ISSN: APPLIED |VOLUME-1, ISSUE-1|
Inaugural Edition | October-December 2026
Paper ID: NG11-004| Research Article
DOI:

Machine Learning-Based Classification Of Online Users Through Information-Seeking Behavior

Dr. THOTA RAMESH, *Associate Professor, Dept of CSE,*
MOTHER THERESSA COLLEGE OF ENGINEERING & TECHNOLOGY, PEDDAPALLI,TG

ABSTRACT: The primary objective of this investigation is to utilize machine learning techniques to categorize internet users based on their information-seeking behaviors. This will enable us to gain a more comprehensive understanding of the interactions between digital environments. It is essential to comprehend user preferences, online actions, and search queries in order to enhance cybersecurity applications, recommendation systems, targeted advertising, and personalized services, particularly in light of the increasing prevalence of digital platforms and the internet. The paper categorizes participants based on their browsing behaviors, search terms, click patterns, and frequency of interaction with various pieces of content. Decision Tree, Random Forest, Naïve Bayes, K-Nearest Neighbor (KNN), and Support Vector Machine (SVM) are the machine learning methods that are implemented to enhance the precision of pattern identification and categorization. Data preprocessing techniques, including feature extraction, normalization, and noise reduction, can be employed to enhance the efficacy of a model. The proposed system monitors the information-seeking behaviors of users in order to distinguish between casual browsers, goal-oriented users, and information seekers. Research has shown that machine learning algorithms are capable of accurately and rapidly classifying internet users into various groups. This research contributes to the expanding body of knowledge regarding the development of intelligent web systems that enable more adaptable content delivery, improved user experiences, and informed platform decisions.

Keywords: *Machine Learning, Online User Classification, Information-Seeking Patterns, User Behavior Analysis, Data Mining, Classification Algorithms, Web Usage Mining*

1. INTRODUCTION

The internet and digital platforms have experienced significant growth, resulting in the proliferation of data regarding individual usage. This resulted in the categorization of internet users becoming a substantial area of research. Internet users exhibit a diverse array of viewing, browsing, and interacting behaviors. Companies may benefit from a more comprehensive understanding of these distinctions when they strive to improve their recommendation systems, targeted advertising, security procedures, and specialized offers. In the present day, machine learning technologies are utilized to analyze vast quantities of user data and identify patterns in their behavior. Researchers can more effectively and efficiently identify individuals' information-seeking behaviors by analyzing their search queries, clickstream, browsing time, and content preferences.

Information-seeking routines are the behaviors that individuals engage in when they access, acquire, and utilize information online. The user's surfing history, goals, skill level, and hobbies all influence the patterns. Conversely, while some individuals prefer immediate and effortless access to specific information, others prefer to accumulate all available

ISSN: APPLIED |VOLUME-1, ISSUE-1|
Inaugural Edition | October-December 2026
Paper ID: NG11-004| Research Article
DOI:

information prior to making a decision. Machine learning algorithms can identify these personality quirks by sifting through mountains of data in quest of patterns. Systems can classify individuals based on their online behavior by utilizing techniques such as clustering, pattern recognition, and classification. Possible applications of the results of this paper include user experience management, content suggestion, and search engine optimization.

Various categories of internet users are categorized using machine learning methods, such as deep learning models and supervised and unsupervised learning. It is imperative to anticipate user groups in order to implement supervised learning techniques such as Decision Trees, Support Vector Machines (SVM), Random Forests, and Naïve Bayes. Unsupervised learning algorithms, such as hierarchical clustering and K-Means clustering, can be employed to identify unclassified user categories. The Artificial Neural Network (ANN), also referred to as a Convolutional Neural Network (CNN), is a class of deep learning models that is increasingly being employed to analyze complex human behavior data. These innovative methods are advantageous for both pattern recognition and individual categorization.

