MINING DISCIPLINARY RECORDS OF STUDENT WELFARE AND FORMATION OFFICE: AN EXPLORATORY STUDY TO ENHANCE THE UNIVERSITY SERVICES PORTFOLIO

Paulino H. Gatpandan, Shaneth C. Ambat

School of Graduate Studies, AMA University, Quezon City, Philippines

ABSTRACT

Data mining is the process of analyzing large datasets, understanding their patterns and discovering useful information from a large amount of data. Decision tree as one of the common algorithm of data mining is a tree structure entailing of internal and terminal nodes which process the data to eventually produce a classification. Classification is the process of dividing a dataset together in a high-class set such that the members of each set are nearby as expected to one another, and different groups are as far as expected from one another, where distance is measured with respect to the specific variable(s) you are trying to predict. Data Envelopment Analysis is a technique wherein the productivity of a unit is evaluated by equating the volume/amount of output(s) in relation to the volume/amount of input(s) used. The performance of a unit is calculated by equating its efficiency with the best-perceived performance in the data set. In this study, a model for measuring the efficiency of Decision Making Units will be presented, along with related methods of implementation and interpretation. DEA assesses and evaluates the efficiency of a unit dubbed as Decision-Making Units or DMU. There are many classification techniques and algorithms but the research study used decision tree using CHAID algorithms. Classification decision tree algorithm using CHAID as data mining technique identifies the relationship between the demographic profile of the students and the category of offenses. Cross tabulation is a tool used to analyze categorical data. It is a type of table in a matrix format that shows the multivariate occurrence dissemination of the variables and delivers a basic picture of the interrelation between two variables. Both CHAID algorithm and cross tabulation obtained the same results implying that higher percentage of students commit minor offenses regardless of college, gender, year level, month and course. The CHAID algorithm used in a software application Student Offenses Remediation System (STORES) serves as remediation plan for the university. Further studies should be conducted to identify the effectiveness of the remediation plan by conducting an empirical investigation on the rule set and/or implement another algorithm to determine the program efficiency.

KEYWORDS

Data Mining, Classification, Decision Tree, CHAID Algorithm, Data Envelopment Analysis

1. INTRODUCTION

Data mining is the process of analyzing large datasets, understanding their patterns [1] and discovering useful information from a large amount of data [2]. The common data mining algorithms include a) decision tree method, b) bionic global optimized genetic algorithm and neural network, c) statistical analysis and reports of the exclusive counter example method, among others.

The decision tree is a tree structure entailing of internal and terminal nodes which process the data to eventually produce a classification. The decision tree is capable of preserving the role of dimensionality reduction and at establishing an optimum size classification.
Classification is one of the most common applications for data mining. Classification is the process of dividing a dataset together in high-class sets such that the members of each set are nearby as expected to one another, and different groups are as far as expected from one another, where distance is measured with respect to the specific variable(s) you are trying to predict.

Data Envelopment Analysis is a state of the art benchmarking method which is of great value for multi-criteria benchmarking studies. Efficiency is the ratio between the outputs produced with the number of inputs used. In the Handbook of data envelopment analysis of [3] data envelopment analysis is a data-oriented approach for assessing and evaluating the performance of a set of peer objects called Decision Making Units (DMUs), which transform multiple inputs into multiple outputs. In most recent researches, DEA was used in evaluating the performances of different kinds of entities engaged in different activities such as hospitals, cities, courts, universities, business firms, and others. Several scholarly works used the DEA as measuring efficiency performance of an organization.

Every university maintains academic enablers to emphasize the motivational and cognitive factors for student academic success. In other universities, the academic affairs and student affairs was combined into one reporting unit because of the twofold strengths, it enables student affairs personnel to become familiar with the priorities of the academic division, and it allows them to develop closer relationships with faculty in order to more effectively help students to learn in a holistic and coordinated manner [4].

The Student Welfare and Formation Office (SWAFO) is an entity responsible of carrying out of student discipline policies, rules and code of practice as specified in the Student Handbook. The office is currently facing in a volume of students committing an offense from different colleges.

