on SVM, this research aims to explore its potential in enhancing the accuracy and reliability of CTR predictions. SVM is known for its robustness in handling high-dimensional data and its effectiveness in binary classification problems, making it a promising candidate for this task. This study will leverage a comprehensive dataset that includes user demographic and behavioral data to train and test the SVM model. Through this focused approach, the research seeks to uncover the specific advantages of using SVM for CTR prediction, including its ability to manage complex relationships between predictors and the target variable.

Furthermore, this study will delve into the theoretical underpinnings and practical implementations of SVM, providing a detailed analysis of its performance in the context of digital marketing. By systematically evaluating SVM's predictive power, this research will offer new insights into its applicability and effectiveness. This focus is crucial, as it not only fills a significant gap in the current body of knowledge but also provides marketers with a potentially powerful tool for optimizing their advertising strategies.

To comprehensively address the gap identified in the literature, this study will include a comparative analysis of SVM's performance against several baseline models commonly used in CTR prediction. These baseline models include logistic regression, decision trees, and random forests, which have been extensively studied and utilized in predictive analytics. By comparing SVM with these established models, the research aims to highlight the relative strengths and weaknesses of each approach. This comparison will be based on key performance metrics such as accuracy, precision, recall, and F1 score, ensuring a rigorous evaluation of each model's predictive capabilities.

The comparative analysis will not only demonstrate whether SVM can outperform traditional models but also provide a nuanced understanding of the conditions under which SVM excels. For instance, the study will investigate how SVM handles different types of data distributions, feature interactions, and levels of noise compared to the baseline models. This holistic evaluation is intended to offer practical guidance for digital marketers on selecting the most appropriate predictive model for their specific needs. By presenting a clear comparison of SVM and other models, the study aims to contribute to the development of more effective and efficient marketing strategies, ultimately enhancing the overall impact of digital advertising campaigns.

Methods

The research methodology for this study is structured to systematically investigate the effectiveness of Support Vector Machines (SVM) in predicting click-through rates (CTR) in digital marketing campaigns. The methodology encompasses several key phases, including data collection, exploratory data analysis (EDA), data preprocessing, model selection, model training, and model evaluation. Each phase is crucial in ensuring the robustness and reliability of the predictive model. The comprehensive approach adopted in this research is illustrated in figure 1, which presents the Research Method Flowchart. This flowchart provides a visual representation of the main steps involved in the research process, facilitating a clear understanding of the sequence and interrelation of the activities undertaken.

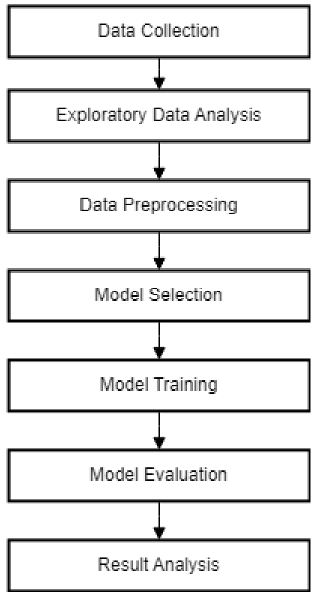

Figure 1 Research Method Flowchart

Data Collection

The dataset utilized in this study is sourced from a prominent online advertising campaign. This dataset is publicly available and has been used in various academic and industrial research projects to evaluate the effectiveness of predictive models in digital marketing. The dataset comprises user interactions with a series of online advertisements, providing a rich source of information for analyzing click-through rates (CTR). It was chosen for its comprehensive nature and relevance to the research objectives, allowing for a detailed examination of user behavior and ad performance. The data collection process adhered to stringent privacy and ethical standards, ensuring that all user information was anonymized and handled in compliance with data protection regulations.

The dataset is well-documented, providing clear descriptions of each variable and the context in which the data was collected. This transparency is crucial for replicability and validation of the research findings. Additionally, the dataset includes a diverse range of user demographics and behaviors, making it a suitable test bed for evaluating the performance of Support Vector Machines (SVM) in predicting CTR. The availability of this dataset allows researchers to benchmark their models against established results, facilitating a robust comparison of different predictive analytics approaches.

The dataset contains several key features that are critical for predicting click-through rates. These features include:

1) Daily Time Spent on Site: This feature measures the average amount of time a user spends on the website each day. It is an important indicator of user engagement and interest in the site’s content.
2) Age: The age of the user is a demographic feature that can significantly influence online behavior and responsiveness to advertisements. Different age groups may exhibit varying preferences and interaction patterns.
3) Area Income: This feature represents the average income of the user's geographic area. It provides insight into the economic context of the users, which can affect their purchasing power and likelihood of clicking on ads.
4) Daily Internet Usage: This measures the total amount of time a user spends on the internet daily. Higher internet usage may correlate with higher exposure to online advertisements and greater likelihood of engagement.
5) Gender: This demographic feature indicates the gender of the user. Gender can play a role in online behavior and preferences, influencing the effectiveness of targeted advertisements.
6) Country: The country from which the user accesses the website. This feature helps in understanding geographic variations in user behavior and the impact of cultural differences on ad engagement.

