

Multiple educational data mining approaches to discover patterns in university admissions for program prediction

Julius Cesar O. Mamaril¹, Melvin A. Ballera²

¹College of Computing Sciences, Pangasinan State University, Lingayen Pangasinan, Philippines

²Graduate Programs, Technological Institute of the Philippines, Manila, Philippines

Article Info

Article history:

Received Nov 25, 2021

Revised Feb 18, 2022

Accepted Mar 2, 2022

Keywords:

Attribute selection

Design algorithm

Educational data mining

Pattern discovery

Program prediction

Supervised learning

University admission

ABSTRACT

This paper presented the utilization of pattern discovery techniques by using multiple relationships and clustering educational data mining approaches to establish a knowledge base that will aid in the prediction of ideal college program selection and enrollment forecasting for incoming freshmen. Results show a significant level of accuracy in predicting college programs for students by mining two years of student college admission and graduation final grade scholastic records. The results of educational predictive data mining methods can be applied in improving the services of the admission department of an educational institution, particularly in its course alignment, student mentoring, admission forecast, marketing, and enrollment preparedness.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Julius Cesar O. Mamaril

College of Computing Sciences, Pangasinan State University

Alvear E, Poblacion, Lingayen, 2401 Pangasinan, Philippines

Email: jmamaril@psu.edu.ph

1. INTRODUCTION

The second half of the 20th century brought swift and unceasing growth in computing power and has had a significant brunt both in the practice of statistical science and computer technology [1]. These advancements in computer technology allowed databases to be capable of handling enormous and more complex datasets augmenting the power of data processing in the fields of artificial intelligence and data mining [2]. Undoubtedly, among the sectors that benefited from these developments in the education industry which now harness the promises of educational data mining due to the fact that the academy is a silent data warehouse that continuously and regularly upsizes its data collection almost every day for those educational institutions with computer systems or every semester for those schools who still employ the good old paper-based and manual records processing and management [3].

This paper is in preparation for the future research of the authors and crucial to the development plans of a custom-built educational data mining (EDM) for program prediction to be deployed in a local state university. Although numerous data mining software is currently available both open-source and proprietary in nature, the by can be observed in utilizing these data mining software: they all require high technical proficiency and experience to operate them which most university admission and student registration employees will fall short; they were developed to be general-purpose DM engines, thus, include more features that are beyond the need for EDM purposes; some have limitations in terms of the number of rows of records to be processed for free; they involve re-uploading/reprocessing of records for different prediction approach or appending records; and the majority cannot be added as a free third-party software module, library or service to a software development endeavor.

This paper sought to answer the by problems: i) There are only a few schools that offer customized data mining systems for admission purposes; ii) There are some studies that historical data never converged for preprocessing in data mining; iii) There are many attributes used, but it is difficult to extract effective attributes to formulate a dynamic rule-based approach in the admission process; iv) There are only a few studies that discover knowledge or pattern based on historical data that has not been used by the university and v) Not all forecasting met the university's expectations.

To bridge the gap and mitigate the problems enumerated above and as has been mentioned awhile regarding the aim of the authors to custom-develop an EDM software, it is of tantamount importance to aspire and limit the following for the meantime in this paper: identify the data mining pre-processing techniques to be used with university admission historical records; enumerate which among university admission datasets are the significant attributes to be used in data mining which can be modeled predict an ideal program for an incoming college freshman; determine how relationship and clustering algorithms can be implemented to discover patterns in university historical records, and identify which patterns discovered can be a significant addition to the knowledge base for the prediction of a college program for freshmen students during university admissions.

2. RELATED WORKS

EDM is concerned with developing, researching, and applying computerized methods to detect patterns in large collections of educational data patterns that would otherwise be hard or impossible to analyze due to the enormous volume of data they exist within [4]. EDM pertains to the growing research and exploration in the sector of education that concerns primarily the combined use of statistics, machine learning, artificial intelligence, and data mining using historical records in the academe. Generally, this sector of education pursues to improve different data exploration methodologies by observing the different levels of significant hierarchies to uncover new information and understand how stakeholders in the educational institution learn, choose, decide, and improve themselves in the context of school settings [2]. Among the most common applications of EDM can be seen in intelligent tutoring systems, student behavior forecasting, university admission forecasting, and student graduation forecasting [3]. The selection of potential students while still in high school using educational historical records has been one of the major applications of EDM among first-world western countries [5].