In this study, DEA will be used to assess the efficiency of the seven (7) colleges to determine the most efficient college which will serve as a benchmark for the inefficient colleges. One of the variables to be considered to determine the efficiency of a certain college is the minimal number of students who commit offenses.

2. **LITERATURE REVIEW**

Charnes, Cooper, Lewin & Seiford [5] introduced Data Envelopment Analysis in their work *data envelopment analysis theory*. DEA has been used for measuring and assessing the relative performances of a set of organizations called DMUs which use a variety of inputs to produce a variation of common output.

Several scholarly works have been identified and done in measuring the efficiency in different fields such as banking industry, education institutions, farming and agricultures, crimes among many others using data envelopment analysis.

Walraven, Koning, Bijmolt & Los [6] proposed data envelopment analysis as a method for benchmarking sponsorship efficiency and demonstrate its effectiveness by applying it to a sample of seventy-two (72) major Dutch sports sponsorship projects. Baran, Wysokinski, Stas, Samolejova & Lenort used DEA model to identify the metallurgical efficiency such as overall technical efficiency, pure technical efficiency and scale efficiency of branches in Poland. Dzemydaite, Dzemyda, & Galiniene [8] measured the efficiency level of innovation systems by applying a nonparametric DEA in Eastern and Central Europe. Tran Thi Thu & Bhayat [9] measured the technical and scale efficiency of Vietnamese Commercial banks using a nonparametric DEA. CCR model was applied to measure the overall technical efficiency. BCC model was applied to evaluate the managerial skills. Dugulea & Dugulea [10] used DEA to determine the technical efficiency of academic departments at Transylvania University during the
academic year 2014 – 2015. The results of DEA model was used as a basis for improving the efficiency of academic departments. Kashim, Kashim, Nadhar Khan, Rahim, & Hassan [11] suggested the best framework for measuring the productivity in Higher Learning Institutions (HLI). Data Envelopment Analysis model was used to analyze efficiency, effectiveness, and productivity of the Higher Learning Institutions. Malhotra, Poteau, & Fritz [12] compared the DMUs organizations’ performance of thirteen thrifts and mortgage finance companies for the year 2008 to 2011 to evaluate the relative efficiency of the firms in terms of financial strengths. Luna, Gil-Garcia, Luna-Reyes, Sandoval-Almazan, & Duarte-Valle [13] proposed a study that assesses the efficiency performance of electronic portals to fully understand the factors that influence the quality of the information and services provided to businesses, citizens and other stakeholders. The study used DEA to calculate how the government’s efficiency in terms of inputs to produce high-quality e-government portals. Dobrea, Ciocoiu, & Dinu [14] applied an output oriented BCC DEA model to measure the sustainable and balanced development in the investment of the regional efficiencies in the 42 countries panel data from the year 2005 to 2010. Dočekalova, & Bočkova [15] utilized the application of the DEA on Research and Development (R&D) to measure the effectiveness of Czech manufacturing industry. Elshamy [16] utilized DEA to investigate and estimate the technical efficiency of Egyptian manufacturing sector. Vierstraete [17] used a nonparametric DEA method to assess the three factors that measure the health, education, and standard of living of the population. Estrada, Song, Kim, Namn, Kang [18] developed a method of stepwise benchmarking for inefficient DMUs based on proximity-based target selection. DEA was used to determine the efficiency score of DMU. Adams [19] computed and evaluate the technical efficiency of public school districts in the State of Arkansas using an output oriented variable returns to scale data envelopment analysis.

Hou, Guthrie, & Rigby [20] theorized that remediation has grown up from developing notions into an extensively accepted new institutional custom to comprise of an activity correct shortages, and include a process for outlining deficiencies and providing resources for improvement towards acceptable performance [21]. The remediation plan is a valuable learning tool that refers to a process of providing planned course of actions in correcting the deficiency or an act of offering improvements. A study using a remediation plan were identified in the scholarly works of [22] [23] [24] and [21]. Mee & Schreiner [22] summarized key findings of the literature about remediation strategies used in nursing programs and policies that impact the student outcomes. Gajewski, & Mather [23] presented an overview of a course-based remediation model developed to enhance student learning in order to increase the rate of student retention. Makhani, Bradley, Wong, Krynski, Jarvis, & Szumacher [24] reviewed remediation in allied health and professional programs. Remediation requires multiple assessors and several assessment tools, feedback and reassessment, and proactive involvement in supporting the identified students. Wu, Siewert, & Boiselle [21] developed a multifaceted approach to early identification and prompt remediation of difficulties to resident evaluation. The study designed a comprehensive remediation program which used resources within radiology department.

