Article

Not peer-reviewed version

Recommender System for University Degree Selection: A Socioeconomic and Standardized Test Data Approach

Enrique DelaHoz-Dominguez and Raquel Hijón-Neira *

Posted Date: 13 August 2024

doi: 10.20944/preprints202408.0905.v1

Keywords: recommendation system; learning analytics; machine learning

Article

Recommender System for University Degree Selection: A Socioeconomic and Standardized Test Data Approach

Enrique J. Delahoz-Domínguez [1] and Raquel Hijón-Neira [2,*]

[1] Universidad del Magdalena, Santa Marta, Colombia.; enriquedelahoz@unimagdalena.edu.co
[2] Computer Science Department, Rey Juan Carlos University, 28933 Móstoles, Spain; raquel.hijon@urjc.es
* Correspondence: Raquel.hijon@urjc.es

Abstract: Recommender systems in education are becoming more widespread, typically focusing on recommending courses or study materials. This study proposes a machine learning approach to recommend a university degree based on high school and university standardised test results, incorporating students' socioeconomic information as input variables. The objective is to develop a tool for students' decision-making, supporting the sustainable development goal of Quality Education by providing a data schema to maximise the likelihood of a successful match between the student's profile and the academic program. With its focus on equity in education, this study provides a data-driven approach to assist students in selecting suitable university degrees, aiming to improve educational outcomes and inspire a more equitable education system.

Keywords: recommendation system; learning analytics; machine learning

1. Introduction

Quality education is vital to the sustainable development goals set by the United Nations for 2030. Specifically, the goal is to eliminate gender disparities in education and ensure equal access to all levels of education. To achieve this, objective tools are needed to aid students in making strategic and operational decisions regarding their education. Our study is committed to this goal, and we believe our machine-learning approach can significantly contribute to its achievement.

Choosing what to study is a pivotal decision in a person's life. Paradoxically, few objective frameworks are available to support this critical decision, and often, career choices are made based on subjective elements. Factors influencing career choices in college students include monetary and nonmonetary costs and benefits, cultural and social capital measures, and major differences between STEM and non-STEM graduates [1]. Various studies have shown that poor career choices can lead to significant issues, such as student dropout, poor academic performance, financial problems, and personal dissatisfaction [2,3].

Moreover, career indecision in college students is predicted by greater identity moratorium and diffusion, less maternal acceptance, and fewer years in college [4]. Consequently, a student satisfied with their professional choice is likely to perform better academically, thereby improving the efficiency of the study group. Data generated by former students regarding their career choices and academic performance can be used to create a decision-making framework that incorporates intrinsic factors of the learning process, such as standardised test results, and extrinsic factors, such as the type of school and parent's education and occupation.

In this context, our study proposes a machine-learning model to recommend university degrees for new students. By applying data science techniques to standardised high school and college test results and incorporating socioeconomic variables, the model predicts students' academic performance on standardised tests at the end of college. This allows for recommendations on the career paths that will likely result in better performance. This work has practical applications for any student deciding which career to pursue and offers universities an advancement in student

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 August 2024 doi:10.20944/preprints202408.0905.v1

counselling by using records from former students to generate helpful information for future professionals.

2. Related Work

In recent years, recommendation systems in the educational field have gained significant popularity, encompassing various stages of the student's relationship with academic institutions. From a high-level perspective, there are systems designed for university enrollment recommendations. These educational recommender systems aim to enhance the relevance of proposed study programs to students’ interests. Specifically, their objectives include mapping students' interests to a broad range of available courses to ensure they find courses that match their learning needs [5]. With the recent developments in the availability of online content, there is a growing need for systems that perform data analysis to find the best match for students based on their unique interests and capabilities [6]. Researchers have conducted numerous studies to develop systems that fill this gap using intelligent information processing techniques, such as recommender systems.

Throughout the literature, various approaches leverage recommender system techniques, including collaborative filtering [7], content-based filtering, and hybrid recommender systems based on fuzzy systems [8] and complex machine learning models [9]. Each approach collects and analyses students' historical behaviours, interests, or performance measurements to provide personalised recommendations. For example, collaborative filtering assumes that measuring the degree of similarity between users’ experiences can help identify items of interest. Suppose two students tend to enrol in similar courses. In that case, it is likely that if student A receives a good grade in a particular course, student B will also perform well, as they have demonstrated similar academic behaviours in past records [10].

Conversely, content-based approaches incorporate textual data from users' previous histories. Unlike collaborative filtering, content-based systems do not suffer from issues related to data matrix sparsity. However, these approaches require mechanisms to transform text features such as tags, keywords, and topics into vectors [11]. For instance, a content-based recommender system effectively assists prospective students in Bangladesh choose top-K private universities based on several parameters, achieving high accuracy [12]. After creating vector representations, content-based approaches use cosine similarity to determine the degree of similarity between users.

