# An Autonomous Courses Recommender System For Undergraduate Using Machine Learning Techniques

**Conference Paper** · March 2020

DOI: 10.1109/ICMCECS47690.2020.240882



**5 authors**, including:

Usman Haruna
Universiti Malaysia Terengganu
**6** PUBLICATIONS **17** CITATIONS
SEE PROFILE

Garba Aliyu
Ahmadu Bello University
**16** PUBLICATIONS **11** CITATIONS
SEE PROFILE

Idris Abdulmumin
Bayero University, Kano
**18** PUBLICATIONS **21** CITATIONS
SEE PROFILE

Shehu Adamu
American University of Nigeria
**6** PUBLICATIONS **8** CITATIONS
SEE PROFILE

# An Autonomous Admission Recommender System Using Machine Learning Techniques

Murtala Isma'il
*Department of Computer Science*
*Yusuf Maitama Sule University*
Kano, Nigeria
murtalaismail@gmail.com

Usman Haruna
*Department of Computer Science*
*Yusuf Maitama Sule University*
Kano, Nigeria
usmancyz01@gmail.com

Garba Aliyu
*Department of Computer Science*
*Ahmadu Bello University*
Zaria, Nigeria
algarba@abu.edu.ng

Idris Abdulmumin
*Department of Computer Science*
*Ahmadu Bello University*
Zaria, Nigeria
iabdulmumin@abu.edu.ng

Shehu Adamu
*Department of Computer Science*
*Federal Polytechnique*
Bali, Nigeria
shehuadamushehu@gmail.com

***Abstract*—A university admission process is a complex process that needs huge amount of time and labour in order to allocate course(s) to their prospective applicants. The main objective of this work is to find the best way of allocating courses to applicants without human intervention. This model can help Universities by providing simple way of allocating admission to their prospective applicants. In reflection of the high degree of prediction accuracy, adaptability is an advantage, as the system can predict more appropriate courses that agree with the students' grades. The system is adaptive, and can be used as a framework for further research on admission recommender system. The proposed system employs Linear Regression Model, Naïve Bayes, Support Vector Machine, Kth-Nearest Neighbour and Decision Tree Classifiers. The research uses the Secondary School results and the average score of UTME and Post-UTME. The data used were obtained from Ahmadu Bello University, Zaria and Bayero University, Kano. The Linear Regression Model has the Root Mean Squared Error (RMSE) to be $2.614\ e^{-14}$. The results showed that Naïve Bayes Classifier and Support Vector Machine have an accuracy of approximately equal to 99.94% which outperforms Decision Tree and Kth-Nearest Neighbor algorithms with an accuracy of 98.10% and 99.87% respectively.**

***Keywords—University, Admission, Recommender, Prediction, Classifiers***

## I. INTRODUCTION

In the current era of information technology, speedy advances in information and communication technologies (ICT) have been turned out to be common if not necessary in various disciplines. The tools and materials that are used in this age are replaced virtually daily. In parallel with these changes, ICT applications have joined the trend of learning and teaching activities [1]. Accordingly, education is accepted in Nigeria as the most powerful instrument of change and for national development. For education to perform its role, the placement of students into the appropriate course of study is paramount [2]. Recommender systems use artificial intelligence approaches to offer users with item recommendations.

Academics and executives identify that recommender systems offer pronounced prospects and challenges for government, education, business and other fields, with more modern positive developments of recommender systems for real-world applications becoming obvious. It is therefore, important that an educational review of the current system be conducted both theoretical research and more importantly the practical developments in the recommender systems [3].

The university education system is the basic part for those aspiring to be researchers and for those preparing for job life. Therefore, admission into the university is of great concerned. Every year, every university prepares several processes for admitting students who apply for the admission into that university. Now, the question is that, how does a student choose a course and how does the university allocates courses to aspiring candidates? This is a complex issue that depends on many factors such as secondary school results, Unified Tertiary Institution Matriculation Examinations (UTME) score, Post-UTME score to mention but a few. The selection and allocation of a course to study that is suitable for a given secondary school graduate can be a difficult decision to make by the students and the universities respectively. This process is not only dependent on the students' UTME scores, it moreover relies upon students' experiences, catchment areas, less educationally developed states and different capabilities weighting criteria that relate to the execution of their tertiary institution. According to [4] "College or school affirmation is a complex choice process that goes past matching test scores and confirmation necessities."

