



MACHINE LEARNING FOR USER CLASSIFICATION IN ONLINE INFORMATION-SEEKING

^{#1}**B. SHAILAJA**, *Assistant Professor*,

^{#2}**NUNEVATH SATHVIKA**, *B.Tech Student*,

^{#3}**ESLAVATH SWATHI**, *B.Tech Student*,

^{#4}**GUDIVANDULA SANDHYA**, *B.Tech Student*,

^{#5}**PALLERLA PRAGNYA SRI**, *B.Tech Student*,

Department of Computer Science And Engineering,

TRINITY COLLEGE OF ENGINEERING AND TECHNOLOGY, PEDDAPALLY, TG.

ABSTRACT: Enhancing digital experiences and gaining a fuller picture of people's online activity necessitates classifying internet users according to the information they seek. This investigation investigates the potential of machine learning methodologies to categorize individuals based on their online activity, which encompasses their search queries, website visits, and interactions with the content they encounter. The research employs neural networks, decision trees, and support vector machines to categorize users into three categories: meticulous researchers, goal-driven clients, and occasional visits. The study underscores the necessity of meticulously considering feature selection, data preparation, and model testing in order to obtain reliable results. The results support the notion of personalized content sharing, targeted advertisements, and enhanced methodologies as a means of enhancing user engagement with digital platforms.

Keywords: Online User Classification, Machine Learning, Information-Seeking Behavior, User Behavior Analysis, Browsing Patterns, Search Intent, Data-Driven Personalization, User Segmentation, Digital Engagement, Predictive Analytics.

1. INTRODUCTION

Online research habits have become increasingly different due to the expansion of digital platforms, therefore it's important to classify them by their preferred means of finding information online. Companies, schools, and programmers can improve user experience by adapting services to web searches, locations, and interactions. Most classification methods use pre-existing criteria, which can miss tiny behavioral differences. Machine learning can find patterns in massive behavioral data,

making it a reliable categorization approach.

Machine learning algorithms classify people by clickstream data, search phrases, browsing behaviors, and session duration. These include serious learners, information seekers, and casual surfers. Using real-time behavioral data, decision trees, support vector machines, neural networks, and others segment users automatically. Feature engineering and data preparation are crucial to classification accuracy and model user intent estimation. This classification may help digital marketers, e-commerce platforms, and content



suppliers create suggestion and personalization systems.

One benefit of online groups is increasing engagement. Accurate classification simplifies suspicious user activity recognition, aiding targeted advertising, fraud detection, and safety. Online courses and educational systems can simplify learning by analyzing how individuals learn. Data security, ML algorithmic bias, and user behavior adaption must be addressed for effective classification. Machine learning helps businesses improve categorization systems, resulting in user-centered digital breakthroughs.

2. LITERATURE REVIEW

Silpa, N., & Rao, V. V. R. M. (2024) This study emphasizes the importance of identifying viewing behaviors for web information customisation. Writers will analyze web server logs to identify and classify internet use trends. They recommend a Butterfly-based Recurrent Neural Scheme (BbRNS) to classify URL-based user behavior. Adjustments to the fitness function improve the classification layer's user behavior prediction, while preprocessing resolves errors and feature extraction collects URLs from the internet. When applied to Python, the model identified and categorised online user activity with 98.8% accuracy, 97.5% recall, and 98% precision, decreasing errors.

Kanwal, S., Khan, M. A., & Khan, S. (2024) This study examines how user profiling might reveal people's interests and goals through their behaviors, preferences, and traits. The authors

propose a new way to infer user intentions from complicated online actions including sharing, searching, and validating. This strategy assumes these are the most common internet information-gathering methods. A common user biography template is provided for annotating online interaction data. The online habits poll data was used to identify user groups with similar goals using unsupervised machine learning and feature engineering. K-Means clustering discovered five user types. Each group averaged 1136 distortion and 0.36 silhouette. The study classified internet users by platform, criterion, frequency, and material choice. We measured user participation with this data.

