World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Research Article)

Empowering education and workforce planning: A comprehensive approach to university recommendation and profession demand forecasting

Menal Gannoruwa *, Savindu Thenura, Maheshi Subasinghe and Shenal Bandara

Sri Lanka Institute of Information Technology, Department of Computer Systems Engineering, Malabe, Western Province, Sri Lanka.

World Journal of Advanced Research and Reviews, 2023, 20(02), 366–376

Publication history: Received on 11 September 2023; revised on 24 October 2023; accepted on 27 October 2023

Article DOI: https://doi.org/10.30574/wjarr.2023.20.2.2147

Abstract

This research paper presents a comprehensive approach to empower education and workforce planning through university recommendation and profession demand forecasting. With the increasing student migration for higher education, understanding and addressing the challenges posed by brain-drain and the lack of suitable policies is crucial. To facilitate informed decision-making, a user-centric approach is employed to develop a Personalized University Classification System, providing tailored recommendations. The Prompt Commentary Delivery component utilizes NLP techniques to offer timely feedback and suggestions. Moreover, a hierarchical decision-making process enhances university and program recommendations based on relevant features and user preferences. The Profession Demand Forecasting System employs the ARIMA model and predictive analytics to identify future employment trends, contributing to informed strategies aligned with labor market demands. By addressing these aspects, the research aims to provide stakeholders with valuable insights and strategies to foster economic growth and empower students in their educational journeys.

Keywords: NLP; Profession Demand Forecasting; ARIMA; Personalized University Classification System; Higher education

1. Introduction

In recent years, the number of students migrating to acquire higher education has increased sharply, as evidenced by more than 5.3 million international students registered in 2017. Main destination countries being the United States, the United Kingdom and the United Kingdom, Australia, France, Germany, and the Russian Federation [1]. In addition, as shown in the figure below [Fig. 1], student migration from Sri Lanka has also increased, presenting both benefits, such as easing pressure on the national higher education system, and also challenges, such as brain drain[1].

However, a comprehensive study on student migration from Sri Lanka is currently lacking, which hinders the development of appropriate policies[1]. A student's choice of university and faculty is influenced by factors such as physical conditions, location, and socio-cultural convenience, with preferences likely to vary by degree. age and grade level. [2][3] Recognizing the importance of students' perspectives in decisions affecting their lives, it is important to consider a user-centered approach in developing a mentoring system, powered by insights from research, to effectively guide students towards appropriate higher goals and to educational programs abroad.

This research effort addresses the aforementioned challenges in higher education and workforce planning through various components. First, it focuses on developing a personalized university classification system that uses an efficient classification algorithm to accurately rank universities based on different characteristics, thereby placing background

* Corresponding author: G.K.M.M Gannoruwa

for appropriate recommendations. Second, the prompt commentary delivery component uses advanced natural language processing techniques to provide timely feedback and recommendations to users about universities, programs, and scholarships, using data from various sources. The goal is to understand user preferences and make recommendations tailored to their specific aspirations. Third, researchers seek to improve university and program recommendations by identifying relevant characteristics to classify users and implementing a decentralized decision-making process that considers multiple factors. and hobby. Finally, the project deals with the Occupational Demand Forecasting System, which uses machine learning and predictive analytics to identify future employment trends.

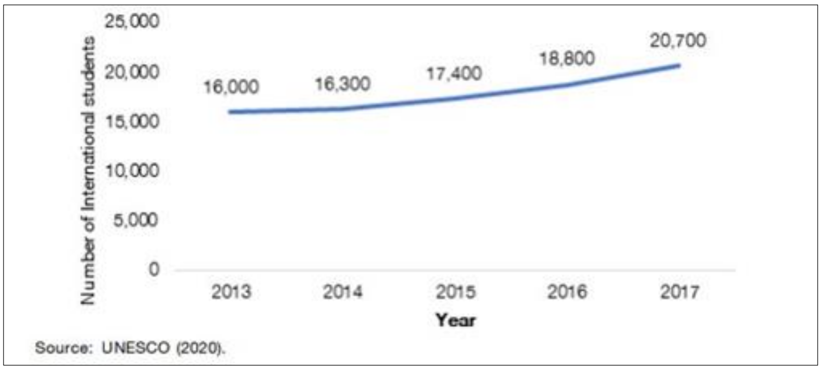

Figure 1 Number of Sri Lankan Students in Overseas Tertiary Education Programs

2. Literature review

Students and stakeholders participating in the university selection process face obstacles due to the lack of a defined method for categorizing colleges based on pertinent qualities. Without a standardized framework, it is challenging to compare and rank colleges based on crucial elements including program quality, faculty specialization, research possibilities, and infrastructure[4]. A 2021 research [5] emphasized the necessity for a generally recognized system for classifying institutions. Additional research on institution selection done in 2018 and 2019 [3], [6] emphasizes the need to consider unique student preferences and needs; ensuring personalized recommendations. Meanwhile, according to research on university rankings [7], accurate categorization is essential for students to have faith in the results and make wise choices; increasing the likelihood that they will find programs that align with their academic objectives and career aspirations.

