International Journal of Computer Applications (0975 - 8887)
Volume 133 - No.2, January 2016

An Admission Decision Support System for Nigerian Universities

Audu Musa Mabu
Department of Computer Science, Yobe State University
KM7 Gujba Road, Damaturu
Yobe State, Nigeria

Farouq Aliyu Muhammad
Department of Computer Science, Yobe State University
KM7 Gujba Road, Damaturu
Yobe State, Nigeria

ABSTRACT

Data mining is an essential step in the process of knowledge discovery. It is a process that analyses large observational data sets to find relationships within them and to summarize this data into useful information. In this paper, a Decision Support System (DSS) is developed using an Iterative Dichotomizer (ID3) algorithm. The system is designed to help Nigerian universities in enrolling students. The proposed system will increase the accuracy and speed of their admission system. The test and evaluation the system has shown promising results with an Accuracy factor of approximately 92%.

General Terms

Artificial Intelligence, Data Mining

Keywords

Decision Support System; Decision tree; Classification; ID3 algorithm

1. INTRODUCTION

Education is the source of civilization and national development. Therefore, it needs to be standardized. One of the ways this standardization can be achieved is by admitting students that best fit university system. Lack of proper admission process leads to low quality education and research. For example, Nigeria has the largest universities in Sub-Saharan Africa. However, there are only 15 engineers and scientists per million persons that carry out research in Nigeria.[22] Recently, the number of applicants seeking admission into Nigerian universities increases exponentially.[8] But they lack adequate facilities to meet the admission challenges.

Currently, candidates submit their applications to Joint Admission Matriculation Board (JAMB).[1] The application consists of the candidates: Date of birth. State of origin. Unified Tertiary Matriculation Exam (UTME) score, which is similar to Scholastic Aptitude Test (SAT) exams in the United States. For those students who have undergone some kind of tertiary education, the final results of their studies is submitted in lieu of UTME, this is known as Direct Entry (DE). Finally, the candidate must also submit a secondary school certificate issued by West African Examinations Council (WAEC), National Examinations Council (NECO) or National Business and Technical Examination Board (NABTEB).[15] JAMB submits the aforementioned information to four institutions selected by the candidate.

At the tertiary institutions, an admission committee manually evaluates every candidate's data against a set of admission requirements. Table 1 shows a typical Nigerian university admission requirement. The committee produces a list of candidates eligible for admission. Another committee verifies the list.

These procedures are too cumbersome, time consuming, prone to lots of human errors and irregularities. Many qualified candidates may lose their admissions to not qualified candidates or they may be assigned to courses they do not deserve.[28] This necessitates automation of the system.

In this paper, a supervised learning system is implemented. The system uses Iterative Dichtomizer (ID3) algorithm to classify the candidates into admitted and non-admitted groups. The classification is based on the criteria mentioned in Table 1. The problems of admission exercise will be a thing of the past if the proposed system is adopted. The remaining sections of this paper are as follow: Section 2 presents detailed review of systems developed by other researchers. Section 3 describes the proposed system. Section 5 evaluates the performance of the proposed system. Finally, Section 6 concludes the paper.

2. LITERATURE REVIEW

Currently, universities tend to use a cocktail of Enterprise Resource Planning (ERP) applications and Intelligent Decision Support Systems (IDSS) in order to arrive at the candidates that will be admitted.[20] This technique is prone to errors that may arise during switching from one application to the other. Hence a single application capable of supporting the universities decide on candidates to admit is necessary.

Several attempts have been made by researchers to develop intelligent decision support systems that can help in university admission processes. Muhammad *et al.* [3] developed a rule based Decision Support System that helps students in selecting a suitable faculty during admission exercise. The application is developed using CLIPS(C Language Integrated Production System) programming language. Moreover, the system consists of a general rule-base and an inference engine. The rule-base engine generates rules base on a training set. These rules are retrieved by the inference engine and are used to solve new problems. The system uses "IF-THEN" conditions which reduces the accuracy of prediction.