Moreover, hybrid frameworks combine complex models to generate personalised recommendations. Machine learning applications incorporating deep neural networks have produced robust and accurate recommendations. Natural language processing techniques, such as latent Dirichlet allocation (LDA) and latent semantic analysis (LSA), are also used to process extensive textual profiles to understand the semantics of the information used in the recommendation process [13]. Despite their advantages, textual features often contain inherent biases and uncertainties. Researchers have widely adopted fuzzy logic systems to handle these issues, which model vagueness logically and provide heuristic rules [14].

In educational recommender systems, fuzzy logic consistently mitigates vagueness by using linguistic variables and can be combined with other approaches, such as clustering, to find concrete user similarities. For instance, fuzzy rules can better interpret user ratings on specific items, offering greater flexibility in handling demographic data [15].

However, collaborative filtering does not perform well in specific domains, particularly in educational contexts where students have diverse capabilities, learning curves, and core strengths. The sparsity of data matrices indicating users' collaboration degrees further degrades the performance of collaborative filtering approaches in educational recommender systems [16]. To address this, researchers have developed hybrid recommendation frameworks that combine multiple methods to enhance robustness and accuracy. For example, a fuzzy logic-based intelligent recommender system measures students' skills using fuzzy rules and a Mamdani fuzzy inference system to identify their interests [17].

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 August 2024 doi:10.20944/preprints202408.0905.v1

A context-aware recommender system based on fuzzy logic was proposed to generate personalised recommendations by assessing each user's performance [18]. This model calculates the required contexts for applying fuzzy logic based on the learners' assessment scores and the time to complete the assessment. Additionally, novel fuzzy-based recommender systems leverage family tree-based fuzzy inference systems to capture critical concepts in user profiles and available courses, improving neighbours' semantic and content similarities for personalised recommendations [19].

Ontology-based approaches have also been explored to develop recommendation engines. These approaches transform human-generated information into machine-readable language, which is then used in machine learning-based solutions. For instance, ontology-based recommender systems have been proposed to better guide students in higher education systems by assessing their strengths, weaknesses, interests, and abilities. These systems collect data explicitly from user profiles and implicitly from surveys conducted with graduate students. Machine learning methods create and cluster profile models of graduate students, which are then used in hybrid recommendation engines [20].

Furthermore, deep learning-based recommendation frameworks have been developed for e-learning platforms. These models use K-Nearest Neighbor (KNN) and collaborative filtering approaches to find similar students with comparable learning capabilities. Other frameworks incorporate sentimental and cognitive analysis and learning style categorisation to generate personalised course recommendations [21].

Overall, intelligent [22] personalised course advising models and hybrid recommendation systems combining collaborative and content-based filtering, optimised with genetic algorithms, have significantly improved recommendation reliability and performance. These advancements in recommendation system technologies offer practical applications for students and educational institutions, enhancing the guidance provided to students in their course registration processes (Tilahun & Sekeroglu, 2020).

3. Theoretical Framework

Academic processes generate much contextual data, such as academic performance and demographic and financial information. Thus, universities and colleges have tutors who carry out accompaniment processes to advise students on the most relevant careers according to their academic profile. However, a tutor can't manage all the information of the thousands of previous students about whom there is information about their chosen career and academic performance; in addition, tutors can generate a bias based on their relationship with the students or their own experiences and beliefs.

For the present study, we articulate a dataset where the academic and sociodemographic information of the students in the final stage of secondary school and university is related. The dataset comprises the results of standardised tests administered to students at the end of their high school and college education. Therefore, a supervised learning model was employed to train various algorithms to predict university exam results based on data from the high school stage. In addition, the key variables that determine the results in the standardised examination of the university stage were determined. To facilitate the application of the model, the results are provided as recommendations that students, tutors, or universities can utilise. This information is intended to aid in making one of the most critical decisions in a person's life: selecting a university career.

The vector $\{X_1, X_2, \ldots X_n\}$ gives the data structure, representing the input variables for the model, and "y" means the student's average result on the university standardised test. The output "y" is a continuous vector from 0 to 300. Hence, the proposed recommendation system can be structured as a forecasting model, wherein the implicit relationship between the model's input and response variables is articulated in Equation 1.

$$y=f(x_1, x_2,\ldots,x_n)+ \text{epsilon}, \tag{1}$$

Where "f" is an algorithm associated with a supervised machine learning model used to make predictions based on the training dataset, and epsilon represents the error arising from the discrepancy between the actual value and the value predicted by the supervised model. Thus, the

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 August 2024 doi:10.20944/preprints202408.0905.v1

recommendation of studies is associated with the degree that presents a higher value of Ypred for a new student "Xnew" due to the function Ypred = f(Xnew).

