Algorithmic Pathfinders: Advancing University Recommendations through Hybrid Models

[1]Advait Mandar Mandke, [2]Kashmira Chetan Shah, [3]Prajwal Annaso Narute, [4]Aishwarya Mane

[1,2,3]Research Scholar, [4]Assistant Professor, Marathwada Mitra mandal's College of Engineering, Pune, India, [1]*advait.mandke@gmail.com*, [2]*kashmirachetan75@gmail.com*, [3]*prajwalnarute02@gmail.com*, [4]*aishwaryamane@mmcoe.edu.in*

Abstract: Recommender Systems are widely used in various domains to provide personalized suggestions to users, particularly for less critical decisions. However, in areas where decisions carry significant implications, such as selecting foreign universities for education, recommendation systems should not only focus on immediate engagement but also on promoting diversity and aligning with long-term preferences. The process of choosing the right foreign institution for education can be complex due to admission criteria and academic resources, potentially leading to applicants wasting time and money by applying to unlikely universities. To address this issue, this study proposes the development of a recommendation system for graduate admissions. This system would leverage historical data of past students and multiple factors to offer tailored university suggestions to applicants. Drawing inspiration from previous papers tackling similar challenges, the research aims to identify the most efficient technique through a comprehensive evaluation process. This involves examining past research, evaluating performance metrics, and comparing methodologies to determine the technique that excels in precision, robustness, and user satisfaction. The findings of this research promise not only an effective recommendation system but also the expansion of possibilities for machine learning and data science in guiding higher education choices. By offering a more informed and personalized academic journey, this research underscores the role of modeling in shaping academic futures and sets the stage for innovative data-driven advancements in university recommendation systems.

Keywords — Web Scraping, Higher education, Ensemble Learning, Intelligent Recommender System (IRS), University Recommendation System(URS.)

I. INTRODUCTION

In today's dynamic scene of data and innovation, machine learning (ML) has developed as a transformative constraint, reshaping decision-making forms over different areas. Established within the thought that frameworks can learn and make strides independently, ML has revolutionized businesses such as healthcare and funds. Eminently, recommender frameworks, a subset of ML applications, have picked up prominence.

Recommender systems analyse client inclinations to supply custom-made recommendations, customarily in divisions like amusement and e-commerce. Be that as it may, their esteem is progressively recognized in the scholarly world, especially in selecting outside colleges for higher instruction[1]. In a period of plenteous data and different scholarly alternatives, finding the perfect college for graduates is a multifaceted challenge.

Utilizing modern approaches, counting data-driven models, gets to be basic to explore the complexities of the higher instruction scene viably. The headway of information analytics, ML, and manufactured insights prepares us with devices to address these complexities. By drawing motivation from pertinent investigation papers, we think about points to investigate the potential of modelling in instruction and distinguish the foremost successful method for college proposal systems[2]. Through an exhaustive assessment handle, centring on metrics such as exactness, accuracy, review, and F1 score, we look to supply an fair evaluation of different methods.

By comparing discoveries and techniques from past investigations, our objective is to pinpoint the strategy that exceeds expectations in terms of accuracy, strength, and client satisfaction. This inquires about not as it were points to disclose an viable method but too extends the skylines of ML and information science in directing higher instruction choices. By cultivating personalized scholarly ventures driven by information, we aim to require a step towards a more informed and user-centred approach to educational decision-making. Ultimately, this comparative study underscores the pivotal role of modelling in shaping the

 DOI : 10.35291/2454-9150.2024.0071

International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-09, Issue-12, Mar 2024

academic futures of individuals and paves the way for innovative data-driven advancements in university recommendations.

II. LITERATURE SURVEY

A. Helping university students to choose elective courses by using a hybrid multi-criteria recommendation system with genetic optimization[1].

The paper presents a hybrid multi-criteria course recommendation system that combines collaborative filtering (CF) using student ratings, grades, and branches with content-based filtering (CBF) using course professors, competences, knowledge area, and contents. A genetic algorithm optimizes the weights for each criteria and configuration parameters like similarity metrics. Advantages include automatically finding optimal criteria weights and configurations to improve recommendations. Limitations include needing large datasets. Overall, it proposes a flexible hybrid recommendation system that combines CF and CBF techniques and uses genetic algorithm optimization to configure the model for optimal performance.

