Machine Learning Model for Recommending Suitable Courses of Study to Candidates in Nigerian Universities

Chapter · September 2021

DOI: 10.1007/978-3-030-87013-3_20

Garba Aliyu
Ahmadu Bello University

Usman Haruna
Universiti Malaysia Terengganu

Idris Abdulmumin
Bayero University, Kano

Shehu Adamu
American University of Nigeria

Machine Learning Model for Recommending Suitable Courses of Study to Candidates in Nigerian Universities

Garba Aliyu[1], Usman Haruna[2], Idris Abdulmumin[1(✉)], Murtala Isma'il[2], Ibrahim Enesi Umar[1], and Shehu Adamu[3]

[1] Department of Computer Science, Ahmadu Bello University, Zaria, Nigeria
{algarba,iabdulmumin,ieumar}@abu.edu.ng
[2] Department of Computer Science, Yusuf Maitama Sule University, Kano, Nigeria
uharuna@yumsuk.edu.ng
[3] School of Information Technology & Computing, American University of Nigeria, Yola, Nigeria
shehu.adamu@aun.edu.ng

Abstract. The diversity of courses and complications of admission requirements are complex tasks particularly in Nigerian Universities where a number of parameters are used during the admission process. These courses may be wrongly assigned to applicants who have not met the minimum requirements. In a previous related work, a model was developed to address this issue. However, the model considered only seven subjects out of the mandatory nine subjects required of every senior secondary school student to register (O'Level). Such a decision may be to the detriment to the candidates because credits may be required from those subjects that were not considered. This paper tends to enhance the existing model to address all these issues. Grade of nine Secondary school subjects, the aggregate score of Unified Tertiary Matriculation Examination (UTME) and post-UTME, and catchment area are used as parameters in this study. The results were obtained when various reference classifiers were trained and tested using the processed dataset of the O'Level and JAMB results of candidates seeking admission into the university. Individual classifiers namely, Logistic Regression, Naive Bayes, Decision Tree, K-Nearest Neighbor, and Random Forest were trained and evaluated using reference performance metrics namely precision, recall, and f1-score. The resulting best classifier, the Random Forest, has shown to be correct 94.94% of the time and is capable of detecting correctly 94.17% of the classes. Since the precision and recall are similar in value, the f1-score tends to favor this classifier also with a value of 93.19%.

Keywords: Admission · Recommender system · Prediction · Classifiers · Machine learning

O. Gervasi et al. (Eds.): ICCSA 2021, LNCS 12957, pp. 257–271, 2021.
https://doi.org/10.1007/978-3-030-87013-3_20

1 Introduction

The rapid evolution of information technology has resulted in prompt developments across innumerable disciplines – education, healthcare, transportation, etc. The tools and resources that are used in this era are substituted nearly daily. Analogous to these changes, various ICT applications are substituting the traditional processes that aid teaching and learning activities [11]. Education has been acknowledged in Nigeria as the most dominant instrument of positive change and for national development [6]. For education to execute its part, recommending an alternative – or perhaps a more suitable – course of study to students is supreme [1]. Many students apply to different courses of their choice in different universities for their undergraduate study with their secondary school results as well as consistent test scores such as Unified Tertiary Matriculation Examination (UTME) and Post-UTME that helps determine their suitability or not for studying such choices. Institutions are expected to offer admission to suitable candidates based on their Senior Secondary Certificate Examination (SSCE or O'Level) results and other relevant test scores. However, in this whole route, while the course selection is the most critical phase for applying to undergraduate admission, it is difficult, time wasting and, in many instances, where students are not qualified for admission in their selected choice of courses, allocating alternatives to such students can be better improved.

Recommender systems are artificial intelligence-based systems that intelligently offer to users appropriate recommendations of items based on the previous interests of other people. They have been shown as very capable of serving users appropriately in online services, for marketing by eBay, Amazon, etc. [14], recommendation of words or phrases on platforms such as Google search, social media, and emailing service, in education, health, among other things. Most of the manufacturing industries have been advancing in many fields of data science by employing a recommender engine as a leading preference in their daily business. Academics have become aware of the tremendous potentialities offered by recommender systems and have adapt these systems to suggest courses especially by online classes administrators, e.g. Coursera, Lynda, etc. It is, consequently, critical that these systems are employed to traditional universities during the processes of admissions to help the students to study courses based on their capabilities. A knowledgeable analysis of the present-day system is need to be conducted to each theoretical study and practical tendencies of the recommendation systems [9]. Data mining techniques are very much useful to determine such kind of hidden knowledge from the important as well as composite data types [4].

