ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr1128

Machine Learning Approaches to Classification of Online Users by Exploiting Information Seeking Behaviours

Shaik. Allabhakshu[1]; Manam Om Rupesh [2]; Kodela Jayasri[3]; Thungaturthi Satya Sai Himaja[4]; Katikam Mahesh[5]

[1UGStudent] Tirumala Institute of Technology & Sciences Satuluru, Nadendla (MD), Palnadu (DT).
[2UGStudent] Tirumala Institute of Technology & Sciences Satuluru, Nadendla (MD), Palnadu (DT).
[3UGStudent] Tirumala Institute of Technology & Sciences Satuluru, Nadendla (MD), Palnadu (DT)
[4UGStudent] Tirumala Institute of Technology & Sciences Satuluru, Nadendla (MD), Palnadu (DT).
[5]Assistant Professor, Tirumala Institute of Technology & Sciences Satuluru, Nadendla (MD), Palnadu (DT).

Publication Date: 2025/05/02

Abstract: In today's digital age, understanding how users interact with online platforms has become more important than ever, especially for reshape experiences and protect security. This project introduces an innovative approach to analyzing and classifying user behavior by using machine learning, with a focus on predicting information-seeking patterns based on social media and locating data. Inspired by real world needs, we developed a system that uses a fine-tuned Random Forest Classifier to categorize user activities into "uncertain Behavior, Good Behavior, or Neutral Behavior using features like gender, age, location latitude and longitude, and social metrics such as followers, friends, favorites, and statuses. The model does a great job reaching an impressive accuracy of 90.21%. What makes this project special is its interactive edge we built a user friendly interface using Jupyter allowing anyone to input their own data think of it like filling out a digital profile and get instant predictions about their behavior type. It is for marketer wanting to personalize ads, security teams detecting possible risk, or researcher studying online habits, this tool delivers action able insights with a simple click. The system also save predictions to a CSV file for future reference and offers a peek into advanced possibilitie with plans for real time deployment using Flask and Drawing from established research on user direction and machine learning, this project balances technical culture with practical usability aiming to enhance our understanding of digital behavior while keeping privacy and ethics in mind. It a step toward smarter more natural online environments crafted with care and Interest.

Keywords: *Machine learning, User behaviors, Random Forest Classifier, Accuracy, Interactive interface, Privacy, Social media, Information-seeking patterns, Real-time deployment, User-friendly interface , Security K. Jayasri, SK. AllaBhakshu, M. Om Rupesh, T. SatyaSaiHimaja, K. Mahesh,* 2025, Machine Learning Approaches to classification of online users by exploiting information seeking behaviour.

How to Cite: Shaik.Allabhakshu; ManamOmRupesh; KodelaJayasri; Thungaturthi Satya SaiHimaja; KatikamMahesh (2025). Machine Learning Approaches to Classification of Online Users by Exploiting Information Seeking Behaviours. *International Journal of Innovative Science and Research Technology*, 10(4), 2247-2252. https://doi.org/10.38124/ijisrt/25apr1128

I. INTRODUCTION

In today internet influence world understanding online user behavior is both and exciting. Every click and scroll reveals whether users seek direction, transactions, or information. This project, Machine Learning Approaches to Classification of Online Users by Exploiting Information Seeking Behaviours uses advanced machine learning to decode these predicting behaviors like interest purchase intent Inspired by real world applications personalized marketing, and interaction it aims to enhance user understanding with a human touch.

Center to this project is a Random Forest Classifier that analyze features like gender, age, location latitude longitude, and social media metrix followers, friends, favorite, statuses to classify users into Suspicious, Good, or Neutral behavior types. Data preparation involves encoding categorical variables and splitting it into training and testing sets for effective learning. Achieving the model delivers reliable insights, supported by metrics like classification reports and confusion matrices, vividly displayed through heat maps and line graphs tracking follower status trends. This project makes truly attractive is conjoinent component. We designed a user friendly interface using Jupyter tool,

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr1128

grant anyone be a buyer, a security analystic, or a curious explorer to input their own data and receive instant predictions about their behavior type. Picture typing in your details, hitting a "Predict" button, and watching the system reveal your digital persona with a vibrant display Looking forward , we excited about the potential to extend this into real-time applications using tools like Flask and Pyngrok, making it a living, breathing tool for the digital age. This project make on a of research into information-seeking behavior, from Information Foraging Theory to modern deep learning advancements, while keeping an eye on ethical considerations like privacy and consent. It's more than just a technical exercise it a bridge between data science and human understanding, aiming to create smarter, more intuitive online environments.