In our study, a novel approach is given to structure a recommendation system for students, using longitudinal information of the student from his high school stage until the completion of the university. Consequently, the paper answers the following problem questions:

- How to generate recommendations of careers to study at the university to new students based on academic and socio-demographic data
- What features have the most significant impact on the recommendation of which career to study?
- Which machine learning model offers the best prediction performance

4. Materials and Methods

It is difficult to anticipate which supervised model will be the one that will achieve a better fit to the data a priori; this condition is known as the No Free Lunch theorem. Since the algorithms with greater theoretical prediction capacity have less or no capacity to explain the relationships between the input and output variables. Thus, considering the difference in the performance of the different algorithms, five algorithms will be used: Random Forest, XGBoost, GLMNET, Linear Regression and Decision Trees.

4.1. Proposed Modelling Structure

This study proposes the creation of five machine learning models to generate suggestions on which professional path a student should pursue. The method is determined by dividing the first data set into 80% training and 20% testing. The training technique was designed using a 10-fold cross-validation scheme, in which the data set is split randomly into ten equal halves. Nine are utilised as the training set, with the remaining one serving as the testing set. The mean value of the predictive outcomes from the ten rounds is used to evaluate the algorithm.

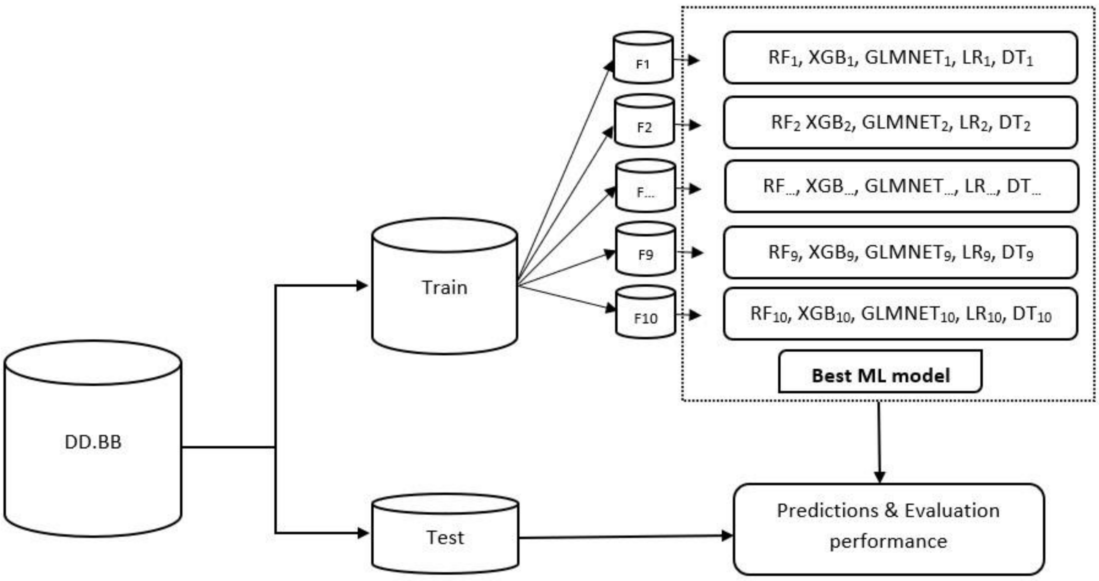

Figure 1. Machine learning development.

The model's input variables represent student information. They are classified into three categories: information on standardised test results, socioeconomic information in high school, and information on standardised test results in college. Table 1 shows the structure, category, and description of the variables.

Table 1. Study variables.

Category	Variable (code)	Type	Levels/Scale
Student Background	Gender (gen)	C	Male, female
	Department of Residence (dep.res)	C	Students' department of residence
	School type	C	Public, private
	School Calendar (sch)	C	Calendar_A, Calendar_B
	Father education (fedu) Mother education (medu)	C	complete professional education, Incomplete professional education, None, does not know, Postgraduate, complete secondary school, incomplete secondary school, complete Technical degree, incomplete technical degree.
	Father's occupation (focu) Mother's occupation (mocu)	C	unemployed, general manager, auxiliary level employee, Domestic employee, businessman, Stay-at-home, day labourer, private company employee, government employee, Other activity or occupation, Little Businessman, Independent professional, Unpaid family worker, Self-employed, Worker without remuneration.
Standardised test at Highschool	Critical Reading (CR)	N	Score in the test (0-100)
	Math (Math)	N	Score in the test (0-100)
	Citizenship Skills (CS)	N	Score in the test (0-100)
	Science (sci)	N	Score in the test (0-100)
	English (ENG)	N	Score in the test (0-100)
Standardised test at Highschool	Spro.result (SPRO)	N	Score in the test (0-300)

In operationalising the model, a new student enters their academic results and social variables into the model and thus receives the recommendation of the careers for which a better performance is projected in the Saber Pro university exam results. Figure 2 represents the operational flow in which the algorithm recommends degree options.

