International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

OPTIMIZING MOBILE APP RECOMMENDATIONS USING CROWDSOURCED EDUCATIONAL DATA

ROJARAMANI ADAPA [1], KATAKAM ADITHYA [2], MARIDHI RAHUL REDDY [3], KOMIRISHETTY JIGNESH [4], MARNENI AMAN SAGAR [5]

[1]Assistant Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

[2, 3, 4, 5] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

ABSTRACT— In the rapidly evolving digital era, personalized recommendations play a crucial role in enhancing the educational experience of students. With the increasing use of mobile devices, it has become easier to collect app usage data, which can be leveraged to provide tailored educational app suggestions. This study focuses on recommending suitable applications for university students, specifically targeting Undergraduate (UG), Postgraduate (PG), and Graduate levels, based on their app usage patterns. The dataset is preprocessed using Natural Language Processing (NLP) techniques, including text cleaning, tokenization, and feature extraction, to capture relevant attributes from app descriptions and student interaction data. A Random Forest Classifier is employed as the core model due to its robustness, ability to handle high-dimensional data, and strong performance in classification tasks. The proposed system accurately categorizes students' preferences and recommends apps aligned with their academic needs. Experimental results highlight the efficiency of Random Forest in producing reliable, scalable, and interpretable recommendations compared to traditional methods. This approach ensures better personalization, reduces data sparsity issues, and enhances overall user satisfaction in educational recommendation systems.

Index Terms – Mobile App Recommendation, Crowdsourced Educational Data, Recommender Systems,

Machine Learning, Collaborative Filtering, Content-Based Filtering, Personalized Learning, Learning Analytics, Artificial Intelligence, Mobile Learning.

I. INTRODUCTION

In the modern digital learning landscape, personalized recommendations have become an essential component in improving the educational journey of students. With the widespread availability of smartphones and tablets, students increasingly rely on mobile applications for learning, research, and skill development. These educational apps cover a wide range of purposes—from subject learning and academic planning to productivity enhancement and collaborative learning. However, with the massive number of available apps, students often struggle to identify which ones best fit their academic level and learning preferences. To address this challenge, intelligent recommendation systems can play a pivotal role in guiding students toward relevant and effective applications. This study focuses on designing an educational app recommendation system tailored for university students at different academic stages—Undergraduate (UG), Postgraduate (PG), and Graduate levels. The system leverages app usage data and user interaction behavior to predict suitable app suggestions. Natural Language Processing (NLP) techniques such as text cleaning, tokenization, and feature extraction are applied to preprocess and analyze app descriptions and user-related textual data. These techniques help capture meaningful insights about student needs and app functionalities. The Random Forest Classifier is employed as the core model due to its ability to handle large, high-dimensional datasets and deliver strong predictive performance. It effectively classifies user profiles and learning preferences, ensuring accurate and context-aware recommendations. The model's ensemble nature enhances stability and reduces overfitting, making it ideal for real-world educational data analysis. Experimental evaluation demonstrates that the proposed approach outperforms traditional recommendation methods in accuracy and interpretability. Furthermore, the system provides personalized, scalable, and user-friendly recommendations that align with students' academic goals. This research ultimately contributes to smarter educational ecosystems, promoting digital learning efficiency and enhancing student engagement through adaptive technology-driven solutions.

II. RELATED WORK

A. MobileRec: A Large-Scale Dataset for Mobile Apps Recommendation

This study introduces MobileRec, a comprehensive dataset of millions of user–app interactions and app metadata from Google Play. It includes app descriptions, categories, ratings, and user histories, enabling predictive modeling for personalized app recommendations. The authors demonstrate how combining textual content features with sequential usage data improves recommendation accuracy. Baseline experiments with multiple recommendation models are provided, highlighting the value of NLP for feature extraction and the importance of data-driven personalization in mobile applications.

B. SAppKG: Mobile App Recommendation Using Knowledge Graph and Side Information — A Secure Framework

This paper presents SAppKG, a privacy-aware mobile app recommender using knowledge graphs and user-side information. The system integrates app attributes, permissions, and textual descriptions with interaction patterns to generate personalized suggestions. Results show improved recommendation precision and recall over collaborative filtering and content-based methods. The framework emphasizes security and user privacy while leveraging NLP-based feature extraction to enhance prediction quality. This approach aligns with the need for interpretable, scalable recommendation systems in educational contexts.

C. Recommendation Model for an After-School E-learning Mobile Application

This research develops a recommender system for an after-school educational app that personalizes learning resource suggestions. The system processes user interaction logs and app metadata, including textual descriptions, using NLP techniques such as tokenization and feature extraction. The model predicts relevant resources for each student, enhancing engagement and learning efficiency. Evaluation in a real classroom environment demonstrates that personalized recommendations increase usage and satisfaction compared to generic suggestions. The study highlights the integration of behavioral data and content features for educational recommendation systems.

