Journal of Computer Science

Original Research Paper

Artificial Intelligence and Mathematical Modelling-Based Techniques to Improve Social Media Marketing

[1]Kiran Chaudhary, [2]Mohammad Naquibur Rahman, [3]Nabeela Hasan, [4]Mansaf Alam and [5]Aakash Punit

[1]*Department of Commerce, Shivaji College, University of Delhi, New Delhi, India*
[2]*Department of Marketing, Alasala College, Dammam, Saudi Arabia*
[3]*Department of Artificial Intelligence and Machine Learning, New Delhi Institute of Management, New Delhi, India*
[4]*Department of Computer Science, Jamia Millia Islamia, New Delhi, India*
[5]*Department of Commerce, Kirori Mal College, University of Delhi, New Delhi, India*

Article history
Received: 03-08-2024
Revised: 03-09-2024
Accepted: 16-09-2024

Corresponding Author:
Nabeela Hasan
Department of Artificial Intelligence and Machine Learning, New Delhi Institute of Management, New Delhi, India
Email: nabeela.hasan@ndimdelhi.org

Abstract: The Businessperson is using social media to do marketing of their product. Almost all people use social media. Most people relate to social media. Social media marketing is beneficial in making the product more popular and easy ways to introduce the product to everyone in economical ways. In this study, we introduce the concept of machine learning and mathematical models to enhance social media marketing for the product. We have used the concept of data analytics in this study to analyze social media data and based on the outcome of the analytics result, we developed strategies for marketing the product. The Bayesian Regularization (BR) Model has achieved the highest accuracy of 93% and the Root mean squire error is as low as 1.2 approximately in decision tree regression (DTR). The Area Under The Curve (AUC) of the model is .57. The model will be beneficial to improving social media marketing. The business personnel will benefit from marketing products on social media networks.

Keywords: Artificial Intelligence, Social Media Marketing, Machine Learning, Mathematical Modeling, Data Analytics

Introduction

Social media is now a way to reach every person in society. Social media plays an essential role in marketing the product. Social media marketing plays an essential role in popularizing the product. There are momentous challenges that stay alive from pessimistic electronic word-of-mouth and an interfering and frustrating online brand appearance. Using social media, individuals may interact and work with one another from afar. A constant stream of adverts from different microblogging and social networking channels influences customers' purchasing attitudes, as illustrated in Fig. (1) (Fuchs, 2008). Marketing professionals use social media to do business research, acquire information about a company, and promote a product. Because of the wide range of topics they explore on social media, the present global population of social media platforms is steadily growing (Drummond and French, 2008).

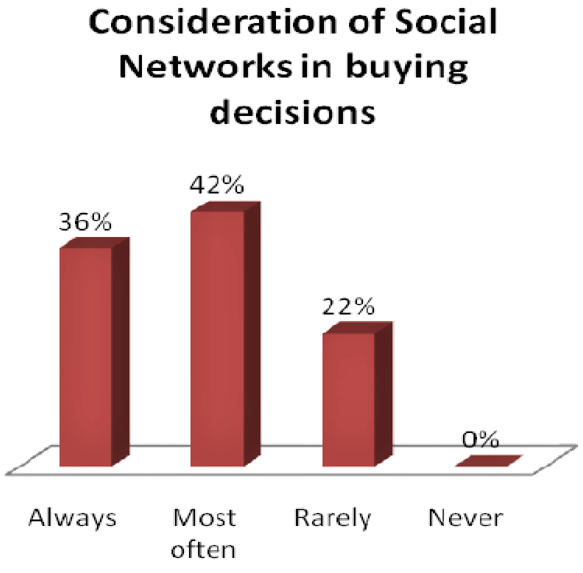

Fig. 1: Social media consideration in buying decisions (Bashar *et al*., 2012)

Several enterprises use analytical techniques, such as sentiment analysis, to further monitor and respond to what customers see about companies and their goods in digital marketing (Fan and Gordon, 2014). Even though there have been several studies on how to understand and predict the changes in the number of customer engagement behaviors between posts (Lee *et al.*, 2018), to the best of our comprehension, there has been no research on how to predict the most probable consumer engagement behavior in response to a marketing post. Based on the results of our analysis, we can determine the most popular kind of customer response to a post by looking at the distribution of votes inside and between posts that solicit responses. Machine learning techniques have an excellent performance in forecasting human choices in many applications (Kleinberg *et al.*, 2017; Mullainathan and Spiess, 2017; Plonsky *et al.*, 2019), which is essential because generating precise and accurate predictions about behaviors and actions is critical for humans and businesses to survive and obtain a competitive edge in changing settings (Nagar and Malone, 2011). For dynamic data analysis, businesses will rely primarily on data mining methods and Machine Learning (ML) tools in the future. Text mining methods such as gender prediction, text categorization, named entity identification, and sentiment analysis are all used in many areas, such as detecting fraudulent activities, share market forecasting, customer relations, and sentiment analysis (Pejić Bach *et al.*, 2019). Today, predictive maintenance incorporates business data to address needs and make choices (Hasan *et al.*, 2021). YouTube is the third most frequently visited website and the biggest platform for distributing video content online and second after Google in terms of search volume (Duffett *et al.*, 2019). Utilizing YouTube to promote products and services generates billions of dollars annually. Using this instance, it's evident that digital platforms play a significant part in Internet marketing. Machine learning technologies and marketing tactics may be used to increase results. Machine learning and statistical statistics often focus on forecasting and are inextricably intertwined (Hasan *et al.*, 2022).