Table 1. Typical criteria for admission into Nigerian Universities

S/N	Attributes	ATTTRIBUTE VALUES
1	Mode of Entry	Type of Application
		UTME
		DE
2	UTME Score	Student's Aggregate Score in UTME
		Above 179
		Below 180
3	Pre-Degree	If student has Attended Pre-Degree Program
		Yes
		No
4	Pre-Degree Points	Student's Grade for Pre-Degree Programme (If Pre-Degree Mode)
		Above 2.49
		Below 2.50
5	DE points	Student's A-Level Grade (If DE Mode)
		Above 2.49
		Below 2.50
6	Post UTME/ Post DE Score (Entry Exams)	Student's Marks obtained in Post UTME/DE
		Above 39
		Below 40
7	No. of O-Level Credits	Number of O-Level Credits in not more than two sittings
		Above 4
		Below 5
8	Catchment Area	If student is with catchment area (If Nigerian) otherwise Nationality
		Yes
		No
9	Admission Status	Admission Status of the Applicant
		Admitted
		Not Admitted

Tanna,[27] developed a decision support system that helps candidates decide which field of engineering suit them. The system takes mean of current and previous Common Entrance Tests (CET) scores of student who applied for that field. The mean is then compared with the candidates score in order classify him as Best bargain, Ambitious or Safe with respect to that field. The author argues that mean serves as a feedback to the system. However, using mean provides a linear feedback which takes the system long time to self-correct errors. In other words, arithmetic mean is very sensitive. Therefore,few large values in a sample may lead to the wrong prediction.[12] Moreover, machine learning techniques a generally better predictors than statistical techniques.[25] [24] [10]

Oladokun *et al.*,[16] pointed out that students ability to study a certain course in a certain university is composed of linguistic variable and therefore it can be solved predicted using fuzzy logic. Salimi *et al.* [23] developed an Multi-criteria Decision Making (MCDM) system that allows firms select the best university for research collaboration. The system is based on Fuzzy Analytic Hierarchy Process (FAHP). The authors used Bellman and Zadeh technique in order to correct the skewness and non-linearity of fuzzy numbers into account.[5] However, the system fails to account for the interdependencies between criteria.

Neural network analysis is another technique that can be used in situations where the predictor-outcome relationship is unknown.[9] [14] [17] Alenezi *et al.*,[2] developed a model that helps decision makers predict whether a candidate can pass premedical stage in medical schools. The authors believe that if a student passes premedical stage, there is a high chance that he can succeed in studying medicine. The model uses Cascade Correlation Network supervised learning algorithm which is trained using Quick-Propagation to predict a given students GPA (Grade Point Average). The authors were able to train their model with student data from 2003/2004 session and it was used to predict that performance of students in 2004/2005 session and 2005/2006 session. However, the model was able to predict 2005/2006 session more accurately. This inaccuracy is because neural networks require large and accurate dataset.[28]

Another machine learning technique that is nearly as accurate as neural networks yet it requires less dataset and proves itself easy to train is Decision Tree technique.[13] [26] Decision tree technique is widely used because of is interpret-ability, ease of use and strength of prediction.[11] Yadav *et al.*,[30] used three decision tree algorithms, namely ID3, C4.5 and CART (Classification And Regression Trees) to predict improvement in performance of engineering student. The authors found that C4.5 has the highest accuracy followed by ID3 algorithm, which is only 5.5% less accurate. In a similar experiment, ID3 gave the highest accuracy, while C4.5 came second with 13% less accuracy.[29] This is due to the fact that the latter experiment used smaller training set than the earlier. Therefore, ID3 algorithm is the most accurate when small dataset is used. Furthermore, it has the fastest execution time and the least inaccurate predictions in both experiments.