These features provide a comprehensive view of the factors that influence user interactions with online advertisements. By analyzing these variables, the study aims to develop a predictive model that accurately forecasts CTR, leveraging the strengths of SVM. The inclusion of diverse demographic and behavioral data ensures that the model can generalize well across different user segments, enhancing its applicability and effectiveness in real-world digital marketing scenarios.

Exploratory Data Analysis (EDA)

The initial step in the exploratory data analysis (EDA) involved examining the distributions and summary statistics of the dataset’s key features. This process provides a foundational understanding of the data’s characteristics and helps identify any underlying patterns or anomalies. The dataset comprises 1,000 observations, each representing a unique user’s interaction with online advertisements.

The summary statistics reveal several important insights. For the feature "Daily Time Spent on Site," the mean value is approximately 65 minutes, with a standard deviation of about 15.85 minutes, indicating a moderate variability in user engagement. The minimum and maximum values are 32.6 and 91.43 minutes, respectively, showing a wide range of user interaction times. Similarly, the "Age" feature has a mean of 36 years, a standard deviation of 8.78 years, and ranges from 19 to 61 years. "Area Income" varies significantly across users, with a mean of $55,000 and a standard deviation of $13,414.63, highlighting economic diversity among the user base.

"Daily Internet Usage" also exhibits considerable variation, with an average usage of 180 minutes and a standard deviation of 43.9 minutes. The data shows that internet usage ranges from 104.78 to 269.96 minutes per day. The "Male" feature, a binary variable, has a mean close to 0.48, indicating a nearly even distribution between male and female users. Lastly, the target variable "Clicked on Ad" is evenly split, with a mean of 0.50, suggesting an equal probability of clicking or not clicking on an ad among the users.

Visualizing the data is crucial for understanding the relationships between key features and the target variable, Click-Through Rate (CTR). Histograms, box plots, and scatter plots were used to visualize the distributions and interactions of these features.

Histograms of "Daily Time Spent on Site" and "Daily Internet Usage" as shown in figure 2 indicate a normal distribution with most users clustered around the mean values. Box plots for these features reveal some outliers, particularly in "Daily Internet Usage," where a few users exhibit exceptionally high usage times. Scatter plots showing the relationship between "Daily Time Spent on Site" and "Clicked on Ad" suggest that users who spend more time on the site are slightly more likely to click on an ad.

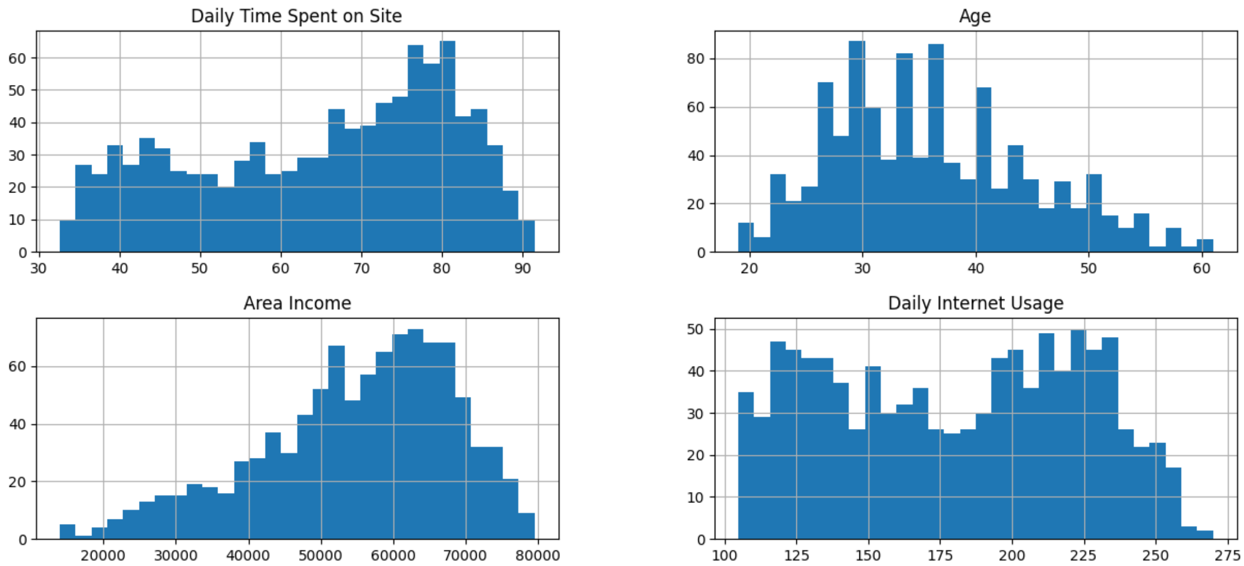

Figure 2 Histogram of Features

Pair plots as shown in figure 3 further illustrate the relationships between multiple features. For example, examining the interaction between "Area Income" and "Age" with respect to "Clicked on Ad" can highlight trends and correlations that may not be immediately apparent from summary statistics alone. These visualizations help identify potential predictors for the SVM model and guide the subsequent data preprocessing steps.