B. Enhancing recommendation stability of collaborative filtering recommender system through bio-inspired clustering ensemble method[2].

The literature survey encompasses various methodologies in recommender systems, particularly focusing on bio-inspired clustering techniques. Proposed approaches demonstrate superiority over recent works in recommendation accuracy and stability, leveraging bio-inspired clustering ensembles and hybrid clustering methods. While the bagging approach aids recommendation generation, the emphasis remains on collaborative filtering, clustering, and bio-inspired techniques. Evaluation on real-world datasets like Yelp and TripAdvisor underscores practical applicability. However, specific limitations are not explicitly addressed, although user satisfaction is highlighted as crucial in evaluation. The integration of bio-inspired clustering ensembles enhances recommendation stability and accuracy, contributing to more efficient collaborative filtering systems.

C. The University Recommendation System for Higher Education[3].

This research delves into recommendation systems in education, particularly focusing on predicting student needs and suggesting suitable universities based on user data. Emphasizing the importance of recommendation systems in education due to data proliferation and the abundance of opportunities provided by the internet, the paper explores various data mining methods to uncover hidden knowledge. While highlighting the significance of recommendation systems in predicting student needs and reducing information overload, the study acknowledges the lack of learning in the education field, which often leads to wrong university choices. Despite enhancing user experience and increasing the chances of additional product purchases, the survey underscores the need for understanding user preferences and addressing potential limitations in educational decision-making processes.

D. Multi-Criteria Review-Based Recommender System—The State of the Art[4].

The literature survey explores the utilization of multi-criteria recommender systems (RSs) to enhance accuracy, particularly through the incorporation of user-generated reviews. By focusing on multi-criteria review-based RSs, the study aims to improve performance and offer personalized recommendations. Integrating user-generated reviews not only enhances RS performance but also mitigates potential issues. The combination of multi-criteria decision analysis (MCDA) with RSs further enhances personalized recommendations based on user behaviour. While collaborative filtering (CF) approaches offer serendipity, they may lead to overspecialization, underscoring the importance of fine-grained analysis in user behaviour. The integration of review elements not only improves the accuracy of RSs but also enhances user and item profiles for more precise recommendations, despite potential limitations such as overspecialization.

E. College Recommendation System for Students Using Datamining With Collaborative Filtering Algorithm–2[5].

This survey examines a college recommendation system designed to aid students in selecting ideal courses and colleges by utilizing the collaborative filtering algorithm. Implemented through the RapidMiner tool for data processing and recommendation, the system ranks colleges based on criteria such as NAAC, NBA ratings, and placements. While addressing student difficulties in college selection due to a lack of information, the study underscores the importance of leveraging collaborative filtering to recommend colleges effectively. Despite assisting students in achieving their educational goals and dreams, limitations may arise from insufficient data, highlighting the need for further research to enhance recommendation accuracy and address student concerns comprehensively.

F. An Effective Recommendation System to Forecast the Best Educational Program Using Machine Learning Classification Algorithms[6].

The paper investigates the recommendation of optimal academic paths for students using machine learning (ML) algorithms. By employing correlation-based feature selection and multiple ML algorithms for forecasting, the study aims to select the best model through cross-validation techniques. Performance evaluation metrics like Log loss and ROC-AUC are utilized to assess model efficacy. Despite focusing on recommending best academic paths

 DOI : 10.35291/2454-9150.2024.0071

and accurately forecasting academic performance, limitations in scope and usability exist within the literature, prompting the need for further research to address these constraints comprehensively.

III. PROPOSED METHODOLOGY

Data Collection Process:

The data collection process in our project utilized web scraping, an automated technique for gathering data from web pages. This method was chosen due to its efficiency in acquiring specific information necessary for our research, such as university names, GRE, TOEFL, and IELTS scores, among other relevant data points. We primarily sourced our data from various online platforms catering to students aspiring for higher education abroad, as these platforms provided comprehensive insights into colleges, courses, admission criteria, and other pertinent information.

To execute web scraping, we began by identifying the data sources and determining the specific data components to be retrieved. This involved pinpointing university names, GRE, TOEFL, IELTS scores, and other relevant details, and selecting multiple websites as our chosen data repository. We employed the Quick Web Scraper tool for this purpose, specifying the data pieces required for extraction.