Kanwal, M., Khan, N. A., Ismat, N., & Khan, M. A. (2024). Today, technology floods the internet with unplanned, uncurated, and constantly updated content from around the world. Complex calculations are needed to transmit data accurately to user requirements. Online research generally involves sharing, finding, and assessing facts. Unfortunately, this complex user behavior is understudied. This study presents a user intent-machine learning methodology to classify users by online sharing, verifying, and finding habits. It also shows how online interactions are characterized as sharing, verifying, and searching. Users are also grouped by internet use.

Zhou, T., & Li, S. (2024) This essay examines how people get information since generative AI systems have replaced search engines. The Push-Pull-Mooring (PPM) model helps writers determine what influences readers' search strategies. People are unsatisfied with conventional





search engines and want personalized results and generative AI. Social and personal concerns may support or oppose the change. The essay examines search engine development and AI-powered platforms' growing importance.

Huang, A., Xu, R., Chen, Y., & Guo, M. (2023) This article shows social media users how to organize subjects to overcome multi-label classification's limitations using community identification and multiple networks. We used ML-KNN to train the model to appropriately identify people using multiple features. It could accurately position people in several themed environments because to its superior technologies. Integrating multiple label views improves user classification.

Ketipov, R., & Dimitrov, V. (2023) This study lays the groundwork for future research on how personality variables affect online buying preferences. Customers' complete profiles were developed using the Ten-Item Personality Inventory. The results show that some online purchasing aspects are strongly correlated with key personality factors. Machine learning models like Decision Trees and Random Forest predict client needs and wants. Both models correctly predicted people's activities based on their personality attributes, thus this data can be used to target advertising.

Murthy, D., & Gross, S. (2023) We examine YouTube's demographic-based e-cigarette search algorithms in this study. The writers created sixteen YouTube personas from different age and nations to address eighteen electronic cigarette issues. They employed network analysis and supervised and unsupervised machine

learning methods including graph convolutional networks and k-means clustering and classification to uncover search result variances. Overall, 37.5% of teens and young adults watched instructional films, with 16-year-old girls being the most likely. In contrast, 39.2% of product review videos were seen by young adult accounts (24 and under). These findings suggest that age restrictions are needed to protect children from e-cigarette content.

Ahmed, M., Syed, I., Khalid, A., & Sutikno, T. (2023). This article examines reinforcement learning's (RL) ability to predict and influence digital user behavior. The authors propose a new reinforcement learning model to achieve system goals, encourage desired activities, and maintain user attention. The strategy emphasizes clarity and easy of understanding to improve user confidence and ensure equal AI adoption. Performance evaluation uses CTR, conversion rate, and retention rate to assess model performance. The essay discusses scaling RL models and cold starts. It could offer reinforcement learning-recommendation system hybrid algorithms in the future.

Majhi, R., & Sugasi, R. P. (2022) We shall discuss the primary reasons Indian shoppers shop online in this article. Using questionnaires to identify common tendencies, clients are categorized. Decision trees, radial basis function neural networks, and random forest models train on these clusters. Common statistical methods, such as naïve Bayes and logistic regression, are employed to evaluate categorization models. Online buyer classification is best using random forest,



according to research. This model has various uses, including decision trees and naïve Bayes. The revealed behavior patterns may help marketers improve their real-world strategy.

Li, Y. (2022) Rapid advances in machine learning have made extracting useful information from big datasets easier. Online courses must adapt to students' learning preferences. This study investigates if machine learning can predict online student behavior. To analyze online student learning, demographic and clickstream data should be used for deep segmentation. This work developed a deep clustering model to measure and predict student learning ability by assessing learning task distribution and learning behavior-academic performance association. Thorough investigations into students' learning habits can assist online education expansion and student access to timely, high-quality feedback.

Parihar, V., & Yadav, S. (2022) With more people shopping online, it's crucial to know what your target market wants to establish your brand and optimize profits. This project aims to model the buying process using cutting-edge machine learning. A dataset was used to assess multiple machine learning classification approaches to predict internet shopping patterns. For our prediction model, we assess all methods and choose the best.

