Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717

Understanding Social Media Behavior in the USA: AI-Driven Insights for Predicting Digital Trends and User Engagement

Muhammad Hasanuzzaman[1], Miraz Hossain[2], Md Majedur Rahman[3], Md Masud Karim Rabbi[4], Medhat Mohiuddin Khan[5], MD Abdul Fahim Zeeshan[6], Babul Sarker[7], Md Nazmul Shakir Rabbi[8], Mohammed Kawsar[9]

Abstract

The swift advancement in social networking platforms has radically shifted the patterns and nature of how people connect with services and brands in America. The utmost objective of this research project was to implement artificial intelligence together with machine learning approaches for creating predictive models that forecast digital pattern development as well as user commitment through social media interactions within the United States. The data used in this analysis are posts aggregated from two leading online platforms, X-Twitter and Reddit, and consist of user-generated material covering a multifaceted set of topics and opinions. X-Twitter posts are current and real-time and give insight into what is currently being discussed and the opinions that are making headlines, whereas Reddit content provides extensive commentary and user engagement on numerous subreddits. To make the dataset more comprehensive and richer in information, extensive engagement metrics like likes, shares, and comments are used to extend its reach and provide insight into the extent to which users engage with the material presented to them. For this research, we used the multi-model approach so that there would be an exhaustive study and strong predictions, by implementing models such as Logistic Regression, Random Forest, and XGB-Classifier. To assess the models properly, we made use of several performance metrics like Accuracy, Precision, Recall, and F1-score. Logistic Regression only manages to achieve a below-average accuracy, signaling an average level of predictive quality. The Random Forest model fares slightly better with a slightly better accuracy rate, which implies that its ensemble method increases its predictive power to classify instances more effectively. In turn, the XG Boost model took the top spot in the comparison with an accuracy rate, projecting its ability to identify complex patterns in the data and showcasing the highest predictive level among the three models. The use of model outputs can greatly maximize real-time content strategy for brands and organizations seeking to maximize engagement in the USA. Based on user behavior patterns and engagement metrics, models can give insight into what should be posted and when to maximize attention. In campaign optimization, predictive modeling assists U.S. brands in making strategic decisions regarding ad spend allocation. Based on an examination of past performance and engagement metrics, brands can see which content is driving the greatest impact and engagement levels and make strategic investments in high-performance content that genuinely resonates with desired audiences. For public communication and policy, predictive models are particularly helpful in projecting how the U.S. public will react to news announcements, policy initiatives, or campaigns. Boosting the effectiveness of predictive models by incorporating transformer-based NLP models like BERT is one direction to explore in the future.

Keywords: *Social Media, User Engagement, Artificial Intelligence, Machine Learning, Digital Trends, Data Analysis, Consumer Behavior, Marketing Strategies, Predictive Modeling.*

Introduction

Background

Al Montaser et al. (2025), indicated that social media operates as a major determinant for public trends, consumer choices, and opinion formation in the United States with an unprecedented force. During the past twenty years, Facebook and Twitter, together with Instagram and TikTok, have brought revolutionary changes to human communication and reshaped society through their influence on social norms and political dialogues as well as consumer purchasing patterns. Through its digital platform, users can share

[1] Master's in Strategic Communication, Gannon University, Erie, PA, USA.
[2] Master of Business Administration, Westcliff University.
[3] Master of Business Administration (MBA), Trine University, Indiana.
[4] Master's in Business Administration, International American University.
[5] Master of Business Administration, Westcliff University.
[6] Master of Arts in Strategic Communication, Gannon University, Erie, PA, USA.
[7] Masters of Science in Business Analytics (MSBA), Trine University, Angola, Indiana, USA.
[8] Master of Science in Information Technology, Washington University of Science and Technology.
[9] MSc, Analytics & Information Science, Duquesne University.

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717

opinions and life experiences at an unparalleled scale, thus building an expansive database that shows the American public's thoughts. The substantial implications emerge from this discovery because organizations and business enterprises now utilize social media analytics to understand market preferences for creating specialized marketing approaches. Success in present-day business demands mastery of social media behavior because effective audience connections through these platforms determine the rise or fall of brands and their services (Mohaimin et al., 2025).

According to Ray et al. (2025), the large quantity of data created through social media interactions creates substantial difficulties for the interpretation of this data. The experimental analysis faces challenges because user behavior shows multiple complications from fast-changing trends and different demographic populations, and varying levels of user engagement. Current data analytics techniques experience difficulty operating effectively with the fast-changing nature of online interactions because they produce a gap between gathered information and valuable understanding. The enormous volume of data available leads decision-makers to experience analysis paralysis, which prevents them from producing meaningful conclusions from their available data. Islam et al. (2025a), posited that the attempt to quantify and analyze user engagement becomes more challenging because social media contains subjective content that includes text-based posts and photographs, and videos. Advanced technologies, along with artificial intelligence and machine learning systems, require immediate implementation to analyze the extensive social media complexity and yield important behavioral insights.

Problem Statement

User engagement pattern interpretation becomes more complex because social media continues to evolve. User behavior changes frequently because it responds to multiple elements between cultural developments and economic standards, and technical innovations. The constant changes in user behavior create significant difficulties for marketers and businesses to build predictive models for engagement levels since these models serve as fundamental tools for developing strategic plans (Gupta & Khan, 2024). Organizations that fail to adapt correctly face important repercussions that result in resource waste as well as insufficient marketing initiatives linked to reduced brand affinity. Social media users encounter substantial difficulties because their environment includes a high amount of misinformation and polarized content that complicates their information navigation. Users participate with content for multiple purposes that include true interest alongside social pressure and basic curiosity, making it difficult for companies to understand their audience correctly (Emma & Tawkoski, 2022).

As per Dang & Van Thich (2025), implementing useful analytic systems represents a necessary step because they properly detect the complex ways users behave. Artificial intelligence and machine learning integrated into social media analytics present an effective solution by processing significant data volumes in real-time while finding hidden patterns not visible without traditional methods. The combination of new technological systems enables researchers to produce predictions that accommodate complicated user behaviors, thus supplying effective recommendations to marketing strategists. AI-driven methods need a strong comprehension of technological foundations as well as social patterns that exist in the online space to achieve successful execution (Das et al., 2025). The main objective of this study analyzes possible methods to leverage machine learning and AI, which help resolve data overload problems while managing behavioral intricacy to improve digital trend and user engagement prediction abilities.

Research Objective

The core goal of this analysis is to implement artificial intelligence together with machine learning approaches for creating predictive models that forecast digital pattern development as well as user commitment through social media interactions within the United States. The study utilizes modern analytical tools to discover valuable knowledge that helps marketers, alongside influencers and tech platforms, create strategic choices about their marketing operations. Multiple essential parts form this target, which demands precise determination of metrics expressing user engagement and creation of adaptable prediction models alongside assessing marketing consequences from obtained results. The fundamental objective of this research depends on understanding social media patterns that emerge from artificial

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717

intelligence-based operations to design user-engaging strategies, enhancing brand loyalty.

To this effect, the research will examine the user-generated content on various social media platforms, in both quantitative and qualitative senses of engagement. This aspect includes looking at trends of user interactions, including likes, shares, comments, and also content creation, and exploring what sentiments are expressed by user posts. The study will use machine learning algorithms to identify patterns and correlate them with the predicted behavior of future engagement. Additionally, the research will examine the ethical concerns in the use of AI in social media analytics, making sure that the insights will be actionable as well as humane towards user privacy and data integrity. What we are aiming for is to come up with a framework wherein the insights that the marketers and content creators utilize AI powers to make their decision-making process and social media engagement better.

Scope & Significance

While on one level, this is an issue of interest in any country where social media is dominant, the focus of this research is on U.S.-based social media users, who will be unique due to their uniquely shaped cultural and demographic factors. Among the diverse population of America, the factors of age, race, economic status, and geographical location should be taken into consideration while focusing on social media interactions and their patterns of engagement. With this demographic in focus, the research strives to offer such nuanced insights to be highly applicable for the American market and hence, increase the usefulness and relevance of the findings for American digital marketers and influencers in the relevant space. While this research is not isolated in terms of its ability to help stakeholders understand user behavior, its implication goes beyond gaining that knowledge; it aims to provide tools and information to help stakeholders adjust to the swiftly evolving digital realm, leading to more effective communication and engagement strategies.

At the same time, the implications of this research are extensive, as applying these insights will help industries and sectors that are engaged with social media to engage with their customers. But such information can be useful within retail and entertainment, as well as healthcare and education. The way businesses leverage AI-driven insights will help them enhance their marketing campaigns, imbibe their content, which can resonate with the target audience better and is more profitable. Moreover, this research contributes to the larger discussion concerning how to ethically use AI to enhance analytics by advocating for ethical and responsible usage of the data that we uncover through analytics. This process will not only help to improve the business output but also contribute to ensuring a more equitable and transparent digital ecosystem where users also perceive that they are being empowered and respected in the digital realm.

Literature Review

Evolution of Social Media Behavior

Akshay & Mohaneshwar (2024), argued that, for the past decade, the evolution of social media behavior in the United States has seen dramatic shifts in user interactions, wanted technologies, and other such things. In the beginning, social media network sites were made as a connective media or communication space on the web, and individuals could share their updates and manage relationships. But as the land has grown up, people have moved towards a more dynamic, more content-consuming, curation, and dissemination nature of engagement. As users are getting accustomed to engaging with more content types such as images or videos, live streams, and ephemeral stories, visually rich and interactive experiences are being favored. While part of this is due to changing preferences, this is also a refinement to the capabilities of social media platforms that have evolved to integrate more features to push user engagement through gamification, personalized feeds, and interactive community elements (Amoo et al., 2024).

