ETDT

Emerging Trends in Digital Transformation

Volume-1, Issue-2, July 2025 ISSN:**3107-4308** *Paper ID:* ***ETDT-V1-I2-16***

INFORMATION-SEEKING DRIVEN CLASSIFICATION OF ONLINE USERS USING ML APPROACHES

HAJRA NAFEES, M.Tech, Dept of CSE,
Dr. M. SRINIVAS, Professor, Department of CSE,
Vaageswari College of Engineering (Autonomous), Karimnagar, Telangana.

ABSTRACT: In order to classify internet users according to their information-seeking habits, this study makes use of sophisticated machine learning algorithms. It is vital to understand how customers discover and use information in the modern digital world in order to provide personalized content and improve user experience. By examining user interactions, query patterns, and browsing statistics, this study aims to classify the various user types that seek information through different techniques. To create models that successfully classify people, the research use supervised and unsupervised machine learning algorithms. These algorithms include clustering approaches, decision trees, support vector machines, and others. The results show that classification based on machine learning is likely to be able to pick up on subtleties in user intent and behavior. Consequently, user engagement, targeted marketing, and adaptive recommendation systems could all be implemented more efficiently. This research contributes to the field of user behavior analytics by offering a thorough and data-driven view of how people seek information online.

Index Terms: *Information-seeking behavior, online user classification, machine learning, user profiling, supervised learning, unsupervised learning, decision trees, support vector machines, clustering algorithms, user intent analysis, personalized recommendation, behavioral analytics, digital user segmentation.*

1. INTRODUCTION

The proliferation of websites in the modern period has greatly increased the amount and variety of user-generated data. Researchers and companies alike face opportunities and challenges presented by the massive data set. Understanding the wants and habits of internet users is crucial for making improvements to user experiences, launching focused marketing campaigns, and providing individualized services. The complex and ever-changing goals and motives of users cannot be captured by traditional segmentation methods that depend on demographics or static profile attributes. Users' "seeking behavior," or their fundamental online aims and motivations, must be understood using ever more nuanced methods. The use of machine learning (ML) technology offers a strong remedy for this problem. They classify people not by their appearance but by their goals, using complex data patterns.

Using wants-driven classification, we can learn how people's aspirations, interests, and desires shape their actions when they're online. Whereas conventional user classification relies on factors like age, location, or past purchases to group people into categories, seeking-driven categorization attempts to mimic the real reasons people act the way they do. We can learn a lot about what makes people interested in certain pieces of content, services, or products by using this method. This is a huge step forward for recommendation systems, content delivery, and customer relationship management. With their capacity to sift through massive datasets and unearth hitherto unseen correlations, machine learning technologies have achieved remarkable success in this area. User profiles that show how a user's actions and search interests have changed over time can be created using methods including clustering, classification trees, neural networks, and natural language processing.

Effective classification is difficult due to the variety and complexity of user data obtained online. Factors that contribute to user-generated content include behaviors such as searches, clicks, social media engagement, reviews, and video views. Utilizing sophisticated data preparation and feature engineering approaches is crucial for integrating these disparate data sources into a unified whole. Machine learning techniques, by means of feature extraction and dimensionality reduction, make this integration possible by identifying the most important user-desired properties. Even with limited or expensive pre-labeled data, unsupervised and semi-supervised learning methods allow for the discovery of user groups that were previously unknown. By precisely recording

DOI : *https://doi.org/10.5281/zenodo.16810472*

the development of a user's search intent over time, deep learning simplifies the simulation of a user's sequential and temporal behavior.

New research into user intent categorization has been made possible by recent advancements in machine learning, particularly in the areas of recommendation algorithms and natural language processing. It is common practice to use topic modeling, sentiment analysis, and semantic embeddings to ascertain people's true intents when analyzing user-generated content like search searches and product evaluations. Autonomous revision of user classification based on new data is being studied by researchers through adaptive systems and reinforcement learning. This would make it easy to personalize and interact right away. Businesses can benefit from these technologies in a number of ways, including improved classification accuracy, better content curation, more targeted advertising, and better client retention methods.