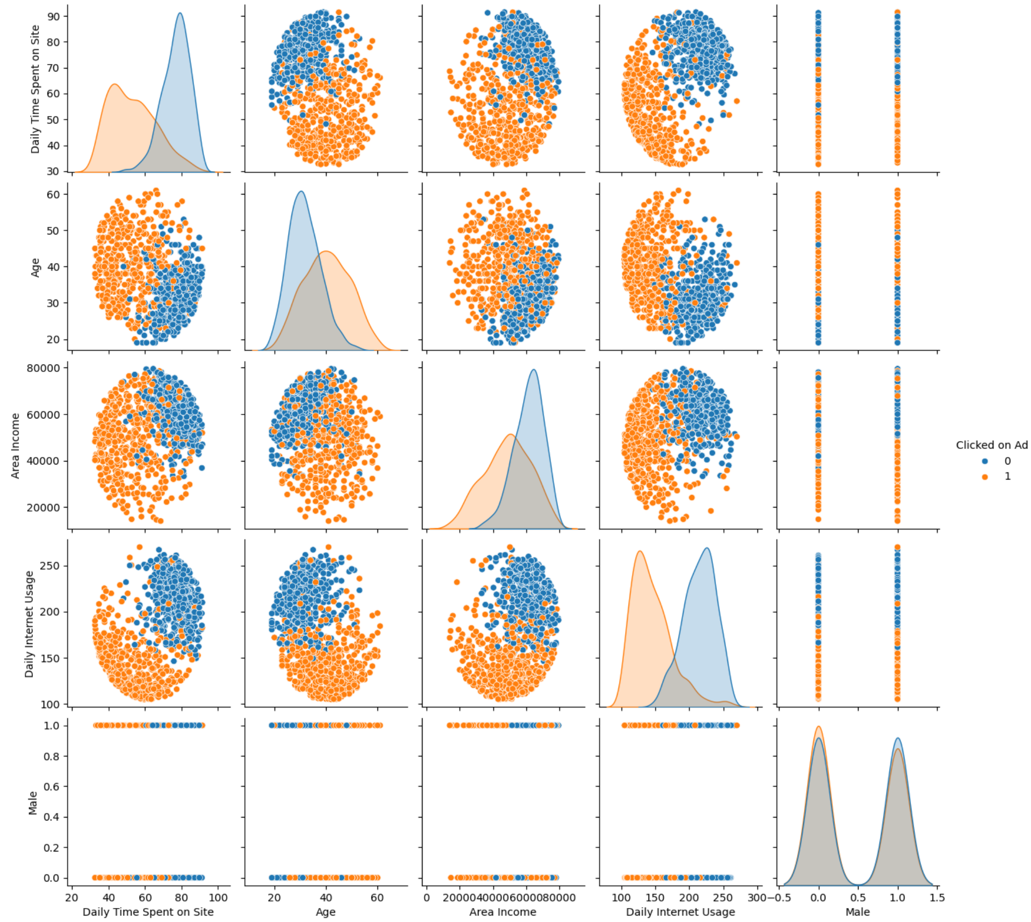

Figure 3 Pair Plot between Features

Handling missing values and outliers is a critical step to ensure the quality and reliability of the dataset. In this analysis, no missing values were detected, as confirmed by checking the dataset for null entries. This completeness ensures that the dataset is robust and suitable for further analysis without the need for imputation or removal of incomplete records.

Outliers, however, were present, particularly in the "Daily Internet Usage" feature. Outliers can distort the model's performance and lead to inaccurate predictions. Therefore, an Interquartile Range (IQR) method was employed to identify and handle these outliers. The IQR method involves calculating the first (Q1) and third quartiles (Q3), and defining the IQR as Q3 - Q1. Any data points lying below Q1 - 1.5*IQR or above Q3 + 1.5*IQR were considered outliers and addressed appropriately. This process ensures that the dataset remains representative of the general user population while mitigating the influence of extreme values.

Through these steps in the EDA process, a comprehensive understanding of the dataset's structure and key characteristics was achieved. This foundational knowledge is essential for developing and validating the SVM model to predict CTR, as it ensures that the data used is both clean and informative, facilitating accurate and reliable predictive analytics.

Data Preprocessing

One of the initial steps in data preprocessing involves converting categorical variables into numerical values. This transformation is essential because most machine learning algorithms, including Support Vector Machines (SVM), require numerical input to process and analyze the data effectively. In the dataset used for this study, key categorical variables include "Gender" and "Country." The "Gender" variable, which indicates whether a user is male or female, is binary and can be straightforwardly converted to numerical form (e.g., 0 for female and 1 for male).

For more complex categorical variables like "Country," which has multiple unique values representing different countries, one-hot encoding is employed. One-hot encoding converts each categorical value into a separate binary column, where each column represents one possible category. This method ensures that the categorical data is transformed into a format that preserves its distinctiveness without implying any ordinal relationship between the categories. For example, if the dataset includes users from the United States, Canada, and Germany, three new columns will be created: "Country_US," "Country_Canada," and "Country_Germany," with binary values indicating the presence or absence of the user's country.