Following data source identification, the subsequent step was configuring the web scraper. Leveraging the Instant Web Scraper tool, we configured our scraper by inputting the website URLs, selecting the desired data elements, and defining the scraping parameters. Additionally, we chose the export format for the collected data, opting for a CSV file for easy handling and analysis.

Once the web scraper was configured, we initiated the scraping process to extract data from the websites. The Instant Web Scraper tool navigated through the sites, locating the specified data elements, and retrieving the necessary information. Throughout this process, we closely monitored the scraping operation to ensure the reliability and accuracy of our data collection efforts.

Data Preprocessing:

Data preprocessing constituted a crucial stage in our methodology, aimed at ensuring the quality and relevance of the dataset. We initiated this phase by performing thorough data cleaning to address inconsistencies and remove outliers, thereby enhancing the integrity of the data. Subsequently, we undertook data transformation to convert raw data into a structured format amenable to analysis.

Feature scaling techniques were applied to standardize variables such as GRE scores, ensuring uniformity and facilitating comparison across different data points. Moreover, we conducted feature selection to identify the most informative attributes, thereby refining the dataset for model training.

Correlation coefficient : Corr(x,y) = cov(x,y)/std(x)*std(y)

Model Training:

In the model training phase, we leveraged ensemble learning techniques to develop robust predictive models for recommendation systems. Specifically, we integrated stacking, a form of ensemble learning, which involved combining collaborative-based filtering and multicriteria-based filtering methodologies.

Collaborative-based filtering utilized user interactions to make recommendations, while multicriteria-based filtering considered multiple dimensions to refine the recommendation process. By fusing these complementary approaches within a stacking framework, we aimed to capitalize on their strengths and enhance predictive performance. Model training involved optimizing model parameters and hyperparameters to achieve superior predictive accuracy.

User-Item Rating Prediction =

$$\hat{r}_{u,i} = \frac{\sum_{v \in N(u,i)} \text{sim}(u,v) \times (r_{v,i})}{\sum_{v \in N(u,i)} |\text{sim}(u,v)|}$$

$$\text{TOPSIS Score} = \frac{\sqrt{\sum_{j=1}^{n} (w_j \times \text{best}_j - x_{ij})^2}}{\sqrt{\sum_{j=1}^{n} (w_j \times \text{best}_j - \text{worst}_j)^2}}$$

Model Evaluation:

Model evaluation served as the final stage in our methodology, where we rigorously assessed the performance of the developed models. We utilized comprehensive evaluation metrics such as accuracy, precision, recall, and F1 score to quantify the predictive accuracy and robustness of the models across diverse scenarios.

Cross-validation procedures were employed to ensure the reliability and generalizability of the results. Comparative analyses were conducted to benchmark the performance of our approach against existing methods or baseline models. The findings from model evaluation provided valuable insights into the effectiveness and practical implications of the proposed methodology for recommendation systems and personalized services.

Performance of Stacking Techniques:

The integration of stacking techniques, leveraging collaborative-based filtering and multicriteria-based filtering, demonstrated promising results in enhancing predictive accuracy. Through the fusion of these complementary methodologies within a stacking framework, we observed notable improvements in predictive performance compared to individual techniques.

The collaborative-based filtering approach leveraged user interactions to make recommendations, while the multicriteria-based filtering considered multiple dimensions to refine the recommendation process. Our experiments highlighted the synergistic effects of combining these

 DOI : 10.35291/2454-9150.2024.0071

approaches, resulting in enhanced predictive performance across various evaluation metrics.

By combining collaborative-based and multicriteria-based filtering within a stacking framework, our approach synergistically improved predictive accuracy, leveraging user interactions and considering multiple dimensions for refined recommendations.

Evaluation Metrics:

The performance of the predictive models was rigorously evaluated using a range of evaluation metrics, including accuracy, precision, recall, and F1 score. Cross-validation procedures were employed to ensure the reliability and robustness of the results. Our experiments demonstrated consistently high performance across these evaluation metrics, indicating the effectiveness of the proposed methodology in accurately predicting relevant outcomes.

Comparative Analysis:

Furthermore, we conducted a comparative analysis to benchmark the performance of our approach against existing methods or baseline models. Our experiments revealed that the proposed methodology outperformed baseline models in terms of predictive accuracy and robustness.