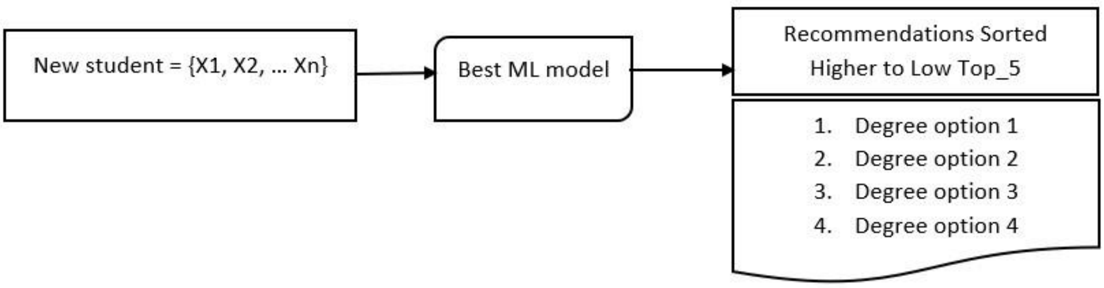

Figure 2. Predictions structure.

4.2. Data

Data were collected from the Colombian Education Institute (ICFES) [22,23] and included the records of undergraduate students who took the Saber11 (high school) and SaberPro (University)

 doi:10.20944/preprints202408.0905.v1

exams from 2006 to 2019. These data are made publicly available by ICFES for research and analysis. The comprehensive database contains 921,041 results from high school and university standardised tests. The dataset encompasses 102 focus options covering various categories such as engineering, literature, science, and art. Additionally, it includes socioeconomic information such as parent's education and occupation, type of school, academic calendar, gender, and department of residence. Consequently, the recommender system will be represented by fourteen variables, as outlined in Table 1. Eight variables are categorical and derived from demographic information and the student's background. This information was sourced from the ICFES repository, with the standardised test result designated as the target variable at the university level.

4.2. Descriptive Data Analysis

Table 2 shows the statistical summary for the 921,041 records adjusted by gender in the database for the fourteen years of the study. As shown in Table 2, the gender distribution is 58% female and 42% male. Besides, the median and mean values for all modules evaluated are higher for men than women.

Table 2. RSME Results.

Variable	gender	n	Min	q1	Median	Mean	q3	Max	sd	IQR
CR	F	548046	0	47.1	53.0	53.5	59.3	100	9.5	12.1
	M	372995	0	47.3	53.2	54.1	60.0	100	9.6	12.7
	all	921041	0	47.2	53.0	53.7	59.4	100	9.5	12.2
MATH	F	548046	0	45.0	51.0	51.8	58.1	100	10.8	13.2
	M	372995	0	47.7	55.0	56.0	63.0	100	12.1	15.4
	all	921041	0	45.4	53.0	53.5	60.0	100	11.5	14.6
SCI	F	548046	0	45.6	51.0	52.0	57.8	100	9.6	12.2
	M	372995	0	47.5	53.2	54.5	61.0	100	10.7	13.5
	all	921041	0	46.3	51.9	53.0	58.6	100	10.1	12.3
CS	F	548046	0	45.9	52.0	52.4	58.2	100	9.7	12.3
	M	372995	0	47.7	54.0	54.1	60.3	100	10.3	12.6
	all	921041	0	46.1	53.0	53.1	59.8	100	10.0	13.6
ENG	F	548046	0	43.0	49.0	52.4	58.4	100	13.6	15.4
	M	372995	0	43.5	50.9	54.5	61.7	100	14.6	18.2
	all	921041	0	43.5	50.0	53.3	59.7	100	14.0	16.2
SPRO	F	548046	0	10.6	131.2	99.6	155.4	265.0	69.3	144.8
	M	372995	0	10.9	135.4	103.3	162.0	268.8	72.0	151.1
	all	921041	0	10.7	132.6	101.1	158.0	268.8	70.4	147.3

4.3. Performance Metrics

Metrics and comparison structures are needed to identify improvement patterns in the regression model that presents the best performance in different situations. These metrics are called performance indicators; they will give an idea of the general performance of the prediction models of the standardised exam results—a well-fitted regression model results in predicted values close to the observed data values. The base model, which uses the mean for each predicted value, would generally be used without informative predictor variables. Therefore, the fit of a proposed regression model should be better than the fit of the base model to be considered as a model capable of generating information beyond the average value.