Retrospectively, Bag et al. (2022), articulated that social media behavior started changing when mobile applications helped people socialize and create content in real time. Because about 90 percent of U.S. adults have smartphones, individuals are using social media as the time basis for connecting with other people and consuming content wherever and whenever they want. It has been happening alongside the short-form

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717

content boom with Tik Tok and Instagram Reels, among others, offering to the users' short attention spans a combo of upping their creativity and self-expression. These platforms have also evolved by using machine learning to create algorithms to cringe at based on what a user responds to or interacts with, thereby feeding content results based on personal interests. The most important thing about using this algorithmic curation is that it does not just determine what users see; it determines how users will see content, usually leading to echo chambers and the reinforcement of preexisting beliefs (Choksey et al., 2023).

Predictive Analytics in Social Media

In recent years, there has been a growing interest in the application of predictive analytics in social media, giving rise to the desire to understand and predict social users' engagement and trend cycles successfully. There is prior research that reveals how artificial intelligence (AI) and machine learning (ML) models can analyze the huge amounts of data produced from social media interactions and can extract useful insights that supply valuable inputs about marketing strategy and content creation (Dahhiya et al., 2025). One can think of many models to predict user engagement from historical data, demographic factors, and behavioral patterns, from logistic regression to more complex neural networks. The predictive models that include user activity rate, content types, and temporal variables allow these marketers to predict engagement levels and design strategies accordingly (Ho et al, 2021).

Predictive analytics in social media have seen drastic changes with the advent of real-time processing and analysis of data. As the latter advances and uses technologies that can process and analyze data in real time, models can almost instantly respond to changing user behaviors. It is particularly helpful in the context of fast-moving topics or virus content, and gaining a competitive advantage can be done by understanding user sentiment and engagement dynamics (Kobsa, 2017). To build robust models that can predict not only the engagement metrics but also the lifespan of trends, researchers have used different algorithms such as decision trees, support vector machines, or deep learning techniques. These models have further integrated the power of Natural Language Processing (NLP) that permits these models to understand textual data to extract sentiment and theme relevance from it, resulting in ever more refined predictions of user sentiment and another related engagement pattern (Okeleke et al., 2024).

However, the techniques of predictive analytics in social media are still not adequate. However, most of the present models are inclined towards particular levels of generation like share, etc, without accounting for user interaction in a holistic manner, which might give rise to incomplete or misleading analysis. Moreover, online social media platforms continue to evolve rapidly, and the associated algorithms make the modeling of social media interactions a harder task (Patil et al., 2024). However, with user preferences continually changing and new platforms arising, such models have an immediate need to adapt to the changes. The literature points out the fact that there is an increased recognition of the need to integrate multiple datasets and refine predictive techniques for improvement in the accuracy and application of engagement forecasting in the expanding digital context (Rana et al, 2023).

Sentiment and Engagement Modeling

According to Rathee (2025), the intersection of engagement modeling and sentiment analysis on social media has become an important area to investigate, with many existing studies exploring how text-based sentiment and metadata can be used to make predictions regarding user engagement. Sentiment analysis to date has consisted of some form of computational assessment of user-generated text to identify the emotional tone and has been used to measure audience response and predict levels of engagement with some success. Researchers have used an array of methods to examine sentiments mentioned in posts, comments, and reviews, and include supervised and unsupervised learning methods among these techniques. Quantifying emotional response allows these studies to identify correlations between sentiment and patterns of interaction among users and offers important insights into the role and impact of emotional resonance on engagement levels (Sizan et al., 2025).

Salonen & Karjaloto (2020), established that the use of metadata, including timestamps and platform attributes, alongside keyword analysis enables sentiment and engagement modeling to attain its best

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717

potential. Engagement analysis through research indicates posts will perform best when published at specific times during particular days since reaching optimal interaction rates depends on timing. Research studies evaluate the content-sharing platforms because each platform has distinct user behaviors that must be taken into consideration. Specially selected keywords together with trending topics become vital elements because content synched with mainstream dialogues or well-liked subjects attracts increased attention from users. Research combining the multiple-dimensional analysis of user engagement has resulted in advanced precision models that help marketers, together with content creators, obtain important insights (Sharma et al, 2025).

Nevertheless, this domain faces several challenges even as sentiment and engagement modeling have made some progress. A limitation of sentiment analysis is that sentiment analysis is inherently subject to cultural and contextual classification, which can depend on the demographics of users. In addition, the rapidity of social media and the nature of sentiment meanings, which can shift swiftly in response to external events or changes in public opinion, impart additional challenges to building stable predictive models (Sizan et al., 2025). Besides, most of the existing studies pay little attention to how sentiment, context, and user behavior interact. To address these challenges, we need to keep researching to improve the sentiment analysis technique and build up the complete model of engagement that can adapt to the dynamics of social media interactions and improve its predictive capabilities, plus its understanding of user behavior (Teepapal, 2025).

Research Gaps and Need for Regional Context

Notwithstanding, Dang et al. (2025), argued that the industry lacks the comprehension or need for country/geographical-specific user behavior modeling in the United States. Most of the literature available today tends to take the same direction and use similar findings with varying demographics and contexts without considering the particular cultural, demographic, or behavioral characteristics of American social media users. There is, of course, a stark difference in the landscape of the U.S. as a result of its diverse population and its culture, varying access to technology, and so on, which can profoundly impact the way people engage in social media. Thus, the absence of such studies focused on the distinctive characteristics of social media behavior within the American context is a matter demanding more localized studies designed to shed away the insightful dimension of how social media is being utilized by businesses and marketers in that sphere (Jui et al., 2023).

Moreover, Al Montaser et al. (2025), held that there is a lack of comparative research in terms of the assessment of the accuracy of machine learning models in predicting engagement for different social media platforms. There are many proposals for different predictive models, but most of the studies make use of only one methodology, which may not be a preferable method to use in combination with others. The absence of a comparative analysis limits the ability to determine what works in predicting user engagement, a key component for creating impactful marketing strategies. Dahhiya et al. (2025), suggested that a comparative analysis of the performance of various models for the prediction of engagement enables researchers to disclose best practices for the use of AI and ML in social media analytics to enhance the predictive capabilities of the marketing and content creation.

With the existence of these gaps, future research should focus on developing region-specific models that are in line with the richness of the U.S. social media landscape. Using localized data, researchers can use prediction methodologies that predict the customized behaviors of American users regarding national cultural factors, demographics, and technological access. In addition, such comparative analysis of machine learning techniques should fuel more accurate engagement forecasts and help marketers refine their tactics and more efficiently engage with their audiences. This research gap is geared towards addressing these social media behavior research gaps within the U.S. to not only enhance the understanding of the behavior on social media in the U.S. and help fill the digital marketing and analytics field but also contribute to more informed decision-making within the increasingly complex digital environment.

Data Collection and Preprocessing

Data Sources

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717

The data used in this analysis are posts aggregated from two leading online platforms, X-Twitter and Reddit, and consist of user-generated material covering a multifaceted set of topics and opinions. X-Twitter posts are current and in real-time and give insight into what is currently being discussed and the opinions that are making headlines, whereas Reddit content provides extensive commentary and user engagement on numerous subreddits. To make the dataset more comprehensive and richer in information, extensive engagement metrics like likes, shares, and comments are used to extend its reach and provide insight into the extent to which users engage with the material presented to them. The combination of these sources not only adds to the richness and depth of the dataset but is also used to provide better insight into user behavior and user engagement patterns and, thus, into predictive modeling that can identify and forecast online trends and user interaction.

Preprocessing Steps

The preprocessing operations on the dataset are important in establishing the quality and potency of the data to be analyzed. Firstly, text cleaning takes place by doing the tokenization to split the posts into single words or phrases and then removing stopwords—words that are frequently used and are not vital to the message, like "and," "the," and "is." Tokenization and stopwords removal make the dataset more streamlined by concentrating on more significant material. Sentiment scoring can be employed to quantify the posts' emotional tone and provide context for engagement predictions. Another important operation is the selection of features, where particular aspects like hashtags, when the posts are made, and user categories are selected and extracted. Hashtags are used to show what is popular and trending and make the content more discoverable with them. The posts' timings can show patterns of user engagement from the week to the day and vice versa. Segmentation by user categories allows demography or interest-wise grouping and further refines the analysis and makes predictive modeling more focused. In general, these preprocessing operations are intended to make the dataset ready to be analyzed and modeled effectively and eventually benefit from the insights drawn from the social media data.

Key Features Selection

S/No.	Key Features	Description
01.	**Post Content**	The genuine text or material uploaded by members in posts themselves that may include opinions, questions, or facts related to a particular topic.
02.	**Engagement Metrics**	Quantitative indicators of user interaction with posts in the form of likes, shares, and comments that reflect the interest and interaction level of the audience.
03.	**User ID**	A distinctive identifier is given to every user on the site so that individual user behavior and engagement patterns can be monitored and measured over a period.
04.	**Timestamp**	The timestamp when the post was created, which is vital to assess user engagement patterns by period, like peak activity windows.
05.	**Hashtags**	Phrases or keywords that include a '#' prefix to group posts and make them searchable under themes or topics that are currently popular or trending.
06.	**User Category**	Demographic or interest-based group classifications of users, including age, location, or individual interests, are used to segment an audience to provide focused analysis.
07.	**Sentiment Score**	A quantitative measure of the emotional tone in a post based on sentiment analysis methods that identify the content as being positive, negative, or neutral.
08.	**Post Type:**	The type of material being presented, i.e., text, image, video, or link, may affect engagement level and user interactions.

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717

09	**Reply Count**	The quantity of replies to one's particular post signifies how controversial or engrossing the material is and reflects on user interaction patterns.
10.	**Share Count**	The overall count of how many times users have shared a given post to show its reach and level of popularity within the social environment.