A decision tree is a data mining technique for presenting the classifiers and regressions model. It consists of some nodes and branches. In the classifying decision tree, the leaves specify classes. Decision tree describes all content alone independent of an expert to interpret the output. The graph method may be simpler than other classifying methods due to explanation and analysis. On the other hand, big numbers of nodes may create decision tree graphic representation as harder. Several scholarly works were identified using a classification data mining technique. Abbaszadeh Afshar, Ayoubi, Besalatpour, Khademi, & Castrignano [25] estimated the soil clay content in two depths using the geophysical technique (Ground Penetration Radar and Electromagnetic induction) and ancillary variables (remote sensing and topographic data) in an arid region in Southeastern Iran. CHAID model showed greater potential in predicting soil clay content from geophysical and ancillary data while traditional regression model did not perform as well. Eyduran, Keskin, Erturk, Dag, Tatliyer, Tirink, & Tariq [26] used CHAID algorithm to obtain a
more flexible prediction of fleece weight from some physical wool characteristics of sheep which provides a tree-based decision. Kurt, Duru, Canbay, & Duru. [27] applied a classification data mining to predict the magnetic susceptible of the soil. Jimenez-Perez, & Mora-Lopez [28] applied and utilized a two data mining techniques using clustering and classification techniques for the predicting hourly values of global radiation for the next day. Mistikoglu, Gerek, Erdis, Mumtaz Usmen, Cakan, & Kazan [29] demonstrated the utility of decision tree method in fall accidents experienced by roofer. Algorithms such as C5.0 and CHAID were used in the study to produce decision trees and to mine rules that demonstrate the associations among the input and output variables for roofer fall accidents. Zhang, Fu, Peng, & Li [30] investigated the classification of developing and non-developing tropical disturbances in the Western North Pacific of a decision tree using a C4.5 algorithm. Althuwaynee, Pradhan, Park, & Lee [31] used CHAID method to perform the best classification fit for each conditioning factors and combined it with logistic regression to find the corresponding coefficient of best fitting function. The study showed the efficacy and consistency of the collaboration DT and LR model in conquest vulnerability plotting. Miller, Fridline, Pei-Yang, & Marino [32] generated a model for the early detection of Metabolic Syndrome (MetS) in young adults. The model was derived using CHAID decision tree analysis with waist circumference user-specified to detect MetS in young adults using records of the National Health and Nutrition Examination Survey (NHANES) year 2009-2010. Cho & Kurup [33] utilized a decision tree approach using two different tree models such as C4.5 and CART for the classification and dimensionality reduction of Electronic Nose (EN). Kaur, Singh, Garg & Harmanpreeet [34] studied about classification algorithm for farm Decision Support System. Several classification algorithms used are limited search, ID3, CHAID, C4.5, improved C4.5 and One VS all decision tree and applied to one common data set of the crop with specified class.

3. METHODOLOGY

The study is a quantitative method of research that uses the statistical method to quantify and analyze the data to generalize results from a sample population. It appears the discussion in tables containing data in the form of numbers and statistics. During the development of the remediation plan, the proponent will use the framework for remediation plan phases.

3.1. RESEARCH DESIGN

Data Envelopment Analysis is a technique wherein the productivity of a unit is evaluated by equating the volume/amount of output(s) in relation to the volume/amount of input(s) used. The performance of a unit is calculated by equating its efficiency with the best-perceived performance in the data set. In this study, a model for efficiency measurement of Decision Making Units will be showed, along with related methods of implementation and interpretation. DEA assesses and evaluates the efficiency of a unit dubbed as Decision Making Units or DMUs using the illustrations below:

\[
\text{Performance} = \frac{\text{virtual output}}{\text{virtual input}} = \frac{u_1 y_{1o} + \ldots + u_s y_{so}}{v_1 x_{1o} + \ldots + v_m x_{mo}} \quad \text{Eq. (1)}
\]

where:

- \(x\) and \(y\) are resp. the input and output vectors and
- \(u_s\) output's weight, \(v_m\) output's weight.