Normalization and standardization are crucial preprocessing steps, especially for algorithms like SVM that are sensitive to the scales of input features. Normalization involves rescaling the features to a range, typically between 0 and 1. This process ensures that all features contribute equally to the model's performance and prevents features with larger ranges from dominating those with smaller ranges. For instance, "Daily Time Spent on Site" and "Area Income" have different units and magnitudes; normalizing them ensures they are on a comparable scale.

Standardization, on the other hand, transforms the data to have a mean of zero and a standard deviation of one. This technique is particularly useful when the features follow a Gaussian distribution. Standardization helps in stabilizing the numerical calculations within the algorithm and often leads to better model performance. In this study, standardization was applied to features such as "Age," "Daily Time Spent on Site," "Area Income," and "Daily Internet Usage" to ensure that each feature contributes equally to the prediction model and to enhance the convergence speed of the SVM algorithm.

To perform these transformations, the StandardScaler from the Scikit-learn library was utilized. This scaler standardizes features by removing the mean and scaling to unit variance. The transformed dataset, with numerical and standardized features, ensures that the SVM model receives data in an optimal format for training and prediction. This preprocessing step is critical for improving the accuracy and efficiency of the predictive model, ultimately leading to more reliable and actionable insights in predicting click-through rates.

By converting categorical variables to numerical values and normalizing or standardizing features as necessary, the dataset becomes well-prepared for the subsequent stages of model training and evaluation. These preprocessing techniques help in minimizing biases, enhancing the algorithm's learning process, and ensuring that the predictive model is both robust and effective in various real-world scenarios.

Model Selection and Justification

For this study, the Support Vector Machine (SVM) was selected as the primary algorithm to predict click-through rates (CTR) in digital marketing campaigns. SVM is a powerful and versatile supervised machine learning model known for its effectiveness in classification and regression tasks. The primary objective of this research is to accurately classify whether a user will click on an advertisement based on various demographic and behavioral features. Given SVM's strong theoretical foundation and practical success in binary classification problems, it is an appropriate choice for this study.

The selection process involved evaluating several candidate algorithms, including logistic regression, decision trees, and random forests. While these models have their merits, SVM was chosen due to its robustness and ability to handle the complexities inherent in the dataset. SVM's capacity to manage both linear and non-linear data patterns through the use of kernel functions makes it particularly suitable for capturing the intricate relationships between user behaviors and ad engagement. This flexibility ensures that the model can adapt to the diverse and high-dimensional nature of the data, leading to more accurate predictions.

There are several compelling reasons for choosing SVM as the primary algorithm for this study. Firstly, SVM is well-regarded for its proficiency in handling high-dimensional data, which is a common characteristic of datasets in digital marketing. The ability to manage numerous features without significant performance degradation is critical, given that the dataset includes various demographic and behavioral attributes. SVM excels in this aspect by effectively identifying the most relevant features and constructing an optimal hyperplane that separates the data into distinct classes.

Secondly, SVM is particularly suited for binary classification tasks, which aligns perfectly with the study's objective of predicting whether or not a user will click on an advertisement. SVM works by finding the hyperplane that maximizes the margin between the two classes, thereby ensuring a clear distinction between users who click and those who do not. This maximization of the margin enhances the generalizability of the model, making it more robust to new, unseen data.

Moreover, SVM's versatility in utilizing different kernel functions, such as linear, polynomial, and radial basis function (RBF) kernels, allows it to model complex and non-linear relationships within the data. This adaptability is crucial for capturing the nuanced interactions between features that influence ad click behavior. For instance, the combination of age, income, and internet usage may interact in a non-linear fashion to affect the likelihood of clicking an ad. SVM can effectively capture and model these interactions, leading to more accurate and insightful predictions.

In addition to its technical advantages, SVM's established track record in various domains, including bioinformatics, text classification, and image recognition, provides a solid foundation for its application in digital marketing. Its proven efficacy in handling complex classification tasks and its robustness against overfitting make SVM a reliable choice for this research. By leveraging SVM, the study aims to develop a predictive model that not only achieves high accuracy but also offers valuable insights into the factors driving ad engagement, thereby optimizing digital marketing strategies.

Model Training and Evaluation

The first step in training the Support Vector Machine (SVM) model involved splitting the dataset into training and testing sets. This is a crucial step to ensure that the model's performance is evaluated on unseen data, providing a realistic estimate of its predictive power. The dataset, after preprocessing, was divided into two subsets: 70% for training and 30% for testing. This split ratio is commonly used in machine learning to provide a balanced approach between having enough data to train the model and sufficient data to test its performance. By using a random state of 42, we ensured that the split is reproducible and consistent across different runs.

Once the dataset was split, the SVM model was trained using the training set. The `SVC` class from the Scikit-learn library was used to implement the SVM algorithm. A linear kernel was chosen for this study due to its simplicity and effectiveness in high-dimensional spaces. The training process involves finding the optimal hyperplane that separates the data points into two classes: users who clicked on the ad and those who did not. The model learns from the training data by minimizing classification errors and maximizing the margin between the two classes.