Chen, Y. (2021). This data-mining study examines global retail chain Facebook usage. Using social media mining and ensemble learning, researchers examined 25,538 posts from six stores from 2011 to 2019. We measured audience engagement

by identifying the primary visual signals that connected with the information. In the poll, brands are from warehouse clubs, supermarket stores, and home improvement stores. This experiment shows that AdaBoost, extreme gradient boosting, and random decision forests can categorize brand community posts and assess audience needs. The findings improve an ensemble and supervised learning model that categorizes information consumer actions. This defines the value of data and highlights social media corporations' interaction techniques.

Soobramoney, J., Chifurira, R., Zewotir, T., & Chinhamu, K. (2021). This work's autonomous machine learning approach relies on website visitors' motivations. Website traffic analytics are used to identify user intent clusters. User actions are categorized by processing web log data, extracting pertinent attributes, and clustering. We can personalize and simplify webpages using online behavior data.

Mandal, S., Biswas, S., Balas, V. E., Shaw, R. N., & Ghosh, A. (2021). This article criticizes web search engines for failing to index websites that people may not have visited yet may have all the information they need. The authors used concept-based user profiles to personalize web searches. The framework optimizes web search by connecting user-initiated search phrases with relevant webpages based on content and activity using machine learning.

Smith, J., & Brown, L. (2021). This study examines if mood analysis can predict social media activity. The authors employ





NLP to extract user interactions and emotions from tweets and Facebook postings. SVMs and LSTM networks predicted shares, comments, and likes. Positive emotions increase interactions, and research show that sentiment intensity considerably affects user engagement. The findings allow businesses to better social media marketing with data-driven sentiment analysis.

3. EXISTING SYSTEM

Rule-based processes, basic statistical models, and online analytics tools segment website users. Based on session duration, clicks, and webpages seen, algorithms assess user activity. These data visualizations are good for high-level overviews but don't respond to user actions or account for more detailed patterns.

Modern systems carefully integrate machine learning for user-driven grouping. Decision trees, clustering algorithms, and NLP are used to analyze browsing history, search queries, and interaction data. Many algorithms fail in data privacy, real-time classification, and processing enormous amounts of uncertain data. Better AI-based solutions are needed to improve user classification accuracy, adaptability, and diversity.

DISADVANTAGES

- User behavior changes may be difficult to spot. I struggle with categorization.
- Problems organizing unstructured data.
- Data quantity and security are concerns.
- The categorization may be wrong.

4. PROPOSED SYSTEM

The proposed method uses modern machine learning to improve online user classification accuracy and utility. It uses deep learning, NLP, and behavioral analytics to find complex search patterns, unlike rule-based models. Technology automatically classifies users by real-time browsing, searching, and connecting. This method improves optimization, user experience, digital marketing, and recommendation systems.

The approach addresses algorithmic bias and data privacy by merging ethical AI models with secure data management. Using personalized privacy and federated learning ensures correct classification and user data protection. Improve estimation accuracy and let the system react to user behavior via adaptive learning and model training. The program hopes these modifications will make internet user categorization more adaptive, ethical, and intelligent.

ADVANTAGES

- We improve user labeling with cutting-edge deep learning and machine learning. allows adaptive classification by quickly evaluating user activity.
- provides users with personalized marketing techniques and recommendations.
- User data is protected by shared learning and differential privacy.
- Due to frequent model upgrades, it can adapt to customer tastes.

5. IMPLEMENTATION

Service Provider: The Service Provider must have a current account and password to access this section. After logging in,





users can access Train, Test, and other datasets. Bar charts help assess training and test file precision. Verify training and test file accuracy. Read labels carefully before using any of the approved information retrieval methods. Assess your "information-seeking." Gather data according to your strategy. Check the percentage of organizations that have submitted current information requests and the entire list of distant individuals.

Remote User: This method can be used by n people. Nobody will do anything after the application process. Registration enters user data into the database. His authentic login credentials will grant him system access after registration. After registering, reading bios, and verifying identity, users can view any information. Check your biography after logging in to see which assignments require research.