In every new academic session, each university formulates diverse rules for admitting students who apply for admission. The challenge to that is determining which of the courses can be given to a particular student based on the admission requirements fulfilled. The recommended program may be the same as what the candidate applied for or contrary. The complexity of this is subject to many admission requirements criteria in SSCE, UTME score, Post-UTME, among other things. According to [6], beside each candidate's results, other requirements such as 45% strictly admitted based on merit, 20% from less

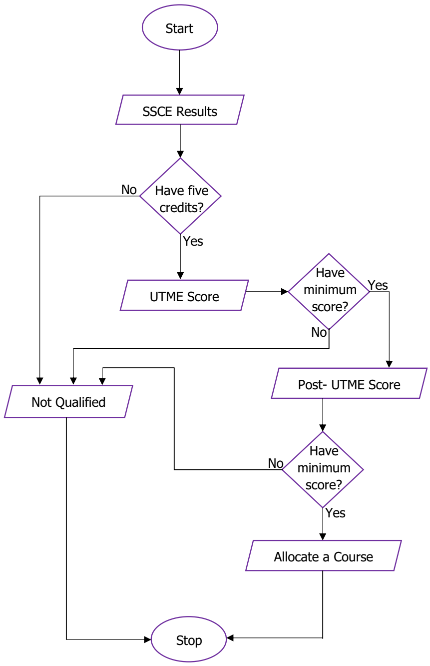

Fig. 1. Existing admission process.

educationally developed states (LEDs) to ensure equity, 20% from catchment areas, etc. further complicates the admission process. According to [13], university admission is a difficult decision process that goes beyond matching test scores and validation necessities. Most of the investigations carried out for university admission are based totally on the point of view of the universities who decide the most appropriate course to give to the candidate, and no longer at the interest of applicants who are either capable or not to study the program to administered and, therefore, making the choice of the course study to administer to the prospective students by the university a trivial problem every year [2].

The present way of assigning course of study to applicants in Nigerian universities – see illustration in Fig. 1 – needs reconsideration as allocating a suitable course is very sensitive to the candidates' performances in the higher institution of learning. In the start, the candidates are required to write for UTME. During the registration, each candidate indicates three subjects combination that are requisite for the course of study of their choice, with English being the fourth and compulsory. After obtaining the score in the UTME, the candidates are then requested to provide the grades, in each subject, they obtain in their SSCE. These subjects must include that which were registered in the UTME and, also, compulsorily passing Mathematics and English language. Most students sit for, and provide the grades of, nine subjects in their SSCE. As at the time of this study, the minimum UTME score to qualify for securing admission into any

Nigerian university is 160, based on the guidelines set by the Joint Admission and Matriculation Board (JAMB). However, most universities increase the score depending on the program the candidate applied and the availability of vacancies of each course [7]. Furthermore, some universities require a candidate to write a post-UTME examination, an examination that is conducted to further screen the candidates, provisional to the standards required by the administration of the university. After the scores of the UTME and post-UTME are aggregated and the SSCE is considered, the universities then assign courses to candidates first according to their choices and then according to the availability of slots in other courses, and this usually results in various problems.

The aim of this study is, therefore, to develop a model for recommending suitable and appropriate courses of study for prospective undergraduate students in Nigerian universities. This is despite the fact that some of these candidates may be recommended with a course different from what the they applied and may fail to accept the offer. This may then result in wasted slots. To reduce this, the model, therefore, takes into consideration and gives priority to the intended programs the candidates wished to study as long as they have met the minimum requirements and provides, alongside, a list of recommendations for other suitable programs. We explored the use of various classifiers such as Logistic Regression, Decision Trees (DT), K^{th} Nearest Neighbor (KNN), Naive Bayes (NB) and Random Forest (RF). Although previous work [6] have presented a tremendous contribution towards developing the course recommender system, the work focused only on the use of a relatively small number of dataset and few number of features (parameters) to automatically predict a suitable course to the student despite the fact that researches have shown that the more the number of data the better the accuracy [15,18]. Therefore, we extended the work in Isma'il et al. [6] by providing more datasets and using more features to automatically recommend a more appropriate course of study for the candidate. Nine subjects were considered in this work as opposed to seven subjects used in Isma'il et al. and, also, priority is given to the candidates' choice aside from other recommended courses that the candidate has met the requirements.

Thus, this work aims to address the following problems that are associated with admissions recommendation in Nigerian universities:

– courses given to candidates who may not best be suited for such courses or may even have some deficiencies in their SSCE,
– wrong assignment of candidates across disciplines, for example, candidates who applied for a particular discipline say in Science may be allocated courses in Engineering due to the similarity of requirements, Social Sciences to Art, etc.
– wastage of vacancy as students refuse to register into some courses allocated to them while others are looking for same.
– priority usually not given to the applicants' choice of courses which they have met the minimum requirements before exploring to recommend other relevant programs.