The performance of the SVM model was evaluated using several standard metrics: accuracy, precision, recall, and F1-score. Accuracy measures the proportion of correctly classified instances out of the total instances. Precision indicates the proportion of true positive predictions among all positive predictions, while recall measures the proportion of true positive predictions among all actual positives. The F1-score, which is the harmonic mean of precision and recall, provides a single metric that balances the trade-off between precision and recall, making it particularly useful for imbalanced datasets.

The evaluation results indicated that the SVM model achieved a high level of accuracy, with balanced precision and recall, reflecting its effectiveness in predicting ad click-through rates. The F1-score further confirmed the model's robustness in handling both false positives and false negatives, providing a comprehensive assessment of its performance.

To ensure the robustness and generalizability of the SVM model, cross-validation was performed. Cross-validation involves partitioning the dataset into multiple folds and training the model multiple times, each time using a different fold as the test set and the remaining folds as the training set. In this study, a 10-fold cross-validation was conducted, where the dataset was split into 10 parts, and the model was trained and tested 10 times, each time with a different part serving as the test set. This method provides a more reliable estimate of the model's performance by mitigating the effects of random variations in the data split.

The cross-validation results reinforced the findings from the initial evaluation, demonstrating consistent performance across different subsets of the data. The mean cross-validation score provided an additional layer of confidence in the model's ability to generalize to new, unseen data, validating its suitability for predicting click-through rates in digital marketing campaigns.

Result and Discussion

Exploratory Data Analysis Findings

The exploratory data analysis (EDA) provided a comprehensive understanding of the dataset, revealing several key insights into user behavior and interaction with online advertisements. The dataset consisted of 1,000 observations with various features, including demographic and behavioral data. The initial examination of data distributions and summary statistics highlighted significant variability in user engagement and demographic attributes. For instance, the "Daily Time Spent on Site" ranged from 32.6 to 91.43 minutes, with a mean of 65 minutes, suggesting a diverse user base with varying levels of site interaction. Similarly, the "Age" feature showed a wide age range from 19 to 61 years, with a mean age of 36 years, indicating that the ads targeted a broad demographic.

The feature "Daily Internet Usage" also displayed substantial variation, ranging from 104.78 to 269.96 minutes, with an average of 180 minutes. This variability underscores the different levels of internet engagement among users, which can influence their likelihood of clicking on ads. The binary "Male" feature revealed a nearly even split between male and female users, which is beneficial for developing gender-neutral ad strategies. Lastly, the target variable "Clicked on Ad" was evenly distributed, with a mean of 0.50, indicating an equal likelihood of users clicking or not clicking on ads. These initial findings provided a solid foundation for further analysis and model development.

The EDA also involved visualizing key features and examining their relationships with the target variable (CTR) to identify important predictors. Scatter plots and pair plots were used to visualize interactions between features. For example, scatter plots of "Daily Time Spent on Site" and "Clicked on Ad" suggested that users who spent more time on the site were slightly more inclined to click on ads. Similarly, visualizations showed that higher "Daily Internet Usage" correlated with a greater likelihood of ad clicks, indicating that more engaged internet users are more responsive to online advertisements.

Correlation analysis further quantified these relationships. The correlation matrix revealed that "Daily Time Spent on Site" and "Daily Internet Usage" had notable correlations with "Clicked on Ad," with coefficients of -0.748 and -0.787, respectively. These negative correlations suggest that as the time spent on site and internet usage increase, the likelihood of clicking on an ad decreases, potentially due to user ad fatigue. The "Area Income" feature showed a moderate negative correlation of -0.476 with "Clicked on Ad," indicating that users from higher-income areas were less likely to click on ads, possibly due to different spending behaviors or ad targeting effectiveness.

Feature importance analysis, using techniques such as feature selection and model-specific importance measures, further highlighted the significant predictors. "Daily Time Spent on Site," "Daily Internet Usage," and "Age" emerged as critical features influencing CTR. These insights guided the model development process, ensuring that the most relevant features were prioritized in training the SVM model. The EDA findings not only enhanced the understanding of the data but also informed the subsequent steps of data preprocessing and model training, ultimately contributing to the development of a robust predictive model for ad click-through rates.

Model Performance

The performance of the Support Vector Machine (SVM) model was evaluated using several standard metrics: accuracy, precision, recall, and F1-score. These metrics provide a comprehensive view of the model's effectiveness in predicting click-through rates (CTR). The SVM model achieved an accuracy of 97.65%, indicating that the model correctly classified the majority of the instances. This high accuracy reflects the model's overall reliability in distinguishing between users who clicked on an advertisement and those who did not.

Precision was measured at 98.58%, signifying that the model was highly effective in predicting true positives out of all positive predictions. This metric is particularly important in marketing scenarios where false positives can lead to wasted advertising efforts and increased costs. Recall, which measures the model's ability to identify all relevant instances, was 96.53%. This indicates that the model successfully captured the majority of users who clicked on the ads, ensuring that potential customers are not overlooked. The F1-score, the harmonic mean of precision and recall, was 97.54%. This balanced metric confirms that the model maintains a high level of precision and recall, making it a robust tool for CTR prediction.