Data mining is defined as the process of extracting useful information from vast amounts of data. The author noted that many other terms are being used to interpret data mining, such as knowledge mining from databases, knowledge extraction, data analysis, data archaeology, and data dredging [6]. There are four goals of EDM: i) Establishing a student model that integrates the measurement of the overall learning satisfaction of students, detailed student information and characteristics to predict future learning behavior; ii) Improving education domain frameworks to discover new methodologies to support learning and improve previous student models; iii) Investigating the results and causality of educational support through the implementation of electronic learning systems; iv) Expanding the breadth and reach of scientific information regarding the system of learning by implementing and improving student models through computer software [7].

The use of high school report cards, admission test results, and final GWA from historical records of graduate students as attributes for a rule-based classification technique that uses university admission policies for course/career alignment as data mining rules and looks for similar patterns to predict when students drop out of college or transfer to other schools [8]. Furthermore, historical data per course and department, as well as variables extracted from the admission form, are used as input parameters for a forecasting technique to predict the total number of "would-be" enrollees and assess preparation and enrollment capacity [9].

2.1. Phases of educational data mining

The task of discovering patterns in the educational sector involves the same data mining algorithms and approaches in the business sector although the intentions of the latter are for the betterment of learning in general among its stakeholders in the academe and has little to do with monetary gains. These educational data mining tasks, similar to general-purpose data mining, are usually divided also into phases [10].

The first phase of EDM is uncovering relationships among historical educational data. Information submitted by students from enrollment to graduation is among the richest and most detailed relevant information that can be used as models for prediction, analytical and quantitative research. Because these data are required to be given by each student, relationships that will be discovered can be validated and double checked for consistency, frequency, and integrity. The pre-processing phase of EDM usually involves identifying which data are related to one another and what other attributes that are available on record can be utilized to create or sustain a stronger attribute relationship. To identify all possible or relevant relationships,

several algorithms can be utilized such as classification-based approaches, sequential pattern mining clustering, regression, association rule mining, and factor analysis. Under this phase, a huge part of the historical data is used as a training set to create the model [11].

After identifying the relationships among attributes, phase 2 of EDM involves relationship validation to get away with overfitting. The infield of statistics, overfitting refers to an analysis that relates too narrowly or almost exactly to a specific dataset and may consequently fail in predicting reliably future or upcoming data. Usually, a portion of historical data is set aside as test data. By adjusting criteria, parameters, factors, or thresholds, a significant level of accuracy can be attained but will still provide leeway for upcoming data [12].

Phase 3 of EDM involves the application of validated relationships or the result of phase 2 to make reliable predictions in the chosen educational context in which educational historical data, attributes, and relationships are selected and aimed at [13]. Under this phase, the rate of prediction accuracy is monitored before it can be used to render decisions. There are usually two main approaches for EDM under this phase: discovery with models and distillation of data for human judgment. The discovery with models approach uses the high rate predictions made under phase 3 to be utilized as a component in another analysis either by predicting another variable for next-level analysis or the use of predicted variable as it relates to attributes used in the prediction. This approach requires that the predictions made under phase 3 have proven general ness across applicable contexts [11]. Meanwhile, a distillation of data for human judgment literally applies human inferences or higher-level human intelligence, intuition, and mental grasp which the computerized EDM process may lack. Data distillation by humans is purposed either to classify or identify. Predictions, data frequency, or patterns that humans can classify can aid immensely in the improvement of the prediction model while those that data relationships and patterns that humans can identify are purposed for interpretation as computers may provide a visual representation of predictions or results, but they lack the ability to interpret these visual data and the sublime reasons why such visual presentation was arrived at or formed [12].