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is an important stage in the data analysis pipeline that entails summarizing and visualizing data to reveal patterns, relationships, and findings before employing official modeling methods. It focuses on the application and utilization of numerous statistical and visual tools to identify the underlying structure of the data set so that analysts can extract patterns, outliers, and correlations among the variables. In EDA, methods like descriptive statistics, data visualizations (such as histograms, scatter plots, and box plots), and correlation measures are used to develop an informed comprehension of the characteristics and distribution of the data. Not only does EDA support the evaluation of data quality and completeness, but it also informs subsequent actions in the study, driving the selection of suitable modeling approaches and ensuring that vital factors are not missed. Finally, EDA is an initial step that improves the efficacy of subsequent data analysis stages.

Distribution of Respondents by Age Group

The Python code uses the seaborn package (imported as sns) to make a count plot illustrating the distribution of respondents by age group from a pandas DataFrame df. It begins by importing required libraries: seaborn to make statistical visualizations and matplotlib.pyplot (imported as plt) to provide plotting capabilities, and plotly.express (imported as px) without being called in this current snippet. The code then defines a "whitegrid" style and "Set2" colour palette in the seaborn plot and sizes the figure with matplotlib. The main code block is the sns. countplot function with the Data Frame df and with the 'What is your age group?' column used to define the x-axis to plot the counts within each unique age group. The .value_counts().index ensures the age group on the x-axis is plotted in ascending order by count in the dataset. The code concludes with the title being defined on the plot, rotating x-axis text to make it more readable, labeling the x and y axes, resizing the plot to avoid having the labels overlay one another, and presenting the created count plot with plt.show().

Output:

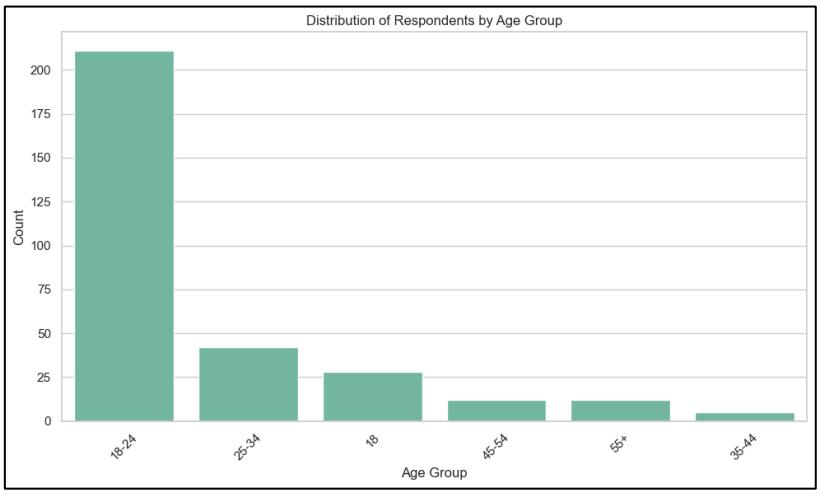


Figure 1: Distribution of Respondents by Age Group

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717

The age group chart illustrates respondents' distribution by age group and presents an overwhelming presence of 18-24-year-old individuals, with more than 200 respondents, the largest group in the dataset. The group dominates the responses with high participation from younger users. The following age ranges are conspicuously underrepresented with far fewer participants: 25-34 years with about 50 respondents, and 35-44 and 45-54 years with around 20 respondents each. The 55+ group has the lowest representation, with fewer than 10 respondents. The data showcases the overwhelming skew towards younger ages, with an indication that this group is primarily responsible for driving social media participation in this instance, and the far lower participation among older age ranges. The implication could extend to the area of directed advertising and content streams, with the necessity to adapt strategies to appeal to younger consumers.

Used by Different Occupations

The implemented Python code is used to plot a stacked bar plot to show the association between the 'What is your occupation?' and 'What device do you use to access the internet?' columns in a pandas Data Frame called df. It starts by employing pd.crosstab() to form a contingency table called device_occupation that counts the instances of each device used under each category of occupation. Then it employs the .plot() function on this crosstab table with the kind='bar' argument to generate the bar plot. The stacked=True keyword is used to stack the bars by occupation type and display the count of each type in the form of segments in one bar. The colormap='viridis' is used to set the colors used by the various types of devices. At the end, the code adds the title to the plot, labels the x and y axes, rotates the x-axis tick marks to make them more readable, sets the layout to avoid the text from being on top of one another, and shows the final stacked bar plot using plt.show().

Output:

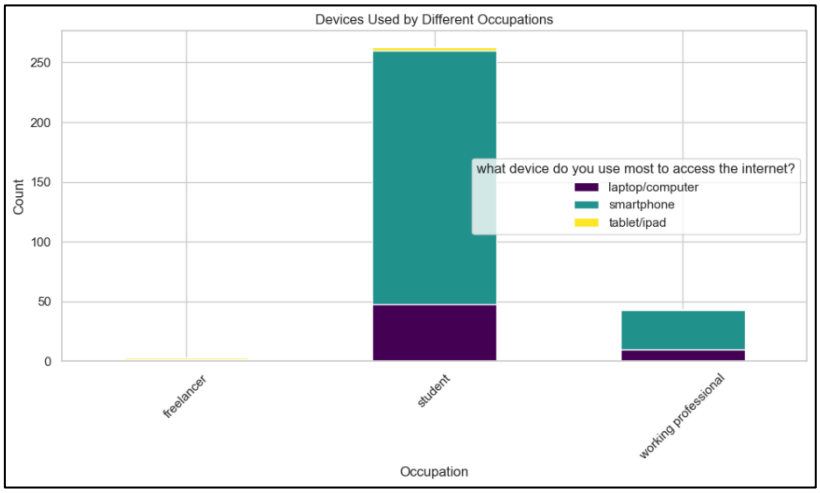


Figure 2: Devices Used by Different Occupations

The graph showing devices used by occupations marks an interesting tendency in internet access patterns among occupations. Students decisively lead the set, with more than 250 respondents stating that they mostly access the internet on a laptop or computer. Quite by contrast, the remaining occupations—complete with "freelancer" and "working professional"—have much lower figures, with freelancers relying on smartphones more so than laptops and working professionals having a more evenly spread distribution but still far from the student level. The evidence points to students having a strong reliance on laptops or computers, probably because of their academic needs, whereas freelancers and working professionals show more differentiated use, perhaps an indication of the nature of work and the nature of mobile devices. The observation highlights the significance of being aware of the preferences in devices to address the particular demands made by specific occupational segments through digital content and services.

Most Used Social Media Platform

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717

The Python code visualizes the distribution of the most frequently used social media platforms from the pandas DataFrame df. It first calculates the count of each unique social media platform in the column 'what is your most used social media platform?' and assigns it to platform_counts using .value_counts(). It then uses seaborn's barplot function to display the bar plot with the x-axis being the social media platforms (accessed from the index attribute of platform_counts) and the y-axis being the count of users for each platform (accessed from the values attribute of platform_counts). The palette "mako" option is used to define the bar colors. Last, the script configures the plot title, sets the y-axis title to "Number of Users", rotates the x-axis tick labels to make it more readable, rescales the plot to avoid overlapping, and shows the created bar plot with plt.show().

Output:

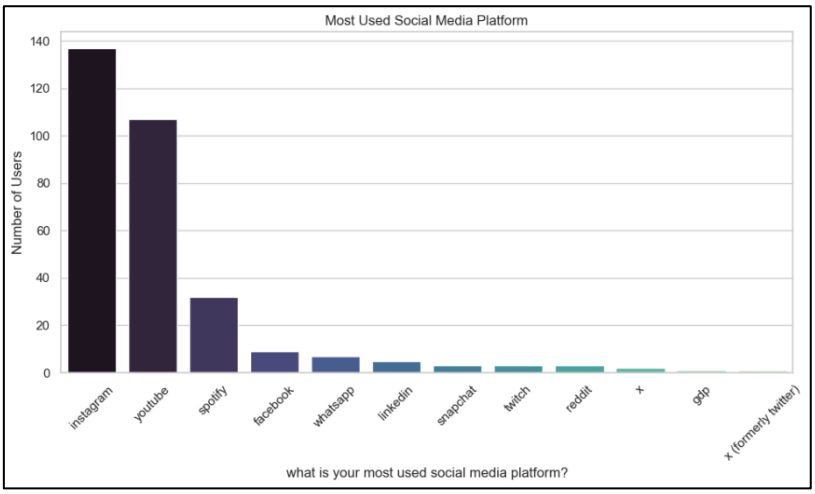


Figure 3: Most Used Social Media Platform

The graph above shows the most popular platforms used, demonstrating an overwhelming preference by users in favor of Instagram as the leading platform, with more than 140 million users. The high figure attests to the appeal of Instagram, whose users mostly consist of younger audiences who prefer visual material. After Instagram comes the second-rated platform, which is YouTube, and it has far fewer users, with about 50 users, and it demonstrates an equally strong engagement with video material. Spotify ranks high despite being mainly used to stream audio material and has an equally strong following with about 30 million users acting in its capacity as a social platform. Traditional platforms like Facebook, WhatsApp, LinkedIn, Snapchat, and Reddit are far lower in ranking, with fewer than 30 users, and reflect declining user engagement on these platforms. The near-zero value recorded on lesser-known platforms like Grip and the previous Twitter reflects that these platforms will not receive significant user attention based on this survey. Generally, the evidence points to Instagram's leading role in social platforms and reflects evolving user preferences towards more interactive and visual platforms.