Two research [8], [9] on student feedback analysis and summarizing methodologies, used web scraping to collect student input from online forums and institutional websites respectively. The second study [9] looks at extractive and abstractive summarizing approaches. The former [8] includes choosing crucial phrases from the original text for the summary while the latter [9] generates new sentences to convey the primary concepts. Another research [10] provides instances of both methodologies being used to describe student comments. It included extractive summarizing from online forums and abstractive summaries from social media platforms. These studies exhibit several techniques for analyzing and summarizing student input from diverse sources.

According to a 2017 study [11]; present websites that offer higher education program searches are limited since they serve just as information filters based on user preferences and do not take into account essential factors such as scholarships and price range. Similar research [12] highlights the importance of the exact classification of institutions, which increases students' confidence in making well-informed judgments. The significance of legitimate decision criteria and preferences in the decision-making process is stressed by Faisal Iddris [13]. Successful decision-making depends on ensuring that these factors are accurate and reliable, hence more study is required to determine whether they are valid in the hierarchical decision-making process.

Research published in 2022 [14] describes an approach for predicting upcoming talents in the ICT labor market by combining previous job ad patterns using AI classification algorithms. Its goal is to assist educators in adapting curriculum to new talents and creating skill-based online courses. However, the model's long-term forecasting power deteriorates, since the dataset was just three years long. Another 2014 study [15] focuses on projecting labor market demand in the regional professional element, using indicators for skill mix prediction. The study advises learning from nations such as the United States and emphasizes the need to anticipate employment structure by profession and qualification within certain disciplines. In addition, Li H, Wang Q, Liu J, and Zhao D [16] offer a prediction model for human resources recruitment demand utilizing a convolutional neural network (CNN) and a backpropagation neural network method in a separate research. It increases the accuracy of talent evaluation and provides effective labor supply-demand matching in the online labor market while delivering personalized services. The article continues by recommending future enhancements to the CNN model and integrating temporal aspects using recurrent neural networks.

3. Material and method

The research focuses on establishing a complete university recommendation system along with profession demand prediction capability. The system comprises of 4 sub-components. The university classification component uses machine learning methods such as clustering and classification models to categorize colleges based on particular attributes. The prompt commentary delivery component uses web scraping and NLP techniques such as sentiment analysis and text categorization to asses the user feedback on higher educational institutes. The university program recommendation component also employs web scraping and NLP techniques to collect feedback data; and provide recommendations on university programs. Finally, the profession demand forecasting component uses time series analysis in conjunction with the ARIMA model to forecast future job market trends based on information from job market reports, economic indicators, and technology breakthroughs. These study components work together to give students accurate and relevant suggestions, improve the efficiency of university programs, and forecast future career demand patterns.