However, [5] is in the opinion that most present investigations of university admission are based on the point of view of universities who are to admit the new incoming students, and not on the interest of students who decide which course to administer to persevere higher education. This study proposes a novel design of a recommender system that can give recommendations regarding which course a prospective applicant should apply for and be proffered by the university taking into cognizance the student's ordinary level subjects combination and the average of UTME and Post-UTME scores. Selecting the most appropriate amongst the several candidates that apply to a university each year is not a trivial problem [5].

In Nigeria, the admission system is based on three major criteria that is merit which carries 45%, catchment areas with 35% and educationally less disadvantage states have the 20%

[6]. This work is aimed to be achieved by using some algorithms such as Linear Regression, Kth-Nearest Neighbor, Support Vector Machine (SVM) and Decision Tree Algorithm.

## II. Statement of the Problem

The contemporary way of allocating courses to the students in our higher institutions need to be reconsidered since course allocation is a sensitive aspect of student education in the higher institution being the crux through which every student must experience. At the beginning, students are required to sit for UTME. But before then each student has to register with the Joint Admission and Matriculation Board (JAMB) indicating the subjects offered at the secondary school with grades obtained per each subject and specifying the subjects to write during the UTME. Some Universities also conduct Post-UTME depending on the criteria imposed by the management of the university. According to JAMB, the minimum score in UTME that will grand students opportunity to secure admission into the Nigerian university is 160, but some universities reduce or increase the score depending on the course that applicant applied for and the available slots per course [7]. Universities allocate courses to students randomly, and this results to various problems some of which are:

- Course is allocated to students who have deficiency in his senior secondary school certificate examinations (SSCE) that is necessary for the study of the course given to him/her
- There is wrong placement of students to a particular course, for example a student who applied for Medicine finds himself in Engineering
- Wastage of vacancy as students refuse to register into some courses allocated to them while others are looking for same.

But with the coming up of an automated based system, that looks into students' secondary school subjects' combination, UTME and Post-UTME scores, (that is taking the computed average) the system will provide a solution based on recommendation to which course should be allocated to students.

## III. Literature Review

Several studies have been conducted on the recommender system in various fields like e-commerce, hospital, university recommendation and admission recommendation systems by different researchers. In addition, so many researches have been carried out on the admission recommender system in different countries such as China, India, Iraq, Saudi Arabia, Turkey among others. The Method uses in building recommender system is very crucial and it cannot work accurately if the model is not well built. The system expected to get as much information as possible from the dataset to gather for experience to provide a reasonable recommendation right from the onset [8].

In [9], the authors have developed a University Recommender System for Graduate Studies in the USA by employing Kth Nearest Neighbors, Random Forest and Support Vector Machines proffer a solution to certain problems by weighing the target university's perspective to judge whether a student's profile is contentious sufficient to be admitted. Hence, the students could get a better idea of where they hold and can secure an intelligent well-formed decision. But this model can only predict for which university the students may choose for graduate studies and vise-vasa.

The work of [10] have proposed a recommendation engine that recommends a list of universities to apply for postgraduate admission to persevere higher degrees amidst funding. In this study, applied research was employed on planning and developing a recommender system for graduate admission seekers that can aid students to determine graduate school rivalling their whole educational profile. The system improved a method to transform relational databases for students of all sorts of related information into a universal database format applying academic data of prosperous students who have already gotten a chance to study overseas. The K-nearest Neighbor algorithm was employed for determining top N related students for the test and suggest Top K universities from N related students..

In [4], another college admission system utilizing hybrid recommender based on the data mining techniques and knowledge discovery rules for managing college admissions prediction difficulties has transpired. The advanced system includes two cascaded hybrid recommenders operating unitedly with the help of institute predictor, for attaining great performance. The system was designed as a prototype which was implemented and tested with live data available in the On-Demand University Services (ODUS-Plus) database resources, at King Abdulaziz University (KAU), Jeddah, Saudi Arabia. The algorithm uses previous historical GPA scores for predicting the probability of admission.

In [11], the authors have conducted similar research for Iraq universities in which they used the University of Human Development (UHD), Sarchia, Sulaimaniyah, KRG, Iraq as the case study. Their system used a hybrid method of Neural Network (NN), Decision Tree (DT) and Our Proposed Algorithm (OPA). They also used DT to classify the applicants into 10 groups with each group with special properties, NN used to apply the applicant to the available courses and their algorithm mixed with NN were used to find the best suitable course. Besides, their work considered GPA, test score, candidates' interest and their desire jobs as decision parameters.

In [12], the authors conducted a study on the prediction of child delivery mode. The model employed data mining classification models which included K-Nearest Neighbor (KNN), Naïve Bayes (NB), Support Vector Machine (SVM), and Decision Tree models to predict the mode of delivery in obstetrics by considering both maternal and fetal factors.