Mathematical intelligence blends statistics and Artificial Intelligence (AI). ML is a branch of AI that learns from input data and generates insights as a model for making wise judgments when faced with unexpected test data (Bashar *et al.*, 2012). There are new social data analysis and unstructured data analysis levels in the microblogging age (Kottursamy *et al.*, 2016) that may identify recurring actions. The primary analytical tools are AI and ML. AI, Computer vision, and remote sensing (Jain *et al.*, 2000) are just a few of the many fields in which automatic (machine) pattern identification, description, categorization, and grouping are the latest research topics.

Our study led us to pose the research issues as follows:

1) What characteristics might be discovered to indicate the distinctions across postings in distinguishing behavioral responses?
2) How accurately might consumers' most probable engagement behavioral reaction options be predicted if these characteristics were used?

This research helps fill a gap in the literature on social media marketing and predicting consumer behavior. Companies might use our research to compare various advertising approaches. Companies should have well-defined marketing objectives for their content and use suitable design tactics to elicit the desired reactions from their target audiences. Some prediction work might be done before the marketing materials are released to the public for more excellent knowledge. This research focuses mainly on the multiple machine-learning approaches that may be used to anticipate client purchasing preferences through social media.

Literature Review

The article (Dwivedi *et al.*, 2021) has discussed the combined impending from more than a few of the most important digital and social media marketing experts. The motive of the paper (Arasu *et al.*, 2020) is to investigate social media data analytics with the help of machine learning tools and this novel comes up for mounting a social media marketing approach. Several marketers prefer to use AI to change data into precious customer insights. Information gathering is an art (Kietzmann *et al.*, 2018) that involves identifying the benefits of online marketing for improving information gathering and feedback. Most of the users are using social media platforms. The world's most used social media platform is given in Fig. (2) and Table (1) for their comments on the product, such as their interest and requirements.

Table 1: Contribution of various social platforms

S. No	SM platform	Based on monthly active users, active user accounts, advertising audience, or unique monthly visitors (in Million)
1	Facebook	2449
2	YouTube	2000
3	WhatsApp	1600
4	FB Messenger	1300
5	Wexin/WeChat	1151
6	Instagram	1000
7	Douyin/Tiktok	800
8	QQ	731
9	Qzone	517
10	SNA Weibo	497
11	Reddit	430
12	Snapchat	382
13	Twitter	340
14	Pinterest	322
15	Kuaishou	316

DOI: 10.3844/jcssp.2025.78.87

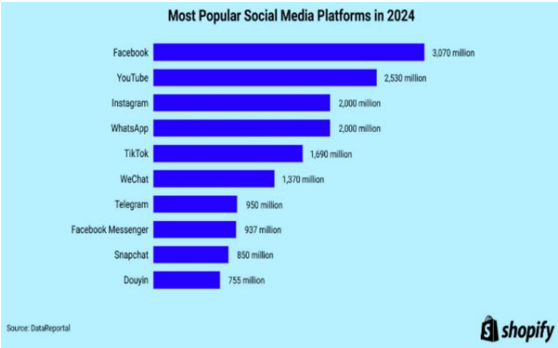

Fig. 2: The world's most popular social platform (Staff, 2024)

The researcher (Basri, 2020) studied the artificial intelligence-assisted impact of social media marketing on start-up businesses of small and medium enterprises' performance in Saudi Arabia. This research work is exclusive as it examines how Artificial Intelligence social media marketing practices have a significant role in enhancing the performance of small and medium enterprises. It is concluded that effective management plays an essential role as a moderator. The research work (Nair and Gupta, 2021) in which the authors identify various AI applications for social media and digital marketing experts and agencies. This study further studies that collaboration and creativity must be maintained to get a better return on investment. A smart city's ability to manage its resources and its residents' quality of life relies on its ability to communicate with one another, analyze its data and, in some instances, use artificial intelligence (Alam *et al*., 2022).

Mustak *et al.* (2021), discussed the dominant research topics related to Artificial Intelligence in marketing. They applied two matching analytical approaches to inspect the advancement and constitution of the research domain. They observed four contributions in their research work. Firstly, they observed through text analytics techniques, highly developed topic modeling algorithms, and scientometric analysis. Secondly, they explained that the data-driven approach helps reveal leading Artificial Intelligence research topics in marketing. Further, they classified them as consumer or business and strategy–related research. Thirdly, they did the scientometric analysis that produces information maps to co-citation clusters, landmark articles, and conceptual and theoretical foundations. Lastly, they observed an all-embracing discussion of research gaps, producing a vigorous outline for enhancing the intensity and span of Artificial Intelligence research in marketing. The article suggests investing in smart urbanization and boosting trade integration and collaboration with other Asian nations in the Information and Communication Technology (ICT) sector (Rao *et al*., 2023).