After approaching the prediction results of the EDM process with either the discovery of models or distillation of data through human judgment, phase 4 will now take place which involves using the EDM process predictions to support decision-making procedures and in crafting decision policies. The decision-making process reflects the actual addition of obtained information to the knowledge base [14].

2.2. Data mining pre-processing techniques

Nowadays, there exist several definitions and characterizations of data mining and its corresponding tasks. However, data mining can be simply viewed as a four-step process: data pre-processing, data mining, pattern evaluation, and knowledge presentation. Among these four steps, the pre-processing step involves the most crucial tasks to gain relevant results, as the old cliché in the computer world: garbage in, garbage out. Aside from being the most crucial part of the process, being the foundation of the whole DM system, it involves dealing with actual, physical raw data, thus, a substantial amount of effort and time are usually allotted for this task [10].

Data mining pre-processing is a process consisting of an iterative sequence of the following steps: i) Data cleaning which involves tasks of removing noise and inconsistent data; ii) Data integration wherein multiple data sources may be combined; iii) Data selection which involves database retrieval of data relevant to the analysis task, and iv) Data transformation which pertains to transforming or consolidating data into forms appropriate for mining by performing summary or aggregation operations [15].

Depending on the nature or context of the data mining goal, the pre-processing of data may involve simple or sophisticated choice methods or algorithms in order to identify the most significant attributes and determine the difficulty and the general quality of models that can be taken into account in the actual data mining stage [16].

2.3. Educational data mining attributes selection

Data mining attributes refer to the fundamental qualities or characteristics of individual data to be mined. In terms of raw data, a filled-up form refers to a record while each item in the form pertains to a data attribute. Technically, when viewed in a column-row relationship or a two-dimensional dataset, a filled-up form or records becomes the row while the individual composition or items that make up a record becomes the equivalent unique identifier of each column. The more rows you have means the more records to process. The more columns you have means the more attributes or characterization you can have for a record. For simplification purposes and to relate with the current subject of this paper, a student record refers to a row of records while items such as name, age, gender, grades, refer to the attributes of the record [17].

In data mining, an attribute is directly equivalent to a variable in a model. Thus, the fewer variables you have and the more records or samples you can aggregate, the more interpretable the result of the model

can be. Meanwhile, the more variables you have but fewer samples could mean less significant accuracy and would be difficult to interpret [18].

In order to efficiently process voluminous data, the important and relevant variables are often selected which is called feature selection which is a dimension reduction methodology that aims to reduce the number of original predictors (usually denoted as p predictors) to be operated in a model by choosing the best predictors (denoted as d predictors) [11].

Figure 1 exhibits the graphical representation of feature selection where X refers to an attribute and n to the total number of samples or data points. In the left portion of the image, the eight attributes pertain to the original predictors or p predictors which will require $R^{n \times 8}$ operations to come up with a contextual result R . Meanwhile, the right portion of the image shows that it only requires three variables (d predictors) and $R^{n \times 3}$ operations to come up to the same contextual result R . This dimension reduction directly impacts and benefits computational time and resources. It also avoids overfitting and increases prediction performance [16].

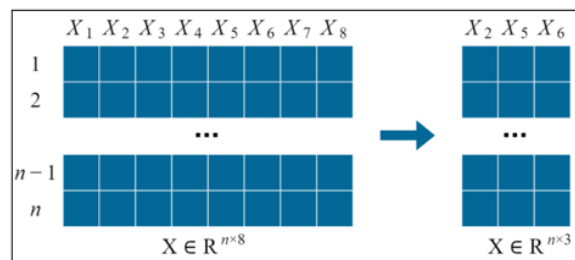


Figure 1. A graphical representation of feature selection as a data mining dimensional reduction methodology of reducing multiple variables to only a few important variables to be used by a model

2.3.1. Approaches in feature selection of data mining

There are three approaches to feature selection in data mining. These are filter, wrapper, and embedded approaches [9]. Figure 2 visually represents the filter approach of feature selection in data mining. This method chooses the features to be used in a model which is independent of the type of classifier algorithm to be utilized. It is simple and only requires to be performed once, however, it does not take into account the feature dependencies and relevant interaction with the classifier [14].