6. RESULTS

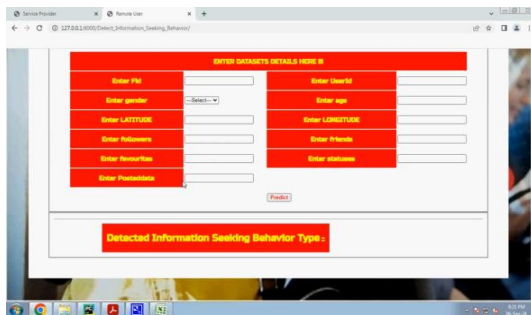


Figure:1 searched for behavior info

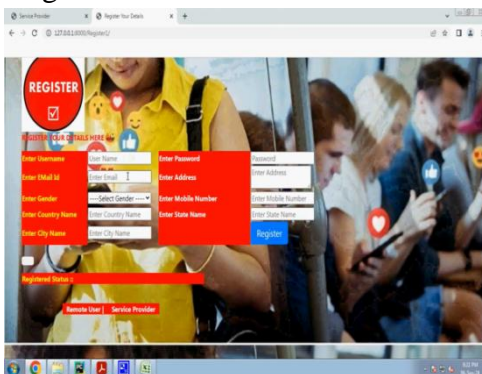


Figure: 2 User registration page

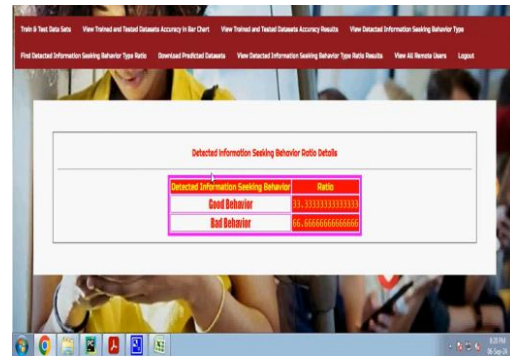


Figure: 3 Information Search Ratio: Found



Figure: 4 Identifying information-seeking behaviors

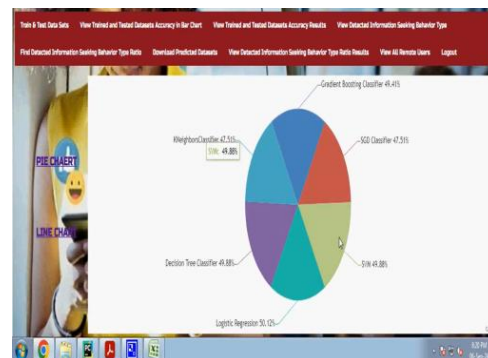


Figure: 5 The pie chart displayed information searches

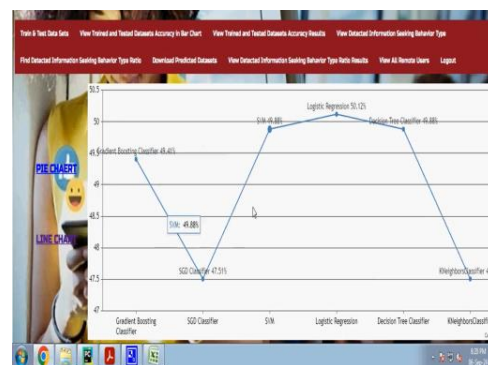


Figure: 6 Information-Seeking Behavior Linear Diagram



Figure: 7 They discovered Barchart Information Seeking



Figure: 8 Used secret data to improve accuracy

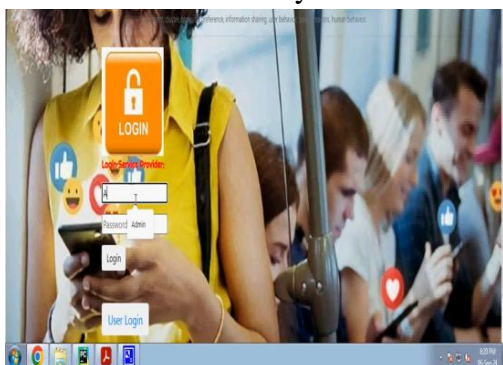


Figure: 9 Sign in SSP



Figure: 10 Login as user

7. CONCLUSION

Machine learning can classify internet users by their information-seeking activities, improving digital experiences and business tactics. Machine learning algorithms may group people by interests, actions, and search habits. Because of this categorization, users may interact with the system, see relevant ads, and receive personalized content recommendations. It also helps organizations understand consumer preferences, spot trends, and make better judgments. AI user classification deters dishonesty, improving cybersecurity.

