JDEBM
Journal of Digital Engineering and Business Management

Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/15
DOI: https://doi.org/10.5281/zenodo.20593604

CLASSIFICATION OF ONLINE USERS USING MACHINE LEARNING AND INFORMATION-SEEKING PATTERNS

[#1]**Chilumula Indu,** *Department of MCA,*
[#2]**Mr. K. Nagendra Prasad,** *Assistant Professor, Department of MCA,*
Vaageswari College of Engineering(Autonomous), Karimnagar, TG.

ABSTRACT: In order to improve comprehension of interactions within digital contexts, the primary objective of this project is to classify internet users based on information-seeking behaviors and machine learning methodologies. It is imperative to understand user preferences, online behaviors, and search queries in order to improve cybersecurity applications, recommendation systems, targeted advertising, and personalized services, given the rapid proliferation of the internet and digital platforms. The research classifies individuals based on a variety of online behaviors, including the frequency of content interactions, browser history, search queries, and click patterns. Machine learning methodologies, such as Naïve Bayes, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree, and Random Forest, are implemented to identify patterns and improve classification precision. The efficacy of a model is improved by data preprocessing methods, including feature extraction, normalization, and noise reduction. By analyzing their information-seeking behaviors, the proposed system can distinguish between casual browsers, goal-oriented users, and information seekers. Machine learning algorithms are capable of accurately and rapidly classifying internet users into multiple categories, as evidenced by experiments. This research contributes to the development of intelligent web systems that improve user experiences, facilitate the delivery of content more flexibly, and assist users in making informed decisions on digital platforms.

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

1. INTRODUCTION

The significant expansion of digital platforms and the internet has resulted in a substantial quantity of data pertaining to individual usage. As a result, the classification of internet consumers has become an essential area of research. The activities of Internet consumers in terms of viewing, searching, and interaction are characterized by substantial variability. Organizations can improve their personalized services, security measures, recommendation systems, and focused advertising by comprehending these distinctions. Currently, machine learning (ML) methodologies are advantageous for the evaluation of extensive user data and the identification of logical patterns in behavior. Researchers can more accurately and quickly classify individuals into distinct groups by analyzing information-seeking activities, such as search queries, clickstreams, browsing duration, and content preferences.

The means by which individuals locate, access, and employ information online are referred to as information-seeking routines. The patterns are influenced by the surfing history, level of competence, objectives, and interests of each user.

JDEBM
Journal of Digital Engineering and Business Management
Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/15
DOI: https://doi.org/10.5281/zenodo.20593604

However, some individuals prefer to evaluate facts from multiple sources before making a decision, while others desire immediate and direct access to specific information. By analyzing prior user activity and revealing concealed patterns within extensive datasets, machine learning algorithms can identify these behavioral traits. Systems can classify individuals according to their online activities by utilizing techniques such as clustering, classification, and pattern recognition. The insights gained from this research have the potential to improve search engine optimization, content suggestion, and user experience management.

Machine learning methodologies, such as supervised learning, unsupervised learning, and deep learning models, are employed to categorize online users into distinct categories. When employing supervised learning methodologies, such as Decision Trees, Support Vector Machines (SVM), Random Forests, and Naïve Bayes, it is frequently necessary to anticipate which user groups will be employed. Hierarchical clustering and K-Means clustering are examples of unsupervised learning methods that enable the identification of unclassified user categories. Artificial Neural Networks (ANN) and Recurrent Neural Networks (RNN) are gaining popularity as deep learning models for managing complex data related to human behavior. These sophisticated methods facilitate the identification of patterns and improve the accuracy of individual classification.

The categorization of internet users is advantageous to numerous industries, such as cybersecurity, social media, e-commerce, e-learning, and healthcare. The presentation of relevant products is facilitated by user classification on e-commerce platforms, which is based on the perusing and purchasing behaviors of users. Educational institutions can establish personalized classrooms and provide content that fulfills the requirements of each pupil by comprehending the most effective learning methods. Machine learning is implemented by social networking platforms to identify fraudulent accounts, identify false information, and determine user preferences. In order to identify suspicious activities, internal threats, or hostile users, cybersecurity systems analyze information-seeking behaviors. By combining behavioral analysis with machine learning, the system's comprehension is improved, and the decision-making process is optimized.