There are several classification techniques and algorithms but the research study used decision tree using CHAID algorithms. In the scholarly works of [32], decision tree analysis can visualize
the relationship pathways between the binary object variable and the associated uninterrupted and/or categorical predictor variables with a tree image.

Other scholarly works of [31] stated that Multivariate statistical CHAID method is a tree that has branches which represent the predictors and discriminate sample groups that produced an estimated significant chi-square level of interaction.

The proponents opted to use classification decision tree algorithm using CHAID as data mining technique to identify the relationship between the demographic profile of the students and the category of offenses. The methodology that the proponents used for the Remediation Plan process is shown below.

![Remediation Plan Process](image-url)

**Figure 1.0 Remediation Plan Process**

The process is divided into three (3) phases namely: identify, analyze, and mitigate. Identification phase involves gathers pertinent information of students’ profile, offenses committed, and complaints. Alerts will be identified based on the number of recurring offenses. Analysis phase involves set of rules or algorithm using the demographic method to filter and extracts data profiling and classify results. The mitigation phase responds and remediates a planned course of action as a student remediation program.

The use of efficient, effective, and precise remediation plan techniques is very important for a system to provide a useful course plan of actions such as a student remediation program.

![Remediation Process Diagram](image-url)

**Figure 2.0 Remediation Process Diagram**

Demographic Filtering techniques classify the information into a demographic category based on student profile details and offense category.
3.2 Other Tools Used in the Research Study

MaxDEA software was used to obtain measures of productivity and efficiency to conduct data envelopment analysis.

SPSS was used to generate the decision tree for classification model using CHAID algorithm.

Microsoft.net 4.5 was used as the environment framework, Visual Studio 2013 as the integrated development environment, Visual C# as programming language, ASP.NET mv5 as development framework for single page application in the web-based environment, SQL Server 2012 for data store/persistence, Internet Information Services as the web server, jQuery 1.10 for JavaScript framework, jQuery EasyUI for front-end user interface framework, DotNet.Highcharts for charting and visualization, Entity Framework 6 for domain entities/models and Dapper Micro-ORM for access.

4. Results and Discussion

The efficient colleges were determined through Data Envelopment Analysis. The proponents used the model of Radial Measure of Efficiency using an input-oriented orientation and a Rate-to-Scale (RTS) using Scale efficiency. The proponents also used another model to test the efficiency of the colleges such as Charnes, Cooper, and Rhodes (CCR) model using input-oriented and output-oriented and Hybrid Measure of Efficiency model and NonRadial Measure of Efficiency model using an input-oriented orientation and a Rate-to-Scale (RTS) using constant returns to scale (CRS). The three models mentioned obtained the same results.

Table 1.0: The variables used in determining the efficient and inefficient Colleges

<table>
<thead>
<tr>
<th>DMUs</th>
<th>Inputs</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSCS, CLAC, CTHM, COEd, CCJE, CEAT, CBAA</td>
<td>Budget, Personnel, Radio, Computer, Laptop, Average Population</td>
<td>Number of Students without Offenses, Number of Students with Offenses</td>
</tr>
</tbody>
</table>

Table 1.0 presents the seven (7) colleges as Decision-Making Units (DMUs), inputs such as budget, personnel, radio, computer, laptop, average population, the number of students with offenses, and the output which is the number of students without an offense.