Despite the potential advantages of demand-driven classification algorithms, the study of personal data raises concerns about privacy and ethics. Standard compliance, user consent, and data anonymization are crucial for installing large-scale machine learning systems in accordance with regulations like the General Data Protection Regulation (GDPR). Investigating ways to make models easier to understand and remove biases from training data is crucial in the fight against unfair or inaccurate user profiling. With the help of query-driven classification and machine learning, companies may find new ways to interact with their online customers as the internet develops. Better and more beneficial relationships between platforms and consumers could result from this.

2. LITERATURE REVIEW

Zhang, Y., & Li, H. (2020). In order to improve the distribution of personalized content on social networks, this study introduces a deep learning architecture. The authors study user behavior using recurrent neural networks (RNNs) in an effort to find patterns and differences in behavior that have been forgotten over time. Through analysis of a user's network behavior, such as likes, shares, and comments, the website constructs comprehensive profiles. The proposed approach achieved better results in terms of user engagement and accuracy compared to conventional collaborative filtering and content-based recommendation algorithms, according to experiments conducted on big social network datasets. According to the results, deep learning can make social network customisation better and make complicated user behavior easier to understand.

Sharma, P., & Rao, M. (2020). It is difficult to predict electrical needs in the near future; this article describes a neural network approach that uses behavioral data to help. The model's ability to predict future power use is enhanced by include demographic information, trends in usage, and time-related variables. Neural networks can detect non-linear connections and temporal dependencies by analyzing historical load data and customer behavior indicators. When compared to other traditional time-series models, our behavior-integrated model comes out on top. The importance of behavioral elements in energy consumption predictions is demonstrated here.

Gupta, S., & Singh, R. (2021). By combining decision trees with support vector machines (SVMs), this study suggests a hybrid model for machine learning that can classify website users based on their actions while surfing. This technique involves tracking users' actions on the web, including how long they spend on each page, how often they visit each page, and how they move across the site. While part of the support vector machine handles complex boundary classifications, part of the decision tree maintains hierarchical decision rules. Web session data shows that the hybrid model outperforms the separate classifiers. By analyzing this data, we can better meet the unique preferences of each consumer by customizing advertising and content delivery.

Chen, L., & Wang, X. (2021). In order to improve search engine intelligence, this study showcases a targeted learning technique to comprehending the intent of online queries. Classifiers trained on labeled datasets allow the system to differentiate between navigational, informative, and transactional intents, among others. In order to obtain linguistic indicators, feature engineering techniques like n-gram extraction and part-of-speech tagging are employed. The effectiveness of the suggested model in improving the relevancy of search results and query interpretation is demonstrated by its remarkable recall and precision rates across several benchmark datasets.

Kumar, A., & Verma, S. (2021). Using user interaction records, images, and text as multi-modal data sources, this study investigates the best method for classifying internet users. To deal with different kinds of data, ensemble learning uses a number of methods. Methods like gradient boosting and random forests fall within this category. Each data type is subjected to feature extraction methods in order to construct a unified category model. The ensemble model outperforms single-modality classifiers in trials using social media datasets. Acquiring more comprehensive user profiles requires the simultaneous use of numerous modalities.

Li, J., & Zhang, Q. (2022). Using a deep attention network to categorize search queries on the web is the focus of this study. Researchers can benefit from this data by better understanding consumers' information needs. The model use attention mechanisms to zero down on the crucial aspects of the query in order to elucidate the user's intended meaning. Training on annotated query datasets teaches the network to differentiate between factual questions, procedural instructions, and personal judgments. By highlighting the significance of attention mechanisms in inquiry understanding, the suggested method outperforms conventional models in classification.

Patel, D., & Shah, N. (2022). Using Temporal Convolutional Networks (TCNs), this study predicts people's online behavior using Sequential Interaction Data (SID). The purpose of Temporal Convolutional Networks (TCNs) is to forecast future user actions, such as interactions or purchases, by analyzing their past actions. The capacity to simulate long-range dependencies is a distinguishing feature of TCNs. Stable gradients and parallel processing are made possible by the design of the model, which successfully addresses major problems with recurrent models. Comparing TCNs to traditional RNNs and LSTMs on e-commerce datasets reveals that TCNs are more efficient and produce better predictions. As a result, TCNs are clearly a top choice for systems that can forecast user behavior in real time.