Social Media Use Right After Waking Up

The Python program makes a pie chart out of the responses to the question "the first thing you do when you wake up is scroll through your social media account," from pandas DataFrame df. It starts by computing the count of each unique answer (presumably 'Yes' and 'No') in the column given by .value_counts() and assigns it to wakeup_sm. It then uses matplotlib.pyplot.pie to make the pie chart with the counts as the sizes of the wedges. The labels are assigned to the index of the wakeup_sm Series (the 'Yes' and 'No' responses), and autopct='%1.1f%%' to format the percentage on each wedge to one decimal place. The startangle=140 makes the beginning point of the first wedge be 140 degrees and colors=sns.color_palette("pastel") provides the coloring palette to the wedge slices with seaborn's "pastel" ordering. Finally, the code adds the title to the pie chart, scales the layout so the figure and axes fit well together and are not overlapping, and presents the chart with plt.show().

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717

Output:

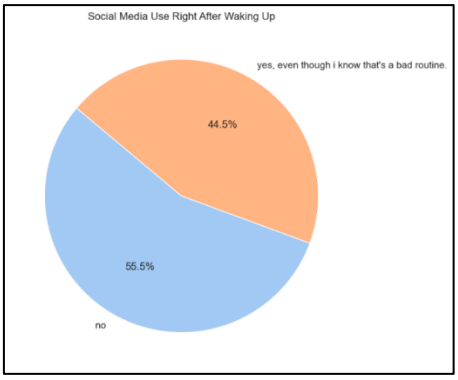


Figure 4: Social Media Use Right After Waking Up

The pie chart showing the use of social media on waking demonstrates a significant division in user behavior. Around 55.5% of participants state that they don't go on social media immediately on waking, which indicates that there could be an important group of people who begin with alternative morning routines or behaviors. In contrast, 44.5% say that they do access social media in the morning, first thing, even if it is something that they know is not so healthy to do. What this finding does show is that there is a strong inclination towards accessing social media immediately on waking by almost half of the participants, and this might reflect similar patterns in digital addiction and the place of social networks in daily life. The evidence does open up some questions regarding the consequences of these behaviors on well-being and productivity and how users cope with the negative aspects of beginning the day through the use of social media.

Frequency of Distraction Due to Social Media

The script employs the seaborn package (imported and aliased to sns) to create the count plot to display the frequency of distraction by using the 'how often do you find yourself distracted while working or studying because of social media?' column of DataFrame df. The sns.countplot() function is employed with y= to display the levels of distraction frequency on the y-axis and use the bar length to display the count with which each level appears. The script then titles the plot "Frequency of Distraction Due to Social Media", labels the x-axis to read "Count" and the y-axis to read "Distraction Frequency", rescales the layout so the plot fits with its components without any overlaps by using plt.tight_layout(), and ends by presenting the plot with plt.show().

Output:

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717

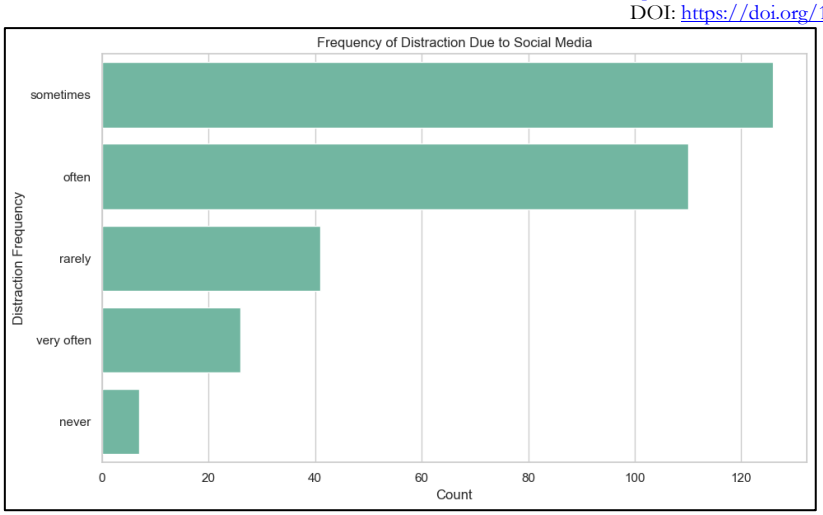


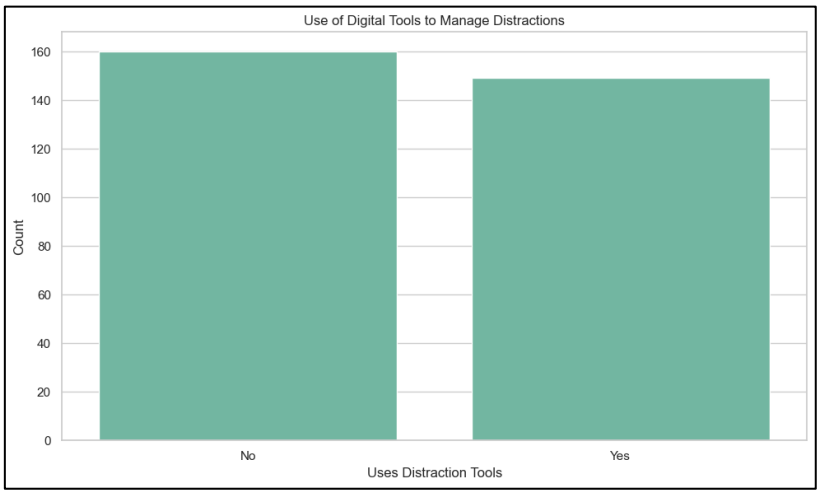
Figure 5: Frequency of Distraction Due to Social Media

The graph portraying the rate of distraction through social media reveals an overarching problem among users, with most reporting being "often" and "sometimes" distracted by it. In particular, roughly 120 respondents are often distracted, and some 80 users admit to being sometimes distracted. The high proportion reflects that social media is an ordinary cause of interruption to daily routines. Conversely, fewer participants admit being seldom (approximately 40) or very frequently (approximately 20) distracted by social media, and very few respondents say they are never distracted by it. The revelations, therefore, point towards the effect that social media has on focus and productivity and call for ways to control its influence, mainly by people who are constantly being sidetracked by duties or tasks. Generally, the findings show that distractions through social media are universal in people's lives.

Use of Digital Tools to Manage Distractions

The script makes a count plot with seaborn (sns) to display the distribution of the responses to the question "Do you use tools or apps to manage distractions?" from df. The sns.countplot() is called with x= to state that the various responses to this question will be on the x-axis and the height of the bars will reflect the count of these responses. The script then allows the plot title to be set to "Use of Digital Tools to Manage Distractions", the x-axis to be labeled "Uses Distraction Tools", and the y-axis to be labeled "Count". It further specifically sets the x-axis tick positions and labels to [0, 1] and ['No', 'Yes'], respectively, probably expecting binary responses. The script finally scales the plot layout to avoid overlay with plt.tight_layout() and displays the created count plot with plt.show().

Output:



Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717

Figure 6: Use of Digital Tools to Manage Distractions

The graph showcasing distraction management using digital tools reveals nearly an even split among participants, with 160 participants reporting that they do not use them and approximately 140 reporting using them. The close split indicates there is an important split in user behavior toward distraction control strategies. The larger group not using these kinds of tools may reflect either reliance on legacy methods or unfamiliarity with existing resources used to control distractions. In contrast, the high rate of users who are using distraction control tools indicates recognition of the problem presented by digital distractions. The findings point to an important intervention area since increased awareness and education on the value of these tools to control distractions may support improved productivity and attention among the majority who are not currently utilizing these tools. As a whole, the evidence points to an increased need to make people more aware and better educated on the value of the tools to effectively control distractions.

Word Cloud of Regularly Used Social Media Platforms

The Python code creates a word cloud to represent the most frequently used social media platforms in a column in DataFrame df titled 'Which social media platforms do you use regularly?'. It commences by installing and importing Word Cloud and counter from the module collections. The data is processed by initially removing any missing values in the column in question and subsequently splitting comma-separated strings with more than one platform chosen into separate platform name strings, removing leading/trailing whitespace, and flattening the list into platform_list. A counter object platform_freq is created to count the frequency of each platform. A WordCloud object is then created with given width and height and background colour and colormap, and it makes the word cloud from the frequencies. The word cloud so created is then plotted with matplotlib using the given figure size and interpolation and turning off the axes, and creating the title "Word Cloud of Regularly Used Social Media Platforms".

Output:

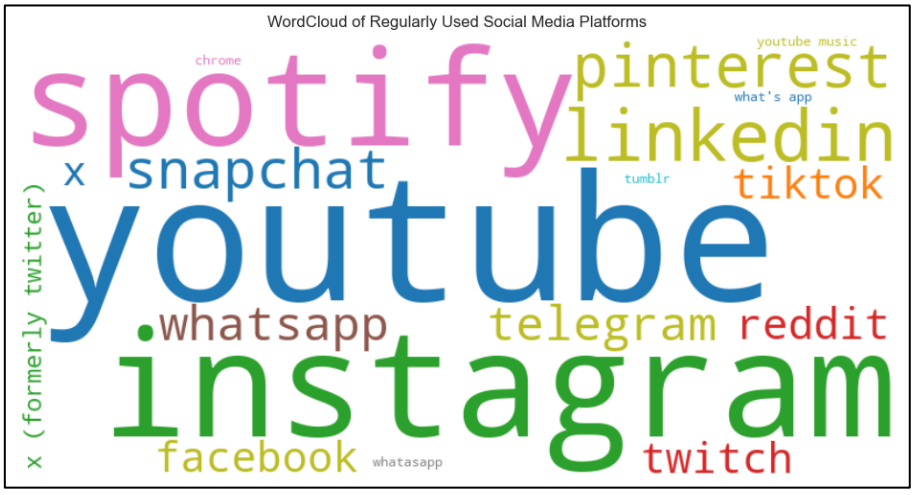


Figure 7: Word cloud of Regularly Used Social Media Platforms

The word cloud visualizing commonly used social networking sites emphasizes the differences in popularity among the individual services, with size indicating the frequency with which the service is mentioned. "Spotify" and "YouTube" are prominently featured, with the implication being that these are among the most popular platforms because of the multimedia aspects of these services. "Instagram" is featured prominently as well, with high rates of visual networking involvement being indicated. Other services like "Snapchat," "TikTok," and "WhatsApp" are visible on the chart, with these indicating relevance in modern-day social networking patterns, particularly among the younger generation. "Facebook" and "LinkedIn" are smaller by comparative measures, with this showing declining informal use by their peak usage levels. The presence of "x (formerly Twitter)" indicates continued awareness despite recent platform modifications. In

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717

general, the word cloud is an illustration of user choice by way of visual representation, with an emphasis on platforms with heavy visual and audio offerings and traditional social networking sites potentially undergoing some level of decline in user interest.