In their study (Chaudhary *et al*., 2021), the researchers applied big data technology to process and analyze data to forecast consumer behavior on social media platforms. They further stated that consumer behavior prediction is based on some parameters. They have used the regression model to predict consumer behavior on the social media platform. Researchers in Ameen *et al*. (2022) to aid researchers and industry professionals, provide a taxonomy of the core competencies essential to marketing creativity and the effect of various AI competencies on these competencies. This article (Peleshchyshyn and Bandrovskyi, 2019) is helpful because it lays the groundwork for future studies on how artificial intelligence may be used to market innovation. Building a formal model is to examine information dependency and dissemination and this study focuses on identifying the components of informational impact in online social networks. Linear algebraic equations are used to characterize the informational dynamics effect.

Proposed Model to Enhance Social Media Marketing

The model will be helpful for social media marketing. We have used the concept of AI in this model to improve online shopping and this model is given in Fig. (3). This model is based on questions with parameters such as segment type, segment description, answer, count, and percentage, which are briefly discussed in Table (2).

Materials and Methods

After getting data based on the various parameters of the question, this model performs preprocessing using various preprocessing techniques to make quality data. This model's cleaned data component will store the quality data and send it for Testing and training. This model component sent 80% of the data for training and 20% for Testing. Finally, the model gives the result and will be validated to check whether the result is accurate. This model is mathematically represented as, let *Q* represents questions and parameters like Segment Type Segment Description, Answer, Count, and Percentage are represented as *ST*, *SD*, *A*, *C*, *P*. Therefore, Eq. (1):

$$Q = \{ST, SD, A, C, P\} \qquad (1)$$

PP represents the preprocessing, so mathematical representation is given in Eq. (2):

Table 2: Features description

Feature	Explanation	Feature type
Question	Questions asked from the influencers	Text
Segment type	The type of audience surveyed	Text
Segment description	Details of targeted audience	Text
Answer	The audience gives answers	Text
Count	Number of responses of each type	Numerical
Percentage	Percentage of responses of each type	Numerical

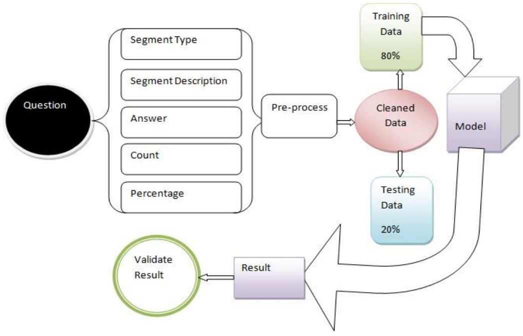

Fig. 3: Social media marketing model

$$PP = \sum(ST + SD + A + C + P) \quad (2)$$

The cleaned data is represented with *CD*, so the relationship between *PP* and *CD* is *CD* = *PP*. Let the model represent *m*, training, and testings are *x* and *y*. The *CD* is given in Eq. (3):

$$CD = \sum(.8x + .2y) \quad (3)$$

The mathematical representation of the model is given in Eq. (4):

$$M = \int \sum([(ST + SD + A + C + P)]) + ([\sum(.8x + .2y)]) \quad (4)$$

Q = *f* (*ST*, *SD*, *A*), the *ST*, *SD*, and *A* are input variables based on it. We can find how the social media platform influences the product's sales. Let *C* be the output, based on the variable *A*, then *C* = *f* (*A*) and *P* is output based on *C*, so *P* = *g*(*C*). The composite function is *P* = *g* (*f*(*A*). The derivates of percentage *P* concerning *A* $are\ \frac{\mathrm{dP}}{\mathrm{dC}} = \frac{\mathrm{dp}}{\mathrm{d(A)}} * \frac{\mathrm{d(A)}}{\mathrm{dP}}$. Now *C* is output, based on *SD* and *A* variables the *C* = *f* (*SD*, *A*) and *P* is output based on *C*, therefore *P* = *g*(*C*). In this case, the composite function is *P* = *g* (*f* (*SD*, *A*). The derivatives of percentage *P* concerning (*SD*, *A*) is $\frac{dP}{dC} = \frac{\mathrm{dP}}{\mathrm{d(A)}} \cdot \frac{\mathrm{d(A)}}{\mathrm{dC}} + \frac{\mathrm{dp}}{\mathrm{d(SD)}} \cdot \frac{\mathrm{d(SD)}}{\mathrm{d(C)}}$.

Data Preprocessing

We have removed the missing values from our dataset and we use NaN to represent the missing value in the dataset using mathematical calculation involving NaNs propagating NaNs through to the final appropriate result. We have removed NaNs from the data before performing statistical computations. The outlier is removed from the dataset using mean and standard deviation for all parameters, given in the equation below:

$$M_{\mathrm{count}} = \frac{\sum_{i=1}^{m}(count)_i}{m}$$

where, M_{count} is the mean of the count parameter in our dataset:

$$\mathrm{S}_{\mathrm{count}} = \sqrt{\frac{\sum((count) - (\frac{\sum_{i=1}^{m}(count)_i}{m}))^2}{m-1}}$$

$$S_{per} = \sqrt{\frac{\sum((count) - (\frac{\sum_{i=1}^{m}(percentage)_i}{m}))^2}{m-1}}$$

Let O_c represent outliers in count and O_p in percentage parameters in our dataset:

$$O_c = abs(count - (\frac{\sum_{i=1}^{m}(count)_i}{m})(ones(m,1)) > n^*)$$

$$(\sqrt{\frac{\sum((count) - (\frac{\sum_{i=1}^{m}(count)_i}{m}))^2}{m-1}})(ones(m,1))$$

$$O_p = abs(percentage - (\frac{\sum_{i=1}^{m}(percentage)_i}{m})(ones(m,1)) > n^*)$$

$$(\sqrt{\frac{\sum((percentage) - (\frac{\sum_{i=1}^{m}(percentage)_i}{m}))^2}{m-1}})(ones(m,1))$$

The *n* is the number of S_{count} and S_{per} computations for parameter count and percentage based on the dataset for our problems.