The rest of the paper is organized into the following sections. Section 2 reports the relevant related literatures. Section 3 describes, in detail, the proposed method for realizing the solutions to the identified problem. In Sect. 4, we discussed the results obtained after evaluating the recommendation systems. Finally, Sect. 5 presents the conclusion of the work and suggests future work.

2 Review of Related Literature

Many kinds of research exist on the recommendation systems in diverse fields such as hospitals, e-commerce, education, government among others. In several countries such as Nigeria, Saudi Arabia, India, China, among others, researches on admission recommender system have been carried out. This section discusses some of the related works in relation to building a recommender system for the university system of education.

In Swaminathan et al. [16], a recommender system was built using Kth Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Machines (SVMs) to recommend Graduate Studies of a university to candidates. They used the candidates' profile and the university's perspective to decide if it is relevant and sufficient to secure admission. This helps the candidate to choose and secure admission from the most appropriate university upon the recommendation made by the model. However, the model was not built to recommend a course to prospective candidates who are seeking admission.

Similarly, Hassan et al. [4] built a recommender engine that lists N universities that matched the candidates' educational profile who are seeking a postgraduate admission. This uses the data of those candidates that have been given a chance to study abroad from different countries in the world to recommend to prospective candidates. Enough data is required to be generated, which has to emanate from students who have previously secured admission, for these models to make an excellent recommendation.

In Ragab et al. [13], a recommender system was built using a combination of knowledge discovery rules and data mining called a hybrid recommender engine to recommend college admission. The model was designed as a prototype, which was implemented and evaluated using active data accessible on the On-Demand University Services (ODUS-Plus) database sources, at King Abdulaziz University.

A similar study to Ragab et al. was conducted by Wakil et al. [17] which used the University of Human Development as the field of study for universities in Iraq. Decision Tree (**DT**) was used to classify the candidates into 10 groups with each group having a unique property, **NN** employed to fit the candidate to the available courses, and to find the most appropriate course. Furthermore, the work regarded grade point average (GPA), test score, candidates' interest, and desire jobs as decision parameters. A hybrid system approach using **NN**, **DT**, and Our Proposed Algorithm (**OPA**) were used.

A recommender system was trained in the study conducted by Ismail et al. The system takes a combination of the grades of seven subjects in their SSCE to recommend a suitable course of study for each candidate. The study, though, did not take into consideration and give priority to the actual choice of each candidate. It does not consider the whole of the nine subjects provided in the

SSCE results. It was also trained on less data with a small number of courses that can be allocated.

This work presents an extension to the work originally presented in International Conference in Mathematics, Computer Engineering and Computer Science by [6]. In this paper, we consider more subjects in the SSCE result of candidates (9 subjects vs the 7 subjects in [6]), more courses from additional faculties, catchment areas, and we give priority to the choice of applicants and then followed it by a list of other recommended courses if minimum requirements have been made. It is observed that some of the subjects that are required for Computer Science are not required for Accounting and vice versa. Furthermore, two or more courses may have the same requirement, for example, Computer Science and Physics, Accounting and Economics, etc. Nevertheless, our recommender engine is capable of recommending those possible courses but giving priority to the course applied if a candidate has met all the requirements. Different classification algorithms are used to build models and we choose the best that outperformed others in terms of standard performance metrics.

3 The Methodology

This section presents the step by step description of the proposed approach, including the admission recommendation process, the conceptual model for building the recommendation systems, data collection and algorithm selection.