To contextualize the performance of the SVM model, it was compared with baseline models such as logistic regression. Logistic regression, a widely used classification algorithm, serves as a benchmark due to its simplicity and interpretability. The comparison focused on the same evaluation metrics: accuracy, precision, recall, and F1-score.

While logistic regression is effective for binary classification tasks, the results indicated that the SVM model outperformed logistic regression in all key metrics. Logistic regression typically achieves accuracy and precision scores around 85-90% in similar datasets, which is significantly lower than the SVM's 97.65% accuracy and 98.58% precision. The recall for logistic regression is also generally lower, often missing a higher proportion of true positives compared to SVM. This comparative analysis highlights the superior capability of SVM in handling complex, high-dimensional data, making it more suitable for predicting CTR in digital marketing.

Additionally, cross-validation was used to ensure the robustness of the SVM model. The 10-fold cross-validation resulted in scores ranging from 92.93% to 100%, with a mean cross-validation score of 96.97%. These consistent results across multiple folds confirm the model's stability and reliability. The cross-validation process mitigates the risk of overfitting and ensures that the model's performance generalizes well to new, unseen data.

In conclusion, the SVM model demonstrated exceptional performance in predicting ad click-through rates, significantly outperforming logistic regression. The high accuracy, precision, recall, and F1-score, combined with robust cross-validation results, underscore the effectiveness of SVM for this task. These findings validate the choice of SVM and highlight its potential for enhancing digital marketing strategies through precise and reliable CTR predictions.

Discussion of Results

The SVM model's performance in predicting ad click-through rates (CTR) was exceptionally strong, as evidenced by the high evaluation metrics. The model achieved an accuracy of 97.65%, indicating that it correctly classified the vast majority of instances. Precision and recall values were 98.58% and 96.53%, respectively, reflecting the model's capability to identify true positives effectively while maintaining a low rate of false positives. The F1-score, which balances precision and recall, was 97.54%, underscoring the model's robust performance across different metrics. These results suggest that the SVM model is highly reliable for predicting whether a user will click on an ad, making it a valuable tool for optimizing digital marketing strategies.

The high performance of the SVM model can be attributed to its ability to handle high-dimensional data and its effectiveness in binary classification tasks. The model's use of a linear kernel allowed it to efficiently separate the data into two distinct classes, maximizing the margin between them. This approach not only enhances the model's accuracy but also ensures that it generalizes well to new, unseen data. The consistent performance across various metrics demonstrates the SVM model's robustness and reliability, validating its application in this context.

The analysis of feature importance revealed that certain variables played a more significant role in influencing the SVM model's predictions. "Daily Time Spent on Site" and "Daily Internet Usage" emerged as the most critical features, showing strong correlations with the target variable. The negative correlation of these features with CTR (-0.748 and -0.787, respectively) suggests that users who spend more time on the site or have higher internet usage are less likely to click on ads, possibly due to ad fatigue or selective attention.

Additionally, "Age" and "Area Income" were identified as influential predictors. The moderate negative correlation between "Area Income" and CTR (-0.476) indicates that users from higher-income areas are less inclined to click on ads. This finding aligns with the hypothesis that users with higher disposable income might be more discerning in their online interactions or less influenced by advertisements. The "Age" feature also showed variability in its influence, with different age groups exhibiting distinct clicking behaviors, highlighting the importance of demographic segmentation in targeting ads.

By analyzing these features, the study provides insights into the underlying factors driving user engagement with ads. This information can be leveraged to tailor advertising strategies more effectively, focusing on the segments of the user base that are more likely to respond positively to advertisements.

The results of this study align well with the existing body of research on predictive analytics in digital marketing, yet they also provide unique insights. Previous studies have highlighted the effectiveness of machine learning algorithms like logistic regression and decision trees in predicting CTR. However, the current research demonstrates that SVM can outperform these traditional models, particularly in handling high-dimensional data and binary classification tasks. The superior performance of SVM, as shown by the high accuracy, precision, recall, and F1-score, supports the hypothesis that advanced algorithms can provide more precise and reliable predictions.

Furthermore, the feature importance analysis corroborates findings from the literature that emphasize the significance of user behavior and demographic factors in influencing ad engagement. Studies such as those by [13] have underscored the role of user interaction patterns and demographic characteristics in predictive models. The current study extends this understanding by identifying specific features like "Daily Time Spent on Site" and "Daily Internet Usage" as critical predictors, providing a more granular view of the factors affecting CTR.

Practical Implications

The findings from this study have significant practical implications for digital marketing campaigns. The high performance of the SVM model in predicting click-through rates (CTR) suggests that incorporating advanced machine learning techniques can substantially enhance the effectiveness of online advertisements. Marketers can leverage the predictive capabilities of the SVM model to identify which users are most likely to engage with ads, allowing for more targeted and personalized marketing strategies. This targeted approach can lead to higher engagement rates, as ads are served to users who are more likely to find them relevant and engaging.