 doi:10.20944/preprints202408.0905.v1

4.3.1. Root Mean Square Value

The RMSE (Root Mean Square Error) is the square root of the residuals' variance, indicating the model's absolute fit to the data by measuring how close the observed data points are to the model's predicted values. As an absolute measure of fit, RMSE can be interpreted similarly to the standard deviation of the unexplained variance, sharing the same units as the response variable. Lower RMSE values denote a better fit. RMSE is an effective measure of the model's predictive accuracy and is the most critical criterion for assessing fit when prediction is the model's primary objective. Generally, to evaluate the quality of the fit, the model's RMSE value should be significantly lower than the standard deviation of the response variable. The RMSE is calculated by Equation (2).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{2}$$

5. Results

In this study, a recommender system based on machine learning algorithms was built using the R programming language within the Caret package. Caret is a sophisticated library that develops supervised learning modelling with tools for preprocessing and cross-validation of data. Random Forest, XGBoost, GLMNET and KNN supported the recommender system. To train the proposed model, ten folds were created to experiment with the predictions and evaluate each algorithm's performance; the training stage results are listed in Table 3. The average value of the ten folds was calculated as the algorithm's overall prediction performance. As shown in Table 3, the mean RMSE values for XGBoost, RF, GLMNET and KNN are 30.1, 45.0, 41.1 and 51.1, respectively.

Table 3. Cross-validation results.

Fold	XGBoost	RF	GLMNET	KNN
1	29,3	45,2	40,6	47,9
2	31,1	43,7	41,5	50,6
3	29,9	44,0	40,9	53,6
4	30,0	45,5	40,1	49,7
5	30,7	47,2	41,1	48,1
6	31,4	44,1	41,3	49,9
7	32,0	42,8	40,7	48,8
8	27,5	44,6	41,1	51,7
9	29,5	44,9	41,4	49,3
10	30,1	47,8	42,7	61,5
mean RMSE	**30,1**	45,0	41,1	51,1

Therefore, considering that the standard deviation of the output variable is 70.4 and comparing the RMSE results shown in Table 3, the model with the best performance in the training stage is the XGBOOST with an RMSE value of 30.1, followed by the GLMNET algorithm with a RMSE value of 41.1. Consequently, it is essential to analyse the model's consistency; again, the XGBoost is the algorithm with the lowest standard deviation value, followed by the GLMNET.

The predictions of the standardised tests at the university in this study are used to recommend a degree to study; therefore, a high capacity for prediction generates confidence that the recommendations created serve as objective support for decision-making.

 doi:10.20944/preprints202408.0905.v1

5.2. Model Testing

The materials and methods section explained that 30% of the data was reserved for the supervised learning models' validation phase. Table 2 shows the RMSE results for data not previously known by the algorithms. In comparing the models, XGBoost has the best performance, with an RMSE value of 30.1.

The RMSE values for the five algorithms implemented in the recommender system are all lower than the standard deviation of the testing dataset used for the respective models. These results demonstrate that these four models perform well in predicting standardised test results. However, the XGBoost algorithm surpasses the other models (XGBoost vs. GLMNET, RF, KNN) in terms of performance.

6. Discussion

Developing and evaluating a recommender system for predicting standardised test results and recommending suitable undergraduate degrees based on academic and sociodemographic data present significant advancements in educational technology. Our study leverages a combination of machine learning algorithms to enhance students' decision-making processes. This discussion compares our findings with those of previous studies and highlights the unique contributions of our research.

To begin with, Bhumichitr implemented collaborative filtering and ALS for elective course recommendations, achieving an accuracy of 86%. While their approach focused on course selection, our system extends to degree recommendations using a broader set of variables, including sociodemographic data, enhancing the recommendations' contextual relevance [24].

Similarly, Khanam and Alkhaldi used a random forest regression model to classify universities based on student preferences and voting data. Although their system focused on graduate admissions, it parallels our use of random forest in predicting undergraduate degree suitability, highlighting the versatility and effectiveness of this algorithm [25].

Moreover, Esteban developed a hybrid recommendation system combining collaborative and content-based filtering with genetic optimisation to recommend courses. Their success in integrating multiple criteria underscores the importance of using diverse data points, a strategy we also employed to capture a comprehensive view of student profiles [26].

In addition, Samin and Azim applied probabilistic topic models for course and supervisor recommendations, focusing on the faculty's past publications and research interests. Our system similarly uses historical academic data but extends it to include sociodemographic factors, thus broadening the scope and potential impact of the recommendations [27].

Furthermore, Slim proposed a framework for recommending academic programs, courses, and instructors using collaborative filtering and non-negative matrix factorisation. While their system focused on optimising course schedules, our approach aimed at predicting standardised test results and recommending degrees, thereby supporting long-term academic planning [28].