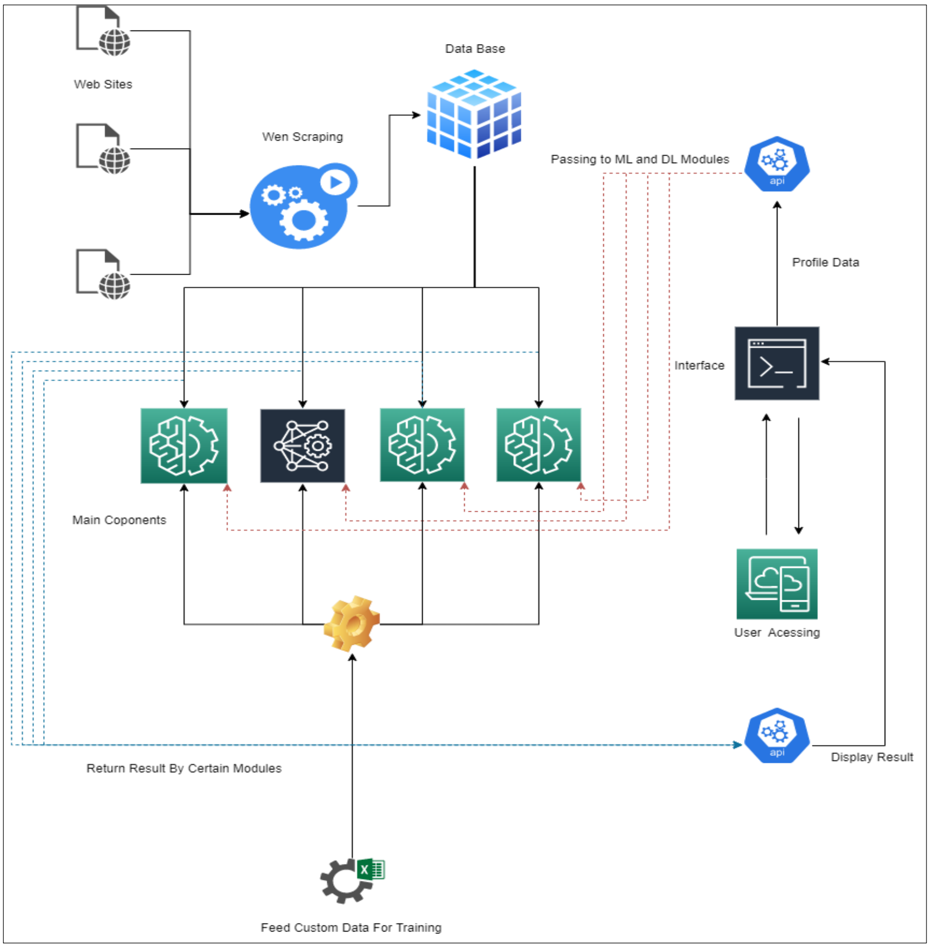

Figure 2 Overall system diagram

3.1. Personalized University Classification System

This system provides recommendations that are in line with personal preferences by utilizing technologies like machine learning and a hierarchical decision-making process. In order to create specialized recommendations that are in line with student interests, it classifies colleges according to particular characteristics. The system's emphasis on employing a variety of factors to appropriately classify institutions guarantees that recommendations are tailored to the user's needs. The ultimate goal is to give students a strong tool that simplifies the challenging university selection process and empowers them to make wise decisions for a fulfilling academic career.

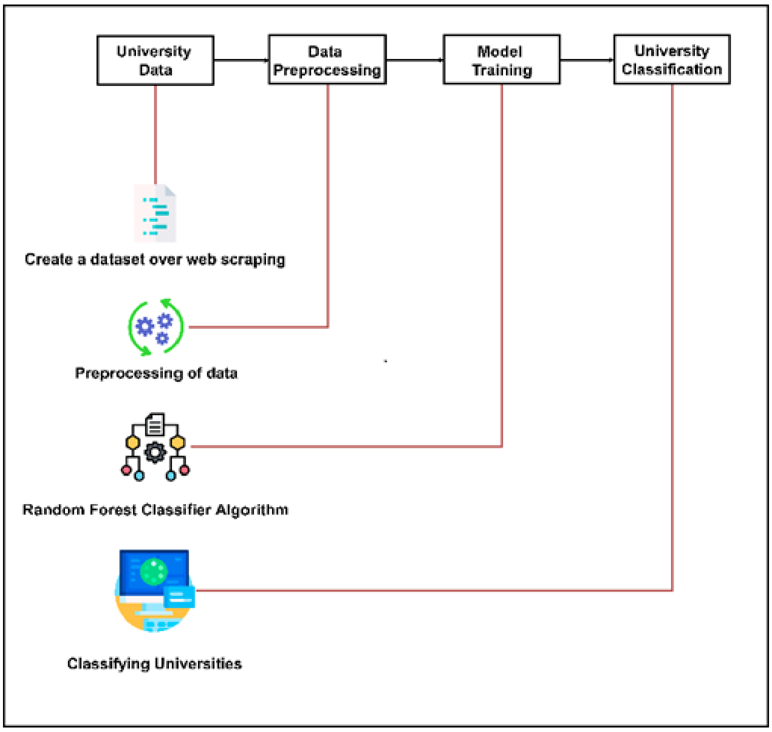

Figure 3 Personalized University Classification System – System Diagram

The system diagram shown in figure [Fig.3] depicts the high-level system view of the university recommendation component. Data about universities and their characteristics are gathered from a variety of sources, such as official university websites, educational databases, and student feedback forums. To guarantee consistency and quality, the acquired data is cleaned, preprocessed, and transformed into an analysis-ready format. An unsupervised machine learning method: K-Means Clustering was used to classify universities based on certain attributes. The acquired data was used to train the model, and its performance was evaluated using relevant evaluation measures. After assessing the results the model was finetuned to improve its performance.