This comparative analysis underscores the efficacy of the proposed methodology in addressing the challenges inherent in recommendation systems and personalized services.

Discussion and Implications:

The results of our study hold significant implications for various domains, including education, e-commerce, and personalized services. By leveraging ensemble learning techniques and meticulous data preprocessing, our approach offers a promising avenue for improving the effectiveness and efficiency of recommendation systems. Moreover, the insights gained from our experiments provide valuable guidance for practitioners and researchers seeking to develop more accurate and robust predictive models in diverse application domains.

IV. RESULT AND DISCUSSION

Findings

In our study, we employed a novel approach by combining stacking methods that integrate collaborative-based filtering and multicriteria-based filtering. This unique fusion strategy resulted in significant improvements in predictive accuracy compared to using each method independently. Specifically, we observed that collaborative-based filtering achieved an accuracy of 96.5%, while multicriteria-based filtering reached an accuracy of 99.2%. These findings indicate the efficacy of our methodology in accurately predicting outcomes, as evidenced by the high level of accuracy achieved by both approaches.

Moreover, when evaluating precision, recall, and F1 score as additional evaluation metrics, our methodology consistently demonstrated strong performance across various scenarios. With precision recorded at 0.88, recall at 0.84, and F1 score at 0.86, our approach showcases reliability and effectiveness in predicting outcomes across diverse datasets and conditions. These metrics collectively reinforce the robustness of our proposed methodology and its ability to deliver accurate predictions.

In comparative analyses against baseline models, our approach consistently outperformed them in terms of predictive accuracy and robustness. The methodology showcased a notable 12% improvement in accuracy over these baseline models, highlighting its effectiveness in addressing challenges inherent in recommendation systems and personalized services. These findings underscore the significance of our approach in advancing the field of predictive modeling and recommendation systems.

Discussion and Implications

The results of our study carry significant implications across various sectors, including education, e-commerce, and personalized services. By leveraging ensemble learning techniques and rigorous data preprocessing steps, our approach presents a promising opportunity for enhancing the efficiency and effectiveness of recommendation systems. The fusion of collaborative-based filtering and multicriteria-based filtering not only improves predictive accuracy but also enhances the reliability of recommendations made to users.

Furthermore, the insights gained from our experiments offer valuable guidance for professionals and scholars working towards creating dependable predictive models in diverse fields of application. The methodology outlined in our study provides a framework for developing more accurate and robust recommendation systems, thereby improving user experience and satisfaction in domains such as online education platforms, e-commerce websites, and personalized service providers.

Overall, our findings contribute to advancing the state-of-the-art in predictive modeling and recommendation systems, offering practical solutions to challenges faced by industries reliant on data-driven decision-making processes. By implementing our methodology, organizations can enhance their recommendation systems, leading to more personalized and relevant user experiences, ultimately driving greater engagement and satisfaction.

V. CONCLUSION

In conclusion, our research has presented a comprehensive methodology for enhancing recommendation systems through meticulous data preprocessing, feature selection, and ensemble learning techniques. Leveraging the Scrapy framework, we constructed a robust dataset and conducted thorough preprocessing steps to ensure data quality and

 DOI : 10.35291/2454-9150.2024.0071

relevance. Feature selection further refined our models, resulting in enhanced predictive accuracy.

The integration of stacking techniques, including collaborative-based and multicriteria-based filtering, showcased significant improvements in predictive performance. Our experiments consistently demonstrated the efficacy of the proposed methodology, as evidenced by high evaluation metrics across diverse scenarios. These findings underscore the potential of ensemble learning techniques in recommendation systems, offering practical implications for personalized services and user-centric applications.

Our study contributes to the ongoing discourse in machine learning by providing insights into the development of more accurate and efficient recommendation algorithms. Moving forward, further research could explore additional ensemble techniques and incorporate real-time data to enhance the adaptability and scalability of recommendation systems. Overall, our research paves the way for advancements in personalized services and recommendation algorithms, driving innovation in user-centric applications across various domains.

ACKNOWLEDGMENT

We are grateful to our external project guides Mr. Gaurav Prakash Saraf (srf.gaurav@gmail.com Synechron Technologies pvt. ltd.), Dr. Aarti Amod Agarkar (pratibha26@gmail.com) and Ms. Priya Ajit Natu (priya.a.natu@gmail.com Citius Tech) for their mentorship for our research.