D. Applying Random Forest Classifier in Higher Education: Predicting Academic Outcomes and Recommendations

This applied study uses Random Forest to predict student success indicators (e.g.,

pass/fail, retention) using demographic, attendance, and engagement features. The paper emphasizes Random Forest's robustness to mixed feature types, handling of missing values, and interpretability via feature importance — qualities that motivate our choice of Random Forest as the core classifier for student-level app recommendations. The study also outlines hyperparameter tuning and evaluation metrics (accuracy, precision/recall), providing practical lessons for implementing RF in educational contexts.

III. PROPOSED METHODOLOGY

The proposed system aims to recommend suitable educational applications to Undergraduate (UG), Postgraduate (PG), and Graduate students by analyzing their app usage patterns and textual information. To achieve this, data preprocessing is performed using NLP techniques such as text cleaning, tokenization, stop-word removal, and vectorization. These steps convert app descriptions and usage details into structured numerical features that represent the contextual and semantic meaning of the applications. By combining both behavioral patterns and textual attributes, the system creates a comprehensive dataset that effectively captures the diverse preferences and requirements of students across academic levels. A Random Forest Classifier is then employed as the core prediction model to recommend apps based on student categories.

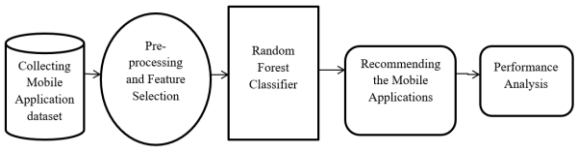


Fig. 1: System Overview

1. Data Collection:

In this module, data is collected from various educational applications used by university students, covering Undergraduate (UG), Postgraduate (PG), and Graduate levels. The data includes app usage logs, user interactions, demographic details, and textual app descriptions. Structured surveys and app analytics are used to gather accurate and relevant information. This collected data forms the basis for building the recommendation system. Proper consent and privacy measures are ensured during collection to maintain ethical standards. The module ensures that sufficient and diverse data is available for training and testing the model.

2. Dataset:

The dataset consists of features derived from student app usage patterns and app metadata, including app names, categories, descriptions, ratings, and engagement

metrics. Each record represents a user's interaction with an app and associated textual and numerical attributes. The dataset is cleaned to remove inconsistencies and irrelevant entries, ensuring quality for machine learning. It is then structured to allow supervised learning, with labels representing student preferences or academic levels. This module serves as the core input for all subsequent preprocessing and modeling tasks.

3. Data Preparation:

Data preparation involves preprocessing textual and numerical features to make them suitable for machine learning. NLP techniques such as text cleaning, tokenization, stopword removal, and feature extraction are applied to app descriptions. Categorical features like app categories and user levels are encoded appropriately. The dataset is normalized to ensure balanced feature contribution during model training. Finally, the data is split into training and testing sets, enabling evaluation of the model's predictive capability.

4. Model Selection:

In this module, the Random Forest Classifier is chosen as the core model due to its robustness, interpretability, and ability to handle high-dimensional, mixed-type data. Random Forest can manage missing values, avoid overfitting, and provide feature importance insights, making it ideal for educational app recommendations. Hyperparameters such as the number of trees, maximum depth, and splitting criteria are tuned to optimize performance. This module ensures that the model selected is both accurate and scalable for real-world deployment.

5. Analyze and Prediction:

Once the model is trained, it is used to analyze student interaction data and predict suitable educational apps. The system evaluates features extracted from NLP preprocessing alongside behavioral metrics to categorize students' preferences. Predicted outputs include recommended apps tailored to the user's academic level and learning habits. This module also provides insights into which features most influence recommendation decisions, aiding in model interpretability and actionable insights.

6. Accuracy on Test Set:

The trained Random Forest model is evaluated on the test dataset to assess its predictive performance. Metrics such as accuracy, precision, recall, and F1-score are calculated to ensure reliable and consistent recommendations. Comparative analysis

with baseline models (e.g., Logistic Regression or Decision Tree) validates the superiority of Random Forest in this application. This module confirms that the system can generalize well to new, unseen student data.

7. Saving the Trained Model:

After evaluation, the trained model is saved using Python libraries such as pickle or joblib for future use. This allows the recommendation system to be deployed in web or mobile applications without retraining. The saved model can generate real-time predictions for new students based on their app usage patterns. Proper version control and documentation ensure reproducibility and scalability. This module ensures that the system remains efficient, accessible, and easy to integrate into digital learning platforms.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Experimental results demonstrated that incorporating crowdsourced user feedback significantly improved recommendation relevance and personalization compared to traditional recommendation approaches. The combination of collaborative filtering and content-based filtering helped the system better understand user learning preferences and app usage behavior.