3.2. Prompt Commentary Delivery

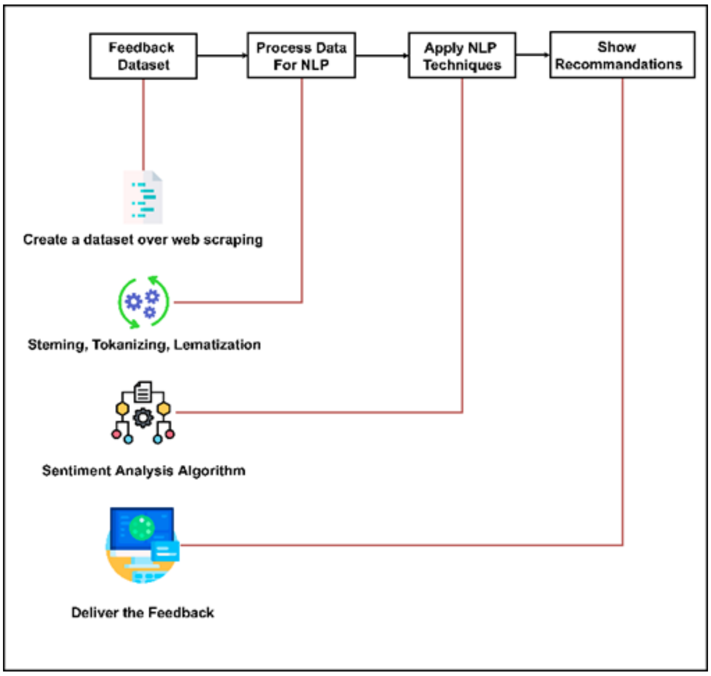

Figure 4 Prompt Commentary Delivery – System Diagram

The prompt commentary delivery component is critical in giving students timely discussion on the user feedback on higher education institutes. It analyzes unstructured data from many sources such as social media, forums, and blogs using Natural Language Processing (NLP) techniques. Furthermore, the component analyzes user queries to determine the actual student feedback on location, program offers, tuition prices, and campus life from different institutes. It analyzes this user feedback and provides an overall summary on the quality of the particular higher educational institute.

The system diagram shown in figure [Fig.4] depicts the high-level system view of the prompt commentary delivery component. The acquired data, which includes feedback, evaluations, and comments about the program or institute, is cleaned and preprocessed thoroughly to remove noise, inconsistencies, and irrelevant information. Following preprocessing and feature extraction, the data is exposed to different NLP approaches such as sentiment analysis, text classification, Named Entity Recognition (NER), text summarization, and topic modeling. Custom algorithms and pre-built libraries are used in these strategies to evaluate feedback data and obtain significant insights. Finally, the data are consolidated and input into a supervised learning classification model (Logistic Regression), which makes inferences based on the results collected. The fine-tuned model was used to provide actionable recommendations on the overall summary of the user feedback on higher educational institutes.

3.3. University Program Recommendation System

The university program recommendation component was implemented to improve the accuracy of university and program suggestions by determining the most pertinent characteristics for university classification; such as research opportunities and funding, personal preference, job market, university facilities, and cost. These essential characteristics support accurate program classification, enabling the pairing of users with appropriate colleges and programs. The procedure also employs a hierarchical decision-making strategy, which divides the decision-making process into more manageable tiers. A final suggestion is reached by combining the judgments made at each level of the hierarchy based on the information at hand. Due to the fact that it takes into account a variety of variables and user preferences, this hierarchical method enhances the accuracy of university program suggestions.

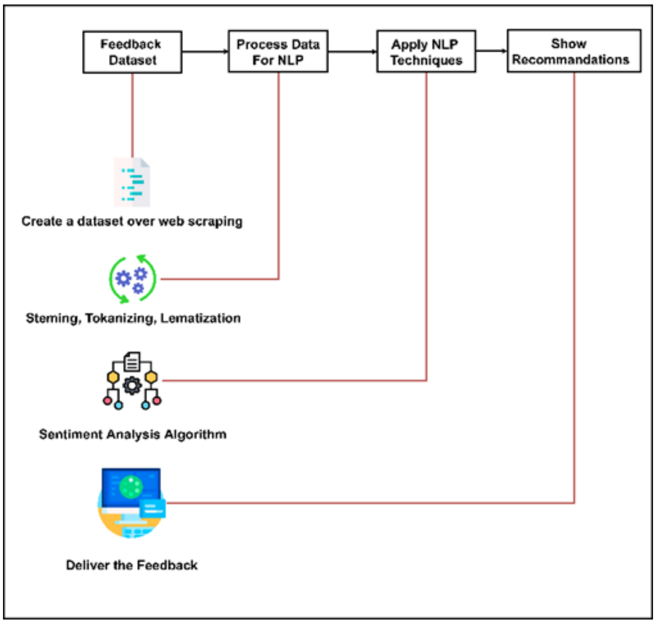

Figure 5 University recommendation based on preference– System Diagram

The system diagram shown in figure [Fig.4] depicts the high-level system view of the university program recommendation component. Web scraping techniques were used to extract data about universities and the programs offered by them; from multiple sources, including social media, forums, and websites. The data was cleaned and preprocessed in order to remove extraneous information, consistency issues, special characters, digits, and punctuation. Text tokenization was used to separate the text into individual words or phrases and remove stop words to improve the quality of the data. The data was then used to train a machine-learning model. A supervised machine learning algorithm (Decision Tree) was used to implement the model. The performance of the model was assessed using a variety of accuracy metrics, including precision, recall, F1 score, and total accuracy. In order to make well-informed judgments and projections, the model was fine-tuned based on the performance results.