We have handled the missing value in our dataset by calculating the mean. The mean is calculated for a feature containing a missing value and replaced with the result for the missing value. The equation used to calculate the mean of features is given below:

$$M_C = \frac{\sum_{i=1}^{n} count_i}{n}$$

$$M_p = \frac{\sum_{i=1}^{n} percentage_i}{n}$$

Where, M_C and M_p are the mean for feature count and percentage, respectively, in our dataset.

We have deliberated superfluous repetition data across our dataset for features. Various techniques are used to make uniqueness in our data. The data uniqueness is shown in Table (3).

Table 3: Data uniqueness

Description	Number
Question	6
Segment type	5
Segment description	305
Answer	24
Count	377
Percentage	646

DOI: 10.3844/jcssp.2025.78.87

Table 4: Data information

S No	Column	Non-null	Count	Dtype
0	Question	5460	non-null	object
1	Segment Type	5460	non-null	object
2	Segment Description	5460	non-null	object
3	Answer	5460	non-null	object
4	Count	5460	non-null	int64
5	Percentage	5460	non-null	float64

The raw data and unorganized facts are necessary to preprocess the data to make it quality. We have processed the data regarding organization structure and presented it in each context to make our data useful. Here is a sample of the data we're most likely to encounter in our job as a preliminary step. The dataset has 2604 rows and 26 columns, given in Table (4). The question description is based on various parameter names, segment types, segment descriptions, answers, counts, and percentages.

Evaluation Parameters

The effectiveness of the suggested model was measured using a few criteria for assessment. The following is a quick explanation of the performance metrics used to assess the quality of the proposed algorithm.

Accuracy: The gold standard for measuring how well machine learning methods perform. The accuracy of a machine learning model may be expressed mathematically as the proportion of correct positive outcomes to correct adverse outcomes:

$$Accuracy = \frac{T_{Pos} + T_{Neg}}{T_{Pos} + T_{Neg} + F_{Pos} + F_{Neg}}$$

Root Mean Square Error (*RMSE*): It is one of the evaluation metrics of anomaly score to identify the distance between input vector A and output vector A' where $A = (a_1, a_2, \dots \dots \dots a_n)$ and $A' = (a'_1, a'_2, \dots \dots \dots . . a'_n)$:

$$RMSE\ (A, A') = \sqrt{\frac{\sum_1^n (a_i - a'_i)^2}{n}}$$

Results and Discussion

The question was asked which social media platform you influence for online shopping. The segment types are mobile, web, gender, university, and custom. The segment descriptions are global results, web, female voters, male voters, various universities, etc. The answer is Facebook, Instagram, Snapchat, Twitter, etc. The highest answer count is 916 and their percentage is 0.342 for Instagram. The shape of the data is given in Table (6).

The consumers who were not influenced by any social media platform were 947 and their calculated percentage is 0.354. We have calculated the count, mean, standard deviation, minimum, maximum, 25, 50, and 75% based on the data set in Table (5).

There is one float64, one int64, and four objects in Dtypes. The memory usage is 256.1+ KB. The segment type of the population and the total number of counts of each segment have been clearly shown in Table (7). It can be quickly concluded from the table that a significant part of the population is contributed by the 'University 'segment type, which counts the highest, i.e., 2724. 'Mobile' is one of the segment types which has contributed the least, i.e., 24. All other population segments have not made any significant contribution, which is illustrated in Fig. (4).

Initially, the misreported values were replaced with Nan entries in the data processing. Because the data were misreported, they must be regarded as missing values. This data was helpful because it transformed all the incorrect data into missing values. The next stage addresses the missing values after clearing up the incorrectly reported data. To figure out how many columns are missing values and, if so, what proportion, an analysis of % missing values per column is conducted. Several specific characteristics were stored numerically. Binary encoding was used to encode these functions. In other cases, too much information, such as economic position and seniority, was crammed into a single column. Features like these were found and either recoded to include more information or separated into two columns so that each feature could be found in its column. Over 20 categories were eliminated or reworked into something more beneficial for users.

YouTube was the most popular social media platform, followed by Facebook, Instagram, LinkedIn and WhatsApp, independent of sexual identity, profession, age, or nation of birth. The usage of other social media is limited or non-existent. Compared to the Web, Gender segment type, university, or custom, the most significant variation in social media usage was seen in mobile devices, as shown in Fig. (5).