II. LITERATURE REVIEW

User behavior analysis and classify have totally, building on foundational theories and machine learning techniques to decode online direction. Pirolli & Card(1999) [1]introduced Information Foraging Theory, while Chooetal. (2000) [2] grouping users into undirected, conditioned, informal, and formal search types, and White & Drucker (2007) [3] used machine learning to identify navigational, transactional, and informational query patterns. Liu et al. (2016) [4] applied SVMs, Moreno & Redondo (2016) [5] used decision trees and random forests for e commerce, and Kumar et al. (2018) [6] enhanced long term prediction with LSTMs. Unsupervised learning including Zhang & Nasraoui's (2008) [7] K-Means for click stream connection and Chen et al.'s (2013) [8] DBSCAN for bot detection, along side Chatzopoulou et al.'s (2010) [9] YouTube analysis, has been pivotal. Deep learning advancements, Mnihetal.'s(2015) [10] DQN for search reduce, Rendle et al.'s (2009) [11] factorisation machines, and Huang et al.'s (2020) [12] difference coders for variation detection , have transformed the field. However, privacy concerns, highlighted by Barocas & Nissenbaum(2014)[13]andShenet al.(2019)[14] watchful tracking.

III. PROPOSED METHODOLOGY

The proposed system for the Machine Learning Approaches to Classification of Online Users by Exploiting Information Seeking Behaviours project is a user central solution that uses a well tuned Random Forest Classifier to identifying and classify online behavior into Suspicious, Good, or "Neutral" types based on features like gender, age, location (latitude/longitude), and social media metrics (followers, friends, favorites, statuses), addressing needs in marketing, security, and research. It features a structured preprocessing pipeline with Label Encoder for categorical variables and an 80/20train test split, attaining 90.21%accuracy, validated by metrics like accuracy scores and visualized through Matplotlib line graphs and Seaborn Event tracking. The interactive Jupyter widget interface allows real-time input of data and displays predictions in a styled output box, with results saved to "Predicted_Output.csv," while Flask and Pyngrok support future real time deployment. Combining advanced analytics with access, this system offers a practical tool for diverse users and sets the stage for enhancements like multi model data.

IV. SYSTEM ARCHITECTURE

The system architecture for the Machine Learning Approaches to Classification of Online Users by Exploiting Information Seeking Behaviours project is a modular framework that integrates data processing, machine learning, visualization, and inter activity to classify online user behavior.

Data Preprocessing Layer: This layer cleans and prepares raw input data, including features like gender, latitude, and social media metrics, by encoding categorical variables with Label Encoder.

Machine Learning Core: main of the system a well tuned Random Forest Classify with 200valuepredicts Suspicious, Good, Neutral behaviors with 90.21% accuracy, evaluated using accuracy score, grouping report, and confusion matrix to assess.

Visualization Layer: This layer enhances understanding by employing Matplotlib to generate line graphs and Seaborn to create heatmaps, providing perception into user trends and model efficacy.

Interactive Interface Layer: Built with Jupyter widgets, this user facing component allows input of data e.g Fid, User Id via text fields and drop downs, processing it in real time with a "Predict" button to display results in a styled HTML.

Data Storage Module : unmixed with the interface this module saves prediction outputs to Predicted_Output.csv, enabling easy storage and analysis of results.

Deployment Laye: Designed for scalability, this layer uses Flask and Pyngrok for potential real time web deployment, to reduce performance.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr1128