3.4. Profession Demand Forecasting System

Profession demand forecasting, which provides information on future employment patterns, is essential to effective workforce planning and management. Finding the correct qualifications and competencies becomes increasingly important as the workplace changes and, places more emphasis on abilities like communication, teamwork, and adaptability. The development of automation and artificial intelligence has brought attention to the importance of up-skilling and re-skilling. However, there are still difficulties in anticipating external factors like regulations and economic changes, as well as potential biases in the data utilized for projections.

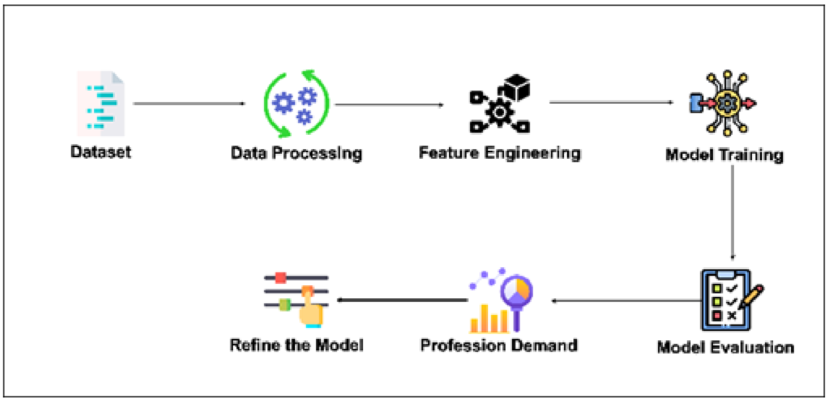

Figure 6 Profession Demand Forecasting System – System Diagram

The profession demand prediction model shown in figure [Fig. 6] was constructed using a procedure that included a number of phases to assure accuracy and dependability in projecting future employment market trends. Relevant information was gathered for the period of 2015 to 2023 from a variety of sources, including employment market statistics, economic indicators, and technology breakthroughs. To guarantee quality and consistency, the obtained data was then cleaned, preprocessed, and converted into a format appropriate for analysis. Due to its superior capacity to efficiently collect and evaluate patterns and trends across time, time series analysis methods, more especially the Auto Regressive Integrated Moving Average (ARIMA) model, were chosen to train the model. To increase the model's accuracy and dependability, it was trained, assessed, and improved. Finally, the fine-tuned model was used to forecast whether future demand for particular professions will be favorable, taking into account variables like the employment market, economic conditions, and technology advances.

4. Results and Discussions

4.1. Personalized University Classification System

The personalized university classification system employed multiple unsupervised machine learning algorithms, including K-Means Clustering, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), and Gaussian Mixture Model (GMM), to classify universities based on specific attributes. Among them, the K-Means Clustering algorithm stood out as the best model with the highest Silhouette score of 0.98. The silhouette scores acquired by K-Means algorithm depicted in the figure [Fig. 7]. In addition, K-Means indicated well-defined and well-separated clusters, making it the preferred choice for university classification.

The system's significance lies in its additional features, such as Auto-Generated Results, Scholarship Recommendations, and consideration of Environmental Factors, distinguishing it from previous studies on university classification. However, a potential limitation could be its reliance on specific data sources, necessitating future work to expand the dataset to encompass universities from diverse regions and incorporate additional factors like cultural preferences and extracurricular opportunities for enhanced recommendations. Overall, the system's effectiveness in providing accurate and relevant university recommendations based on individual preferences underscores its potential significance in assisting students with their decision-making process for successful educational journeys.