However, in addition to more features (variables), the study needs to be innovated in the Nigerian context. Beside the O'level and JAMB results, universities conduct another exam as part of the admission process to test the integrity, quality, and capacity of the applicants before securing admission. All of the existing studies were carried out in different countries that have completely different criteria of admission process when compared to Nigerian universities. The present study is going to explore admission requirements considering all of the above mentioned criteria for the admission process and make a

recommendation on what should be the appropriate course to be given to the students for decision making.

## IV. METHODOLOGY

This section dealt with the research methodology which describes the methodological approaches that were employed to conduct the research which include data used to test the system. It also discusses the variables of the study under sub-headings such as data understanding, data pre-processing and modelling, research design, population, training and test sample size, data collection instruments, scoring procedure, procedures for data collection and data analysis that were adopted to conduct the research.

### *A. Data Understanding*

In order to use any recommender system, the first and most important thing is the data that will be used to train the model. Therefore, data collection is the first and most crucial step for building any recommendation engine. In the course of conducting this research, real dataset was obtained from two different universities which includes Ahmadu Bello University, Zaria (ABU) for the 2015/2016, 2016/2017 and 2017/2018 academic sessions and Bayero University, Kano (BUK) for the 2016/2017 and 2017/2018 academic sessions. A total of 8,700 data instances consisting 4 variables indicating information about the students, including UTME number, UTME score, Post-UTME score and Secondary school results were obtained.

### *B. Data Pre-processing*

In the pre-processing stage, refinement of the data was performed by removing all the instances with some racket values because real-world data tend to be incomplete and inconsistent [12]. The data collected is subjected to "data cleaning" which identifies and fills in the missing values and correct inconsistencies in the data [5]. Unpredictable and unimaginable values were discovered, for example, a string was captured as a UTME or Post-UTME results which were solely integers or an integer in the place of O'level grade which is character as a result of typographical error in the course of data capture. Also, the dataset is divided into two parts, one for training called the training dataset and another for testing and validation called the test dataset. The sanitized and decoded data will be transmitted to the feature selection component in classifier component.

As in the case of secondary school results nine (9) subjects were extracted to test the system which include English Language, Mathematics, Biology, Chemistry, Physics, Agricultural Science, Accounting, Economics and Government.

After the sanitization, before validating the model with the dataset for processing, the target variables were encoded into numeric values. The secondary school subjects were encoded from 1 to 9 as shown in the Table I while the results were encoded as follows: A, B and C grades encoded as 1, while D, E, P are encoded as 2 and F is encoded to be 3 shown in Table II. The grades A1, B2, B3, C4, C5, and C6 are categorized as credit while D7, E8 and P as passed. Having higher grades (A1, B2 and B3) far as the admission requirement is concerned does give priority for securing admission as compared lower grades (C4, C5 and C6).

TABLE I. ENCODING SUBJECTS

| Subjects | Encoded |
|---|---|
| English | 1 |
| Mathematics | 2 |
| Biology | 3 |
| Chemistry | 4 |
| Physics | 5 |
| Agric. Science | 6 |
| Accounting | 7 |
| Economics | 8 |
| Government | 9 |

TABLE II. GRADE ENCODING

| Grades | Encoded |
|---|---|
| A1 | 1 |
| B2 | 1 |
| B3 | 1 |
| C4 | 1 |
| C5 | 1 |
| C6 | 1 |
| D | 2 |
| E | 2 |
| P | 2 |
| F | 3 |

The considered target variables are various courses offered in the university which we will consider some of them for now and for the purpose of the study as shown in Table III.

TABLE III. AVAILABLE COURSES

| Courses | Code |
|---|---|
| **Faculty of Administration** | |
| Local Gov't and Development | 34 |
| Public Administration | 46 |
| **Faculty of Agriculture** | |
| Agriculture and Bio-resource | 2 |
| Agricultural Extension | 3 |
| Agriculture | 4 |
| Fisheries and Aquaculture | 24 |
| **Business School** | |
| Accounting | 1 |
| Actuarial Science | 7 |
| Banking and Finance | 8 |
| Business Management | 9 |
| Marketing | 10 |
| Business Administration | 12 |
| Economics | 21 |
| **College of Health Sciences** | |
| Dental Surgery | 20 |
| Human Anatomy | 29 |
| Human Physiology | 30 |
| Medical Lab. Science | 38 |
| Medicine and Surgery | 39 |
| Nursing Science | 42 |
| Pharmacy | 43 |

According to the JAMB and University, each of the course has requirements for admission. For example, the applicant from Medicine should come from science section in secondary school, the average of UTME and Post-UTME score should be greater or equal to 220 and credit in English, Biology, Chemistry, Physics and Mathematics.