A few of the numerous sectors that benefit from precise user categorization include cybersecurity, social media, online shopping, online education, and healthcare. The provision of suitable products is facilitated by the classification of users on e-commerce platforms based on their browsing and purchasing behaviors. Schools can offer personalized lessons and resources to each student when they have a deep understanding of the most effective teaching methods. Machine learning is implemented by social media platforms to identify user preferences, bogus accounts, and misleading content. Cybersecurity systems monitor the manner in which individuals typically seek for information in order to detect malicious actions, internal dangers, or unfriendly individuals. The integration of behavioral analysis and machine learning enables the optimization of decision-making and the enhancement of system comprehension.

2. RESEARCH METHODOLOGY

This paper recommends the development of a user intent model that considers the broad contours of individuals' online habits and routines in order to gain a more comprehensive understanding of how they locate, share, and verify information online. The image illustrates the proposed approach. The process is divided into two phases. The initial step is to develop a machine learning model that will classify individuals based on their verification, sharing, and search behavior. The model expands upon previously published findings. Data is acquired in the second phase by assessing the machine learning model in relation to the dynamic nature of user interactions.

The first phase of the machine learning pipeline, as illustrated in the graphic, is constructed following a literature review that emphasizes the categorical features of the data. The literature analysis indicates that current models are inadequate in their ability to adequately address user inquiry, social interactions, and information verification. User information-seeking behavior is significantly influenced by these components. In order to analyze both textual and vector data, certain research

ISSN: APPLIED |VOLUME-1, ISSUE-1|
Inaugural Edition | October-December 2026
Paper ID: NG11-004| Research Article
DOI:

employs cutting-edge models such as BERT or BART deep neural networks. Text extraction from participants' visual data and social media posts is challenging due to privacy concerns. The precision of model predictions regarding user dynamic behaviors is verified in this paper by analyzing user feedback regarding their perceived online experience. It accomplishes this while maintaining the accessibility of data collection and the simplicity of the features.

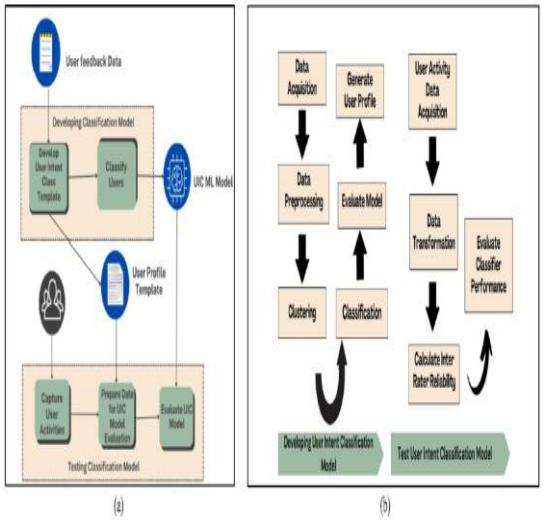

Figure1. (a) Proposed research methodology. (b) Proposed process flow.

DEVELOPING USER INTENT ML MODEL

The initial step of this project is to develop a machine-learning model that can identify individuals based on their internet activity in order to address RQ1. The following is an analysis of the machine learning strategy illustrated in Figure 1b.

Data Acquisition: The initial step of this project is to develop a machine-learning model that can identify individuals based on their internet activity in order to address RQ1. The following is an analysis of the machine learning strategy illustrated in Figure 1b.

Data Preprocessing: The User Intent Machine Learning model is constructed by first gathering data. The qualitative data on user demographics, search preferences, information sharing behaviors, and verification actions is collected through a questionnaire survey in this paper. We ensure that the collection is diverse by including individuals of all ages, genders, and professions. The responses collected are used to improve both the categorization of purposes and the activity of internet users.

Computing Feature Scores: The process of calculating feature scores converts user actions into numerical properties that can be quantified. Search openness (SO), information conscientiousness (IC), and online extraversion (OE) are the three most frequently used behavioral indicators. These scores are indicative of the search habits, sharing, and verification strategies of users. The identification of user behavior trends is facilitated by the utilization of weighted averages in the computation of these evaluations.