During the course of this project, we had made every effort to work with sincerity, dedication, understanding and focus. We were taught and guided by our mentors about doing a thorough research about the project's process flow, it's coding part and the implementation cycle. We were further mentored to have a logical and systematic approach in our thought process. And we are very happy and satisfied to state here that all the check boxes and mandates of our mentors are now ticked off.

REFERENCES

[1] A. Esteban, A. Zafra, and C. Romero, "Helping university students to choose elective courses by using a hybrid multi-criteria recommendation system with genetic optimization," Knowledge-Based Systems, vol. 194, p. 105385, Apr. 2020, doi: https://doi.org/10.1016/j.knosys.2019.105385.

[2] R. Logesh, V. Subramaniyaswamy, D. Malathi, N. Sivaramakrishnan, and V. Vijayakumar, "Enhancing recommendation stability of collaborative filtering recommender system through bio-inspired clustering ensemble method," Neural Computing and Applications, vol. 32, no. 7, pp. 2141–2164, Nov. 2018, doi: https://doi.org/10.1007/s00521-018-3891-5.

[3] A. Nalawade* and B. Tiple, "The University Recommendation System for Higher Education," International Journal of Recent Technology and Engineering (IJRTE), vol. 8, no. 6, pp. 1692–1696, Mar. 2020, doi: https://doi.org/10.35940/ijrte.f7632.038620.

[4] S. M. Al-Ghuribi and S. A. Mohd Noah, "Multi-Criteria Review-Based Recommender System–The State of the Art," IEEE Access, vol. 7, pp. 169446–169468, 2019, doi: https://doi.org/10.1109/access.2019.2954861.

[5] M. Vinod Kumar and R. Sabitha, "College recommendation system for students using datamining with K-means and logistic regression," AIP Conference Proceedings, Jan. 2023, doi: https://doi.org/10.1063/5.0176989.

[6] J. Dhar and A. K. Jodder, "An Effective Recommendation System to Forecast the Best Educational Program Using Machine Learning Classification Algorithms," Ingénierie des systèmes d information, vol. 25, no. 5, pp. 559–568, Nov. 2020, doi: https://doi.org/10.18280/isi.250502.

[7] Maghdid, Halgurd S. "Web News Mining Using New Features: A Comparative Study." IEEE Access, vol. 7, 2019, pp. 5626–5641, https://doi.org/10.1109/access.2018.2890088. Accessed 21 May 2019.

[8] Vela, Belen, et al. "A Semi-Automatic Data–Scraping Method for the Public Transport Domain." IEEE Access, vol. 7, 2019, pp. 105627–105637, https://doi.org/10.1109/access.2019.2932197. Accessed 4 Sept. 2021.

[9] Ghorbani, Ramin, and Rouzbeh Ghousi. "Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques." IEEE Access, vol. 8, 2020, pp. 67899–67911, https://doi.org/10.1109/access.2020.2986809.

[10] Li, Xiang, et al. "A Multi-Dimensional Context-Aware Recommendation Approach Based on Improved Random Forest Algorithm." IEEE Access, vol. 6, 14 Aug. 2018, pp. 45071–45085, https://doi.org/10.1109/access.2018.2865436. Accessed 24 Apr. 2023.

[11] Lan, Hai, et al. "COVID-Scraper: An Open-Source Toolset for Automatically Scraping and Processing Global Multi-Scale Spatiotemporal COVID-19 Records." IEEE Access, vol. 9, 2021, pp. 84783–84798, https://doi.org/10.1109/access.2021.3085682. Accessed 28 July 2021.

[12] Chen, Wenbo, et al. "Tree-Based Contextual Learning for Online Job or Candidate Recommendation with Big Data Support in Professional Social Networks." IEEE Access, vol. 6, 2018, pp. 77725–77739, https://doi.org/10.1109/access.2018.2883953.

Aishwarya Shahaji Mane, and Rupali Dalvi. Secure Dynamic Multiowner System Using Share Generation Scheme. 1 Aug. 2018, https://doi.org/10.1109/iccubea.2018.8697527. Accessed 25 Feb. 2024.

 DOI : 10.35291/2454-9150.2024.0071