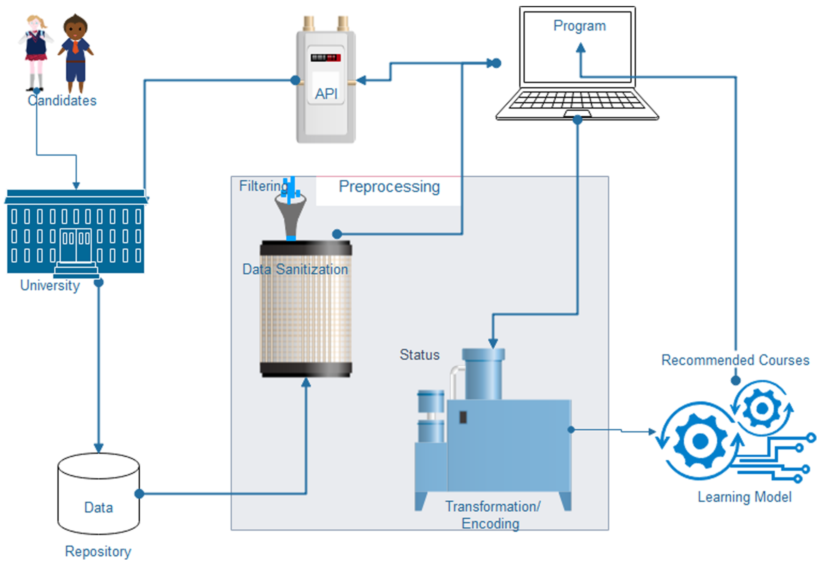

Fig. 2. Enhanced System Model

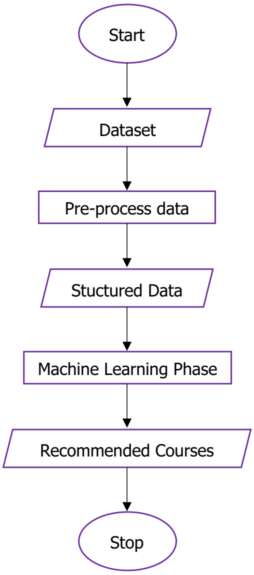

Fig. 3. Proposed admission process.

3.1 The Proposed Approach

The enhanced conceptual model adopted from [6] is shown in Fig. 2. The proposed approach, in contrast to the existing system (Fig. 1), is illustrated in Fig. 3. The first phase of data generation comes to the universities through the JAMB headquarters. The second phase of the data is generated by giving candidates access to the university portal to upload (SSCE results and other details). However, the subjects that every candidate is expected to write as Post-UTME (exam conducted by the university as entrance exam) is automatically added based on the course applied. All the relevant information necessary for the model is fetched as input to the model through an Application Programming Interface (API).

As an extension of the work of [6], this work takes into account the course applied by the candidates, asides from other possible courses that can be assigned. The data obtained is cleaned to eliminate all irregularities before transforming them into the categorical form that it can be fitted and processed by the classification models. The sanitized data then fed into the model for training, testing, and cross-validation to predict the likely courses from unforeseen data that can be given to a prospective candidate. It is worth mentioning that this recommender system make recommendations to the university. It is expected that instead of the existing process of randomly assigning available courses to prospective students, the proposed system make intelligent recommendations based on the candidates' abilities as presented in their academic records. The final decision to either accept the recommended course and assign to the students lies solely on the prerogative of the university based on the availability of admission vacancies, which varies every year.

This study is implemented in three phases in a pipeline fashion. This implies that the output of one phase serves as the input of the next phase. The first phase involves data collection and pre-processing. This is the most important phase because the outcome of this phase will determine the effectiveness of the subsequent phases. In the second phase, selected classifiers are trained, tested, and evaluated. Evaluation metrics used are precision, recall, accuracy, and F1-score. In the final phase, we compare these metrics. The following subsections describe each phase.

3.2 Data Collection

Generalization is a crucial feature required of a machine-learning model. Therefore, the resulting model must generalize well. To achieve that, sufficient data needs to be collected. More so, the data collected need to be representative of the cases we want to generalize. We identified various sources of data but settled on two prominent universities: Ahmadu Bello University (ABU) Zaria, Kaduna State and Bayero University Kano (BUK), Kano State both in Nigeria. Both universities are among the largest and most popular universities in Nigeria and West Africa. They both offer a wide variety of courses and have been doing a good job of managing their collection of data. The demography of students is highly diverse ranging from students from all parts of Nigeria to international students from all over the world. In ABU, we collected a dataset of three (3) sessions; 2015/2016, 2016/2017, and 2017/2018 while in BUK we were able to collect a dataset of 2 sessions; 2016/2017 and 2017/2018. This amounts to a total of 8,700 data instances. The dataset collected comprises 4 variables forming the information about the students. These are the UTME number and score, post-UTME score, and SSCE results.

Real-world datasets are highly affected by the presence of noise, missing values, inconsistent data, and duplicate records, and redundant features. We employed several data pre-processing techniques to eliminate these negative factors. Some of the pre-processing were done using spreadsheet application and python. We also prepared the data so that it is suitable for each of the selected classification algorithms. We then obtain a reduced representation of the data where necessary. Since we gathered our data from different sources, there is the need to merge the data into a single dataset. Fortunately, the dataset used for this research came in the same format (CSV). During data integration and exploration, redundant attributes were detected and eliminated from the resulting dataset. We achieved this by using the well-known Pearson's product-moment coefficient for numeric attributes and correlation test for nominal attributes. We then move forward to cleaning the dataset. During the early stage of pre-processing, the data was split into training and testing. 20% of the data was reserved for evaluation purposes.