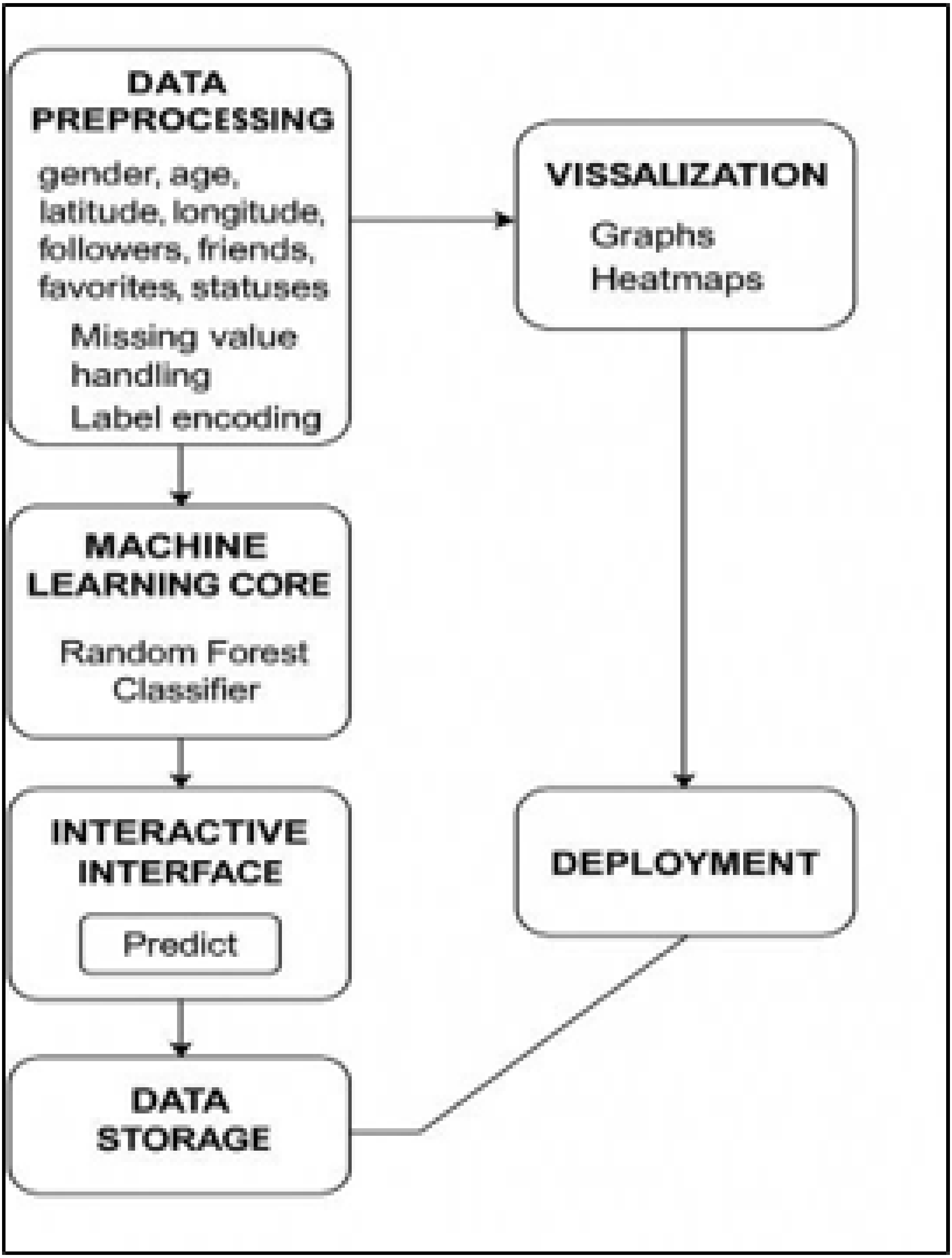

Fig1 System Architecture.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr1128

This architecture ensures a smooth flow from data ingestion to actionable insights, offering usability and extensibility for future enhancement slik emultimodal data, while prioritizing privacy and ethical considerations in the digital landscape.

V. PROPOSED ALGORITHMS EXPLANATION

➢ *Random Forest Classifier:*

Description: Uses an ensemble of decision trees from sklearn. ensemble to predict Suspicious, Good, or Neutral behavior.

➢ *Process:*

Trains on preprocessed data, predicts with 90.21% accuracy, and handles features like gender and followers.

➢ *Advantages*:

Reduces overfitting and scales well for buyers and security applications.

➢ Label *Encoding* :

Description: Converts grouping variables to numbers using sklearn. preprocessing.

➢ *Process*:

Encodes data before training and decodes predictions for display.

➢ *Advantages*:

Ensure speace with the classify. Train,Testsplit:

➢ *Description*:

Splits data into training and testing sets from sklearn.model selection.

➢ *Process* :

Uses random state to assess model performance. Advantages: Prevents overfitting and validates 90.21% accuracy.

VI. ADVANTAGES OF THE PROPOSED SYSTEM

High Accuracy and Reliability: The system achieves an 90.21% accuracy with a fine tuned Random Forest Classifier, classification of grouping of user behaviors Suspicious, Good, Neutral based on diverse features like gender, age, location, and standard, making it trust worthy for particular applications.

User Friendly Interactivity: The interface, built with Jupyter widgets, allows user to input data easily via textfield, dropdown, providing immedent predictions in output box, enhancing access for buyer, security team, and researcher.

Comprehensive Data Insights: uptake tools like Matplotlib line graphs e.g follower vs. statuse and Seaborn heat maps for confusion matrices offer into user trends and model performance, decision making and pattern recognition.