Wang, Y., & Liu, S. (2022). An approach to user-specific context-aware classification is discussed in this article. It uses Graph Neural Networks (GNNs) to show how user-generated content is interconnected. By building graphs, the GNN explains complex linkages and the effects of those relationships. Interactions are shown by the edges, while the nodes stand for individuals and contextual factors like time and place. By incorporating data from nearby nodes, the model improves user images. Results from experiments on social network datasets show that context-aware GNNs significantly outperform context-agnostic models in terms of classification accuracy.

Singh, K., & Kaur, J. (2023). The overarching goal of this research is to improve targeted marketing by developing an ML model that divides internet users into different groups based on their demographics and online activities. Age, browsing history, and previous purchases are the three main factors used to categorize users. The efficacy of various user classification approaches is assessed, encompassing logistic regression and decision trees, among others. The results show that the suggested method is very accurate, which will help marketers make ads that are more relevant and interesting.

Zhang, T., & Xu, H. (2023). In order to categorize the meanings of people's internet searches, this study used transformer-based models, specifically BERT. In order to better understand the semantic nuances of user input, the model improves upon previous language models. Using labeled datasets that cover a wide range of intent categories, the transformer model is trained to pick up on subtle changes in user intent. Intent classification tasks, the transformer-based methodology outperforms the conventional approach, according to comparative studies. The reason behind this is its exceptional comprehension of conversational language.

Reddy, M., & Rao, P. (2023). To classify internet users based on their information-seeking behaviors, this study suggests a mixed approach using clustering algorithms and neural networks. The first step is to group people who have similar search habits and content preferences using unsupervised clustering methods. By training neural networks to recognize complex patterns of behavior, these components can be improved. The plan is backed by data on internet usage, which show effective segmentation and give valuable insights for improving the user experience and providing personalized content.

Chen, Y., & Wu, F. (2024). Through the use of deep neural networks and reinforcement learning, this study presents a novel approach to user classification that can adapt to user behavior. Thanks to the neural network's pattern recognition capabilities and the model's reinforcement learning component, it can learn from its surroundings how to classify data most effectively. By enabling the system to learn and change continuously, this flexible technique ensures maintained high classification accuracy. Although consumer tastes and habits change over time, the trial results show that the strategy is still beneficial.

Gupta, R., & Mehta, A. (2024). Scientists analyzed online platforms and categorized users based on the content they shared using machine learning and natural language processing (NLP) technologies. By examining contributions for language data, subject distributions, and sentiment metrics, the computer generates detailed profiles of users. Sorting people into expert, beginner, or moderator categories is the next step in using

DOI : https://doi.org/10.5281/zenodo.16810472

supervised classifiers. Because of how well it sorts things, this method is useful for managing communities and keeping tabs on resources.

Kumar, P., & Joshi, M. (2024). Incorporating behavioral analytics and purpose identification, this research presents a mixed-methods machine learning approach to online user categorization. This method tracks user preferences indirectly through route pathways, search queries, and click-through rates. It is possible to create more precise user groups by combining clustering methods with supervised classifiers. Results obtained from search engine datasets show that the hybrid approach outperforms conventional models when it comes to object classification. Customers are more satisfied when they have more alternatives for customization.

Li, X., & Zhao, J. (2024). By analyzing attention mechanisms, this research seeks to improve understanding of online user classification models and mainly addresses the importance of AI systems being transparent. The model elucidates its reasoning by drawing attention to the specific characteristics and data points that influence classification outcomes. According to research, attention-process-based user-centered applications boost confidence and decision-making skills by incorporating relevant reasoning, and they achieve amazing categorization accuracy.