Table 5: Description of data

	Count	Percentage
Count	1450.000000	1450.000000
Mean	35.013793	0.199313
STD	95.055604	0.274989
Min	0.000000	0.000000
25%	0.000000	0.000000
50%	1.000000	0.063500
75%	20.000000	0.323750
Max	947.000000	1.000000

Table 6: Shape of data

	Question	Segment type	Segment description	Answer	Count	Percentage
0	What social platform has influenced your online...	Mobile	Global results	Facebook	548	0.205
1	What social platform has influenced your online...	Mobile	Global results	Instagram	916	0.342
2	What social platform has influenced your online...	Mobile	Global results	Snapchat	86	0.032
3	What social platform has influenced your online...	Mobile	Global results	Twitter	179	0.067
4	What social platform has influenced your online...	Mobile	Global results	None	947	0.354
5	What social platform has influenced your online...	Web	Web	Facebook	0	0.000
6	What social platform has influenced your online...	Web	Web	Instagram	0	0.000
7	What social platform has influenced your online...	Web	Web	Snapchat	0	0.000
8	What social platform has influenced your online...	Web	Web	Twitter	0	0.000
9	What social platform has influenced your online...	Web	Web	None	2	1.000

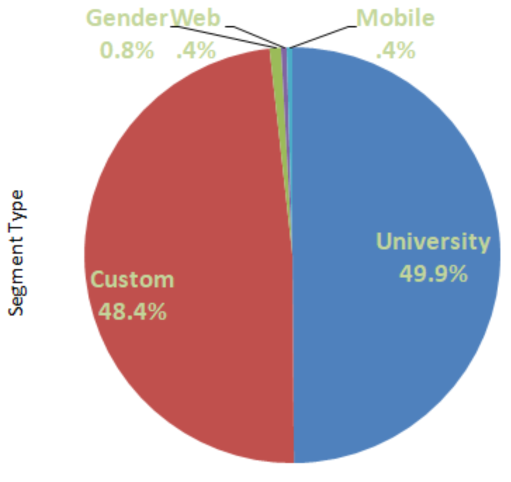

Fig. 4: Contribution of different segments of the population

The connection between a company and its customers, the source of its profits, is of utmost significance to its success, but dealing with a vast list of customers doesn't make life simple, particularly for business retailers concerned with all the transactions they create daily. This is where the need for a new paradigm based on a cutting-edge scientific approach becomes critical. This research has designed a model for segmenting customers based on their preferences. The segment description is given in Table (8).

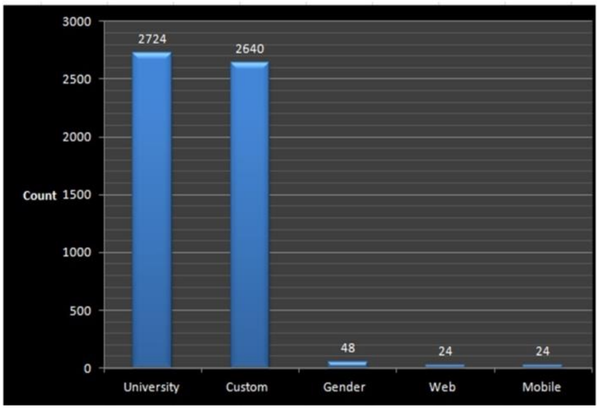

Fig. 5: Contribution of segment type

Table 7: Segment description and their count

Segment description	Number count
What's your major? Other	48
I'm in. Other	24
Your zodiac sign? Gemini (May 21-Jun 20)	24
your zodiac sign? Scorpio (Oct 23-Nov 21)	24
Are you pursuing? Technical / engineering	24
Eldorado High School	4
Hofstra University	4
Oneonta State University	3
Bloomsburg University of Pennsylvania	3
University of California Riverside	3

Table 8: Count of segment answers

S. No	Answer	Number count
1	Celebrities	272
2	Friends	272
3	Funny meme posters	272
4	People who do the things you like	272
5	They're famous so people like to follow them	263
6	They aren't on social media & and focus on actual things to change the world	263
7	They're a leading voice in the culture	263
8	They have a great body and frequently take pics of it	263
9	Snapchat	251
10	Instagram	251
11	Twitter	251
12	Facebook	251
13	None	251
14	Somewhat	247
15	No, it's whatever	247
16	Yes, very much so	247
17	Someone with a lot of social media followers	179
18	Popular friends in your network	179
19	A famous person	179
20	All the above	179
21	Get that money!	152
22	Other (comment)	152
23	Is this product cool?	152
24	This is lame	152

Data for most models must be in the form of numbers, no strings or other non-numerical data types and these numbers may be floats or integers. Therefore, categorical data has been transformed to a more readable integer format, which models may use to forecast better and improve outcomes, known as label encoding. We must employ a quantitative value between zero and the whole number of classes minus one to replace the category value in column Question, Segment type, and Answer shown in Table (9).

The distribution of percentage consumers on social media platforms is graphically shown in Fig. (6).

The distribution of user counts on various social media platforms is graphically represented in Fig. (7).