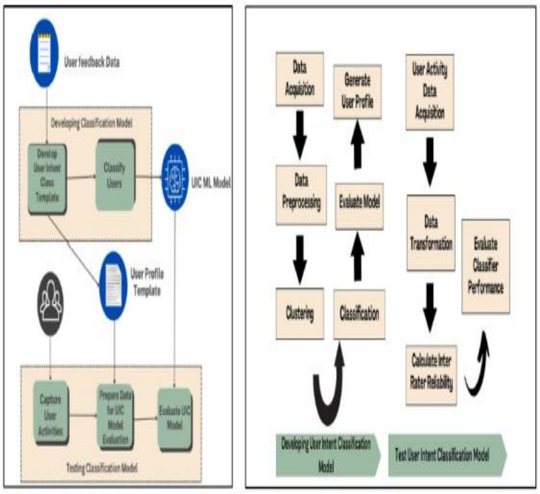
2. RESEARCH METHODOLOGY

This research intends to develop a user intent model that encapsulates the general traits of users regarding online behavior and practices by examining the online behaviors and preferences employed in the processes of locating, sharing, and validating information. The illustration illustrates the proposed procedure. The process consists of two steps. The initial step involves constructing a Machine Learning (ML) model to categorize users according to their online search, sharing, and verification behaviors. The model elaborates on the previously reported work. Phase II entails collecting data through the assessment of the machine learning model in relation to users' dynamic interactions.

Upon reviewing pertinent literature, the Phase 1 machine learning pipeline—

Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/15
DOI: https://doi.org/10.5281/zenodo.20593604

illustrated in the figure—is constructed, emphasizing the categorical attributes of the data. The literature review indicates that user search, social interactions, and information verification are critical components of user information-seeking behavior that existing models inadequately address. Other studies employ cutting-edge models such as BERT or BART deep neural networks and encompass many data kinds, including textual and vector formats. Extracting text from social media and visual data from participants is challenging due to privacy concerns. This project aims to test the accuracy of model predictions on user dynamic behaviors by utilizing user input on their perceived internet experience, while maintaining simplicity in features and accessibility in data collection.



(a) (b)

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

DEVELOPING USER INTENT ML MODEL

The initial phase of this project entails the creation of a machine-learning model capable of categorizing individuals according to their online behavior and addressing RQ1. The machine learning procedure illustrated in Figure 1b is detailed in the subsequent sections.

Data Acquisition: The initial phase in constructing the User Intent Machine Learning model is data collection. The research gathers qualitative data on users' demographics, search tendencies, information-sharing behaviors, and verification activities through a questionnaire-based survey. To guarantee diversity in the collection, individuals of all ages, genders, and professions are incorporated. The analysis of online user behavior and purpose classification relies on the gathered responses.

Data Preprocessing: The aggregated categorical data is sanitized, structured, and converted into a machine-readable format via data preparation. Textual responses are distilled into concise categorical values, amalgamated skewed categories, and extraneous materials are eliminated. One-hot encoding is employed to transform categorical data into binary vectors appropriate for machine learning techniques. This procedure enhances data quality and readies it for feature analysis and model training.

Computing Feature Scores: The calculation of feature scores quantifies user behavior into measurable numerical attributes. The three primary behavioral metrics offered are Search Openness (SO), Online Extraversion (OE), and Information Conscientiousness (IC). These scores reflect users' search behaviors, sharing practices, and verification methods. These ratings are computed using weighted averages, facilitating the identification of user behavioral trends.

Identifying User Groups by Employing Clustering Technique: Clustering algorithms are employed to identify individuals with analogous online behavioral traits. Following an assessment of diverse clustering techniques through

Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/15
DOI: https://doi.org/10.5281/zenodo.20593604

silhouette scores and elbow curve analysis, the research employs K-Means clustering. Users are categorized into distinct User Intent Classes (UICs) by a clustering technique that represents various types of online behavior. These clusters form the foundation for further categorization efforts and facilitate the comprehension of user similarities.