Likewise, Ognjanovic used institutional data and the Analytical Hierarchy Process (AHP) to predict course selections. Similarly, our study utilises institutional data but expands its use to recommend undergraduate degrees, thus providing more holistic support for academic decisions [29].

Additionally, George and Lal employed a personalised approach to course recommendation using ontology and sequential pattern mining. Our system also emphasises personalisation by incorporating sociodemographic factors, ensuring that recommendations align closely with individual student profiles [30].

Moreover, Zayed investigated supervised machine learning techniques for recommending undergraduate programs using academic history and job market data. Our study similarly uses machine learning but focuses on standardised test predictions and degree recommendations, providing a comprehensive tool for student support [5].

In summary, our study stands out by integrating both academic and sociodemographic data to provide holistic recommendations. We achieved a robust prediction model by employing advanced machine learning algorithms such as XGBoost, Random Forest, GLMNET, and KNN. Including variables such as parental education, school type, and regional data enhances the accuracy and relevance of our recommendations, offering a significant improvement over systems that rely solely on academic performance data.

7. Limitations and Future Research

Firstly, while comprehensive, the dataset used in this study may not fully capture the diversity of student populations across different regions, genders and educational systems. The sociodemographic variables included, such as parental education and school type, are limited to available data and may not encompass all factors influencing students' academic performance and degree choices. The dataset utilized for this research only includes two gender options: Male and Female. This binary classification does not encompass the full spectrum of gender identities, which could lead to an incomplete understanding of gender-related dynamics in the context of our study. Future studies should aim to include a broader and more diverse dataset to enhance the generalizability of the findings.

Secondly, although robust, the prediction models employed in this study depend on the quality and accuracy of the input data. Any inconsistencies or inaccuracies in the standardised test results or sociodemographic information can affect the recommender system's performance. Thus, ensuring high-quality data collection and preprocessing is crucial for the system's reliability.

Thirdly, while the machine learning algorithms (XGBoost, Random Forest, GLMNET, and KNN) have shown solid predictive capabilities, they also require significant computational resources and expertise to implement and maintain. This may pose a challenge for some educational institutions with limited technical resources. Future research should explore more efficient algorithms and methods to reduce the computational burden without compromising accuracy.

Additionally, the temporal scope of the data, which spans from 2006 to 2019, presents another limitation. To articulate the dataset, it is necessary to trace the students' academic performance from their high school years through to their university education. This process requires an average of at least five years to obtain a high school student's university standardised exam results. Consequently, there is an inherent delay in the availability of comprehensive data, which may impact the timeliness and relevance of the study’s findings. Future studies should consider strategies to mitigate this delay, such as employing longitudinal data collection methods, to ensure more timely and applicable insights.

8. Conclusions

This study presents a significant advancement in educational technology by developing and evaluating a recommender system designed to predict standardised test results and recommend suitable undergraduate degrees based on a combination of academic and sociodemographic data. Our approach leverages advanced machine learning algorithms, specifically XGBoost, Random Forest, GLMNET, and KNN, to enhance students' decision-making.

Firstly, integrating academic and sociodemographic variables, such as parental education, school type, and regional data, proved crucial in improving the accuracy and contextual relevance of the recommendations. This holistic approach addresses the limitations of previous systems that relied solely on academic performance data, offering a more comprehensive tool for student support.

Our findings demonstrate that machine learning models, particularly XGBoost and Random Forest, exhibit robust performance in predicting standardised test outcomes and recommending undergraduate degrees. These models significantly outperform simpler algorithms, highlighting the importance of employing sophisticated techniques to achieve high accuracy in educational recommendations.

Furthermore, the comparative analysis with prior studies underscores the unique contributions of our research. While previous works focused on course recommendations or graduate admissions, our system extends its applicability to undergraduate degree recommendations, providing long-term academic planning support. Incorporating diverse data points and utilising hybrid recommendation frameworks enhance the system's effectiveness and reliability.

In conclusion, the proposed recommender system supports students in making informed academic decisions and aligns with the broader educational objectives of reducing gender disparities and ensuring equal access to education. Future research can explore the integration of real-time data and the continuous improvement of recommendation algorithms to refine the system's predictive capabilities further. Expanding the system to include personalised learning paths and career counselling features could offer more excellent value to students and educational institutions.

Overall, this study contributes to the ongoing efforts to harness the power of data science and machine learning in education, aiming to improve educational outcomes and promote equity in access to quality education.

Author Contributions: Conceptualization, E.J.D.D. and R.H.N.; methodology, E.J.D.D and R.H.N.; software, E.J.D.D.; validation, E.J.D.D. and R.H.N; formal analysis, E.J.D.D; data curation, E.J.D.D.; writing—original draft preparation, E.J.D.D. and R.H.N; writing—review and editing, E.J.D.D. and R.H.N; visualization, E.J.D.D.; supervision, R.H.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by research grants PID2022-137849OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by the ERDF, EU.