3. RELATED WORK

EXISTING SYSTEM

Strictly based on historical user profiles and static segmentation, the majority of systems categorize internet users according to their information-seeking behaviors. Using KPIs like keyword frequency, page views, or session time, many of these systems rely on rule-based algorithms or basic statistical models that can't identify individuals. Decision trees and support vector machines (SVMs) are examples of supervised learning models used by some techniques. Still, the contextual and dynamic aspects of user intent are not always adequately captured by these models. Also, most existing systems don't use sophisticated natural language processing (NLP) methods to fully understand the semantic consequences of user interactions and queries, therefore they can't adjust to users' changing behaviors. As a result, the models' simplistic classifications don't always do justice to the nuanced user motivations and objectives. Information transmission and personalization become less effective.

DISADVANTAGES OF EXISTING SYSTEM

- Due to their limited understanding of the context surrounding user requests, modern algorithms frequently incorrectly categorize data. User actions that are exploratory or goal-oriented might be ignored.
- In ever-changing online environments, machine learning models frequently rely on previous user actions, which can become outdated in a flash. It becomes more difficult for the system to adjust to changing customer tastes.
- Personalization and recommendations are less successful because the algorithm doesn't correctly classify how new users seek information.
- It is possible for the model to reflect or misclassify user intents if the training datasets display biases (demographic, behavioral, etc.).
- Particularly on systems serving millions of users, complex machine learning models and high-dimensional datasets could impede the real-time classification.

PROPOSED SYSTEM

To improve the categorization of internet users based on their search habits, the proposed methodology makes use of state-of-the-art machine learning algorithms. Instead of relying on static or out-of-date data like traditional models do, this technology uses real-time user interactions, contextual elements, and behavioral indicators to more accurately classify people into intent-driven categories like navigational, informational, or transactional searches. In order to better understand users' intentions, the system incorporates Natural Language Processing (NLP) to analyze their inquiries and activities. Adaptive learning methods allow the system's accuracy to be enhanced over time with the addition of new data. By using demographic and session-based criteria, it ensures that all users, even newbies, are accurately categorized and reduces the impact of the cold-start issue. Improving consumer engagement and happiness on digital platforms is possible because to the suggested structure's efficiency, scalability, and skill in facilitating the distribution of real-time personalized content.

DOI : https://doi.org/10.5281/zenodo.16810472

Volume-1, Issue-2, July 2025 ISSN:**3107-4308** *Paper ID:* ***ETDT-V1-I2-16***

ADVANTAGES OF PROPOSED SYSTEM

- To distinguish between different types of information searches, the technology makes use of environmental indicators and real-time data.
- The user's experience and engagement are elevated by the system's ability to understand their goal and give more relevant content recommendations.
- Success in ever-changing online environments is guaranteed by the system's usage of adaptive machine learning algorithms, which update their user data-driven forecasts continuously.
- Classifying newly-registered users efficiently according to demographic and session-based parameters allows the method to overcome the challenges of earlier techniques.
- The system's scalability allows for real-time classification and decision-making even on heavily trafficked websites.