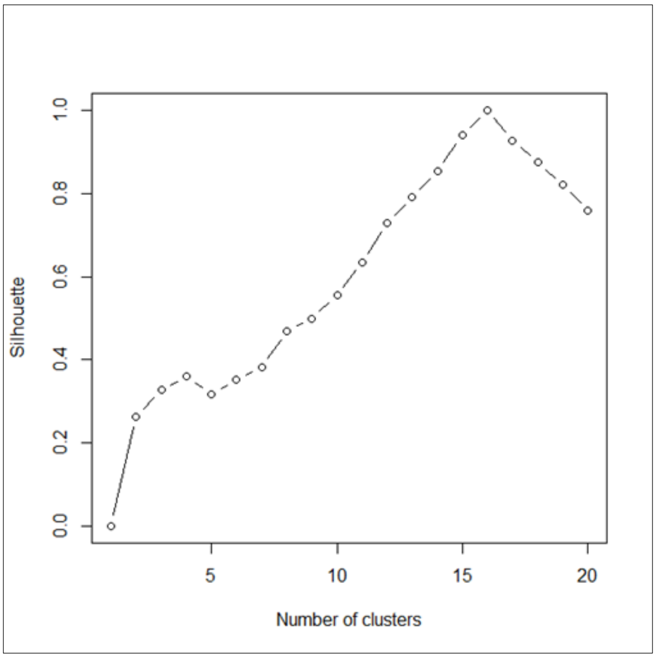

Figure 7 Silhouette scores acquired by K-Means algorithm

4.2. Prompt Commentary Delivery

The prompt commentary delivery component implemented multiple algorithms, including Decision Trees, Random Forest, Logistic Regression, and Support Vector Machine. Their performance was evaluated using precision, recall, F1 score, and overall accuracy metrics. The Logistic Regression model achieved a Train Accuracy of 0.98 and a Test Accuracy of 0.9625, surpassing the performance of the other algorithms. Comparison between test and train accuracies of the logistic regression model can be seen in the figure [Fig. 8]. The Logistic Regression algorithm emerged as the best-performing model with the highest overall accuracy, precision, and recall scores, enabling it to provide actionable recommendations for enhancing the recommended higher educational program or institute.

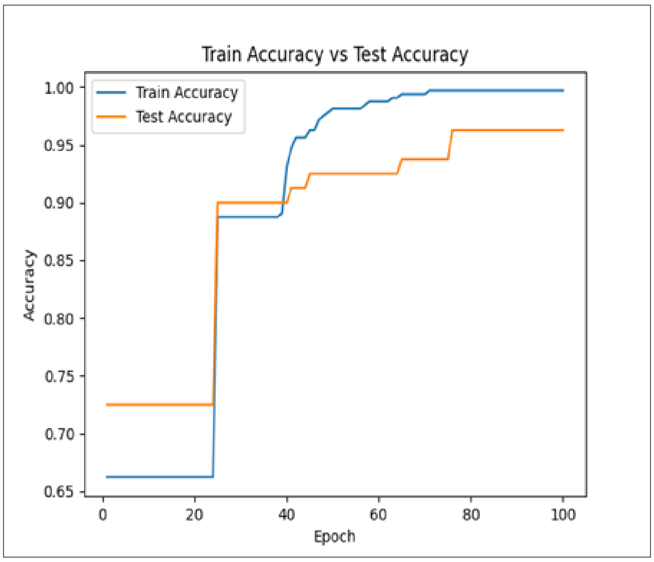

Figure 8 Test accuracy vs train accuracy of Linear Regression model

This research, in comparison to previous studies, introduces a sophisticated approach combining NLP techniques and supervised learning classification algorithms for personalized commentary delivery, adding significant value to the user experience in university selection and program recommendations. However, potential limitations include dataset quality and biases resulting from web scraping. Future work should focus on refining NLP techniques, incorporating user feedback for optimization, and expanding the dataset to improve the system's accuracy and relevance.

4.3. University Program Recommendation System

The research component employed multiple machine learning algorithms for university recommendations based on user preferences. The algorithms used included Decision Trees, Random Forest, Logistic Regression, and Support Vector Machine. The performance of each model was evaluated using various accuracy metrics, including precision, recall, F1 score, and overall accuracy. Among the implemented algorithms, the Decision Trees algorithm emerged as the best-performing model, achieving the highest overall accuracy, precision, and recall scores. The hierarchical decision-making process contributed to better classification accuracy by considering multiple factors and user preferences, leading to tailored and relevant university recommendations. Comparison between test and train accuracies of the decision tree model can be seen in the figure [Fig. 9].