The most challenging was encoding the SSCE results, where the transformation is done in a way that, for example, if a candidate is having a minimum of credit (C4, C5, C6, B2, B3, and A1) in say English Language, then a unique code is given to the grade and the subject as 21. The first digit refers to the subject code while the second (or last digit) corresponds to the grade code. This process continues in the same manner for all the nine-subjects considered in this work. Courses that candidates applied with other possible courses that can be given based on the course's requirements are coded in the same fashion as SSCE results.

3.3 Algorithm Selection

Machine learning algorithms are used in recommender systems to deliver reliable recommendations. There are numerous machine learning algorithms to choose from when building a recommender system. According to [12], the types of machine learning algorithms used in building recommender systems were listed in order of most commonly used to the least commonly used. It was observed that the majority of studies reviewed were based on the first seven machine learning algorithms. We then based our choice of algorithm on these sets. Specifically, in this study, we developed models for our recommender system using five algorithms - Naive Bayes, Linear regression, Decision tree, Random Forest, and K-Nearest Neighbour. We then compared their performance using various performance metrics.

Every algorithm has some unique capabilities and drawbacks thus, the need to explore and choose the best algorithm that suits the model [8,16]. According to Isinkaye et al. [5], DT fits regression and classification while selecting a tree structure. A decision tree is formed from a given dataset that is split into a smaller subset, which makes it eminent in decision-making problem. Naive Bayes is a machine learning algorithm that uses in a diversity of probabilistic nature of knowledge. Varying the value of one feature, which seems not undeviatingly alter or change the value of any of the rest of the features used in the algorithm, it is a powerful algorithm [10]. The classification applied in developing a recommendation system is critical that the system will not give very good recommendations if the model is not thoroughly built. The more the availability of data, the best for the model to learn and gather more experience to provide better, meaningful and more accurate information right from the onset [5].

It is important that these algorithms are trained with a sufficient amount of data that are representative of the problem domain. Luckily the data obtained were provided by Universities where such a solution will be of most importance. Even though the amount of data obtained from these universities is fairly large enough to get meaningful outcome from training with the selected algorithms, better result could be obtained if more data were available.

3.4 Model Training and Evaluation

In the previous subsections, we presented the process of dataset processing, feature selection, and the process of splitting the dataset for training and testing. In this subsection, we address the problem of selecting the best model for use in

Table 1. Comparison of model performances with respect to accuracy, precision, recall and F1 scores for various values for K-Fold. Values indicated in percentages (%). KEY: CV – *Cross Validation*, LR – *Logistic Regression*, NB – *Naive Bayes*, DT – *Decision Tree*, KNN – K^{th} *Nearest Neighbor*, RF – *Random Forest*.

Score	CV	LR	NB	DT	KNN	RF	Score	CV	LR	NB	DT	KNN	RF
Accuracy	5	89.11	83.46	90.42	93.11	94.12	*Recall*	5	89.11	83.05	90.42	93.10	94.15
	10	89.25	83.46	90.28	93.23	94.12		10	89.25	83.45	90.30	93.22	94.17
	20	89.21	83.46	90.23	93.11	94.22		20	89.20	83.45	90.20	93.34	94.17
	avg	89.19	83.46	90.31	93.15	94.15		avg	89.19	83.32	90.31	93.22	94.16
	s.d	7.21	0.00	9.85	6.93	5.77		s.d	7.09	23.09	11.01	12.00	1.15
Precision	5	87.40	88.85	90.56	92.75	94.91	*F1*	5	87.49	84.06	89.45	92.61	93.16
	10	87.23	88.85	90.46	92.80	94.93		10	87.62	84.06	89.45	92.72	93.10
	20	87.42	88.85	90.37	92.94	94.94		20	87.61	84.06	89.22	92.84	93.19
	avg	87.35	88.85	90.46	92.83	94.93		avg	87.57	84.06	89.34	92.72	93.15
	s.d.	10.44	0.00	9.50	9.85	1.53		s.d.	7.23	0.00	11.50	11.50	4.58

the recommender system. The main objective of this section is to train and test the individual reference classifiers mentioned in the research framework section with the same dataset. The implementation of this process was performed using Jupiter Notebook. Model evaluation was performed on the test data set aside earlier on. Each of the models is tested with the test set using cross-validation. GridSearch was also employed to determine the best hyper-parameters for each model. The performance was measured base on precision, recall, accuracy, and f1-score.