In this paper, a decision tree support system that helps Nigerian universities in screening students during admission process is developed using ID3 algorithm.[19] [7] The system is developed using

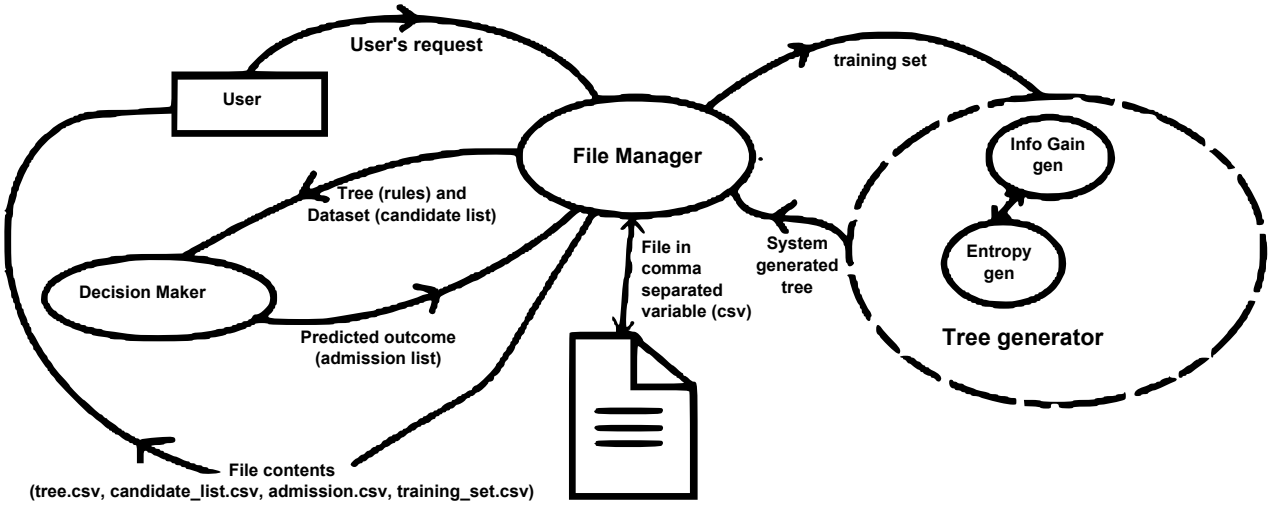

Fig. 1. Data Flow Diagram for the proposed system.

C++ programming language. This programming language is used in order to maximize flexibility.

3. PROPOSED SYSTEM

Figure 1 is a data flow diagram.[6]. It depicts the flow and transformation of data in the proposed system. From the figure, it can be seen that the system is divided into five (5) major components, these are;

(1) User

(2) File Manager

(3) Tree generator

(4) Decision Maker

(5) Comma Separated Variable (CSV) files

The user is an agent, either human or machine (another software) who uses the system. The user can view any generated information by directly opening the csv file generated.
File manager is a group of subroutines that collect data from a csv file given by the user and save them in a variable. Furthermore, the file manager collects the processed information in the form of decision tree or an admission list and presents it to the user in the form of a CSV (Comma Separated Variable) file. A CSV is a file format that uses comma as a delimiter. [21] It is compatible with most spreadsheet documents.

$$Entropy(S) = \sum_{s} -p(I)\log_2^p(I) \quad (1)$$

$$Gain(S,A) = Entropy(S) - \sum_{v}(\frac{|S_v|}{|S|} \times Entropy(S_v)) \quad (2)$$

Where:

S = Sample set

$\sum_{v}$ is each value v of all possible values of attribute A

S_v = subset of S for which attribute A has value v

$|S_v|$ = number of elements in S_v

$|S|$ = number of elements in S

Tree generator is an engine that generates decision trees from a given training set. It consists of two very important subprograms; Entropy generator and Information Gain generator. Entropy generator uses Equation 1 to calculate the entropy of any given value. Information Gain generator uses Equation 2 to calculate information gain of any given attribute. From Equation 1 and 2 is can be seen that the Information Gain generator is dependent on Entropy generator.
Decision Maker (DM) is an engine that deduces future out comes base on the generated decision tree. In other words, it generates admission list from candidates list base on the given decision tree.