Table 9: Conversion of categorical data

S. No	Question	Segment type	Segment description	Answer	Count	Percentage
0	2	2	55	8	249	0.226
1	2	2	55	21	348	0.449
2	2	2	55	6	261	0.247
3	2	2	55	11	91	0.077
4	2	4	210	8	0	0.000
5	2	4	210	21	0	0.000

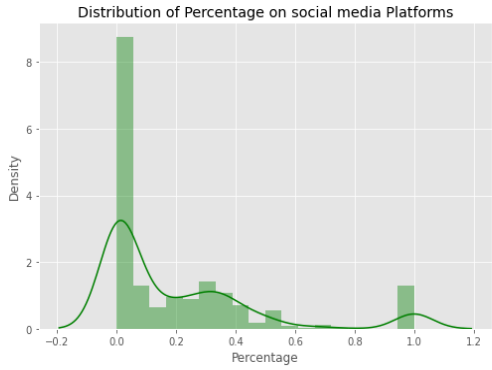

Fig. 6: Distribution of users percentage

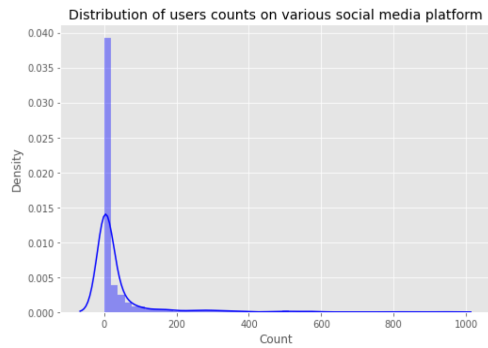

Fig. 7: Distribution of user count

The length of the segment description is 305 and the dtype is int64. The answers to each segment questionnaire have been recorded and the count of each of the answers given by the consumers has been shown in Table (8).

The Receiver Operating Characteristics (ROC) curve is given in Fig. (8) and the value of *AUC* is 0.57. In this curve, the true positive rate is at the X-axis and False Positive is at the Y-axis. *TPR* represents the True positive rate and *FPR* represents the false positive rate. The equation of *TPR* and *FPR* is given below. $TPR = Sensitivity = \frac{TP}{TP+FN}$, where *TP* is true positive numbers and *FN* is False Negative numbers. $FPR = 1 - Specificity = \frac{FP}{FP+TN}$, where *FP* is False Positive numbers and *TN* is True Negative numbers.

The social media dataset has been evaluated using various regression algorithms using a standard split of 70% for training data and 30% for testing data. In this analysis, the root Mean Square Error (RMSE) and accuracy have been taken into consideration for understanding the influence of social media on the shopping decisions of customers shown in Table (10).

Table 10: Comparison of models

S. No	Model	RMSE	Accuracy
1	DTR	1.2323	0.935
2	RFR	23.491	0.670
3	ETR	19.493	0.501
4	ABR	61.510	0.506
5	BR	3.008	0.931

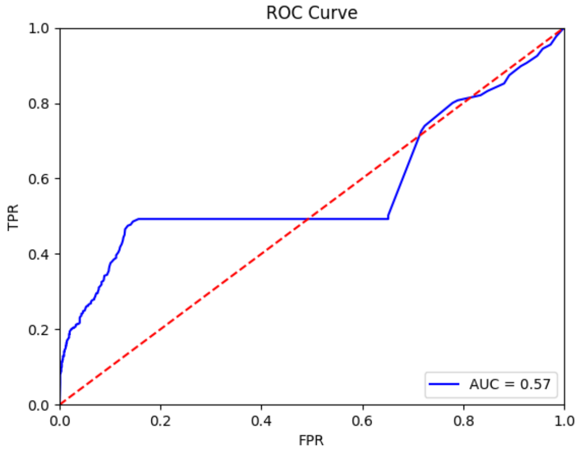

Fig. 8: RoC curve

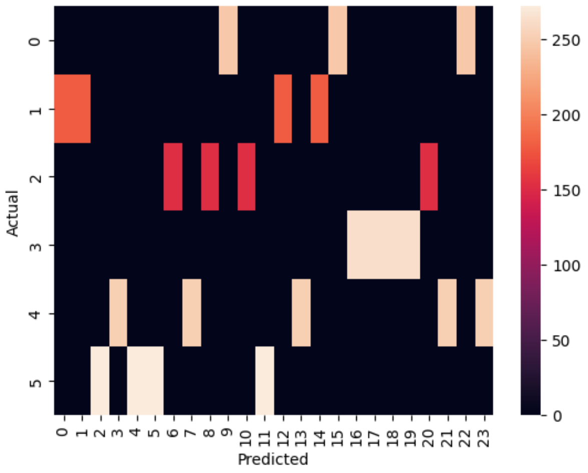

Fig. 9: Confusion matrix

DTR has shown an accuracy of 93%, which is comparatively higher than other models like random forest regression, ensemble tree regression, Ada boost regression, and Bayesian regression. Also, it is observed that the decision tree model has the lowest RMSE compared to other models, which signifies that DTR has performed the best among all others. The confusion matrix is given in Fig. (9).