Table 2.0: 3 Years Scale Efficiency Results of the Seven (7) Colleges

<table>
<thead>
<tr>
<th>DMU</th>
<th>2013 Technical Efficiency Score</th>
<th>2013 Scale Efficiency Score</th>
<th>2014 Technical Efficiency Score</th>
<th>2014 Scale Efficiency Score</th>
<th>2015 Technical Efficiency Score</th>
<th>2015 Scale Efficiency Score</th>
<th>Average Efficiency Score</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSCS</td>
<td>0.92</td>
<td>0.92</td>
<td>1.00</td>
<td>1.00</td>
<td>0.81</td>
<td>1.00</td>
<td>0.94</td>
<td>2</td>
</tr>
<tr>
<td>CLAC</td>
<td>1.00</td>
<td>1.00</td>
<td>0.87</td>
<td>0.87</td>
<td>1.00</td>
<td>0.87</td>
<td>0.93</td>
<td>3</td>
</tr>
<tr>
<td>CTHM</td>
<td>0.75</td>
<td>0.75</td>
<td>0.69</td>
<td>0.69</td>
<td>0.72</td>
<td>0.69</td>
<td>0.71</td>
<td>6</td>
</tr>
<tr>
<td>COEd</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1</td>
</tr>
<tr>
<td>CCJE</td>
<td>1.00</td>
<td>1.00</td>
<td>0.87</td>
<td>0.87</td>
<td>0.77</td>
<td>0.87</td>
<td>0.89</td>
<td>5</td>
</tr>
<tr>
<td>CEAT</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1</td>
</tr>
<tr>
<td>CBAA</td>
<td>0.97</td>
<td>0.97</td>
<td>0.89</td>
<td>0.89</td>
<td>0.93</td>
<td>0.89</td>
<td>0.92</td>
<td>4</td>
</tr>
</tbody>
</table>
Table 2.0 presents the 3-year Scale Efficiency Results of the seven (7) colleges. CTHM got an average efficiency score of 0.71 and ranked 6th being the lowest among all colleges. CCJE got an average efficiency score of 0.89 and ranked 5th inefficient DMU. CBAA got an average efficiency score of 0.92 and ranked 4th inefficient DMU. CLAC got an average efficiency score of 0.93 and ranked 3rd inefficient DMU. CSCS got an average efficiency score of 0.94 and ranked 2nd inefficient DMU. COEd and CEAT got an average efficiency score of 1.00 for three (3) years and ranked 1st being the efficient DMU. COEd and CEAT was determined as efficient DMUs due to less number of students without and offense in relation to average population.

Table 3.0: Average Population, Students with Violation and Percentage of Students with Violation for 3 years

<table>
<thead>
<tr>
<th>School Year</th>
<th>1st Semester</th>
<th>2nd Semester</th>
<th>Average Population</th>
<th>Students with Violation</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012 - 2013</td>
<td>12,864</td>
<td>11,936</td>
<td>12400</td>
<td>6921</td>
<td>55.81%</td>
</tr>
<tr>
<td>2013 - 2014</td>
<td>13,208</td>
<td>12,301</td>
<td>12755</td>
<td>5236</td>
<td>41.05%</td>
</tr>
<tr>
<td>2014 - 2015</td>
<td>13,744</td>
<td>12,977</td>
<td>13361</td>
<td>6065</td>
<td>45.40%</td>
</tr>
</tbody>
</table>

Table 3.0 presents the average population, number of students with violation and percentage of students with the violation. SY 2012-2013 has 6921 students with violation equivalent to 55.81%. SY 2013-2014 has 5236 students with violation equivalent to 41.05%. SY 2014-2015 has 6065 students with violation equivalent to 45.40%.

4.1 CROSS TABULATION

Cross tabulation is a tool used to analyze categorical data. It is a type of table in the matrix format that shows the multivariate occurrence dissemination of the variables and delivers a basic picture of the interrelation between two variables.

Table 4.0: Case Processing Summary

<table>
<thead>
<tr>
<th></th>
<th>Valid</th>
<th>Missing</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Percent</td>
<td>N</td>
</tr>
<tr>
<td>College * Category_Offense</td>
<td>6065</td>
<td>100.0%</td>
<td>0</td>
</tr>
<tr>
<td>Gender * Category_Offense</td>
<td>6065</td>
<td>100.0%</td>
<td>0</td>
</tr>
<tr>
<td>Year_Level * Category_Offense</td>
<td>6065</td>
<td>100.0%</td>
<td>0</td>
</tr>
<tr>
<td>Course * Category_Offense</td>
<td>6065</td>
<td>100.0%</td>
<td>0</td>
</tr>
<tr>
<td>Month * Category_Offense</td>
<td>6065</td>
<td>100.0%</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4.0 presents the case processing summary. The result of cross tabulation shows 100% valid data and 0% missing data.