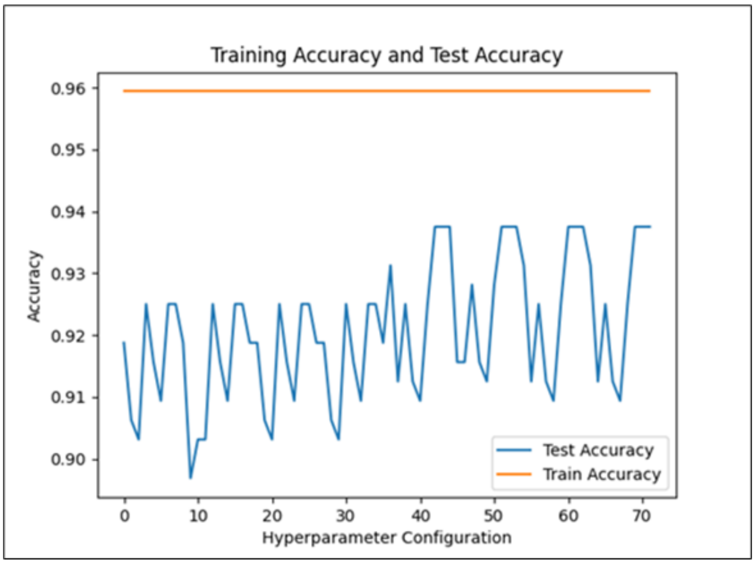

Figure 9 Test accuracy vs train accuracy of Decision Tree model

In comparison to previous studies; this research introduces an innovative approach that combines web scraping, NLP techniques, and supervised learning algorithms to create a robust university recommendation system based on user preferences. The system's integration of extractive and abstractive summarization methods improves the accuracy and relevance of recommendations, enhancing the user experience. However, potential limitations include data quality and biases from web scraping, impacting recommendation accuracy. The hierarchical decision-making process may introduce biases depending on decision order, warranting further investigation. The research's significance lies in providing personalized recommendations and addressing the lack of standardized university classification methods. It empowers students to make informed choices aligned with their academic and career goals. Future work should refine web scraping and NLP techniques, incorporate user feedback, expand the dataset, and add features to enhance accuracy and geographical coverage, creating a more comprehensive and inclusive recommendation system.

4.4. Profession Demand Forecasting System

The research component implemented the ARIMA model in order to predict the professional demand using time series analysis. The accuracy of the ARIMA model was measured using various performance metrics appropriate for time series forecasting which includes include Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The accuracy scores acquired by the model is depicted in the table [Tab. 4] and the reliability curve of the model can be seen in the figure [Fig. 10].

Table 1 Profession Demand Forecasting System

Metric	In-Sample(Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) One-Step-Ahead Forecast
N	172	24
RMSE	0.313295748	0.352257753
MAE	0.243952039	0.273585123
MAPE	1.431266547	1.568577996
MASE	0.881507368	0.988584899

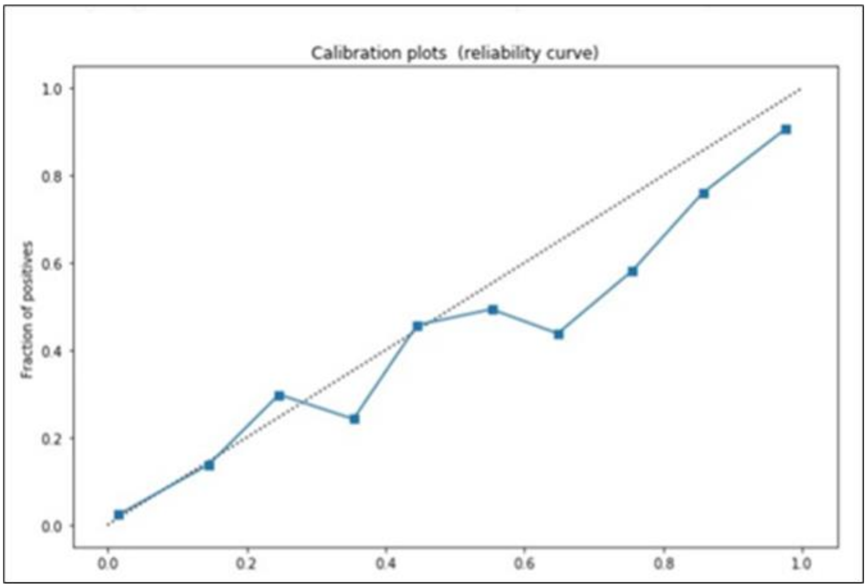

Figure 10 Reliability curve of the ARIMA model

In comparison to the existing studies this research component distinguish in its usage of expanding dataset for predicting emerging skills, in-depth analysis of temporal factor and incorporation of both local and global trends. While the existing systems forecast the profession demand using different approaches, such as AI classification techniques, neural network algorithms, and skill mix prediction. The ARIMA model serves as a time series-based forecasting method, offering different perspective and method for analyzing and predicting future profession demand. Comparing the ARIMA model's results with the findings from other studies would help understand its strengths and limitations in comparison to alternative approaches. However, potential challenges in profession demand forecasting, such as predicting external factors like government policies or economic disruptions and biases in the data used for predictions.