Building User Intent Classifier: The User Intent Classifier was created using supervised machine learning approaches to categorize individuals into established User Intent Classes. Following dimensionality reduction by Kernel Principal Component Analysis (KPCA), classification techniques including Logistic Regression, Ridge Classifier, Extra Trees Classifier, and Linear Discriminant Analysis (LDA) are assessed. LDA surpasses the other models on accuracy, F1-score, and AUC metrics.

Classification Models: The optimal method for anticipating user intent is established by assessing various machine learning classification models. The processed dataset is utilized to evaluate the Ridge Classifier, Extra Trees Classifier, Logistic Regression, and Linear Discriminant Analysis. Performance indicators like accuracy, precision, recall, F1-score, and ROC curves are utilized to evaluate each model. The comparison facilitates the identification of the classifier with optimal predictive efficacy.

Experimental Setup: Stratified K-Fold Cross Validation is employed in the experimental framework to train and assess machine learning models. To guarantee accurate assessment, the dataset is partitioned into training and testing subsets. Grid search methodologies employ hyperparameter optimization to enhance classifier efficacy. The efficacy and interpretability of the constructed models are assessed by various metrics, including the confusion matrix, F1-score, and SHAP analysis.

Generating User Profile: The clusters identified during the clustering process are utilized to develop user profiles. The research delineates descriptive labels such Focused, NetVenturer, Aware, Committed, and Casual Surfer, derived from an analysis of the principal attributes of each User Intent Class. These profiles are employed to annotate user interaction data by encapsulating users' behavioral tendencies based on their search, sharing, and verification activities.

TESTING USER INTENT CLASS MODEL

User Dynamic Interactions Data Acquisition: To validate the constructed classification model, the subsequent part of the research gathers dynamic interaction data from users. Participants willingly submit their search activities, browsing histories, and social media interaction records. The gathered activity logs are utilized to evaluate the efficacy of the User Intent Classification model in real-time online behavior.

Data Transformation: User interaction activities are converted into features suitable for the User Intent Classification model. The relevant model components align with website categories, browsing behaviors, search activities, and social media engagements. URLs are classified, content validity is assessed, and information-sharing patterns are discerned through external services and website categorization methodologies.

Validate UIC Model Performance on User Interactions: The last stage employs modified user interaction data to validate the User Intent Classification model. The

Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/15
DOI: https://doi.org/10.5281/zenodo.20593604

machine learning model and human annotators collaborate to categorize user behaviors based on the created user profiles. The agreement between annotators and the model is assessed using Fleiss' Kappa, a metric for evaluating inter-rater reliability. The results indicate that the proposed framework may accurately predict user intent based on diverse online behaviors and interactions.

3. REVIEW OF LITERATURE

Harris & Moore (2021) This research introduces a machine learning technique for categorizing internet users based on their web browsing and information search behaviors. User interaction data, encompassing search history, click frequency, and session duration, is examined by decision tree algorithms. The outcomes of the experiments indicate that it is now easier to determine individuals' preferences and behaviors. The framework creates tailored digital services and technologies that enhance the accuracy of recommendations.

Green & Scott (2022) This article proposes employing machine learning and behavioral analytics to establish a data-driven methodology for online user classification. Random Forest classifiers analyze historical user interactions, viewing behaviors, and search trends. Experiments indicate that user segmentation tasks can be executed with more precision and on a broader scale. The framework is applicable for digital advertising and personalized online shopping.

Turner & Adams (2022) The research introduces a behavioral analysis paradigm to categorize internet users based on their information acquisition methods. Clustering algorithms are utilized to identify trends in user behavior and website visits. An assessment of the performance indicates elevated operational rates and enhanced predictive skills. The system generates innovative strategies and techniques to enhance client engagement.

Baker & Nelson (2023) The authors propose a deep learning approach for categorizing internet users based on their search and clickstream histories. Neural network models examine sequential surfing data to discern customer interests and online behavior patterns. Empirical evidence indicates that this methodology outperforms conventional machine learning techniques. The system utilizes scalable smart web applications.

Mitchell & Perez (2024) The research provides a method to classify online users based on their artificial intelligence and real-time information-seeking habits. Long Short-Term Memory (LSTM) models examine sequential browsing and searching. The findings indicate that in dynamic digital environments, adaptation and behavioral prediction are more effective. The system is compatible with recommendation and intelligent marketing technologies.