4.2 CHAID ALGORITHM

The CHAID (Chi Square Automatic Interaction Detection) is a form of analysis that delineates how variables best combine to explain the outcome in a given dependent variable. CHAID
analysis builds a predictive model, or tree, to help define how variables best unify to describe the outcome in the given dependent variable. CHAID generates all potential cross tabulations for each predictor category until the greatest result is attained and no further splitting can be accomplished.

**CHAID Rule Set**

/* Node 1 */.
IF (Month = "April")
THEN
Node = 1
Prediction = 4
Probability = 0.888608

/* Node 2 */.
IF (Month != "April" AND Month != "May" AND Month != "Dec" AND Month != "Nov"
AND Month != "Oct" AND Month != "August" AND Month != "June" AND Month !=
"March" AND Month != "Sept" AND Month != "Jan")
THEN
Node = 2
Prediction = 4
Probability = 0.964831

/* Node 16 */.
IF (Month = "May") AND ((Age <= 17)) AND (Gender = "Female")
THEN
Node = 16
Prediction = 4
Probability = 0.686747

/* Node 17 */.
IF (Month = "May") AND ((Age <= 17)) AND (Gender != "Female")
THEN
Node = 17
Prediction = 4
Probability = 0.849057

/* Node 8 */.
IF (Month = "May") AND (Age > 17)
THEN
Node = 8
Prediction = 4
Probability = 0.946578

/* Node 18 */.
IF (Month = "Dec" OR Month = "March") AND (Year_Level = "Fourth Year" OR Year_Level = "Second Year") AND (College != "CLAC" AND College != "CSCS" AND College != "CTHM")
THEN
Node = 18
Prediction = 4
Probability = 0.980000

/* Node 19 */.
IF (Month = "Dec" OR Month = "March") AND (Year_Level = "Fourth Year" OR Year_Level = "Second Year") AND (College = "CLAC" OR College = "CSCS" OR College = "CTHM")
THEN
Node = 19
Prediction = 4
Probability = 0.900943
/* Node 10 */.
IF (Month = "Dec" OR Month = "March") AND (Year_Level = "Third Year")
THEN
Node = 10
Prediction = 4
Probability = 0.907801
/* Node 11 */.
IF (Month = "Dec" OR Month = "March") AND (Year_Level != "Fourth Year" AND Year_Level != "Third Year" AND Year_Level != "Second Year")
THEN
Node = 11
Prediction = 4
Probability = 0.973684
/* Node 12 */.
Based on the rule set of CHAID algorithm using category offense as the dependent variable, prediction of node 1 is 4 referring to minor offense with a probability of 0.89; prediction of node 2 is 4 referring to minor offense with a probability of 0.96; prediction of node 16 is 4 referring to minor offense with a probability of 0.69; prediction of node 17 is 4 referring to minor offense with a probability of 0.85; prediction of node 8 is 4 referring to minor offense with a probability of 0.95; prediction of node 18 is 4 referring to minor offense with a probability of 0.98; prediction of node 19 is 4 referring to minor offense with a probability of 0.90; prediction of node 10 is 4 referring to minor offense with a probability of 0.91; prediction of node 11 is 4 referring to minor offense with a probability of 0.97; prediction of node 12 is 4 referring to minor offense with a probability of 0.91; prediction of node 20 is 4 referring to minor offense with a probability of 0.78; prediction of node 21 is 4 referring to minor offense with a probability of 0.93; prediction of node 14 is 4 referring to minor offense with a probability of 0.82; and prediction of node 15 is 4 referring to minor offense with a probability of 0.93.
referring to minor offense with a probability of 0.91. All the nodes have a prediction value of 4 referring to the minor offense. Both CHAID algorithm and cross tabulation obtained the same results implying that higher percentage of students commit minor offenses regardless of college, gender, year level, month and course.