4. EXPERIMENT

The training set (S) is collected from admission office of Yobe State University, Damaturu, Nigeria. The training set consists of some selected fields (see Table 1) which were obtained from candidate's application forms. The training set also contains the status of admission which tells the system whether the candidate has been admitted or not. From the training set, a classifier model is built by applying data mining techniques to enrollment data. The implemented model classifies the applicants based on applicant background at the time of admission. ID3 algorithm is used and the splitting crite-

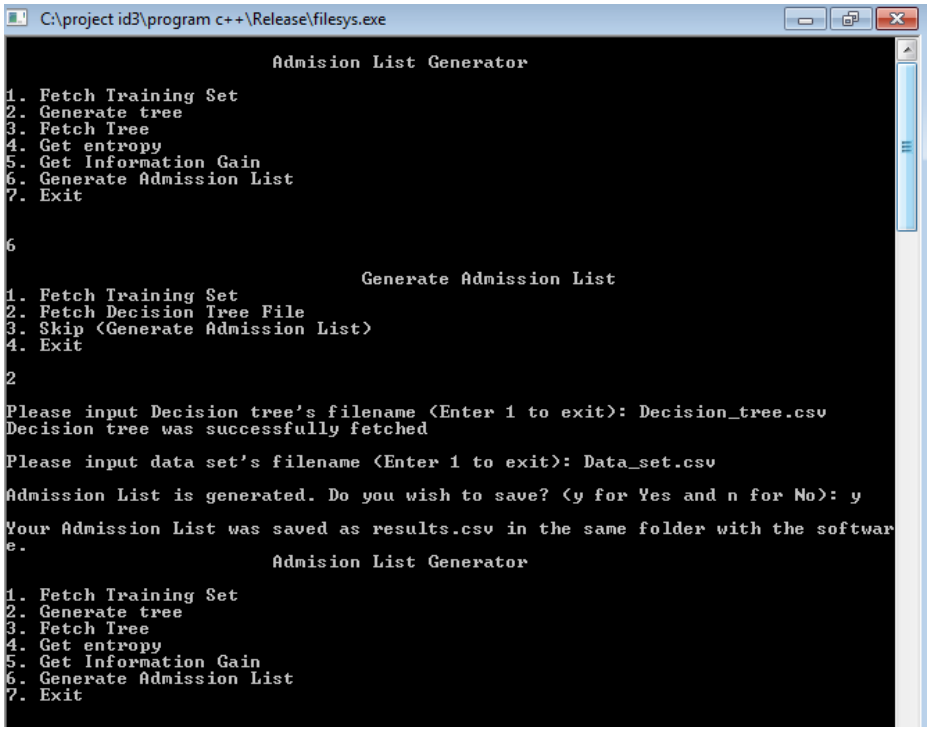

Fig. 2. Snapshot of the proposed system.

Table 2. Confusion matrix showing the results obtained from the proposed system.

	PREDICTED CLASS		
TRUE CLASS	ACCEPTED	REJECTED	
ACCEPTED	639	96	735
REJECTED	57	1085	1142
	696	1181	1877

ria used over the input space is entropy. The implementation of the system is done using C++ programming language. Figure 2 is the snapshot of the proposed system.

5. RESULTS

Table 5 is a confusion matrix for the performance of the proposed system.[4] It shows the confusion matrix for a two class classifier. From this table, other performance indicators are calculated:[18] Equation 3 shows the accuracy (ACC) of the system. It shows the total number of predictions those were correct. Equations 4 and 5 calculate the True Positive Rate (i.e. the proportion of positive cases that were correctly identified) and True Negative Rate (i.e. the proportion of negatives cases that were classified correctly) respectively. Equations 6 and 7 calculate the precision of the predictions. Finally, Equations 8 and 9 shows the F score which is a harmonic mean of precision and recall.

$$ACC = \frac{639 + 1085}{1877} = 0.915 \quad (3)$$

$$R_Accepted = \frac{639}{735} = 0.8694 \quad (4)$$

$$R_Rejected = \frac{1085}{1142} = 0.9501 \quad (5)$$

$$P_Accepted = \frac{639}{696} = 0.9182 \quad (6)$$

$$P_Rejected = \frac{1085}{1181} = 0.9187 \quad (7)$$

$$F_accepted = \frac{2 \times 0.9812 \times 0.8694}{0.9182 + 0.8694} \quad (8)$$

$$F_Rejected = \frac{2 \times 0.9187 \times 0.9501}{0.9187 + 0.9501} \quad (9)$$

From equations 3 through9 is can be deduced that the proposed classifier achieves a high recall at the cost of high precision. Furthermore, the same measures in case of Rejected class are about 0.9501 and 0.9187 for Recall and precision respectively. These level values show that the classifier's performance is above average. The F score shows very promising values that also reaffirmed that the classifier performance is of high-quality and hence can be used for sorting the admission.