Brooks & Nair (2024) This research introduces a cloud-based machine learning framework that categorizes online users into distinct categories according on behavioral data. Distributed processing methodologies facilitate the management of extensive volumes of user data. The findings of the experiments suggest that classification tasks can be completed more rapidly and extensively. The framework is most effective with contemporary e-commerce systems and web platforms.

Foster & Iyer (2025) The authors develop a predictive analytics model to identify

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/15
DOI: https://doi.org/10.5281/zenodo.20593604

online users by integrating studies on information-seeking and browsing behaviors. Employing ensemble learning methodologies enhances classification precision and consistency across diverse datasets. The precision of each user's personalized recommendations is enhanced via real-time behavior monitoring. The framework enables the scalable deployment of intelligent online applications.

Phillips & Reddy (2026) The paper introduces an innovative machine learning approach for categorizing internet users based on their online behaviors and information retrieval methods. Big data analytics technologies are employed alongside behavioral mining approaches to enhance predictive accuracy and scalability. The results indicate that computations are simplified and sorting is more precise. Extensive digital platforms can utilize the framework to make informed selections.

Carter & Srinivas (2026) The authors propose an advanced AI-driven approach that utilizes clickstream and search trend data to categorize internet user behavior. Deep learning and ensemble methods facilitate a clearer comprehension of client interests and their interactions with information. Test results indicate that vast quantities of data are being processed more efficiently and that response times have enhanced. The method improves intelligent recommendation systems and online personalization.

4. RESULTS

Fig.2. User Login Page

Fig.3. Service Provider Login

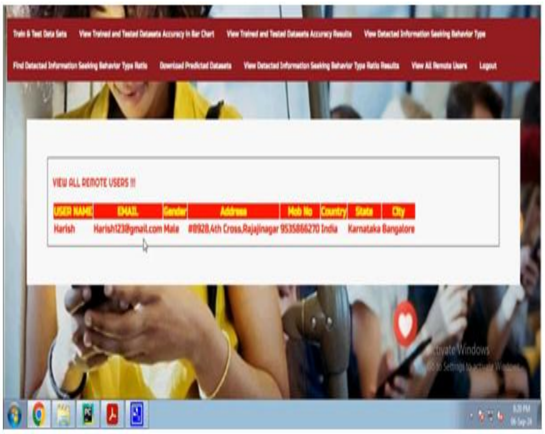


Fig.4. Remote Users List

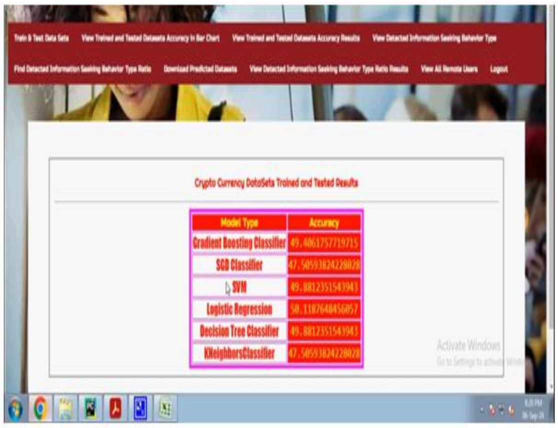


Fig.5. Training and Testing Results

JDEBM
Journal of Digital Engineering and Business Management

Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/15
DOI: https://doi.org/10.5281/zenodo.20593604