Premium Subscriptions vs. Support for Content Creators

The formulated Python code makes a stacked bar plot to show the connection between paying premium subscriptions on social platforms and having ever paid to support a creator from the pandas DataFrame df. It initially makes a crosstab table called premium_creator_ct with pd.crosstab() to tally up the instances of the combination of responses to the 'do you buy premium subscriptions on social platforms? (such as Spotify or YouTube Premium)' Has anyone ever paid to support a creator? (such as membership, subscription)' columns. Then it makes an area stacked bar plot from this crosstab table using .plot(kind='bar', stacked=True), with defined bar colors. The plot contains an added title, "Premium Subscriptions vs Support for Content Creators", x-axis ("Pays for Premium Subscription"), y-axis ("Number of Users"), and custom x-axis tick labels used ('No', 'Yes'). In the top right corner, the script adds a legend called "Paid to Support Creator" to identify the segments on the bars. The script concludes by modifying the plot layout and showing the ensuing stacked bar chart with plt.tight_layout() and plt.show().

Output:

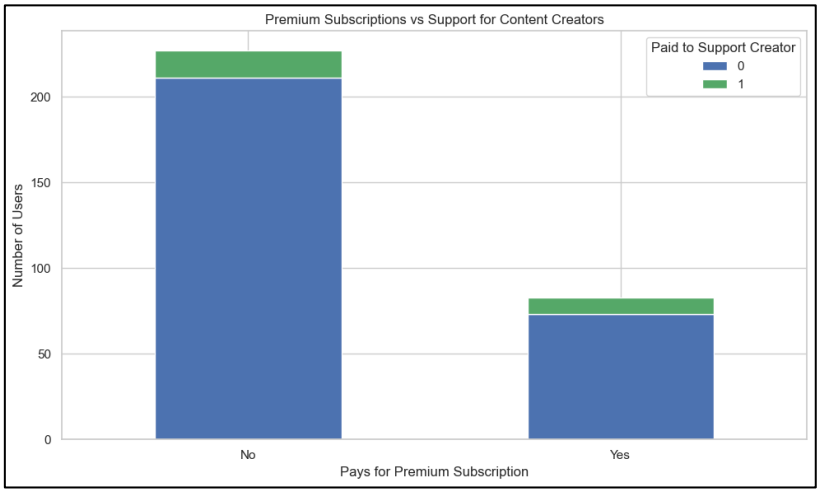


Figure 8: Premium Subscriptions vs. Support for Content Creators

The graph showing support against premium subscriptions to support creators shows an important divergence in user behavior. More than 200 respondents say that they don't subscribe to premium services to support creators, showing an indication that users might want free material or don't want to spend on premium services. Meanwhile, fewer than 50 users say that they subscribe to support creators, so support by way of subscriptions must be relatively low among these users. It highlights some problems facing creators who earn money from subscriptions because the high demand for free access could influence revenue streams. Also, the findings could show an overall attitude to consuming content more broadly in which users value access more than support and will need to consider alternative means to support creators beyond premium subscriptions. Holistically, the graph shows the value of exploring user motivations and actions in the digital content landscape.

Distraction Frequency vs. Use of Management Tools

The Python code constructs a grouped bar plot to show the association between distraction frequency and the use or lack of use of tools or apps to control distractions based on the data in the pandas DataFrame df. It starts by generating a crosstab table called tool_distraction_ct with pd.crosstab() to count the number

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717

of occasions on which each combination of 'how often do you find yourself distracted by social media?' and 'do you use tools or apps to control distractions?' questions is recorded. It then makes a grouped bar plot from this table of crosstabulations with .plot(kind='barh'), with colormap='coolwarm' being the bar colormap to use to represent the use or lack of use of tools and apps to control distractions. The plot has added to it the title "Distraction Frequency vs Use of Management Tools", x-axis title ("Number of Users"), and y-axis title ("Distraction Frequency"), and legend with title "Uses Management Tools" on the lower right-hand side to separate the bars by tool use category. The script concludes with the reformatting of the plot layout and the display of the resultant grouped horizontal bar plot with plt.tight_layout() and plt.show().

Output:

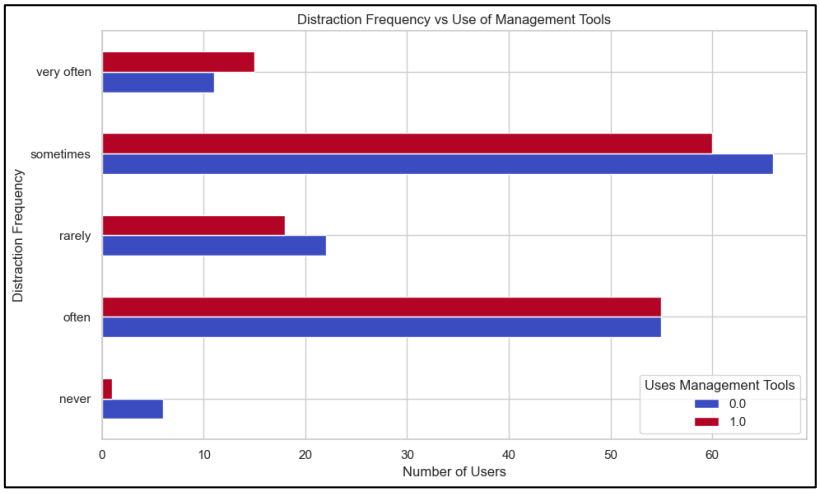


Figure 9: Distraction Frequency vs. Use of Management Tools

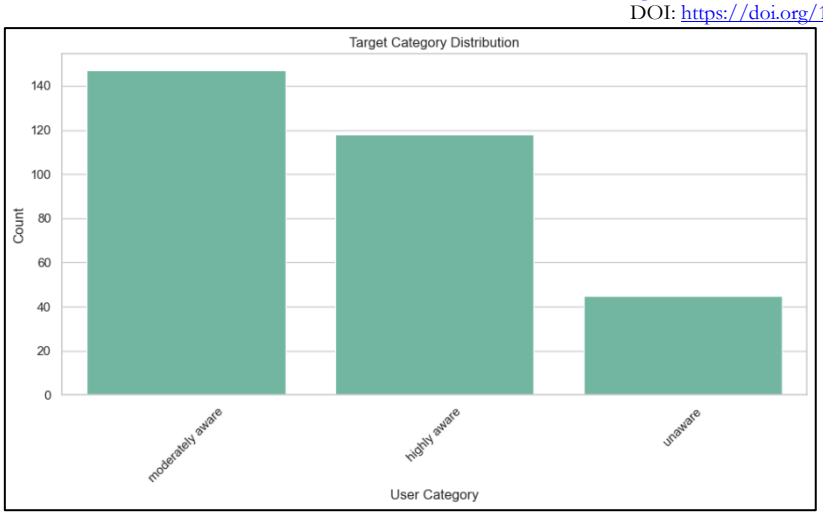
The graph compares distraction frequency and the utilization of management tools, demonstrating clear user patterns. Many respondents claim to encounter distractions "sometimes" (approximately 50 users), and "often" (about 60 users), with the majority in these categories not making use of management tools, represented by the blue bars. In turn, individuals who use tools primarily inhabit the "rarely" and "never" ranks, which suggests that users who utilize these tools are those with fewer distractions. The data reflects that numerous users are frequently distracted; however, there are fewer who try to solve this through the utilization of management tools and/or resources, perhaps indicating an awareness gap for knowledge on how these tools benefit productivity and attention. This discrepancy points to an area to reach out to with distraction management tools and techniques, specifically among those who report being frequently sidetracked, underlining the lack of education on efficient tools and methods to maintain focus and productivity.

Target Category Distribution

The Python code employs the seaborn package (aliased to sns) to create a count plot showing the distribution of one categorical variable called 'category' in the DataFrame df. The function sns.countplot() is used with x='category' to instruct that the categories should be plotted along the x-axis. The order keyword is assigned df['category'].value_counts().index so the categories on the plot are presented in the correct order based on how frequently they occur in the DataFrame from most to least frequently. The code then sets the plot title to "Target Category Distribution", assigns the title "User Category" to the x-axis and "Count" to the y-axis and rotates the x-axis tick marks by 45 degrees so that there is better readability, resizes the plot so that there are no overlaps by calling plt.tight_layout(), and finally presents the plot created by calling plt.show().