In order to investigate the the accuracy of the proposed system with increase in complexity of the training set, another experiment is carried out. The involves a set of training sets generated base on simple boolean expression. The complexity is increased by increasing the number of variables in the boolean expression. Furthermore, the

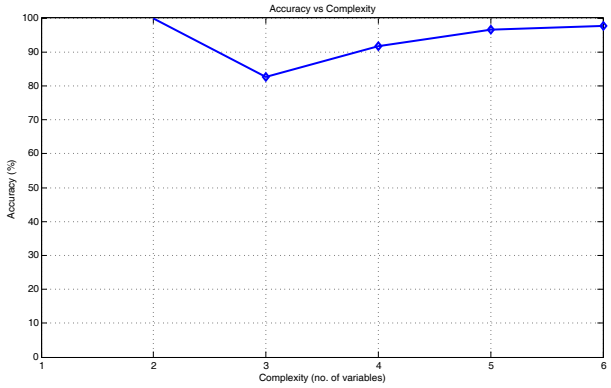

Fig. 3. Accuracy of the proposed system with complexity.

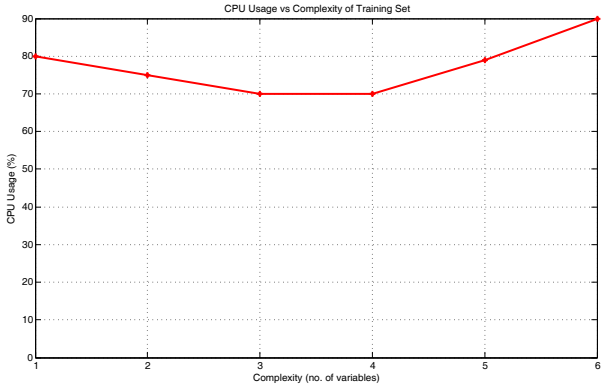

Fig. 4. CPU Usage with increase in complexity

load on the computer as it processes this data is also investigated through Visual Studio Performance Analysis tool.
Figure 3 and 4 show the accuracy and CPU usage with increase in complexity respectively. It can be seen that although the accuracy is high, the CPU usage is at 90%. This is due to the fact that the learning section of the proposed system is at a complexity of O^3.

6. CONCLUSION

This paper presents a Decision Tree Model for Nigerian University admission system. The proposed system learns the admission rules with high recall and precision. This allows the system to fish out the best candidates in a vast number of applicants. The system filters out the candidates such that the admission committee can focus on selecting the best candidates. As such, the workload on the committee as well as the administrative staff is reduced.
The learning section of the proposed system has a complexity of (O^3) which means the systems performance degrades exponentially with increase in training set size. In addition ID3 algorithm requires pre-processing of the training set and the data set. Hence, solution to this two key problem will be focused on in the future work.

7. REFERENCES

[1] Joint Admissions and Matriculation Board. Board history, 2015.

[2] Jawaher K Alenezi, Mohammed M Awny, and Maged MM Fahmy. Effectiveness of artificial neural networks in forecasting failure risk for pre-medical students. In *Computer Engineering & Systems, 2009. ICCES 2009. International Conference on*, pages 135–138. IEEE, 2009.

[3] Muhammad Zaheer Aslam, Abdur Rashid Khan, et al. A proposed decision support system/expert system for guiding fresh students in selecting a faculty in gomal university, pakistan. *arXiv preprint arXiv:1104.1678*, 2011.

[4] Gustavo EAPA Batista, Ronaldo C Prati, and Maria Carolina Monard. A study of the behavior of several methods for balancing machine learning training data. *ACM Sigkdd Explorations Newsletter*, 6(1):20–29, 2004.

[5] Richard E Bellman and Lotfi Asker Zadeh. Decision-making in a fuzzy environment. *Management science*, 17(4):B–141, 1970.

[6] Yu-Liu Chen et al. Data flow diagram. In *Modeling and Analysis of Enterprise and Information Systems*, pages 85–97. Springer, 2009.