Furthermore, each group of course code is further encoded according to the faculty as follows:

- Faculty of Administration [34,46] = 2
- Faculty of Agriculture [2,3,4,24] = 4
- Business School [1,7,8,9,10,12,21] = 6
- College of Health Sciences [20,29,30,38,39,42,43] = 8

## V. System Model

Fig. 1 describes the system model where each component aids in achieving the said objectives. The inputs to the system are the average score of UTME and Post-UTME and the secondary school subjects' combinations with their grades. These inputs are retrieved from universities' data repository through an application programming interface (API). The retrieved data is now sanitized to remove all anomalies before encoding the categorical data into a way that it will be compatible and processed by the classification model. The system uses these inputs to train the model and make prediction of the likely courses that can be given to a prospective candidate.

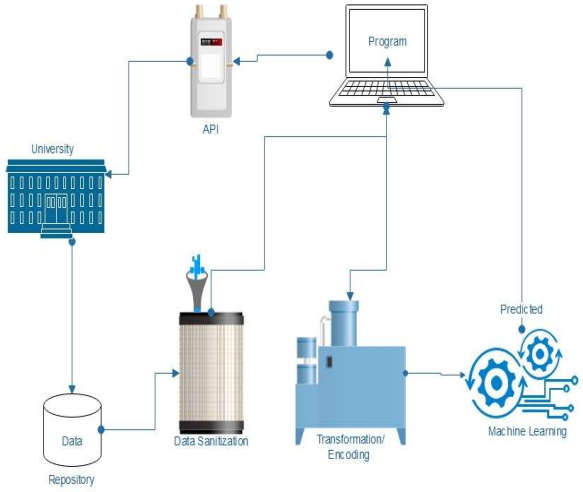


Fig. 1. System Model

## VI. Results and Discussion

The work was implemented using the Linear Regression Model and other four (4) classification models which comprised of Naïve Bayes, Support Vector Machine, Kth-Nearest Neighbor and Decision Tree Classification Models. All the classification models used were found to be efficient as none of them has the efficiency of less than 90%.

Fig. 2 shows the result of the Linear Regression Model where the Root Mean Squared Error (RMSE) is found to be $\mathbf{2.614\ e^{-14}}$. The lower the RMSE the better for the result. Thus, the result suggested how best our model is good enough for decision making when employing Linear Regression Model. The predicted values from the linear model are in the ordered pair as indicated in Fig. 2. For Example, (0, 8.0000000000000284) indicated that the predicted value is approximately equals to 8 which corresponds to a set of courses as detailed explained and encoded in section IV (B).

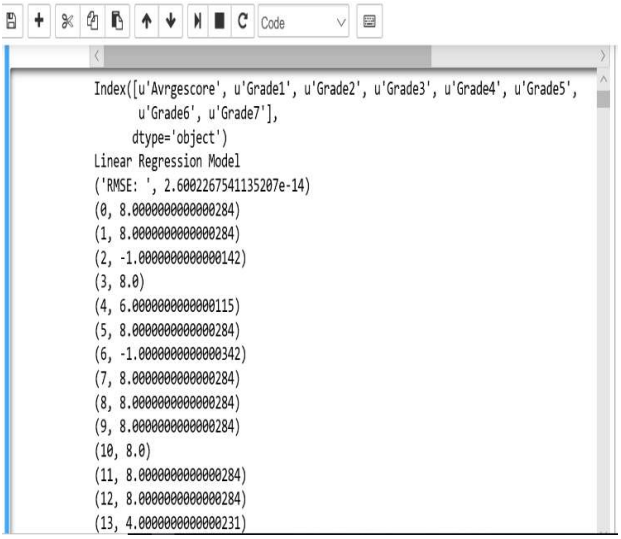


Fig. 2. Linear Regression Model

The Linear regression model is best used when modeling problems that are continuous as opposed to other classification models which are best for modeling discrete problems. Thus, the reason for the decimal points in the predicted result is because of the continuous nature of the Linear Regression Model.

The accuracies of the four different classification models were recorded as depicted in Fig. 3. The results showed that Naïve Bayes Classifier and Support Vector Machine have an accuracy of approximately equal to 99.94% which outperformed Decision Tree and Kth-Nearest Neighbor algorithms with an accuracy of 98.10% and 99.87% respectively. However, the result from Kth-Nearest Neighbor is not too far from the SVM and NB as the difference is just about 0.07%.