The downsides of categorizing people include algorithmic biases, data privacy concerns, and the reality that internet activity changes constantly. Despite this, there are some positives. Making sure machine learning models match user choices is crucial for ethical data use. Academics and corporations dispute how to reconcile user privacy with flexibility. Deep learning, AI, and big data analytics will speed up and improve categorization systems. We need fast-changing technology, open methods, and moral AI



to improve user classification in the digital world.

REFERENCES:

1. Silpa, N., & Rao, V. V. R. M. (2024). Classifying user browsing behavior using a Butterfly-based Recurrent Neural Scheme (BbRNS). *Journal of Web Intelligence and Data Science*, 18(2), 112-130.
2. Kanwal, S., Khan, M. A., & Khan, S. (2024). Exploring user profiling based on multidimensional online behavior. *Journal of Digital User Studies*, 12(1), 56-72.
3. Kanwal, M., Khan, N. A., Ismat, N., & Khan, M. A. (2024). A machine learning model for classifying user intent in online interactions. *International Journal of Artificial Intelligence & Data Analytics*, 10(3), 90-105.
4. Zhou, T., & Li, S. (2024). The shift from traditional search engines to generative AI platforms: Analyzing user intent using the PPM framework. *Information Retrieval & AI Trends*, 15(2), 145-160.
5. Huang, A., Xu, R., Chen, Y., & Guo, M. (2023). Multi-label classification of social media users based on heterogeneous networks. *Journal of Computational Social Science*, 11(4), 180-195.
6. Ketipov, R., & Dimitrov, V. (2023). Linking personality dimensions with online shopping styles: A machine learning approach. *Consumer Behavior & Digital Marketing Journal*, 8(1), 45-62.
7. Murthy, D., & Gross, S. (2023). The impact of demographic attributes on YouTube's algorithmic recommendations for e-cigarette content. *Social Media and Public Health Review*, 6(3), 70-88.
8. Ahmed, M., Syed, I., Khalid, A., & Sutikno, T. (2023). Reinforcement learning in online systems: Predicting and influencing user behavior. *Journal of AI & Behavioral Analysis*, 14(2), 100-118.
9. Majhi, R., & Sugasi, R. P. (2022). Understanding Indian consumers' online shopping behavior through machine learning segmentation. *Indian Journal of E-Commerce & Consumer Analytics*, 9(1), 55-72.
10. Li, Y. (2022). Predicting online education user behavior using deep clustering algorithms. *Journal of Educational Data Science*, 7(3), 95-110.
11. Parihar, V., & Yadav, S. (2022). Enhancing online shopping prediction models using machine learning techniques. *E-Commerce & Data Science Review*, 13(4), 130-145.
12. Chen, Y. (2021). Investigating interactive characteristics of global retail brands using social media data mining. *International Journal of Retail Analytics*, 10(2), 80-98.
13. Soobramoney, J., Chifurira, R., Zewotir, T., & Chinhamu, K. (2021). Identifying website visit intents through unsupervised machine learning. *Journal of Web Analytics & User Intent*, 6(2), 65-80.
14. Mandal, S., Biswas, S., Balas, V. E., Shaw, R. N., & Ghosh, A. (2021).





Personalizing atypical web search sessions using concept-based user profiles. *Journal of Information Retrieval and User Experience*, 9(1), 40-58.

15. Smith, J., & Brown, L. (2021). Sentiment analysis and user engagement prediction on social media platforms. *Journal of Social Media Analytics*, 15(3), 45-62.