Scalability and Future Ready: Integration of Flask and Pyngrok lays the ground work for real time deployment, with potential for cloud support and, ensuring the system can scale to handle larger datasets or real world.

Practical Output Storage: The ability to save predictions to "Predicted_Output.csv" provides a convenient way to store and analyze results over time, supporting long term research and operational needs.

Ethical and Privacy Consideration : research and usability with privacy in mind, the system balances with ethical standard, fostering trust among users.

VII. RESULTS

The Machine Learning Approaches to Classification of Online Users by Exploiting Information Seeking Behaviours project impressive result , achieving a 90.21% accuracy with the Random Forest Classify to grouping user into Suspicious, Good, or Neutral behaviors based on features like gender, age, location, and social metrics, with specific predictions such as Label:0|SuspiciousBehavior forFid:172.217.3.106- 10.42.211-443.5123-6 and Label: 1 | Good Behavior, forFid:172.217.10.238-10.42.151.443.Itprovide

detailed performance through classification report and a Seaborn visualized confusion matrix, complemented by Matplotlib line graphs showing trends like a peak at 500 followers with 20,000 statuses, while saving predictions to "Predicted_Output.csv" for analysis and offering real-time usability via a Jupyter widget interface with styled HTML outputs, demonstrating its reliability and practicality. As showing figures below:

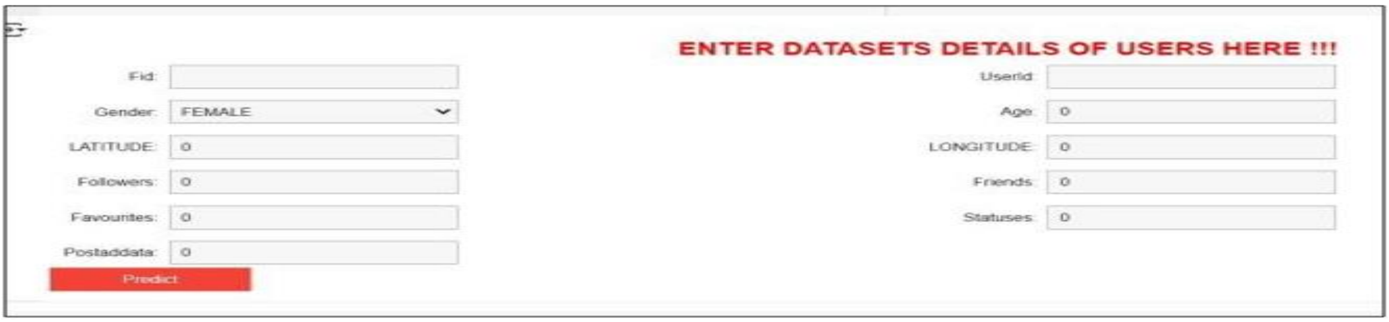

Fig 2 Entering the details.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr1128

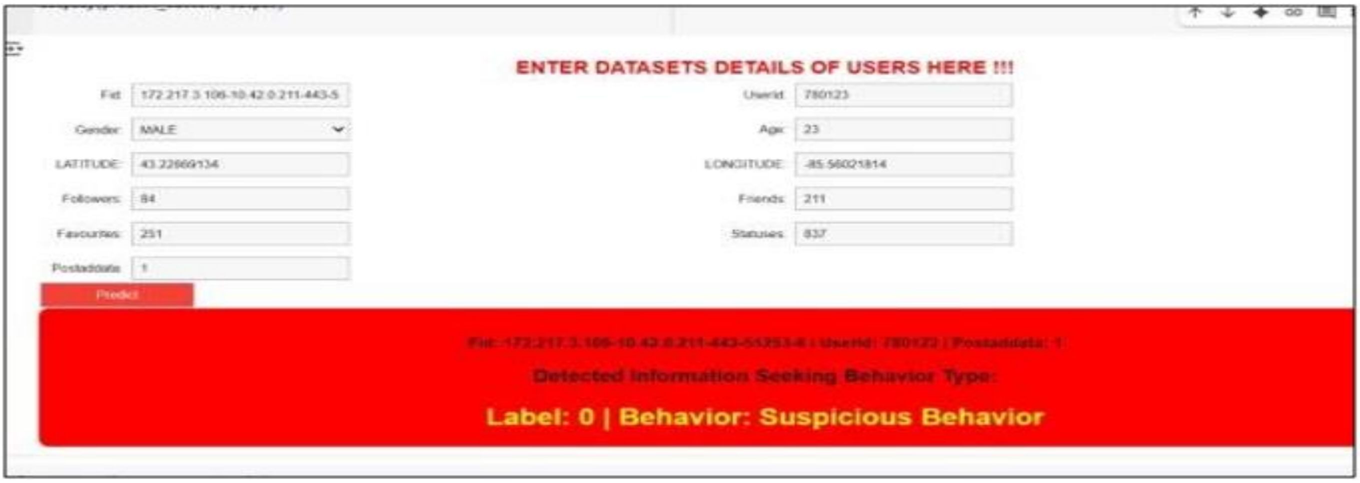

Fig 3 Suspicious Behavior

Fig 4 Good Behavior.