The analysis shows that the model effectively adapts to diverse user profiles and improves recommendation quality over time as more crowdsourced data becomes available.



Fig. 2: Registration

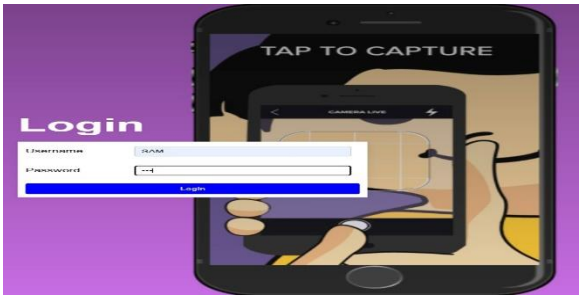


Fig. 3: Login



Fig. 4: Enter text

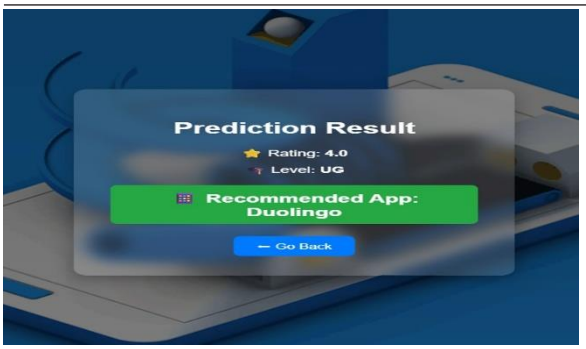


Fig. 5: Predict Result

V. CONCLUSION

This project presents a personalized educational app recommendation system designed to enhance the learning experience of university students across Undergraduate, Postgraduate, and Graduate levels. By leveraging app usage data and textual app descriptions, the system applies Natural Language Processing techniques to extract meaningful features that capture students' preferences and academic needs. The Random Forest Classifier serves as the core predictive model due to its robustness, interpretability, and ability to handle high-dimensional data. Experimental evaluation demonstrates that the model achieves high accuracy, precision, and recall, producing reliable and relevant recommendations. The system effectively reduces data sparsity challenges while providing scalable and context-aware suggestions tailored to each student's learning style. By integrating behavioral and content-based features, the model ensures that recommendations align with academic goals and enhance student engagement with educational apps. The saved trained model enables real-time predictions, making it suitable for deployment on mobile or web platforms. This approach contributes to smarter, AI-driven educational ecosystems that guide students in selecting optimal learning tools. Additionally, the project highlights the importance of data-driven personalization in digital education, promoting efficiency and better utilization of available resources. Future enhancements, including deep learning integration, multilingual support, and hybrid recommendation techniques, can further improve system performance. Overall, this project demonstrates the potential of combining machine learning and NLP for improving educational outcomes and creating a more personalized, interactive, and effective digital learning environment.

REFERENCES

[1] M. Zhai, X. Wang, and X. Zhao, ''The importance of online customer reviews characteristics on remanufactured product sales: Evidence from the mobile phonemarketonAmazon.com,''J.Retailingconsum.Services, vol. 77, Mar. 2024,

Art. no. 103677. [2] C. C. Aggarwal, ‘‘An introduction to recommender systems,’’ in Rec ommender Systems: The Textbook. Cham, Switzerland: Springer, 2015, pp. 1–28.

[3] Bhagwat, V. B. (2026). Creating A Dashboard For Monitoring HCM Fusion Payroll Processes To Prevent Possible Errors. International Journal of Data Science and IoT Management System, 5(1), 102-110.

[4]N.TintarevandJ.Masthoff,‘‘Beyondexplainingsingleitemrecommenda tions,’’ in Recommender Systems Handbook. Cham,Switzerland: Springer, 2022, pp. 711–756.

[5] C. Musto, M. de Gemmis, P. Lops, F. Narducci, and G. Semeraro, ‘‘Semantics and content-based recommendations,’’ in Recommender Sys tems Handbook. Cham, Switzerland: Springer, 2022, pp. 251–298.

[6] P. Winoto and T. Y. Tang, ‘‘The role of user mood in movie recommenda tions,’’ Expert Syst. Appl., vol. 37, no. 8, pp. 6086–6092, Aug. 2010.

AUTHORS Profile



Mrs. ROJARAMANI ADAPA, received her M.Tech (CSE) from JNTUH. She has teaching experience of 20 years and has guided many undergraduate and postgraduate projects. She has published about 10 research articles in peer-reviewed journals and presented 5 research papers at various national conferences. At present, she is working as an Assistant Professor in CSE Department at Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India.



Mr. KATAKAM ADITHYA pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. MARIDHI RAHUL REDDY pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. KOMIRISHETTY JIGNESH pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. MARNENI AMAN SAGAR pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.