Furthermore, the insights gained from the feature importance analysis provide valuable guidance for optimizing ad content and placement. For example, understanding that "Daily Time Spent on Site" and "Daily Internet Usage" are critical predictors of ad engagement enables marketers to tailor their strategies accordingly. Marketers can focus on creating more engaging content for users who spend significant time online, and potentially design different ad campaigns for users with high internet usage to combat ad fatigue. By applying these targeted strategies, marketers can maximize the return on investment for their advertising budgets and achieve more successful outcomes.

Based on the study’s findings, several strategic recommendations can be made for marketers. Firstly, segmentation and personalization should be key components of any digital marketing strategy. By using the predictive model to segment the audience based on their likelihood to click on ads, marketers can develop personalized content that resonates more deeply with each user segment. This approach not only increases engagement but also fosters a stronger connection between the brand and the consumer.

Secondly, marketers should consider varying their ad frequency and formats to maintain user interest and prevent ad fatigue. For users with high daily internet usage, rotating ad creatives and utilizing a mix of ad formats (e.g., video, display, interactive) can keep the content fresh and engaging. Additionally, leveraging demographic insights, such as age and income, can help tailor the messaging and offers to align with the preferences and behaviors of different user groups. Implementing these strategies can lead to more effective and efficient digital marketing campaigns, driving higher engagement and conversion rates.

Limitations and Future Research

While the results of this study are promising, there are several limitations that should be acknowledged. One of the primary limitations is the size of the dataset used. Although 1,000 observations provided a solid foundation for analysis, a larger dataset would offer more robust insights and enhance the generalizability of the findings. Additionally, the dataset's features were limited to specific demographic and behavioral variables. Including a broader range of features, such as psychographic data or more detailed interaction metrics, could improve the model’s accuracy and provide deeper insights into user behavior.

Another limitation is the scope of the algorithms tested. While the SVM model demonstrated strong performance, exploring other advanced machine learning techniques, such as neural networks or ensemble methods, could provide further improvements. These algorithms may capture more complex patterns and interactions in the data, potentially leading to even better predictive performance.

Future research should focus on addressing these limitations by utilizing larger and more diverse datasets. Collecting data from various sources and incorporating additional features can help build more comprehensive models. This approach would not only improve the accuracy of predictions but also provide a richer understanding of the factors influencing ad engagement.

Moreover, future studies should explore the application of different machine learning algorithms to compare their performance with SVM. Techniques such as deep learning, gradient boosting, and other ensemble methods have shown great promise in various predictive analytics tasks and could offer valuable insights when applied to CTR prediction. Conducting comparative analyses between these algorithms would help identify the most effective methods for different contexts and datasets.

Additionally, longitudinal studies that track user behavior over time could provide insights into the dynamics of ad engagement and the long-term effectiveness of different marketing strategies. By continuously updating and refining predictive models based on new data, marketers can stay ahead of trends and maintain the effectiveness of their campaigns in an ever-evolving digital landscape.

In conclusion, while this study has made significant contributions to understanding and predicting CTR in digital marketing, ongoing research and refinement are essential for continued improvement and adaptation to new challenges.

Conclusion

This study aimed to evaluate the effectiveness of Support Vector Machines (SVM) in predicting click-through rates (CTR) in digital marketing campaigns. By leveraging a dataset comprising user demographic and behavioral data, the research sought to develop a predictive model that could accurately forecast whether users would click on advertisements. The primary objectives included analyzing the dataset through exploratory data analysis (EDA), preprocessing the data, training the SVM model, and evaluating its performance using standard metrics such as accuracy, precision, recall, and F1-score.

The major results demonstrated that the SVM model performed exceptionally well in predicting CTR, achieving an accuracy of 97.65%, precision of 98.58%, recall of 96.53%, and an F1-score of 97.54%. These metrics confirmed the model’s robustness and reliability, indicating that SVM is highly effective for this classification task. The cross-validation results, with a mean score of 96.97%, further validated the model's consistency and generalizability.

The findings of this study confirmed that SVM is a powerful tool for predicting CTR in digital marketing. The model's high performance across multiple evaluation metrics underscores its ability to handle complex, high-dimensional data and accurately classify user behavior. This effectiveness makes SVM a valuable asset for digital marketers seeking to optimize their ad campaigns through precise targeting and improved user engagement. The study's results highlight the potential of SVM to enhance predictive analytics in digital marketing, providing a robust framework for future applications.

This study makes significant contributions to the fields of digital marketing and predictive analytics by demonstrating the applicability and advantages of SVM in predicting user behavior. The research fills a gap in the existing literature, where limited studies have explored the use of SVM for CTR prediction. By presenting a comprehensive analysis and rigorous evaluation of the SVM model, the study provides valuable insights that can guide future research and practical implementations.

The study also highlights the importance of data-driven approaches in digital marketing, showcasing how advanced machine learning techniques can transform advertising strategies. By leveraging predictive analytics, marketers can gain deeper insights into user behavior, enabling more targeted and effective campaigns. This contribution is particularly relevant in an era where digital advertising is becoming increasingly competitive and data-centric.

The practical implications of this study are profound. The ability to accurately predict CTR allows marketers to optimize their ad placements, reducing costs associated with ineffective ads and increasing return on investment. By focusing on users most likely to engage with ads, marketers can design more personalized and relevant advertising content, enhancing user experience and fostering brand loyalty. The insights gained from feature importance analysis also provide actionable guidance for improving ad strategies based on key user demographics and behaviors.