Output:

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717



Methodology

Model selection

For this research, we used the multi-model approach so that there would be an exhaustive study and strong predictions by implementing models such as Logistic Regression, Random Forest and XGB-Classifier. Our first-choice model was Logistic Regression, owing to its interpretability at the baseline level. The model elucidates the independent variable and dependent variable's relationship clearly and allows stakeholders to interpret easily how individual features impact the outcome. The coefficients are directly interpretable and are very important to stakeholders who need to interpret the model's choices.

Then, we used a Random Forest classifier to take advantage of its superior ability to assess feature importance and robust classification performance. Random Forest is an ensemble learning approach that uses several decision trees to maximize predictive power and avoid overfitting. From the importance scores produced by this model, we can observe which features are most influential on predictions and provide more insights into the underlying data.

Finally, we made use of the XG-Boost Classifier, which is known to be highly accurate and efficient when working with larger datasets. The XG-Boost uses gradient boosting methods that optimize learning and reduce errors. The model is well suited to identify complex patterns within the data and, therefore, ideal to provide better predictive performance. In combining these three models, we intended to balance interpretability, significance in features, and predictive accuracy.

Model Training and Evaluation

The model training procedure started with a train-test split to partition the dataset into training and testing sets to assess model performance on unseen data. A commonly used 70:30 split for training and testing was used with the added benefit of ensuring a strong evaluation framework. Cross-validation methods were also used to make the performance metrics more reliable by partitioning the training data into several subsets several times, using differing combinations to train the model and validate it on the rest of the portions. This serves to reduce the risk of overfitting and ensure a more general assessment of model performance.

To assess the models properly, we made use of several performance metrics like Accuracy, Precision, Recall, F1-score, and ROC-AUC. Accuracy assesses the model's overall correctness, whereas Precision and Recall offer insight into the performance of the model concerning predictions on the positive class. The F1-score is a harmonic mean of Precision and Recall and presents an average balance between these two. The ROC-AUC measure assesses the discriminative ability of the model in classifying classes and presents a complete overview of its discrimination capabilities.

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717

Lastly, we performed a confusion matrix evaluation to provide deeper insight into model prediction reliability. The confusion matrix breaks down true positives and negatives and false positives and negatives into explicit categories to assess how well instances are classed by individual models. Not only does this evaluation enable us to locate the need to improve, but it is also used to interpret the trade-off among performance metrics to make correct model choice and optimization choices.

Results and Analysis

Model Performance

XBG-Classifier Modelling

The Python code was classified using an XG-Boost model. It is called XGB-Classifier from the xg-boost module and Randomized-Search-CV for finding the best set of hyperparameters. A pipeline object is created consisting of an assumption made elsewhere (preprocessor) and an XGB-Classifier with initial parameters. The dictionary param_dist_xgb specifies the distribution of values for some important hyperparameters such as n-estimators, max_depth, learning_rate, subsample, and colsample_bytree. It uses RandomizedSearch CV to find the best set among these hyperparameters by fitting the xgb_pipeline to the training data (X-train, y-train) through n_iter=20 iterations and cv=5 cross-validation folds. After it identifies the best model, it predicts on the test data (X-test) and assesses its performance by printing out the found best hyperparameters, the accuracy measure, and the classification report consisting of precision, recall, F1-score, and support for all classes.

Output:

Table 1: Showcase XGB Results

Classification Report:

	precision	recall	f1-score	support
highly aware	0.45	0.54	0.49	24
moderately aware	0.43	0.31	0.36	29
unaware	0.42	0.56	0.48	9
accuracy			0.44	62
macro avg	0.43	0.47	0.44	62
weighted avg	0.43	0.44	0.43	62

The above classification report includes the important performance metrics for three categories of user awareness: "highly aware," "moderately aware," and "unaware." Precision for the "highly aware" category is 0.45, revealing that fewer than half the predictions are correct among this group. The recall rate is slightly better at 0.54 and appears to say that 54% of instances are correctly identified by the model. Precision is 0.43 and recall is 0.31 for "moderately aware," revealing difficulties with correctly classifying this group and implications of misclassifications. Precision is 0.42 and recall is 0.56 with the "unaware" category, revealing slightly better performance with detection of instances but still considerable room to improve on this. The overall model precision and recall rate is 0.44, and macro and weighted precisions and recalls round 0.43 to 0.47, revealing a general middle performance rate among all classes. The support values represent the count of instances with this category having the highest value at 29 and implying that increased model performance on this group will probably make the biggest impact.

Random Forest Classifier Modelling

The Python script classifies by employing the Random Forest model. It imports the Random-Forest-Classifier from sklearn. Ensemble and instantiates a pipeline with a preprocessor (simulated to be defined somewhere else) and a Random-Forest-Classifier with a hardcoded random-state. It sets up a parameter grid param_grid_rf with varied values to be tested on the hyperparameters n-estimators, max_depth,

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717

min_samples_split, and min_samples_leaf. Grid-Search-CV is employed to obtain the ideal set of these hyperparameters by tediously going through the grid and assessing the model on the performance using cross-validation (cv=5) on the train data (X-train, y-train). Following fitting the grid search, the best model is employed to predict on the test data (X-test), and its performance is assessed by printing the ideal found hyperparameters, the accuracy score, and the classification report that provides an explicit breakdown of the precision, recall, F1-score, and support by class.

Output:

Table 2: Portraying Random Forest Results

Classification Report:

	precision	recall	f1-score	support
highly aware	0.43	0.42	0.43	24
moderately aware	0.44	0.55	0.49	29
unaware	0.67	0.22	0.33	9
accuracy			0.45	62
macro avg	0.52	0.40	0.42	62
weighted avg	0.47	0.45	0.44	62

The classification report specifies the performance measures against three user awareness types: "highly aware," "moderately aware," and "unaware." The "highly aware" group has a recall and precision of 0.43 and 0.42, respectively, and this group appears to be very difficult to identify correctly by the model with only moderate capabilities to capture correct positives. The "moderately aware" group has slightly lower recall and precision scores with 0.44 and 0.55, and this group appears to be slightly better at detecting correct instances with still sub-ideal performance. The "unaware" group has the highest recall and precision scores with 0.67 and 0.33, with high recall performance and correct predictions, but very poor recall on instances it should observe. The model's overall accuracy is 0.45 with macro and weighted-average recall and precision ranging from 0.44 to 0.52 based on the three categories, and the model appears to perform moderately on all scores on average. The support levels from the classification report show that the "moderately aware" group has the highest level of representation in the set, and thus, improving the model's performance on this group may potentially yield high-quality increases in general classification performance.

Logistic Regression Modelling

Table 3: Displays Logistic Regression Results

Classification Report:

	precision	recall	f1-score	support
highly aware	0.38	0.42	0.40	24
moderately aware	0.41	0.45	0.43	29
unaware	0.50	0.22	0.31	9
accuracy			0.40	62
macro avg	0.43	0.36	0.38	62
weighted avg	0.41	0.40	0.40	62

The classification report provides the performance rates on three user awareness categories: "highly aware," "moderately aware," and "unaware." The "highly aware" category demonstrates 0.38 in precision and 0.42 in recall, revealing that the model detects an acceptable proportion of correct ones but records an appreciable rate of false positives. The "moderately aware" category performs slightly better with precisions

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717

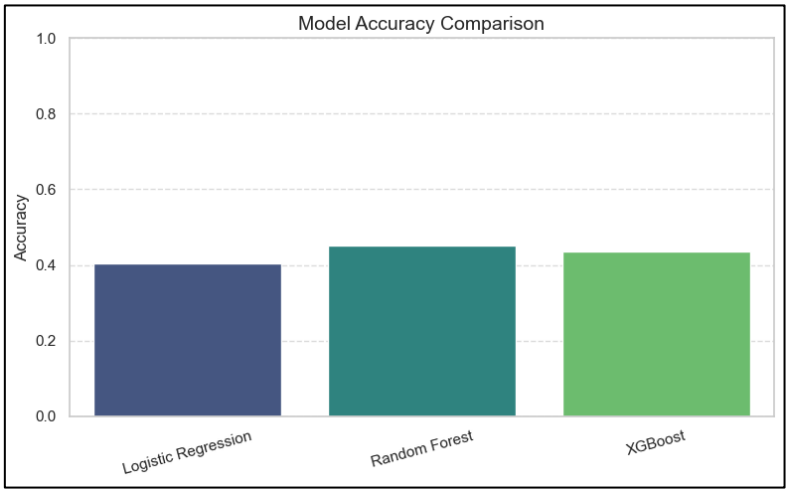
and recalls of 0.41 and 0.45, indicating an improved performance that still shows room for improvement. The "unaware" category demonstrates 0.50 in precision, revealing moderate predictive accuracy, whereas its recall is markedly low at 0.22, indicating that the model does not detect many instances that are there. Generally, the model has an accuracy rate of 0.40 with macro and weighted average precisions and recalls ranging from 0.36 to 0.43. The facts point to the need for improvement, especially on the "unaware" category with the highest precision that performs poorly on recall, highlighting the model's inability to discern well among the categories. The support values are indicative that "moderately aware" is represented the most in the dataset and that this category could be focused on in model improvement to yield positive outcomes.

Comparison of All Models

The Python code compares the performance of three classifying models: Logistic Regression, Random Forest, and XG-Boost. It imports evaluation metrics required (accuracy-score), and tools to plot confusion matrices (though not explicitly used within this snippet). It computes the accuracy of the three models by comparing their predictions (y_pred_lr, y_pred_rf, y_pred_xgb) with the actual test labels (y-test) and saves these scores in a dictionary labeled accuracy_scores. The code then loops through this dictionary and prints the accuracy value of each model. Subsequently, it employs matplotlib and seaborn to plot these scores as a bar plot to easily compare them. The plot has an added title "Model Accuracy Comparison", y-axis with the title "Accuracy" and with a limit set to 0 to 1, rotated x-axis tick marks with the model names used to identify the models on the plot, and uses a grid to make it more readable and adjust the layout to make more room on the plot to display the values before it displays the plot.