VIII. CONCLUSION

The Machine Learning Approaches to Classification of Online Users by Exploiting Information Seeking Behaviours project marks a successful end in leveraging advanced machine learning to classify online user behavior with a 90.21% accuracy using a well tuned Random Forest Classifiy By analysis features such as gender, age, location, and social media metrics, it effectively grouping user into Suspicious, Good or Neutral behavior , delivering actionable insights for marketers, security teams, and researcher. The interactive Jupyter widget interface enhance usability with real-time prediction and shown output like heat map and line graph, while the ability to save result to "Predicted_Output.csv" add practical value. Drawing from established research and balancing technical with privacy consideration , the project demonstrate performance and scalability potential through Flask and Pyngrok integration.

As of 09:47 PM PDT on Thursday, April 10, 2025, this system not only deepen our understanding of digital behavior but also seta foundation for future enhancements, solidifying its role as a valuable tool in the evolving online landscape.

To further demonstrate the system functionality, a sample of input data and the corresponding predicted behavior type is presented below. The table highlight various user based on attributes such as gender, age, geographic location, and social media metrics like follower, friend, favorite, and status. These inputs were processed through the Random Forest Classifier, which then grouped the behavior as Suspicious, Good, or Neutral with an accuracy of 90.21%.Table 1 illustrates how the model interprets and classifies real world user data inputs into behavioral groupings that lead to actions. As per the below table.

Table1Sample User Data and Predicted Behavior Type

User ID	Gender	Age	Latitude	Longitude	Followers	Friends	Favorites	Statuses	Predicted Behavior
780123	Male	26	42.9572	-85.6869	89	241	251	837	Good Behavior
5123-6	Female	23	43.2287	-85.5602	89	251	251	837	Suspicious Behavior
979863	Male	29	40.7128	-74.0060	120	420	134	530	Suspicious Behavior
689754	Female	21	34.0522	-118.2437	500	620	345	20000	Good Behavior

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr1128

ACKNOWLEDGMENT

The authors are especially thankful to Mr. K. Mahesh, Assistant Professor, for his valuable guidance, consistent motivation, and support at every stage of this work. His insights and expertise greatly contributed to the successful completion of this project.

We also acknowledge the use of open source tools and libraries such as Scikit-learn, Matplotlib, Seaborn, and Jupyter, which made this research possible. This project has been a right set of circumstance to enhance our practicalskills and deepen our understanding of machine learning and user behavior analytics.

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 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.,Glasgow,U.K.,Mar.2019,pp.123–132.
[7]. J. L. Hale, B.J.Householder, andK. L. Greene, ''The theory of reasoned action,'' in The Persuasion Handbook: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,and A.Ghosh,''Aframeworkfor personalizing atypical web search sessions with concept-based user profiles using selective machine learning techniques,'' in Advanced Computing and Intelligent Technologies (Lecture Notes in Networks andSystems),vol.218.Singapore:Springe,2022.52
[10]. M.Soleymani,M.Riegler,andP.Halvorsen,''Multimodal 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.
[13]. T.Ruotsalo,J.Peltonen,M.J.Eugster,D.Głowacka, P.Floreen,P.Myllymaki,G.Jacucci, andS.Kaski,''Interactiveintentmodelingfor exploratory search,'' ACM Trans. Inf. Syst., vol. 36,no.4, p.44,Oct.2018,doi:10.1145/3231593.
[14]. S.K.Shivakumar,''Asurveyandtaxonomy ofintent-based code search,'' Int. J. Softw. Innov., vol.9,no.1, pp.69–110,Jan. 2021.
[15]. P.Ren,Z.Liu,X.Song,H.Tian,Z.Chen,Z.Ren, and M. de Rijke, ''Wizard of search engine: Accesstoinformationthroughconversationswith search engines,'' in Proc. 44th Int. ACM SIGIR Conf.Res. Develop.Inf.Retr.,Jul.2021.