Data Availability Statement: The final compilation of the data utilized in this study is available in the Mendeley Data repository with the DOI: 10.17632/fsvt9dr3ww.1. Researchers and interested parties can access the dataset for further analysis and replication studies.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Xu, Y.J. Career Outcomes of STEM and Non-STEM College Graduates: Persistence in Majored-Field and Influential Factors in Career Choices. *Res. High. Educ.* **2013**, *54*, 349–382. https://doi.org/10.1007/s11162-012-9275-2.
2. Akkermans, J.; Paradniké, K.; Van der Heijden, B.I.J.M.; De Vos, A. The Best of Both Worlds: The Role of Career Adaptability and Career Competencies in Students' Well-Being and Performance. *Front. Psychol.* **2018**, *9*. https://doi.org/10.3389/fpsyg.2018.01678.
3. Lent, R.W.; Brown, S.D.; Talleyrand, R.; McPartland, E.B.; Davis, T.; Chopra, S.B.; Alexander, M.S.; Suthakaran, V.; Chai, C.-M. Career Choice Barriers, Supports, and Coping Strategies: College Students' Experiences. *J. Vocat. Behav.* **2002**, *60*, 61–72. https://doi.org/10.1006/jvbe.2001.1814.
4. Guerra, A.L.; Braungart-Rieker, J.M. Predicting Career Indecision in College Students: The Roles of Identity Formation and Parental Relationship Factors. *Career Dev. Q.* **1999**, *47*, 255–266. https://doi.org/10.1002/j.2161-0045.1999.tb00735.x.
5. Zayed, Y.; Salman, Y.; Hasasneh, A. A Recommendation System for Selecting the Appropriate Undergraduate Program at Higher Education Institutions Using Graduate Student Data. *Appl. Sci.* **2022**, *12*, 12525. https://doi.org/10.3390/app122412525.
6. Yan, L.; Liu, Y. An Ensemble Prediction Model for Potential Student Recommendation Using Machine Learning. *Symmetry* **2020**, *12*, 728. https://doi.org/10.3390/sym12050728.
7. Han, J.; Jo, J.; Ji, H.; Lim, H. A Collaborative Recommender System for Learning Courses Considering the Relevance of a Learner's Learning Skills. *Clust. Comput.- J. Netw. Softw. Tools Appl.* **2016**, *19*, 2273–2284. https://doi.org/10.1007/s10586-016-0670-x.
8. Gogo, K.O.; Nderu, L.; Mwangi, R.W. Fuzzy Logic Based Context Aware Recommender for Smart E-Learning Content Delivery. In Proceedings of the 2018 5th International Conference on Soft Computing & Machine Intelligence (ISCMI); November 2018; pp. 114–118.
9. Yanes, N.; Mostafa, A.M.; Ezz, M.; Almuayqil, S.N. A Machine Learning-Based Recommender System for Improving Students Learning Experiences. *IEEE Access* **2020**, *8*, 201218–201235. https://doi.org/10.1109/ACCESS.2020.3036336.
10. Delahoz-Dominguez, E.; Mendoza-Mendoza, A.; Visbal-Cadavid, D. EFFICIENCY OF ACADEMIC ENGINEERING PROGRAMS IN COLOMBIA: AN APPROACH THROUGH DATA ENVELOPMENT ANALYSIS. *J. Eng. Sci. Technol.* **2022**, *17*, 1105–1118.
11. Akbar, Z.; Sopandi, E.; Badruzzaman, B.; Khalik, M.F. The Role of Artificial Intelligence-Based Recommendation Systems in Selection of Courses for Students. *J. Soc. Sci. Util. Technol.* **2023**, *1*, 249–260. https://doi.org/10.55849/jssut.v1i4.671.