Output:

Table 4: Comparison of Model Accuracy



The model comparison chart demonstrates the performance level of three distinct classification models: Logistic Regression, Random Forest, and XG Boost. Logistic Regression only manages to achieve an accuracy rate of around 0.40, signaling an average level of predictive quality. The Random Forest model fares slightly better with an accuracy rate of about 0.45, which implies that its ensemble method increases its predictive power to classify instances more effectively. In turn, the XG Boost model takes the top spot in the comparison with an accuracy rate of around 0.46, projecting its ability to identify complex patterns in the data and showcasing the highest predictive level among the three models. Collectively, the models fail to reach high accuracy levels, but the model improvement from Logistic Regression to XG Boost points to the prospects of boosting performance with more advanced algorithms on classification tasks and underlines the importance of further refinement and optimization to maximize performance level.

Applications in the USA

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717

Real-Time Content Strategy

The use of model outputs can greatly maximize real-time content strategy for brands and organizations seeking to maximize engagement in the USA. Based on user behavior patterns and engagement metrics, models can give insight into what should be posted and when to maximize attention. As an example, if predictive analytics point to some topics being more popular with audiences at particular times of the day or week, brands can plan to maximize attention by uploading posts at these instances. Models can also identify latent trends so that brands can top existing topics or discourses efficiently. In this way, by leading the way instead of following, marketers can place the content so that it can maximize moments when the public will be most interested and, therefore, maximize its probability to go viral and reach more people.

Furthermore, predictive analytics can be used to identify probable viral patterns so that trend-jacking becomes an achievable option for brands. Models can identify rising trends before peak popularity by tracking conversation and engagement metrics on social media. If some topic or hashtag starts to gather steam, brands can immediately produce related content to ride this wave. Introducing brands into influential topics through this anticipatory approach makes them more visible and relevant to consumers. With the successful application of these findings, marketers can develop topical content that resonates with consumers and incites action and the possibility of going viral.

Campaign Optimization

In campaign optimization, predictive modeling assists U.S. brands in making strategic decisions regarding ad spend allocation. Based on an examination of past performance and engagement metrics, brands can see which content is driving the greatest impact and engagement levels and make strategic investments in high-performance content that genuinely resonates with desired audiences. Additionally, messaging that is customized by audience demographics and forecasted interest levels increases campaign effectiveness. Knowing which portions of the audience react well to desired themes or communication styles allows brands to create highly effective personalization messages that drive increased conversion rates and make advertising more efficient and effective.

Additionally, messaging that is customized based on audience demographics and forecasted interest is critical to maximizing campaign effectiveness. Segmentation by audience preferences and behaviors allows brands to tailor messages directly to specific segments. Predictive models can examine user data to make predictions on what demographics will most engage with particular themes or messages and create an ever-deeper level of customization. A brand targeting younger demographics will use edgy, relevant messaging, whereas messages for an older group will highlight value and dependability. The customization level not only increases user engagement but ultimately increases the rate of conversion and thus makes the campaign more economical and successful.

Platform Personalization and Moderation

Predictive modeling is important in driving personalization and moderation on social media platforms. Analyzing user behaviors and engagement patterns allows platforms to identify and recommend content that closely matches individual user interest and liking patterns, thus raising user satisfaction and retention rates. Not only does this personalized approach enhance user experience, but it also leads to increased engagement rates. Engagement predictions can further be used to better moderate content so that platforms can identify and flag or remove poor-quality or dangerous posts that are about to go viral. Platforms can thus create a healthier online community by keeping high-quality posts on top and promoting constructive user engagement.

Aside from personalization, predictive analytics may also be applied in moderating negative or poor-quality posts. Through engagement prediction models, platforms can identify posts that are most likely to receive negative engagement before going viral. Proactive moderating this way maintains an otherwise healthy online community by marking potentially malicious posts, such as misinformation or hate speech, and enabling prompt intervention. As an example, posts with indicative early warning signals for negative

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717

engagement, like high amounts of dislikes or reports, can be quickly reviewed and eliminated. In this way, by targeting quality control with predictive modeling, platforms can make the

Policy and Public Communication

For public communication and policy, predictive models are particularly helpful in projecting how the U.S. public will react to news announcements, policy initiatives, or campaigns. Organizations can forecast public opinion and adjust their messages by examining past reactions and patterns in opinion trends. Public opinion can thus be prepared for by adjusting the messaging tone and rate to match the expectations and concerns of the population. Using the tools of sentiment analysis, policymakers and communicators can develop messages that work with the audience and build support and comprehension for policies. In the long term, this strategic application of predictive analytics can see more effective public outreach and involvement and make efforts timelier and more pertinent.

Furthermore, the use of sentiment analysis can inform the tone and frequency of communication campaigns. When communicators know the emotional state surrounding some topics, they can adapt their messaging to converge with public opinion so that the tone fits the context. In the event of crises, a more empathic and reassuring tone may be called for, whereas announcements with a celebratory tone might require a more positive and uplifting tone. As public opinion continues to be monitored and reforms are made to communication strategies, an organization can better promote public outreach efforts, build credibility, and sustain community participation in the long run towards more efficient governance and public relations.

Discussion and Future Directions

Interpretation of Model Insights

Knowing the importance of the features inferred from predictive models is important to identify what arouses engagement by U.S. users. Some important factors are usually content type, posting hours, user demographics, and interaction history. As an example, video posts will usually receive greater engagement compared to static images, and posts published during high user activity hours will receive more interactions. Further, user characteristics like age and interests profoundly influence patterns of engagement; younger audiences might prefer current topics and lighter posts with comedy value, whereas older crowds might react more to informative and value-based postings. Having this sophisticated insight into the importance of features can assist marketers and content providers with tailoring their strategies to match the behavior and interests of particular audience segments.

Model interpretability and usability are keys to making sure non-technical stakeholders can effectively drive action from insights created through predictive analytics. Simplifying model outputs into easy-to-use dashboards and visualizations allows organizations to make data-informed decisions without highly technical skills on the part of marketing teams, brand managers, and creatives. Clearly explaining how individual features contribute to engagement can make the modeling black box more transparent and build data literacy within an organization. Not only does this make stakeholders more confident in the insight being generated by the model, but it also promotes collaboration among data scientists and business teams and more successful implementation of predictive analytics-backed strategies.

Limitations

There are some limitations to predictive modeling despite its strong advantages. A major challenge is the decay in model performance with time because of quick changes in user behavior and social media platforms. As user behavior and algorithms, and trends are constantly evolving, models that are trained on past data can be less efficient over time and need to be updated and retrained to reflect current levels of accuracy. Having to adapt constantly can tax resources and make the implementation of predictive analytics in dynamic online ecosystems more difficult. Organizations need to be constantly on the lookout to observe model performance and be ready to adjust strategies when platform dynamics change.

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717

A further important limitation is the challenge of capturing sentiment on multimedia material like video and imagery. Although text-based sentiment analysis is well-established, analyzing the emotional tone in visual material is an added challenge. Composition, color palette, and visual storytelling are aspects that can greatly sway audience opinion but are not well represented by current models. This lack implies that businesses will forego important insight into the effect that visual material will have on user interaction. Upcoming studies need to bridge this gap to better comprehend multimedia sentiment and its impact on user interactions on numerous platforms.

Suggested Directions for Future Research:

Boosting the effectiveness of predictive models by incorporating transformer-based NLP models like BERT is one direction to explore in the future. BERT models are great at picking up on subtle language patterns and context and can greatly aid in more effective sentiment determination and engagement predictions. The application of BERT's contextual embeddings can better capture the nuance used in user-generated text and thus make engagement more accurately predicted. This could potentially make it possible for brands to more effectively tailor messaging and content strategies that better engage users and maximize user interaction and satisfaction.

Scaling up to include short video platforms like TikTok and YouTube Shorts is an important future research suggestion. As these platforms expand in popularity, learning more about user behavior and patterns within this type of content format becomes ever more necessary. Creating models that can examine the text and visual components and the dynamic elements of short video, like pacing, sound, and interactivity, will offer a more complete concept regarding user engagement. It can potentially drive novel strategic ideas in marketing, specifically related to the quick-paced nature of short video content, so that brands can better capture audience attention in a crowded online marketplace.

Broader U.S. Digital Ecosystem Implications

Artificial intelligence-powered user behavior forecasting has far-reaching implications on the wider U.S. digital landscape, specifically concerning ethical platform design, advertising practices, and policymaking. Platforms can develop more user-focused experiences with high engagement without degrading user well-being through the power of predictive analytics. As an illustration, ethical considerations can inform the algorithms that incentivize high-quality content and reduce exposure to negative or deceptive material to the greatest extent practicable. Not only does this promote greater user trust, but it also improves the health of the internet environment.

Furthermore, user behavior forecasting insights can guide marketing ethics by cultivating transparency and accountability. Brands that use predictive analytics with responsibility can engage consumers with transparency and respect towards user privacy and choices, and forge deeper relationships with them. In policy-making, user behavior insights from AI can guide regulations protecting consumers and fostering innovation. Policymakers can use these insights to develop regulations that balance the needs of users, brands, and platforms to ensure the online world continues to develop in everyone's interest.