4.3 REDEMPTION PLAN: STUDENT OFFENSES REMEDIATION SYSTEM (SfOReS)

The CHAID algorithm has been implemented in the software application for Student Offenses Remediation System (StOReS). The following figures are the sample screen shots of the system.

![Student Welfare and Formation Office](image)

**Student Welfare and Formation Office**

Ming Disciplinary Records to Enhance University Services Portfolio

---

**Log in**

Use a local account to log in.

- **Email**
- **Password**

- [ ] Remember me?
- **Log in**

[Register as a new user](image)

---

**Figure 4.0 Login Page**

Figure 4.0 shows the Login Page of the system. Registered users are asked for authentication thus entering a username and password for access.

![Complaint Facility](image)

**Complaint Facility**

**Figure 5.0 Complaint Facility**
Figure 5.0 shows the Complaint Facility of the system. Once the student commits an offense, this module serves as a data entry about student information and the committed offenses. The system automatically detects the category of the offense and its equivalent sanctions. The system also determines the remediation to be given to the student.

![Complaint Facility](image1)

**Figure 6.0 Dashboard**

Figure 6.0 shows the Dashboard of the system which presents the summary of the offenses committed by gender, course, college, and offense category.

![Dashboard](image2)

**Figure 7.0 Data Analytics and Report Visualization – By College**

![Data Analytics and Report Visualization](image3)
Figure 7.0 shows a detailed report visualization of the system which presents the offenses committed by college.

![Figure 7.0](image1.png)

**Figure 8.0** Data Analytics and Report Visualization – By Offense Category

Figure 8.0 shows a detailed report visualization of the system which presents the offenses committed by offense category.

![Figure 8.0](image2.png)

**Figure 9.0** Data Analytics and Report Visualization – By College per Course

![Figure 9.0](image3.png)
Figure 9.0 shows a detailed report visualization of the system which presents the offenses committed by college per course.

5. CONCLUSION

The purpose of the study is to conduct mining on disciplinary records of students to determine the pattern of their offenses using CHAID algorithm and cross tabulation. Based on the pattern derived, a remediation plan was developed to help the seven (7) colleges to lessen the number of student offenses. DEA was used to determine the efficiency of the seven (7) colleges that served as DMUs. The DMUs that has lesser number of students with the offense was determined efficient DMUs, hence efficient DMUs served as a model to inefficient DMUs to lessen the number of student offenses. Based on the three 3-year historical data, two (2) colleges was determined as efficient DMUs identifiably, the COEd and CEAT. The CHAID Algorithm was implemented in a software application for Student Offenses Remediation System (StOReS) which is the proponents recommended remediation plan for the university. The proponents highly recommend that the remediation plan should be implemented in the university. Further studies should be conducted to identify the effectiveness of the remediation plan by conducting an empirical investigation on the rule set and or implement other algorithms to determine the program efficiency.

REFERENCES


Authors

**MR. PAULINO H. GATPANDAN** is pursuing his degree of Doctor in Information Technology at AMA Computer University and currently writing his dissertation. He is a Master in Business Administration degree holder from Philippine Christian University-Dasmariñas, and a Master of Science in Computer Science degree holder from AMA Computer University-Makati. He completed his academic baccalaureate degree of Bachelor of Science in Computer Science from the Philippine Christian University-Dasmariñas, Cavite. He is a fulltime faculty of Computer Studies Department at De La Salle University-Dasmariñas. He has been in the academe in the teaching profession for 17 years. He is well-versed in Database Management leading to an IBM Certified Designer-Cognos 10 BI Reports, and IBM Academic Associate DB2-certified.

**DR. SHANETH C. AMBAT** received her Doctor of Philosophy in Engineering major in Information Technology in Hannam University, South Korea in 2009. She received her Bachelor of Science in Computer Science in 1995 and Master of Science in Computer Science in 2004 respectively at AMA Computer University, Makati City. She is currently the Dean of the School of Graduate Studies at AMA Computer University, Quezon City. Her research interest includes data envelopment analysis, data mining, SOM, reinforcement learning algorithm, and fuzzy logic.