Generally speaking, all the classification models employed in this research work have shown how our model is well built. This is achieved based on the dimensionality reduction of the dataset which consequently contributed to reducing the time of the training of the model. This is not limited to that but also helped to avoid overfitting of the model.

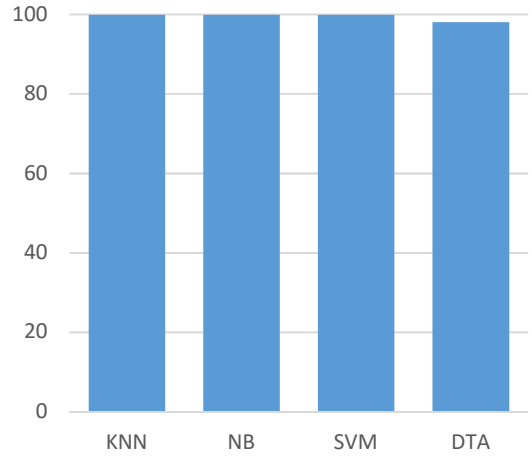


Fig. 3. Accuracy of Classifiers

## VII. Conclusion and Future Direction

Apart of Linear Regression Model which has RMSE of almost zero ($\mathbf{2.614\ e^{-14}}$), four (4) different machine learning classification models were applied on the real data set and it found out that each of the classifiers proved to be efficient in predicting the course allocation as none of them has achieved less than 90% accuracy. The NB and SVM were found to have highest recommendation accuracy among all the classifiers, where both have 99.94%, then followed by KNN with an accuracy of 99.87%, while DTA being the least but still with reasonable accuracy of 98.01%.

This research can serve as a framework for further studies on the admission recommender system by adding more features (especially increasing the O'level results, catchment areas where applicable, among others) and more data.

## References


[1] A. Muhammad, "Undergraduate Students Project Supervision Recommendation System," Federal University, Dutsen-Ma, Katsina, 2017.

[2] A. S. Bulkachuwa, "Availability and Utilization of Microteaching Laboratories and their Influence on Students' Teaching Practice Performances in Colleges of Education,," Bayero University, Kano, Kano, 2017.

[3] J. Lu, D. Wu, M. Mao, W. Wang and G. Zhang, "Recommender System Application Developments: A Survey," Decision Support Systems, vol. 74, no. C, pp. 12-32, 1 July 2015.

[4] A. H. M. Ragab, A. F. S. Mashat and A. M. Khedra, "Design and Implementation of a Hybrid Recommender System for Predicting College Admission," International Journal of Computer Information Systems and Industrial Management Applications, pp. 35-44, 2014.

[5] S. Fong and R. P. Biuk-Aghai, "An Automated University Admission Recommender System for Secondary School Students," in The 6th International Conference on Information Technology and Applications (ICITA), China, 2009.

[6] A. A. Balarabe, Interviewee, Enquiry on the Current University Admission Process. [Interview]. 5 August 2019.

[7] JAMB, "exam/jamb-cut-off-mark/," 20 August 2019. [Online]. Available: https://www.currentschoolnews.com/exam/jamb-cut-off-mark/.

[8] F. O. Isinkaye, Y. O. Folajimi and B. A. Ojokoh, "Recommendation Systems: Principles, Methods and Evaluation," gyptian Informatics Journal, p. 261–273, 2015.

[9] R. Swaminathan, J. M. Gnanasekaran and S. Krishnakumar, "Recommender System for Graduate Studies in USA Suresh," 31 12 2015. [Online]. Available: https://www.semanticscholar.org/paper/Recommender-System-for-Graduate-Studies-in-USA-Suresh. [Accessed 12 12 2019].

[10] M. Hasan, S. Ahmed, D. M. Abdullah and M. S. Rahman, "Graduate School Recommender System: Assisting Admission Seekers to Apply for Graduate Studies in Appropriate Graduate Schools," in International Conference on Informatics, Electronics and Vision (ICIEV), Bangladesh, 2016.

[11] K. Wakil, B. Akram, N. Kamal and A. Safi, "Web Recommender System for Private Universities' Admission in Iraq:UHD Case Study," International Journal of e-Education, e-Business, e-Management and e-Learning, pp. 329-340, 2014.

[12] F. Jauro, B. B. Alhassan, G. Aliyu and S. M. Toro, "Predicting child delivery mode using data mining by considering maternal," Science Forum (Journal of Pure and Applied Sciences), no. 18, pp. 31-35, 2019.

[13] S. Fong and R. P. Biuk-Aghai, "An Automated University Admission Recommender System for Secondary School Students," in in The 6th International Conference on Information Technology and Applications (ICITA), China, 2009.