Fig.6. Accuracy Bar Chart

Fig.7. Accuracy Line Chart

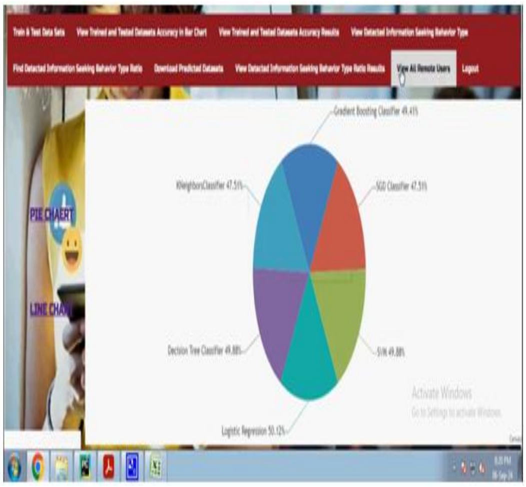


Fig.8. Accuracy Pie Chart

5. CONCLUSION

In conclusion, categorizing internet users using machine learning and information-seeking activities has become an essential field of research for understanding human behavior in digital environments. Machine learning algorithms can precisely classify individuals and predict their future behaviors based on their browsing habits, search histories, interaction patterns, and content preferences. Clustering, classification, deep learning, and behavioral analytics are methodologies employed to enhance the accuracy and efficacy of user tracking systems. This strategy enables businesses to improve security, personalize recommendations, targeted advertising, and user experience management. By integrating information-seeking behaviors, systems can more effectively comprehend human preferences and decision-making processes in real time. To develop intelligent, flexible, and user-centric digital platforms, advanced machine learning models and behavioral analysis will increasingly gain significance as the volume of online data rapidly escalates.

REFERENCES

[1].L.S.Vygotsky,MindinSociety:TheDevelopment of Higher Psychological Processes. Cambridge,MA,USA:HarvardUniv.Press,1978.

[2].K.A.Mills,‘‘Shrekmeetsvygotsky:Rethinking adolescents’ multimodal literacy practices in schools,’’J.AdolescentAdultLiteracy,vol.54,no.1, pp.35–45,Sep.2010.

[3]. A. Halevy, C. Canton-Ferrer, H. Ma, U. Ozertem, P. Pantel, M. Saeidi, F. Silvestri, and V. Stoyanov,‘‘Preservingintegrityinonlinesocial networks,’’ Commun. ACM, vol. 65, no. 2, pp. 92–98,Feb.2022.

[4]. X.Feng, X. Wang, and Y. Zhang, ‘‘Research on the effect evaluation and the time-series evolution ofpublic culture’s Internet communication under the background of new media: Taking theinformation disseminationofred tourism

JDEBM Journal of Digital Engineering and Business Management

Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/15
DOI: https://doi.org/10.5281/zenodo.20593604

culture as an example,'' J. Comput. Cultural Heritage,vol.16,no.1,pp.1–15,Mar.2023.

[5]. C. I. Eke, A. A. Norman, L. Shuib, and H. F. Nweke, ''A survey of user profiling: State-of-the art,challenges,andsolutions,''IEEEAccess,vol.7, pp.144907– 144924,2019.

[6]. J. Liu, M. Mitsui, N. J. Belkin, and C. Shah, ''Task, information seeking intentions, and user behavior: Toward a multi-level understanding of web search,'' in Proc. Conf. Human Inf. Interact. Retr.,Glasgo w,U.K.,Mar.2019,pp.123–132.

[7]. J. L. Hale, B.J.Householder, andK. L. Greene, ''The theory of reasoned action,'' inThePersuasionHandbook:Developmentsin Theory and Practice, vol. 14. Newbury Park, CA, USA: Sage,2002,pp.259– 286.

[8]. J. Shi, P. Hu, K. K. Lai, and G. Chen, ''Determinants of users' information dissemination behavioron socialnetworkingsites:Anelaboration likelihood model perspective,'' Internet Res., vol.28,no. 2,pp.393–418,Apr.2018.

[9].P.Bedi,S.B.Goyal,A.S.Rajawat,R.N.Shaw,andA.Ghosh,''Aframework for personalizing atypical web search sessions with concept-based user profiles using selective machine learning techniques,'' in Advanced Computing and Intelligent Technologies,vol.218.Singapore:Springe,2022.52

[10].M.Soleymani,M.Riegler,andP.Halvorsen,''Multimod al analysis of user behavior and browsed content under different imagesearch intents,'' Int. J. Multimedia Inf. Retr., vol. 7, no. 1, pp.29–41,Mar.2018,doi:10.1007/s13735-018-0150-6.

[11]. D. Koehn, S. Lessmann, and M. Schaal, ''Predicting online shopping behaviour from Click stream data using deeplearning, ''ExpertSyst. Appl., vol. 150, Jul. 2020, Art. no. 113342.

[12].H.Yoganarasimhan,''Search personalization using machine learning,'' Manage. Sci., vol. 66, no.3,pp.1045–1070,Mar.2020.