12. Chang, H.-T.; Lin, C.-Y.; Wang, L.-C.; Tseng, F.-C. How Students Can Effectively Choose the Right Courses: Building a Recommendation System to Assist Students in Choosing Courses Adaptively. *Educ. Technol. Soc.* **2022**, *25*, 61–74.
13. Touimi, Y.B.; Hadioui, A.; Faddouli, N.E.; Bennani, S. Intelligent Chatbot-LDA Recommender System. *Int. J. Emerg. Technol. Learn. IJET* **2020**, *15*, 4–20. https://doi.org/10.3991/ijet.v15i20.15657.
14. Sulaiman, M.S.; Tamizi, A.A.; Shamsudin, M.R.; Azmi, A. Course Recommendation System Using Fuzzy Logic Approach. *Indones. J. Electr. Eng. Comput. Sci.* **2020**, *17*, 365–371. https://doi.org/10.11591/ijeecs.v17.i1.pp365-371.
15. Aguilar, J.; Valdiviezo-Díaz, P.; Riofrio, G. A General Framework for Intelligent Recommender Systems. *Appl. Comput. Inform.* **2017**, *13*, 147–160. https://doi.org/10.1016/j.aci.2016.08.002.
16. Rodpysh, K.V.; Mirabedini, S.J.; Banirostam, T. Employing Singular Value Decomposition and Similarity Criteria for Alleviating Cold Start and Sparse Data in Context-Aware Recommender Systems. *Electron. Commer. Res.* **2023**, *23*, 681–707. https://doi.org/10.1007/s10660-021-09488-7.
17. Aditya, A.; Purwiantono, F.E. The Application of Fuzzy-Analytical Hierarchy Process Method for Majors Selection at Public Universities. *J. Inform. Telecommun. Eng.* **2020**, *3*, 240–251. https://doi.org/10.31289/jite.v3i2.3245.
18. Alghamdi, S.; Alzhrani, N.; Algethami, H. Fuzzy-Based Recommendation System for University Major Selection.; July 22 2024; pp. 317–324.
19. Perumal, S.P.; Arputharaj, K.; Sannasi, G. Fuzzy Family Tree Similarity Based Effective E-Learning Recommender System. *2016 Eighth Int. Conf. Adv. Comput. ICoAC* **2017**, 146–150. https://doi.org/10.1109/ICoAC.2017.7951760.
20. Obeid, C.; Lahoud, C.; El, K.H.; Champin, P.-A. A Novel Hybrid Recommender System Approach for Student Academic Advising Named COHRS, Supported by Case-Based Reasoning and Ontology. *Comput. Sci. Inf. Syst.* **2022**, *19*, 979–1005.
21. Nguyen, V.A.; Nguyen, H.-H.; Nguyen, D.-L.; Le, M.-D. A Course Recommendation Model for Students Based on Learning Outcome. *Educ. Inf. Technol.* **2021**, *26*, 5389–5415. https://doi.org/10.1007/s10639-021-10524-0.
22. Instituto Colombiano para la Evaluación de la Educación - ICFES Saber 11 2022. https://icfes.gov.co/guest/data-icfes
23. Instituto Colombiano para la Evaluación de la Educación - ICFES Saber Pro 2022. https://icfes.gov.co/guest/data-icfes
24. Bhumichitr, K.; Channarukul, S.; Saejiem, N.; Jiamthapthaksin, R.; Nongpong, K. Recommender Systems for University Elective Course Recommendation. In Proceedings of the PROCEEDINGS OF 2017 14TH INTERNATIONAL JOINT CONFERENCE ON COMPUTER SCIENCE AND SOFTWARE ENGINEERING (JCSSE); Chantamunee, S and DoungIn, S and Thanathamathee, P, Ed.; Walailak Univ, Sch Informat; NBTC; DEPA; TOT; Acer; IEEE Thailand Sect; IEEE; NECTEC; ECTI Assoc, 2017.
25. Khanam, Z.; Alkhaldi, S. An Intelligent Recommendation Engine for Selecting the University for Graduate Courses in KSA: SARS Student Admission Recommender System. In Proceedings of the Inventive Computation Technologies; Smys, S., Bestak, R., Rocha, Á., Eds.; Springer International Publishing: Cham, 2020; pp. 711–722.
26. Esteban, A.; Zafra, A.; Romero, C. Helping University Students to Choose Elective Courses by Using a Hybrid Multi-Criteria Recommendation System with Genetic Optimization. *Knowl.-Based Syst.* **2020**, *194*, 105385. https://doi.org/10.1016/j.knosys.2019.105385.
27. Samin, H.; Azim, T. Knowledge Based Recommender System for Academia Using Machine Learning: A Case Study on Higher Education Landscape of Pakistan. *IEEE Access* **2019**, *7*, 67081–67093. https://doi.org/10.1109/ACCESS.2019.2912012.
28. Slim, A.; Hush, D.; Ojha, T.; Abdallah, C.; Heileman, G.; El-Howayek, G. An Automated Framework to Recommend A Suitable Academic Program, Course and Instructor. In Proceedings of the 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService); April 2019; pp. 145–150.
29. Ognjanovic, I.; Gasevic, D.; Dawson, S. Using Institutional Data to Predict Student Course Selections in Higher Education. *Internet High. Educ.* **2016**, *29*, 49–62. https://doi.org/10.1016/j.iheduc.2015.12.002.
30. George, G.; Lal, A.M. A Personalized Approach to Course Recommendation in Higher Education. *Int. J. Semantic Web Inf. Syst. IJSWIS* **2021**, *17*, 100–114. https://doi.org/10.4018/IJSWIS.2021040106.