4 Results and Discussion

In this section, we compared the model performances using various metrics. In [6] – the conference paper submitted regarding this study – accuracy was used as the performance measure. However, it has been demonstrated in [3] that accuracy is not always the preferred metric for measuring the performance of classifiers. We therefore included other concise metrics such as precision, recall and f1-score.

Table 1 compares the accuracy, precision, recall and f1 scores of all reference classifiers used in this study. Corresponding charts of the results obtained are also shown in Fig. 4, 5, 6 and 7. In the table, the average of the scores and standard deviation were computed. It can be observed, from the table above, that the standard deviation values vary significantly among each classifier. This signifies that the choice of classifier is very crucial. Also, we observed from the scores in the table that among all performance measures, the **Random Forest** classifier produced the highest score. Figure 8 illustrates the standard deviation for accuracy and other evaluation metrics obtained from each reference classifier.

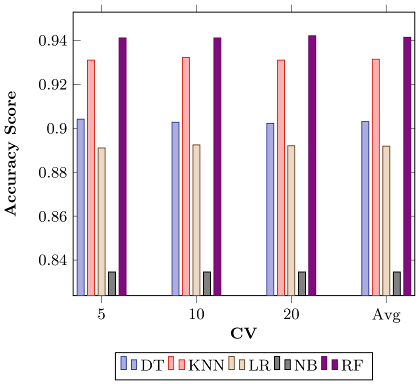

Fig. 4. Accuracy Score. KEY: CV – *Cross Validation*, LR – *Logistic Regression*, NB – *Naive Bayes*, DT – *Decision Tree*, KNN – K^{th} *Nearest Neighbor*, RF – *Random Forest*

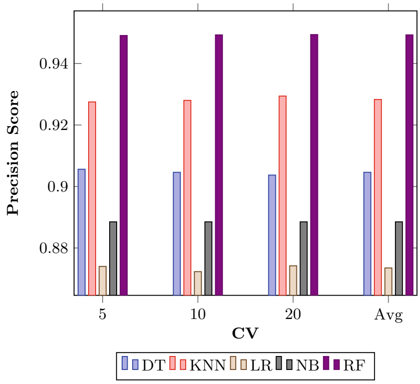

Fig. 5. Precision Score. KEY: CV – *Cross Validation*, LR – *Logistic Regression*, NB – *Naive Bayes*, DT – *Decision Tree*, KNN – K^{th} *Nearest Neighbor*, RF – *Random Forest*

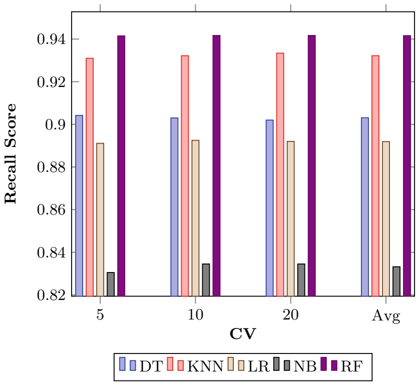

Fig. 6. Recall Score. KEY: CV – *Cross Validation*, LR – *Logistic Regression*, NB – *Naive Bayes*, DT – *Decision Tree*, KNN – K^{th} *Nearest Neighbor*, RF – *Random Forest*

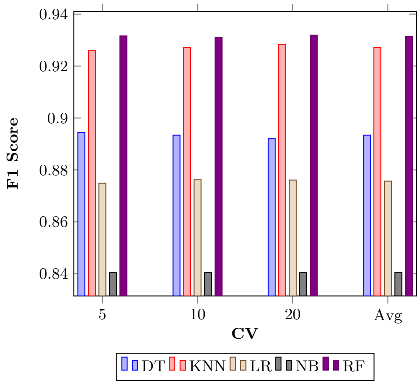

Fig. 7. F1 Score. KEY: CV – *Cross Validation*, LR – *Logistic Regression*, NB – *Naive Bayes*, DT – *Decision Tree*, KNN – K^{th} *Nearest Neighbor*, RF – *Random Forest*

This work increased the number of SSCE results from seven in [6] to nine subjects. Even though universities require candidates to have a minimum of credit in only five relevant subjects, but in exactly which subjects is a major concern, because if, for example, a student registered nine subjects with a university, apart from Mathematics and English, there is maximum of ${}^{7}C_{3}$ – 35 – possible combination to choose the remaining three courses out of the seven subjects. Limiting the number of subjects to seven may be a detriment to the candidates. Some candidates do not register more than seven subjects; as such, the model