[7] Andrew Colin. Building decision trees with the id3 algorithm, 1996.

[8] Diana-Abasi Ibanga Francis. Solving the problem of poor quality of university graduates in nigeria-a proposed holistic approach. *British Journal of Education*, 3(7):52–70, 2015.

[9] Julie M Byers González and Stephen L DesJardins. Artificial neural networks: A new approach to predicting application behavior. *Research in Higher Education*, 43(2):235–258, 2002.

[10] Zaidah Ibrahim and Daliela Rusli. Predicting students academic performance: comparing artificial neural network, decision tree and linear regression. In *21st Annual SAS Malaysia Forum, 5th September*, 2007.

[11] Willis A Jensen. Decision trees for business intelligence and data mining: Using sas® enterprise miner. *Technometrics*, 50(3):409–410, 2008.

[12] J.D. Kelleher, B.M. Namee, and A. D'Arcy. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*. MIT Press, 2015.

[13] S Kotsiantis, Christos Pierrakeas, and P Pintelas. Predicting students'performance in distance learning using machine learning techniques. *Applied Artificial Intelligence*, 18(5):411–426, 2004.

[14] Jing Luan. Data mining and its applications in higher education. *New directions for institutional research*, 2002(113):17–36, 2002.

[15] Toyin D Moyinoluwa et al. Analysing the psychometric properties of mathematics in public examinations in nigeria. *Research on Humanities and Social Sciences*, 5(7):24–30, 2015.

[16] VO Oladokun and DI Oyewole. A fuzzy inference based decision support system for solving the university-course admission choice problem. *International Journal of Computer Applications*, 112(3), 2015.

[17] Mukta Paliwal and Usha A Kumar. Neural networks and statistical techniques: A review of applications. *Expert systems with applications*, 36(1):2–17, 2009.

[18] David Martin Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. 2011.

[19] J.R. Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, 1986.

[20] Vohra Rajan and Narayan Das Nripendra. Intelligent decision support systems for admission management in higher education institutes. *International Journal of Artificial Intelligence & Applications (IJAIA)*, 2(4), 2011.

[21] Dominic J Repici. The comma separated value (csv) file format. *Creativyst Inc*, 2010.

[22] William Saint, Teresa A Hartnett, and Erich Strassner. Higher education in nigeria: A status report. *Higher education policy*, 16(3):259–281, 2003.

[23] Negin Salimi and Jafar Rezaei. Multi-criteria university selection: Formulation and implementation using a fuzzy ahp. *Journal of Systems Science and Systems Engineering*, pages 1–23, 2015.

[24] SJ Sheel, RS Renner, SK Dawsey, SC Conway, and CA Chico. Alternatives to math placement exams: a look at discriminant analysis, neural networks, and ensembles of networks. *Conway (SC): Department of Computer Science, Coastal Carolina University*, 2002.

[25] Stephen J Sheel, Deborah Vrooman, René S Renner, and Shanda K Dawsey. A comparison of neural networks and classical discriminant analysis in predicting students mathematics placement examination scores. In *Computational Science-ICCS 2001*, pages 952–957. Springer, 2001.

[26] Shan Suthaharan. Decision tree learning. In *Machine Learning Models and Algorithms for Big Data Classification*, pages 237–269. Springer, 2016.

[27] Miren Tanna. Decision support system for admission in engineering colleges based on entrance exam marks. *International Journal of Computer Applications*, 52(11):38–41, 2012.

[28] Franklin Wabwoba and Fullgence Mwakondo. Students selection for university course admission at the joint admissions board (kenya) using trained neural networks. *Journal of Information Technology Education: Research*, 10(1):333–347, 2011.

[29] Surjeet Kumar Yadav, Brijesh Bharadwaj, and Saurabh Pal. Data mining applications: A comparative study for predicting student's performance. *arXiv preprint arXiv:1202.4815*, 1(12), 2012.

[30] Surjeet Kumar Yadav and Saurabh Pal. Data mining: A prediction for performance improvement of engineering students using classification. *arXiv preprint arXiv:1203.3832*, 2(2), 2012.