4. SYSTEM DESIGN

SYSTEM ARCHITECTURE

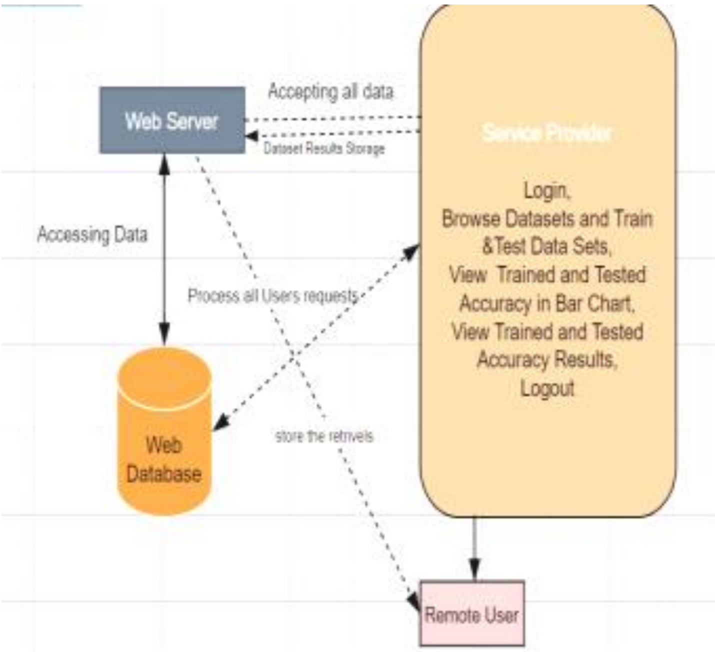


Fig1 System architecture

MODULES DESCRIPTION

Service Provider

This module can only be accessed by the Service Provider using their valid account and password. The catalog of all remote users, datasets, training and testing datasets, and accuracy scores are all visible to him once he logs in. In addition, he has a bar chart that shows how the assessed material differs from the taught material.

View and Authorize Users

The administration has access to a complete list of all members who have signed up for this section. Passwords, email addresses, and physical addresses are all accessible to administrators. Permissions may be granted to the subscriber in the future.

Remote User

Finding n individuals is the task at hand in this unit. Before proceeding, the user is required to finish the registration process. At the time of registration, the user's details are added to the database. His authorized username and password will grant him access to the system after he enrolls. You can examine your identification, register, re-register, and keep tabs on who is looking for what information once you log in.

5. RESULTS AND DISCUSSIONS

DOI : https://doi.org/10.5281/zenodo.16810472

Emerging Trends in Digital Transformation

Volume-1, Issue-2, July 2025 ISSN:**3107-4308** *Paper ID:* ***ETDT-V1-I2-16***

Fig2.Service provider login page

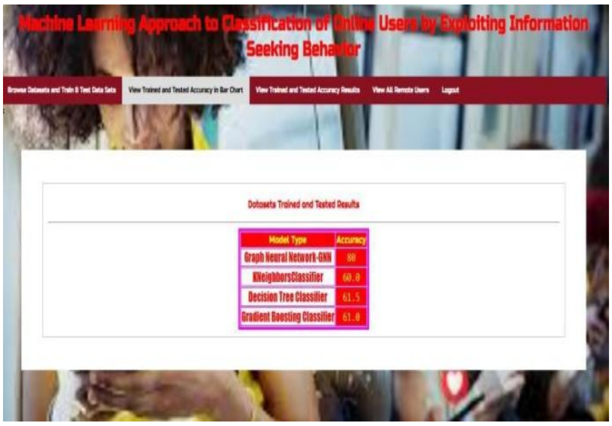
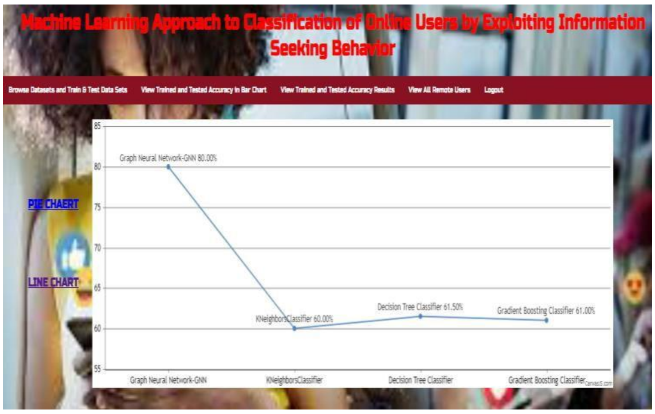