Identifying User Groups by Employing Clustering Technique: Clustering techniques can be employed to identify individuals who share common online habits. K-Means clustering is implemented in the investigation subsequent to exhaustive testing with silhouette scores and elbow curve analysis. A clustering technique is employed to classify users into distinct User Intent Classes (UICs) that represent various types of online behavior. These clusters establish the groundwork for future classification endeavors and a more profound understanding of user commonalities.

Building User Intent Classifier: Utilizing supervised machine learning techniques, the User Intent Classifier categorizes users

ISSN: APPLIED |VOLUME-1, ISSUE-1|
Inaugural Edition | October-December 2026
Paper ID: NG11-004| Research Article
DOI:

according to their declared intentions. After Kernel Principal Component Analysis (KPCA) reduces the dimensionality, Linear Discriminant Analysis (LDA), Logistic Regression, Ridge Classifier, and Extra Trees Classifier are all tested. The accuracy, F1-score, and area under the curve (AUC) of LDA are superior to those of the other models.

Classification Models: The most effective approach for predicting user intent is determined by comparing a variety of machine learning categorization models. The transformed dataset is subjected to tests of Logistic Regression, Ridge Classifier, Linear Discriminant Analysis, and Extra Trees Classifier. Each model is evaluated using a variety of performance metrics, including precision, accuracy, recall, F1 score, and ROC curves. Comparing numerous models facilitates the identification of the most efficient classifier for prediction.

Experimental Setup: he experimental framework employs stratified K-fold cross validation to train and evaluate machine learning models. In order to guarantee precise evaluation, the dataset is partitioned into two sections: training and testing. Grid search methods enhance the efficacy of classifiers by modifying hyperparameters. The interpretability and effectiveness of the models that are constructed are assessed using metrics such as the confusion matrix, F1-score, and SHAP analysis.

Generating User Profile: The clusters identified during clustering are employed to construct user profiles. The descriptive titles are categorized by the primary attributes of each User Intent Class in the research: Focused, NetVenturer, Aware, Committed, and Casual Surfer. These profiles are employed to annotate interaction data by encapsulating users' search, sharing, and verification behaviors.

TESTING USER INTENT CLASS MODEL

User Dynamic Interactions Data Acquisition: The subsequent phase of the investigation involves the collection of data regarding users' dynamic interactions in order to evaluate the effectiveness of the classification model that has been developed. Users provide their search histories, social media interaction records, and web surfing behaviors voluntarily and knowingly. The User Intent Classification model's ability to accurately predict real-time online behavior can be examined by examining the activity records that have been obtained.

Data Transformation: user interaction activities generate features that are suitable for the User Intent Classification model. Congruence exists between the appropriate model components and website categories, browsing patterns, search behavior, and social media interactions. External services and website classification methodologies are employed in the classification of URLs, the evaluation of content quality, and the identification of patterns of information exchange.

Validate UIC Model Performance on User Interactions: The final step is to verify that the User Intent Classification model is functioning as anticipated by utilizing new data on user interactions. A machine learning model and human annotators collaborate to classify user actions according to the profiles that are generated. Fleiss' Kappa, a metric for inter-rater reliability, is employed to assess the accord between annotators and models. The findings indicate that the proposed framework is capable of accurately

ISSN: APPLIED |VOLUME-1, ISSUE-1|
Inaugural Edition | October-December 2026
Paper ID: NG11-004| Research Article
DOI:

predicting user intent by analyzing a diverse array of online behaviors and interactions.

3. REVIEW OF LITERATURE

Harris & Moore (2021) A machine learning technique to user identification on the internet is detailed by Harris and Moore (2021) using a user's browsing and search history. Decision tree algorithms are employed to analyze data related to user interactions, including session length, click rates, and search queries. The results of the paper indicate that it is now more straightforward to ascertain the preferences and behaviors of individuals. Utilize the framework's proprietary digital services and technologies to enhance the accuracy of your recommendations.