Conclusion

An effective AI-based initiative for social media marketing will be able to offer performances in accordance with its advertised capacities. If the results of this three-country, cross-cultural study are accepted, it may pave the way for the creation of AI software for use in social media marketing, which will be driven by financial incentives to disrupt current industry norms. More consumers than ever are purchasing on social media sites (Ryu and Park, 2020). Facebook is a commercial marketplace, one of the most widely used social networking platforms. Users are increasingly turning to Facebook as a marketplace to trade goods with one another (Chen *et al.*, 2016). The results show that Instagram is the social media platform with the highest percentage of influence on online shopping, followed by Facebook, Snapchat, and Twitter. The data also indicate that the University segment type contributed the most to the population, which is around 2724, while the Mobile segment type contributed the least, just 24. There were around 947 consumers who were not influenced by any of the social media platforms.

In data preprocessing, we have replaced all the missing values with NaN, and several missing parameters were replaced with some numerical value. This study compares the performance of various regression algorithms for predicting the influence of social media on customers' shopping decisions. Among all the tested models like random forest regression, ensemble tree regression, ada boost regression, and Bayesian regression, the DTR model has the highest accuracy, around 93.5%, and the lowest RMSE, 1.23. The research indicates that preference-based consumer segmentation may be an effective store strategy. However, the ramifications of the results for companies hoping to use social media to impact online purchases aren't examined in depth. The findings of this study provide light on the impact of social networking sites on e-commerce and point the way toward potential areas of investigation, such as the effect of demographic variables like age, occupation, and country of origin.

This model will help improve social media marketing. Online retailers need to tailor their approaches to meet the needs of a wide variety of customers. One of the significant outcomes of market clustering is the "customization" of marketing for customers. The AUC calculated in our problem is 0.57, which shows the perfect evaluation performance of our model. This study incorporates a blend of conceptual and empirical methodologies to determine how organizations' ethical marketing activities impact consumers' impressions of their brands. We first describe the consumer behavior reaction selection to posts in the context of online marketing as a classification problem and show that the choice can be predicted accurately when human-designed characteristics are combined with machine-oriented high-dimensional post attributes. Market competitive and technological considerations have a negligible effect on the component aspects of consumer brand equity. Because of this finding, customers have minimal control over the marketing activities of these two components involved in the brand awareness process. While there is little attention, it may also give organizations insight into their future growth. The limitation of this study is that we have employed just posting features in our forecast; however, if more characteristics from businesses and consumers could be mined, we may attain higher performance.

Acknowledgment

We want to thank the College of Business, Umm Al Qura University, Makkah, for providing financial support regarding this study. We would also like to thank Jamia Millia Islamia and Shivaji College, University of Delhi, for providing facilities for this study.

Funding Information

The College of Business, Umm Al Qura University, Makkah, financially supports this study.

Author's Contributions

Kiran Chaudhary: Conceptualized the paper and contributed to writing the paper, organizing the contents of the paper, and presenting the result.

Mohammad Naquibur Rehman: Organized the paper and the Literature review and cleaned the data.

Nabeela Hasan: Contributed to data analytics and making the model. Also contributed to improving the quality of the paper by giving more concepts.

Mansaf Alam: Contributed to the mathematical model in the paper and validated the result.

Aakash Punit: Contributes to writing the Introduction part of the paper as well as improving the language of the paper.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and that no ethical issues are involved.

References

Alam, M., Chamoli, A., & Hasan, N. (2022). Smart Cities and the Internet of Drones. In *The Internet of Drones* (1st Ed., pp. 295–322). Apple Academic Press. https://doi.org/10.1201/9781003277491-14

Ameen, N., Sharma, G. D., Tarba, S., Rao, A., & Chopra, R. (2022). Toward advancing theory on creativity in marketing and artificial intelligence. *Psychology & Marketing*, *39*(9), 1802–1825. https://doi.org/10.1002/mar.21699

Arasu, B. S., Seelan, B. J. B., & Thamaraiselvan, N. (2020). A machine learning-based approach to enhancing social media marketing. *Computers & Electrical Engineering*, *86*, 106723. https://doi.org/10.1016/j.compeleceng.2020.106723

Bashar, A., Ahmad, I., & Wasiq, M. (2012). Effectiveness of social media as a marketing tool: An empirical study. *International Journal of Marketing, Financial Services & Management Research*, *1*(11), 88–99.

Basri, W. (2020). Examining the Impact of Artificial Intelligence (AI)-Assisted Social Media Marketing on the Performance of Small and Medium Enterprises: Toward Effective Business Management in the Saudi Arabian Context. *International Journal of Computational Intelligence Systems*, *13*(1), 142–152. https://doi.org/10.2991/ijcis.d.200127.002

Chaudhary, K., Alam, M., Al-Rakhami, M. S., & Gumaei, A. (2021). Machine learning-based mathematical modeling for prediction of social media consumer behavior using big data analytics. *Journal of Big Data*, *8*(1), 73. https://doi.org/10.1186/s40537-021-00466-2

Chen, J. V., Su, B., & Widjaja, A. E. (2016). Facebook C2C social commerce: A study of online impulse buying. *Decision Support Systems*, *83*, 57–69. https://doi.org/10.1016/j.dss.2015.12.008

Drummond, W. J., & French, S. P. (2008). The Future of GIS in Planning: Converging Technologies and Diverging Interests. *Journal of the American Planning Association*, *74*(2), 161–174. https://doi.org/10.1080/01944360801982146

