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 Dichotomizer (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

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

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. Moreover, machine learning techniques are generally better predictors than statistical techniques. 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. 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. 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.[25] Oladokun et al.[16] 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. 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] The system is developed using
3. PROPOSED SYSTEM

Figure 1 is a data flow diagram. 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 a CSV (Comma Separated Variable) file. A CSV is a file format that uses comma as a delimiter. It is compatible with most spreadsheet documents.

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, it can be seen that the Information Gain generator is dependent on Entropy generator.

Decision Maker (DM) is an engine that deduces future outcomes based on the generated decision tree. In other words, it generates admission list from candidates list based 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-
5. RESULTS

Table 2 shows a confusion matrix for the performance of the proposed system. It shows the true positive rate (TPR) of the system. From this table, other performance indicators are calculated. Equation 1 calculates the accuracy (ACC) of the system. Equations 2 and 3 calculate the true positive rate (TPR) and true negative rate (TNR), respectively. Equation 4 calculates the precision (P) of the predictions. Finally, Equations 5 and 6 calculate the F score which is a harmonic mean of precision and recall.

From equations 1 through 6, it 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 reaffirm that the classifier performance is of high-quality and hence can be used for sorting the admission.

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

---

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

<table>
<thead>
<tr>
<th>TRUE CLASS</th>
<th>PREDICTED CLASS</th>
<th>ACC</th>
<th>RAccepted</th>
<th>Rejected</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCEPTED</td>
<td>ACC</td>
<td>0.915</td>
<td>0.8694</td>
<td>0.9501</td>
</tr>
<tr>
<td>REJECTED</td>
<td>REJECTED</td>
<td>0.9187</td>
<td>0.9501</td>
<td>0.9187</td>
</tr>
</tbody>
</table>
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


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

Fig. 4. CPU Usage with increase in complexity.