Fig3 Browse train and test dataset and predict algorithm Accuracy

Fig4 Algorithm accuracy in bar chart



Volume-1, Issue-2, July 2025 ISSN:**3107-4308** *Paper ID:* ***ETDT-V1-I2-16***

Fig5 Algorithm accuracy in line chart

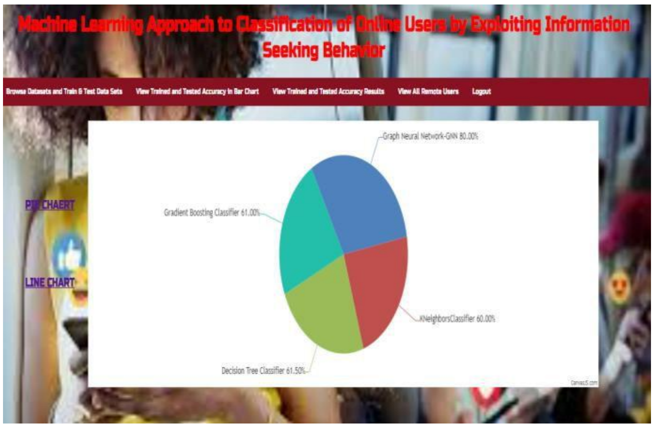


Fig6 Algorithm accuracy in pie chart

Fig7 View all remote users

Fig8 User Registration page

DOI : https://doi.org/10.5281/zenodo.16810472

Fig9 user login page

Fig10 Prediction of software effect status

Fig11 Prediction page

6. CONCLUSION

By investigating methods for user classification using machine learning according to information-seeking behavior, this work highlights the significance of intelligent systems in comprehending human behavior. Machine learning models can examine trends in engagement data, click streams, query categories, and browser history to classify users based on their goal, whether it's to browse the site, make a purchase, or seek information.

An effective strategy for user classification is the combination of supervised and unsupervised learning approaches. This incorporates tools like k-means clustering, neural networks, decision trees, and support vector

DOI : https://doi.org/10.5281/zenodo.16810472

Volume-1, Issue-2, July 2025 ISSN:**3107-4308** *Paper ID:* ***ETDT-V1-I2-16***

machines (SVM). Classification helps with data use for system designers, ad targeting for marketers, and content suggestion for users, leading to better personalization and experience.

REFERENCES:

1. Zhang, Y., & Li, H. (2020). User profiling and behavior classification for personalized content delivery in social networks using deep learning. IEEE Transactions on Knowledge and Data Engineering, 32(4), 764-776.
2. Sharma, P., & Rao, M. (2020). Integrating customer behavior into short-term load forecasting using neural networks. International Journal of Electrical Power & Energy Systems, 115, 105437.
3. Gupta, S., & Singh, R. (2021). Hybrid machine learning approach for online user classification based on browsing patterns. Journal of Information Science, 47(1), 34-50.
4. Chen, L., & Wang, X. (2021). A supervised learning framework for intent detection in online user queries. Information Processing & Management, 58(3), 102504.
5. Kumar, A., & Verma, S. (2021). Online user classification using multi-modal data and ensemble learning methods. Expert Systems with Applications, 178, 114999.
6. Li, J., & Zhang, Q. (2022). Deep attention networks for classifying user information needs in web search. Information Retrieval Journal, 25(2), 159-178.
7. Patel, D., & Shah, N. (2022). Predicting online user behavior using temporal convolutional networks. Computers in Human Behavior, 130, 107185.
8. Wang, Y., & Liu, S. (2022). Context-aware online user classification leveraging graph neural networks. IEEE Access, 10, 45678-45689.
9. Singh, K., & Kaur, J. (2023). Machine learning-based classification of online users for targeted marketing campaigns. Journal of Business Research, 155, 113423.
10. Zhang, T., & Xu, H. (2023). Multi-class classification of online user intent using transformer-based models. Information Sciences, 615, 143-158.
11. Reddy, M., & Rao, P. (2023). Online user segmentation based on information-seeking behavior using clustering and neural networks. Applied Soft Computing, 129, 109712.
12. Chen, Y., & Wu, F. (2024). Adaptive classification of online users using reinforcement learning and deep neural networks. Expert Systems with Applications, 204, 117722.
13. Gupta, R., & Mehta, A. (2024). Information-driven user classification in online forums using natural language processing and machine learning. Knowledge-Based Systems, 269, 109888.
14. Kumar, P., & Joshi, M. (2024). Enhanced classification of online users based on search intent and behavior using hybrid ML techniques. Data & Knowledge Engineering, 141, 101964.
15. Li, X., & Zhao, J. (2024). Explainable AI for online user classification: A case study using attention mechanisms. Information Processing & Management, 61(1), 102716.

DOI : https://doi.org/10.5281/zenodo.16810472