Duffett, R., Petroșanu, D.-M., Negricea, I.-C., & Edu, T. (2019). Effect of YouTube Marketing Communication on Converting Brand Liking into Preference among Millennials Regarding Brands in General and Sustainable Offers in Particular. Evidence from South Africa and Romania. *Sustainability*, *11*(3), 604. https://doi.org/10.3390/su11030604

Dwivedi, Y. K., Ismagilova, E., Hughes, D. L., Carlson, J., Filieri, R., Jacobson, J., Jain, V., Karjaluoto, H., Kefi, H., Krishen, A. S., Kumar, V., Rahman, M. M., Raman, R., Rauschnabel, P. A., Rowley, J., Salo, J., Tran, G. A., & Wang, Y. (2021). Setting the future of digital and social media marketing research: Perspectives and research propositions. *International Journal of Information Management*, *59*, 102168. https://doi.org/10.1016/j.ijinfomgt.2020.102168

Fan, W., & Gordon, M. D. (2014). The power of social media analytics. *Communications of the ACM*, *57*(6), 74–81. https://doi.org/10.1145/2602574

Fuchs, C. (2008). Wikinomics: How mass collaboration changes everything - by Don Tapscott & Anthony D. Williams. *Journal of Communication*, *58*(2), 402–403. https://doi.org/10.1111/j.1460-2466.2008.00391_5.x

Hasan, N., Chaudhary, K., & Alam, M. (2021). Unsupervised machine learning framework for early machine failure detection in an industry. *Journal of Discrete Mathematical Sciences and Cryptography*, *24*(5), 1497–1508. https://doi.org/10.1080/09720529.2021.1951434

Hasan, N., Chaudhary, K., & Alam, M. (2022). A novel blockchain federated safety-as-a-service scheme for industrial IoT using machine learning. *Multimedia Tools and Applications*, *81*(25), 36751–36780. https://doi.org/10.1007/s11042-022-13503-w

Jain, A. K., Duin, P. W., & Mao, J. (2000). Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(1), 4–37. https://doi.org/10.1109/34.824819

Kietzmann, J., Paschen, J., & Treen, E. (2018). Artificial Intelligence in Advertising. *Journal of Advertising Research*, *58*(3), 263–267. https://doi.org/10.2501/jar-2018-035

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2017). Human Decisions and Machine Predictions. *The Quarterly Journal of Economics*, *133*(1), 237–293. https://doi.org/10.3386/w23180

Kottursamy, K., Raja, G., & Saranya, K. (2016). A Data Activity-Based Server-Side Cache Replacement for Mobile Devices. *Artificial Intelligence and Evolutionary Computations in Engineering Systems*, 579–589. https://doi.org/10.1007/978-81-322-2656-7_53

Lee, D., Hosanagar, K., & Nair, H. S. (2018). Advertising Content and Consumer Engagement on Social Media: Evidence from Facebook. *Management Science*, *64*(11), 5105–5131. https://doi.org/10.1287/mnsc.2017.2902

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, *31*(2), 87–106. https://doi.org/10.1257/jep.31.2.87

Mustak, M., Salminen, J., Plé, L., & Wirtz, J. (2021). Artificial intelligence in marketing: Topic modeling, scientometric analysis, and research agenda. *Journal of Business Research*, *124*, 389–404. https://doi.org/10.1016/j.jbusres.2020.10.044

Nagar, Y., & Malone, T. W. (2011). Making business predictions by combining human and machine intelligence in Prediction Markets. *Proceedings of the International Conference on Information Systems, ICIS*, 20.

Nair, K., & Gupta, R. (2021). Application of AI technology in modern digital marketing environment. *World Journal of Entrepreneurship, Management and Sustainable Development*, *17*(3), 318–328. https://doi.org/10.1108/wjemsd-08-2020-0099

Pejić Bach, M., Krstić, Ž., Seljan, S., & Turulja, L. (2019). Text Mining for Big Data Analysis in Financial Sector: A Literature Review. *Sustainability*, *11*(5), 1277. https://doi.org/10.3390/su11051277

Peleshchyshyn, A., & Bandrovskyi, H. (2019). Informational influence in social networks: dynamics modeling based on the system of linear equations. *2019 IEEE 14th International Conference on Computer Sciences and Information Technologies (CSIT)*, 165–168. https://doi.org/10.1109/stc-csit.2019.8929819

Plonsky, O., Apel, R., Ert, E., Tennenholtz, M., Bourgin, D., Peterson, J. C., & Erev, I. (2019). Predicting human decisions with behavioral theories and machine learning. In *arXiv:1904.06866*. https://doi.org/10.48550/arXiv.1904.06866

Rao, A., Talan, A., Abbas, S., Dev, D., & Taghizadeh-Hesary, F. (2023). The role of natural resources in the management of environmental sustainability: Machine learning approach. *Resources Policy*, *82*, 103548. https://doi.org/10.1016/j.resourpol.2023.103548

Ryu, S., & Park, J. (2020). The effects of benefit-driven commitment on usage of social media for shopping and positive word-of-mouth. *Journal of Retailing and Consumer Services*, *55*, 102094. https://doi.org/10.1016/j.jretconser.2020.102094

Staff, S. (2024). *Top 10 Most Popular Social Media Platforms*. DataReportal. https://www.shopify.com/blog/most-popular-social-media-platforms