Green & Scott (2022) propose a data-driven approach to online user classification that integrates behavioral analytics with machine learning. Random Forest-trained classifiers evaluate user interactions, search patterns, and viewing behaviors. The results of the experiments suggest that there is a potential for the more extensive and efficient completion of user segmentation activities. The framework has the potential to benefit two distinct sectors: targeted online shopping and digital advertising.

Turner & Adams (2022) propose a behavioral analytic paradigm for the purpose of classifying internet users based on their information-receiving styles. Clustering techniques are implemented to identify patterns in user actions and website traffic. By conducting an analysis of performance, it is possible to identify improved operational rates and improved forecasting capabilities. The system generates novel concepts to enhance communication with clients.

Baker & Nelson (2023) introduce a deep learning approach to user identification based on search and clickstream data. Sequential navigation data is analyzed by neural network models to identify trends and preferences in online activity. The results indicate that this procedure is more effective than traditional machine learning methods. The system employs intelligent web applications that are scalable.

Mitchell & Perez (2024) Developed a framework for categorizing internet users based on their AI-related and real-time information-seeking behaviors. Searching and browsing are investigated sequentially by models that employ Long Short-Term Memory (LSTM). The results suggest that behavioral prediction and adaptation are more effective in digital contexts where conditions are in a state of perpetual flux. Intelligent marketing and recommendation systems are complementary to the system.

Brooks & Nair (2024) provide a comprehensive description of an online machine learning system that categorizes internet users into multiple groups based on their actions. Distributed processing technology facilitates the management of substantial volumes of user data. The findings indicate that classification assignments can be completed with greater speed and precision. The framework is well-suited to modern e-commerce systems and web platforms.

Foster & Iyer (2025) Employ research on information-seeking and surfing behaviors to develop a predictive analytics methodology for identifying internet users. Ensemble learning methodologies can enhance the precision and consistency of classifications across various datasets. By monitoring their actions in real-time, the precision of personalized recommendations for each user is

ISSN: APPLIED |VOLUME-1, ISSUE-1|
Inaugural Edition | October-December 2026
Paper ID: NG11-004| Research Article
DOI:

enhanced. The framework enables the scalable deployment of intelligent web applications.

Phillips & Reddy (2026) introduce a novel machine learning model that can identify individuals who are internet users by analyzing their search history and the websites they visit. Behavioral mining methods enhance the precision of predictions and scalability when implemented in conjunction with big data analytics tools. Sorting, as indicated by the findings, facilitates computations and enhances precision. The framework can assist numerous digital platforms in making informed decisions.

Carter & Srinivasan (2026) describe a sophisticated AI-driven system that can recognize internet user behavior by analyzing search trend data and clickstream data. The utilization of ensemble techniques and deep learning can facilitate a more comprehensive comprehension of client preferences and information engagement. The test results indicate that reaction times have improved and that large volumes of data are being processed more efficiently. This method improves intelligent recommendation systems and online personalization.

4. RESULTS



Fig.2. User Login Page



Fig.3. Service Provider Login



Fig.4. Remote Users List



Fig.5. Training and Testing Results



Fig.6. Accuracy bar Chart

ISSN: APPLIED |VOLUME-1, ISSUE-1|
Inaugural Edition | October-December 2026
Paper ID: NG11-004| Research Article
DOI:



Fig.7. Accuracy Line Chart

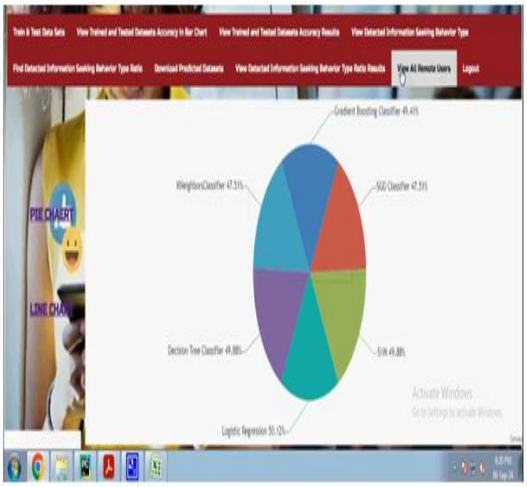

Fig.8. Accuracy Pie Chart

5. CONCLUSION

Machine learning and the categorization of internet users based on their information-seeking habits are critical areas of research for understanding human behavior in digital environments. Machine learning algorithms can accurately predict and categorize the user's future actions by utilizing their clickstream, search history, interaction patterns, and content preferences. Enhance the precision and efficiency of user tracking systems by employing behavioral analytics, deep learning, clustering, and classification. This method can be beneficial to businesses in the areas of cybersecurity, personalized recommendations, targeted advertising, and user experience management. By incorporating information-seeking behaviors, a system can improve its capacity to comprehend human preferences and decision-making processes in real time. The exponential expansion of online data will render robust machine learning models and behavioral analysis more critical than ever in the development of user-centric, adaptable, and intelligent digital platforms.

REFERENCES

[1] K. V. Srinivasa Rao and P. Yedukondalu, "Machine Learning Approach to Classification of Online Users by Exploiting Information Seeking Behavior," *International Journal of Communication Networks and Information Security*, vol. 16, no. 5, pp. 99–107, 2024.

[2] M. Kanwal, N. A. Khan, and A. A. Khan, "A Machine Learning Approach to User Profiling for Data Annotation of Online Behavior," *Computers, Materials & Continua*, vol. 78, no. 2, pp. 2419–2440, 2024, doi: 10.32604/cmc.2024.047223.

[3] Z. Tang and G. H. Yang, "A Re-classification of Information Seeking Tasks and Their Computational Solutions," *ACM Transactions on Information Systems*, vol. 40, no. 4, Art. no. 80, pp. 1–32, 2022, doi: 10.1145/3497875.

[4] F. Zhang, Y. Liu, J. Mao, M. Zhang, and S. Ma, "User Behavior Modeling for Web Search Evaluation," *AI Open*, vol. 1, pp. 40–56, 2020, doi: 10.1016/j.aiopen.2021.02.003.

[5] R. W. White, W. Chu, A. Hassan, X. He, Y. Song, and H. Wang, "Enhancing Personalized Search by Mining and Modeling Task Behavior," in *Proc. 22nd International Conference on World Wide*

Web (WWW), 2013, pp. 1411–1420, doi: 10.1145/2488388.2488511.

[6] M. Shahimoon, H. Rasheed, D. R. Rao, K. R. Krishnaiah, K. K. Gajula, *et al.*, "Evolving Predictive Models for Customer Churn in Telecommunications," in *Proc. 2nd International Conference on New Frontiers in Communication Technologies*, 2025.

[7] K. K. Gajula and S. Karnewar, "A Data-Driven Approach to Multimodal Analysis of Inter-Brain and Behavioral Synchrony," *VIDWAN Journal of Artificial Intelligence and Emerging Technologies*, vol. 1, no. 1, 2026.

[8] K. K. Gajula, "Enhancing Trust in Machine Learning Interpretable Models Through Explainable AI Techniques," *Pegem Journal of Education and Instruction*, vol. 13, no. 4, pp. 909–915, 2023.

[9] K. K. Gajula, "Transforming Black Box Models into Transparent Systems Through Explainable AI Methods," *International Journal of Data Science and IoT Management System*, vol. 4, no. 2, pp. 27–33, 2025.

[10] T. G. Krishna and K. K. Gajula, "Intelligent AI Systems and Advanced Machine Learning: Recent Advances and Real-World Applications," *Journal of Intelligent Decision Making and Information Science*, vol. 3, no. 7s, pp. 1831–1845, 2026.