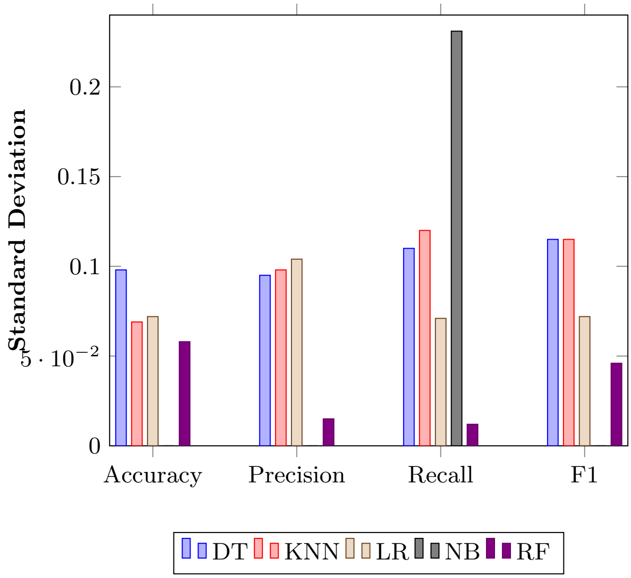

Fig. 8. Standard Deviations for the Accuracy, Precision, Recall and F1 Scores for the various classifiers

Table 2. Best classifier performance

Performance of random forest with CV = 20	
Performance Metric	Score
Precision	94.94%
Recall	94.17%
F1 Score	93.19%

was built with some null values to accommodate nine subjects, but it increases the chances for some candidates.

For each model trained, the performance metrics are presented in terms of precision, recall and F1-score. So we can conclude that random forest performs better than the other reference classifiers considered in this study. More so, the results obtained from the experiments showed that training random forest with a cross validation fold of 20 produces the best f1-score as shown in Fig. 7. We, thereby, extracted and presented the performance of the best classifier in Table 2. Finally, we discussed the result obtained from the comparison. The results obtained for the nine subjects considered in this work, which is 94.15% accuracy even though it is not better than for the seven subjects, which is 99.94% accuracy but because of the null entries for some candidates who do not have up to nine subjects, to some degree, it confuses the model and thus

Table 3. Result comparison between the previous and the current work

Model	Accuracy (%)	
	Previous work [6]	Current work
LR	–	89.25
NB	99.94	83.46
KNN	99.87	93.23
DT	98.01	90.42
RF	–	94.22

wrongly recommended courses to candidates. Consequently, 94.15% of accuracy was achieved as shown in Table 1.

The scores for accuracy obtained when each reference classifier was evaluated with various values for K. Observe that random forest with K=20 produced the highest value for accuracy and the highest average score for accuracy was obtained with random forest classifier. Moreover, the prioritization of candidate's choice reduces the wastage of admissions or blocking some slots that can be given to those who also qualified based on the requirements.

Table 3 shows the performance comparison between the models in [6] and the models trained in our work. It can be observed that the performances in this work under-performed that of the previous work. This is because the data features that were considered in the previous work was fewer than that used in training the models in this work. Despite this, we are confident that our model is able to generate more realistic recommendations based on the additional subjects scores provided – 7 to 9 subjects.

5 Conclusion and Future Work

This study presents a recommender system for helping the university admission process to assign more appropriate courses to prospective students based on prior performances in the O'Level, UTME and post-UTME examinations. Instead of the existing manual system that randomly assign courses based on the subjects combination and the availability of vacancies, it is intended that the proposed system will intelligently assign a more suitable course of study based on the capability of the students. The experimental results in this work were obtained when various reference classifiers were trained and evaluated using the processed dataset of the O'Level and JAMB results of candidates seeking admission into two Nigerian universities. Individual classifiers namely: Logistic Regression, Naive Bayes, Decision Tree, K-Nearest Neighbour, and Random Forest were trained and evaluated with reference performance metrics namely: accuracy, precision, recall, and f1-score. The resulting best classifier (Random Forest) was shown to be correct 94.94% of the time and is capable of detecting correctly 94.17% of the classes. Since the precision and recall are similar in value, the f1-score tends to favour this classifier also with a value of 93.19%.

The nine subjects considered in this study has its limitation because it is possible that an applicant will have five credits but few of which may not be part of those considered and, meanwhile, are part of the requirement for the applied program. Another limitation to this study is the scope of our validation. In future work, we plan to deploy our model in the Nigerian universities whose data was used in this work during the subsequent admission process so as to test and validate the model on new and unseen admission data. This work can be extended by looking at other subjects beyond nine subjects, even though the remaining subjects that are not considered in this study were found to be trivial, as such, they are avoided in order not to confuse the model with null values because few applicants have those subjects. The wastage of slots by candidates who are not usually accepting provisional admission simply because it is not their choice can be further looked into so that this wastage of admissions can further be reduced.