Conclusions

The core goal of this analysis was to implement artificial intelligence together with machine learning approaches for creating predictive models that forecast digital pattern development as well as user commitment through social media interactions within the United States. The data used in this analysis are posts aggregated from two leading online platforms, X-Twitter and Reddit, and consist of user-generated material covering a multifaceted set of topics and opinions. X-Twitter posts are current and real-time and give insight into what is currently being discussed and the opinions that are making headlines, whereas Reddit content provides extensive commentary and user engagement on numerous subreddits. To make the dataset more comprehensive and richer in information, extensive engagement metrics like likes, shares, and comments are used to extend its reach and provide insight into the extent to which users engage with the material presented to them. For this research, we used the multi-model approach so that there would

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717

be an exhaustive study and strong predictions, by implementing models such as Logistic Regression, Random Forest, and XGB-Classifier. To assess the models properly, we made use of several performance metrics like Accuracy, Precision, Recall, F1-score, and ROC-AUC. Logistic Regression only manages to achieve a below-average accuracy, signaling an average level of predictive quality. The Random Forest model fares slightly better with a slightly better accuracy rate, which implies that its ensemble method increases its predictive power to classify instances more effectively. In turn, the XG Boost model took the top spot in the comparison with an accuracy rate, projecting its ability to identify complex patterns in the data and showcasing the highest predictive level among the three models. The use of model outputs can greatly maximize real-time content strategy for brands and organizations seeking to maximize engagement in the USA. Based on user behavior patterns and engagement metrics, models can give insight into what should be posted and when to maximize attention. In campaign optimization, predictive modeling assists U.S. brands in making strategic decisions regarding ad spend allocation. Based on an examination of past performance and engagement metrics, brands can see which content is driving the greatest impact and engagement levels and make strategic investments in high-performance content that genuinely resonates with desired audiences. For public communication and policy, predictive models are particularly helpful in projecting how the U.S. public will react to news announcements, policy initiatives, or campaigns. Boosting the effectiveness of predictive models by incorporating transformer-based NLP models like BERT is one direction to explore in the future.

References

Akshay, M. M., & Mohaneshwar, M. P. The influence and integration of artificial intelligence in social media platforms: a comprehensive examination of user engagement, content creation and future implications for the digital landscape. Digitalization in the marketing and international trade, 91.

Al Montaser, M. A., Ghosh, B. P., Barua, A., Karim, F., Das, B. C., Shawon, R. E. R., & Chowdhury, M. S. R. (2025). Sentiment analysis of social media data: Business insights and consumer behavior trends in the USA. Edelweiss Applied Science and Technology, 9(1), 545-565.

Amoo, O. O., Umoga, U. J., Atadoga, A., Sodiya, E. O., Amoo, O. O., ... & Atadoga, A. (2024). AI-driven personalization in web content delivery: A comparative study of user engagement in the USA and the UK. World Journal of Advanced Research and Reviews, 21(2), 887-902.

Bag, S., Srivastava, G., Bashir, M. M. A., Kumari, S., Giannakis, M., & Chowdhury, A. H. (2022). Journey of customers in this digital era: Understanding the role of artificial intelligence technologies in user engagement and conversion. Benchmarking: An International Journal, 29(7), 2074-2098.

Chouksey, A., Shovon, M. S. S., Tannier, N. R., Bhowmik, P. K., Hossain, M., Rahman, M. S., ... & Hossain, M. S. (2023). Machine Learning-Based Risk Prediction Model for Loan Applications: Enhancing Decision-Making and Default Prevention. Journal of Business and Management Studies, 5(6), 160-176.

Dahhiya, R., Dahiya, V. K., Agarwal, N., Kolluru, V. K., Challagundla, Y., & Chintakunta, A. N. (2025, February). Predictive Analytics in AI Marketing: Transforming Consumer Engagement. In 2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking (CICTN) (pp. 692-697). IEEE.

Dang, P. H., & Van Thich, N. (2025). Factors influencing purchase decisions on social media platforms: The role of explainable artificial intelligence. Edelweiss Applied Science and Technology, 9(2), 1228-1244.

Das, B. C., Sarker, B., Saha, A., Bishnu, K. K., Sartaz, M. S., Hasanuzzaman, M., ... & Khan, M. M. (2025). Detecting Cryptocurrency Scams in the USA: A Machine Learning-Based Analysis of Scam Patterns and Behaviors. Journal of Ecohumanism, 4(2), 2091-2111.

Emma, O., & Tawkoski, J. (2022). Behavioral Analytics for User Engagement: Leveraging AI to Improve Retention Rates.

Gupta, Y., & Khan, F. M. (2024). Role of artificial intelligence in customer engagement: a systematic review and future research directions. Journal of Modelling in Management, 19(5), 1535-1565.

Ho, S. Y., Bodoff, D., & Tam, K. Y. (2011). Timing of adaptive web personalization and its effects on online consumer behavior. Information systems research, 22(3), 660-679.

Islam, M. Z. ., Islam, M. S., das, B. C. ., Reza, S. A. ., Bhowmik, P. K. ., Bishnu, K. K. ., Rahman, M. S. ., Chowdhury, R., & Pant, L. . (2025). Machine Learning-Based Detection and Analysis of Suspicious Activities in Bitcoin Wallet Transactions in the USA. Journal of Ecohumanism, 4(1), 3714 –. https://doi.org/10.62754/joe.v4i1.6214

Islam, M. R., Hossain, M., Alam, M., Khan, M. M., Rabbi, M. M. K., Rabby, M. F., ... & Tarafder, M. T. R. (2025). Leveraging Machine Learning for Insights and Predictions in Synthetic E-commerce Data in the USA: A Comprehensive Analysis. Journal of Ecohumanism, 4(2), 2394-2420.

Islam, M. S., Bashir, M., Rahman, S., Al Montaser, M. A., Bortty, J. C., Nishan, A., & Haque, M. R. (2025). Machine Learning-Based Cryptocurrency Prediction: Enhancing Market Forecasting with Advanced Predictive Models. Journal of Ecohumanism, 4(2), 2498-2519.

Jui, A. H., Alam, S., Nasiruddin, M., Ahmed, A., Mohaimin, M. R., Rahman, M. K., ... & Akter, R. (2023). Understanding Negative Equity Trends in US Housing Markets: A Machine Learning Approach to Predictive Analysis. Journal of Economics, Finance and Accounting Studies, 5(6), 99-120.

Volume: 4, No: 4, pp. 119 – 141
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6717
Kobsa, A. (2017). Privacy-enhanced web personalization. In The adaptive web: Methods and strategies of web personalization (pp. 628-670). Berlin, Heidelberg: Springer Berlin Heidelberg.

Mohaimin, M. R., Das, B. C., Akter, R., Anonna, F. R., Hasanuzzaman, M., Chowdhury, B. R., & Alam, S. (2025). Predictive Analytics for Telecom Customer Churn: Enhancing Retention Strategies in the US Market. Journal of Computer Science and Technology Studies, 7(1), 30-45.

Okeleke, P. A., Ajiga, D., Folorunsho, S. O., & Ezeigweneme, C. (2024). Predictive analytics for market trends using AI: A study in consumer behavior. International Journal of Engineering Research Updates, 7(1), 36-49.

Patil, R., Shivashankar, K., Porapur, S. M., & Kagawade, S. (2024). The role of ai-driven social media marketing in shaping consumer purchasing behaviour: An empirical analysis of personalization, predictive analytics, and engagement. In ITM Web of Conferences (Vol. 68, p. 01032). EDP Sciences.

Rathee, R. (2025). From Businesses-Centric to Consumers-Centric: A Shift of Power in AI-Driven Social Media. In Marketing 5.0 (pp. 89-102). Emerald Publishing Limited.

Rana, M. S., Chouksey, A., Das, B. C., Reza, S. A., Chowdhury, M. S. R., Sizan, M. M. H., & Shawon, R. E. R. (2023). Evaluating the Effectiveness of Different Machine Learning Models in Predicting Customer Churn in the USA. Journal of Business and Management Studies, 5(5), 267-281.

Rana, M. S., Chouksey, A., Hossain, S., Sumsuzoha, M., Bhowmik, P. K., Hossain, M., ... & Zeeshan, M. A. F. (2025). AI-Driven Predictive Modeling for Banking Customer Churn: Insights for the US Financial Sector. Journal of Ecohumanism, 4(1), 3478-3497.

Ray, R. K., Sumsuzoha, M., Faisal, M. H., Chowdhury, S. S., Rahman, Z., Hossain, E., ... & Rahman, M. S. (2025). Harnessing Machine Learning and AI to Analyze the Impact of Digital Finance on Urban Economic Resilience in the USA. Journal of Ecohumanism, 4(2), 1417-1442.

Salonen, V., & Karjaluoto, H. (2016). Web personalization: the state of the art and future avenues for research and practice. Telematics and Informatics, 33(4), 1088-1104.

Sharma, K. K., Kumar, B., Chaudhary, S., & Panwar, R. (2025). Enhancing customer experience through AI-enabled content personalization in e-commerce marketing. Advances in digital marketing in the era of artificial intelligence, 7-32.

Sizan, M. M. H., Chouksey, A., Miah, M. N. I., Pant, L., Ridoy, M. H., Sayeed, A. A., & Khan, M. T. (2025). Bankruptcy Prediction for US Businesses: Leveraging Machine Learning for Financial Stability. Journal of Business and Management Studies, 7(1), 01-14.

Sizan, M. M. H. ., Chouksey, A. ., Tannier, N. R. ., Jobaer, M. A. A., Akter, J. ., Roy, A. ., Ridoy, M. H. ., Sartaz, M. S. ., & Islam, D. A. (2025). Advanced Machine Learning Approaches for Credit Card Fraud Detection in the USA: A Comprehensive Analysis. Journal of Ecohumanism, 4(2), 883 –. https://doi.org/10.62754/joe.v4i2.6377

Teepapal, T. (2025). AI-driven personalization: Unraveling consumer perceptions in social media engagement. Computers in Human Behavior, 165, 108549.