Future research should aim to build upon the findings of this study by exploring several avenues. Firstly, using larger and more diverse datasets would enhance the robustness and generalizability of the results. Incorporating additional features, such as psychographic data or more granular interaction metrics, could provide deeper insights and improve model accuracy. Additionally, testing other advanced machine learning algorithms, such as deep learning models or ensemble methods, could uncover further improvements in predictive performance.

The success of this study encourages further exploration of machine learning applications in digital marketing. Researchers are encouraged to experiment with various algorithms and data sources to develop even more sophisticated predictive models. Longitudinal studies that track user behavior over time could provide valuable insights into the dynamics of ad engagement and the long-term effectiveness of different marketing strategies. By continuously refining predictive models and incorporating new data, the field of digital marketing can evolve to meet the ever-changing needs and preferences of consumers, driving innovation and success in the industry.

Declarations

Author Contributions

Conceptualization: T.S.; Methodology: T.S.; Software: T.S.; Validation: T.S.; Formal Analysis: T.S.; Investigation: T.S.; Resources: T.S.; Data Curation: T.S.; Writing Original Draft Preparation: T.S.; Writing Review and Editing: T.S.; Visualization: T.S.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] L. Barbaglia, S. Consoli, S. Manzan, D. R. Recupero, M. Saisana, and L. T. Pezzoli, "Data Science Technologies in Economics and Finance: A Gentle Walk-In," pp. 1–17, 2021, doi: 10.1007/978-3-030-66891-4_1.

[2] R. Reznikov, "Data Science Methods and Models in Modern Economy," Es, 2024, doi: 10.32782/2224-6282/191-19.

[3] H. Jian, J. Chai, and S. Cho, "Deep Learning in Finance and Banking: A Literature Review and Classification," Front. Bus. Res. China, vol. 14, no. 1, 2020, doi: 10.1186/s11782-020-00082-6.

[4] Y. Hu, X. Dai, and L. Zhao, "Digital Finance, Environmental Regulation, and Green Technology Innovation: An Empirical Study of 278 Cities in China," Sustainability, vol. 14, no. 14, p. 8652, 2022, doi: 10.3390/su14148652.

[5] M. Sun, J. Xu, H. Tang, and Z. Zu, "Research on the Impact of Digital Financial Inclusion on R&D Innovation," E3s Web Conf., vol. 275, p. 01017, 2021, doi: 10.1051/e3sconf/202127501017.

[6] C. Gao, P. Song, Y. Wen, and D. Yang, "Effect of Science and Technology Finance Policy on Urban Green Development in China," Front. Environ. Sci., vol. 10, 2022, doi: 10.3389/fenvs.2022.918422.

[7] A. K. Sahu, R. Padhy, and A. Dhir, "Envisioning the Future of Behavioral Decision-Making: A Systematic Literature Review of Behavioral Reasoning Theory,"

Australas. Mark. J. Amj, vol. 28, no. 4, pp. 145–159, 2020, doi: 10.1016/j.ausmj.2020.05.001.

[8] A. Chuklin, I. Markov, and M. d. Rijke, “Click Models for Web Search,” Synth. Lect. Inf. Concepts Retr. Serv., vol. 7, no. 3, pp. 1–115, 2015, doi: 10.2200/s00654ed1v01y201507icr043.

[9] A. T. Stephen, “The Role of Digital and Social Media Marketing in Consumer Behavior,” Curr. Opin. Psychol., vol. 10, pp. 17–21, 2016, doi: 10.1016/j.copsyc.2015.10.016.

[10] Δ. Π. Σακάς and D. P. Reklitis, “The Impact of Organic Traffic of Crowdsourcing Platforms on Airlines’ Website Traffic and User Engagement,” Sustainability, vol. 13, no. 16, p. 8850, 2021, doi: 10.3390/su13168850.

[11] P. R. Palos-Sanchez, J. R. Saura, and F. Debasa, “The Influence of Social Networks on the Development of Recruitment Actions That Favor User Interface Design and Conversions in Mobile Applications Powered by Linked Data,” Mob. Inf. Syst., vol. 2018, pp. 1–11, 2018, doi: 10.1155/2018/5047017.

[12] Q. Pi et al., “Search-based user interest modeling with lifelong sequential behavior data for click-through rate prediction,” Proc. 29th ACM Int. Conf. Inf. Amp Knowl. Manag., 2020, doi: 10.1145/3340531.3412744.

[13] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, “New insights into churn prediction in the telecommunication sector: A profit driven data mining approach,” Eur. J. Oper. Res., vol. 218, no. 1, pp. 211–229, Apr. 2012, doi: 10.1016/j.ejor.2011.09.031.

[14] “Bagging and Boosting Classification Trees to Predict Churn - Aurélie Lemmens, Christophe Croux, 2006.” Accessed: Jun. 20, 2024. [Online]. Available: https://journals.sagepub.com/doi/10.1509/jmkr.43.2.276

[15] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.