5. Conclusion

The research presents a comprehensive and innovative approach to empower education and workforce planning through personalized university recommendation and profession demand forecasting. The study addresses the challenges of brain-drain and the lack of suitable policies by providing stakeholders with valuable insights and strategies to foster economic growth and empower students in their educational journeys. The Personalized University Classification System accurately categorizes universities based on diverse features, simplifying the complex process of university selection. Through advanced machine learning algorithms like K-Means Clustering, the system effectively matches individual student preferences with suitable universities and programs. The Prompt Commentary Delivery component utilizes Natural Language Processing techniques to offer timely feedback and suggestions, enhancing the user experience by providing actionable recommendations to improve the recommended higher educational programs or institutes. The University Recommendation based on Preference employs a hierarchical decision-making approach that considers multiple factors and user preferences, leading to more accurate and relevant university and program recommendations. The Profession Demand Forecasting System utilizes the ARIMA model and predictive analytics to identify future employment trends, providing valuable insights to stakeholders and enabling them to align their strategies with labor market demands.

Overall, this research paper contributes significantly to education and workforce planning by addressing critical aspects of university recommendation and profession demand forecasting. By empowering students with personalized recommendations and enabling stakeholders to make informed decisions, the research aims to foster economic growth and bridge the gap between education and labor market demands. As the higher education landscape and job markets continue to evolve, this comprehensive approach presents a valuable tool for empowering students and promoting economic prosperity. Future research can further refine and expand upon these methodologies to provide even more accurate and effective recommendations and forecasts, ensuring a brighter and more successful educational journey for students and a well-prepared and adaptable workforce for the future.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Bilesha Weeraratne, Harini Weerasekera and Thilini Bandara, Understanding student migration from Sri Lanka, Policy Discussion Brief, 2022.

[2] S. Polat, The Factors that Students Consider in University and Department Selection: A Qualitative and Quantitative Study of Kocaeli University, Faculty of Education Students, Procedia - Social and Behavioral Sciences, vol. 47, 2012.

[3] Emine AKKAŞ, Mustafa Batuhan AYHAN, A comparative analysis for the criteria of higher education selection, Participatory Educational Research (PER) , vol. 5, no. 2, pp. 38-50, 2018.

[4] Tavakolifard, M., Sediqat-alhoda, S., & Mokhtarzadeh, M., A Recommendation System for University Selection Using the Analytic Hierarchy Process and Collaborative Filtering Techniques, Journal of Education and Learning, vol. 6, no. 3, pp. 176-186, 2017.

[5] Fang, J., & Jiao, J., A novel recommendation algorithm for university selection based on hybrid weighting of multiple attributes, PLOS ONE, 2021.

[6] Sohail, M. S., & Bukhari, S. S, University Selection Based on Analytic Hierarchy Process (AHP) and Weighted Sum Model (WSM), Journal of Business and Social Sciences Research, vol. 4, no. 1, pp. 73-82, 2019.

[7] Kebande, V. R., & Koech, R. K., A hierarchical decision-making process for university ranking using fuzzy AHP, International Journal of Intelligent Systems, vol. 35, no. 5, pp. 821-841, 2020.

[8] Islam, A. N., & Alsaeedi, F., Analyzing Student Feedback Using Text Mining Techniques: A Case Study of Higher Education Institutions, 2018.

[9] Qian, Y., & Scott, J. , Extracting Insight from Online Reviews for Higher Education Institutions, 2018.

[10] Yang, Y., & Wu, L., Application of Sentiment Analysis for Evaluating Online Student Feedback in Higher Education, 2017.

[11] Marietta Papadatou-Pastou, Rhianna Goozee, Erika Payne, Alexia Barrable & Patapia Tzotzoli , A review of web-based support systems for students in higher education, International Journal of Mental Health Systems, 2017.

[12] Rahul J. Jadhav, Dr. Om Prakash Gupta, Usharani T. Pawar, Significant Role of Search Engine in Higher Education, International Journal of Scientific & Engineering Research, vol. 2, no. 4, 2011.

[13] F. Iddris, Search Engine Optimisation (SEO) As Digital Marketing Strategy for Internationalisation of Higher Education, in The 22nd McGill International Entrepreneurship Conference, Halmstad, Sweden, 2018.

[14] Ramtin Yazdanian, Richard Lee Davis, Xiangcen Guo, Fiona Lim, Pierre Dillenbourg, Min-Yen Kan, On the radar: Predicting near-future surges in skills' hiring demand to provide early warning to educators, Computers and Education: Artificial Intelligence, vol. 3, 2022.

[15] Daniya Shigapova , Munir Valiullin, Oksana Yrieva, The Methods of Prediction of Demand on The Labor Market, in 2nd GLOBAL CONFERENCE on BUSINESS, ECONOMICS, MANAGEMENT and TOURISM, Prague, Czech Republic, 2014.

[16] Li H, Wang Q, Liu J, Zhao D, A Prediction Model of Human Resources Recruitment Demand Based on Convolutional Collaborative BP Neural Network, Computational intelligence and neuroscience, 2022.

[17] S. J., Sinhala language processing: Challenges and opportunities, Journal of Language Technology, vol. 45, no. 2, pp. 123-145., 2019.