Acknowledgment. This research was carried out under the support of Ahmadu Bello University, Zaria, Yusuf Maitama Sule University, Kano, and Bayero University, Kano who provided us with datasets.

References

1. Bulkachuwa, A.S.: Availability and Utilization of Microteaching Laboratories and their Influence on Students' Teaching Practice Performances in Colleges of Education. Technical report, Kano (2017)
2. Fong, S., Buik-Aghai, R.P.: An automated university admission recommender system for secondary school students. In: The 6th International Conference on Information Technology and Applications (ICITA), China (2009)
3. Géron, A.: Hands-on Machine Learning with Scikit-Learn Keras, and TensorFlow, 2nd edn. O'Reilly Media Inc., Sebastopol (2019)
4. Hassan, M., Ahmed, S., Abdullah, D.M., Rahman, M.S.: Graduate school recommender system: assisting admission seekers to apply for graduate studies in appropriate graduate schools. In: 5th International Conference on Informatics, Electronics and Vision (ICIEV), p. 503. Institute of Electrical and Electronics Engineers (2016)
5. Isinkaye, F.O., Folajimi, Y.O., Ojokoh, B.A.: Recommendation systems: principles, methods and evaluation. Egypt. Inform. J. **16**, 261–273 (2015)
6. Isma'il, M., Usman, H., Aliyu, G., Abdulmumin, I., Shehu, A.: An autonomous courses recommender system for undergraduate using machine learning techniques. In: International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS). IEEE, Lagos, Nigeria (2020). https://doi.org/10.1109/ICMCECS47690.2020.240882
7. JAMB: Exam/jamb-cut-off-mark/ (2019). https://www.currentschoolnews.com/exam/jamb-cut-off-mark/
8. Jauro, F., Alhassan, B.B., Aliyu, G., Toro, M.S.: Predicting child delivery mode using data mining by considering maternal. Sci. Forum (J. Pure Appl. Sci.) **18**, 31–35 (2019)
9. Lu, J., Wu, D., Mao, M., Wang, W., Zhang, G.: Recommender system application developments: a survey. Decis. Supp. Syst. **74**(C), 12–32 (2015)

10. Machine Learning Plus: How Naive Bayes Algorithm Works? (with example and full code)—ML + (2018). https://www.machinelearningplus.com/predictive-modeling/how-naive-bayes-algorithm-works-with-example-and-full-code/
11. Muhammad, A.: Undergraduate students project supervision recommendation system, Technical report, Federal University, Dutsen-Ma, Katsina (2017)
12. Portugal, I., Alencar, P., Cowan, D.: The use of machine learning algorithms in recommender systems?: a systematic review. Expert Syst. Appl. **97**, 205–227 (2018)
13. Ragab, A.H.M., Mashat, A.F.S., Khedra, A.M.: Design and implementation of a hybrid recommender system for predicting college admission. Int. J. Comput. Inf. Syst. Ind. Manag. Appl. **6**, 35–44 (2014)
14. Said, A., Bellogin, A.: Comparative recommender system evaluation: benchmarking recommendation frameworks. In: Proceedings of the 8th ACM Conference on Recommender systems - RecSys '14, Foster City, Silicon Valley, California, USA, pp. 129–136 (2014). https://doi.org/10.1145/2645710.2645746. http://dl.acm.org/citation.cfm?doid=2645710.2645746
15. Salas, J.: Sanitizing and measuring privacy of large sparse datasets for recommender systems. J. Ambient Intell. Human. Comput. (2019). https://doi.org/10.1007/s12652-019-01391-2
16. Swaminathan, R., Gnanasekaran, J.M., Krishnakumar, S.: Recommender system for graduate studies in USA Suresh (2015). https://www.semanticscholar.org/paper/Recommender-System-for-Graduate-Studies-in-USA-Suresh
17. Wakil, K., Akram, B., Kamal, N., Safi, A.: Web recommender system for private universities' admission in Iraq: UHD case study. Int. J. e-Educ. e-Bus. e-Manag. e-Learn. **4**, 329–340 (2014)
18. Zubairu, M.J., et al.: Text normalization algorithm for Facebook chats in Hausa language. In: 2014 5th International Conference on Information and Communication Technology for The Muslim World (ICT4M), pp. 1–4. IEEE, Kuching (2014). https://doi.org/10.1109/ICT4M.2014.7020605. http://ieeexplore.ieee.org/document/7020605/