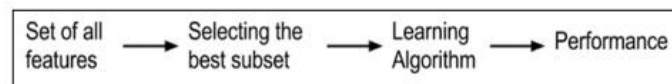


Figure 2. Filter approach of data mining feature selection

Meanwhile, Figure 3 encapsulates the idea behind the wrapper approach of data mining feature selection in which the process of selecting the best subset features refers to the iterative method of generating a subset of predictors (s predictors) from the set of all original p predictors and subjecting it to a learning algorithm to score each attribute in its impact in the total result or prediction accuracy. This approach is totally contrary to the Filter approach as the wrapper approach directly communicates with the classifier algorithm [14].

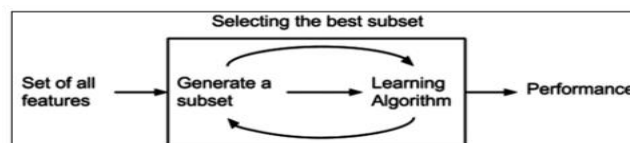


Figure 3. Wrapper approach of data mining feature selection

Finally, Figure 4 summarizes the context of the embedded approach of data mining feature selection which is similar to the wrapper method in iteratively finding the best subset of features, however, the process of feature subset selection is incorporated as part of the classifier algorithm, as such, named embedded approach. This approach offers a reduced computational time, however, requires a lot of technical proficiency to customize the feature selection [18].

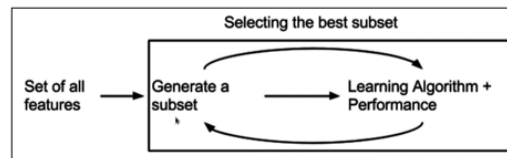


Figure 4. The embedded approach of data mining feature selection

2.4. Data mining pattern discovery techniques

Patterns are everywhere around us. Patterns can be tangible or be naturally present that can be seen by the naked eye or intangible in the sense that it can be perceived or detected by utilizing mathematical solutions through pattern recognition algorithms [19].

In data mining, pattern discovery or pattern recognition is the process of identifying relationships, arrangements, configurations, outlines or repetitions of features from bulk of information contained in a dataset through machine learning algorithms. In this definition, information or knowledge is already there in the form of records and that the search for relationships of among features of one record to another pertains to the process of extracting patterns. The set or subset of features used in pattern recognition are referred to as features vector which is composed of the equivalent numerical values of each feature packed into one multi-dimensional array [20].

Generally, pattern discovery can be summarized to involve two aspects: classification and clustering, which then identifies the type of machine learning implemented, whether supervised or unsupervised learning [21] as presented in Figure 5.

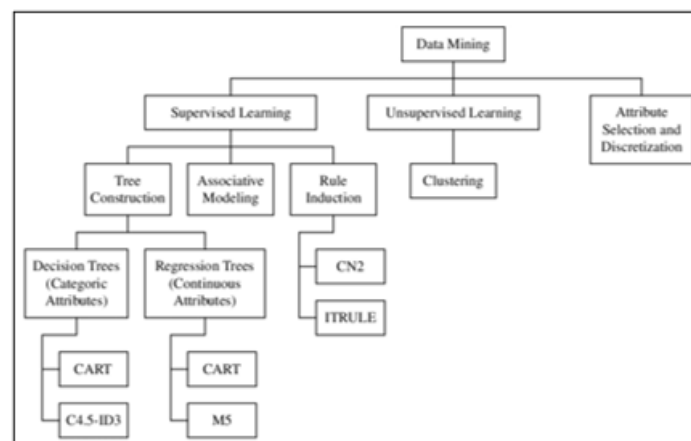


Figure 5. Data mining aspects and underlying techniques

2.4.1. Classification (supervised learning)

Classification is a thorough data analysis task of finding a model that will embody, define, characterize and distinguish data classes and concepts from other data classes and concepts belonging to the same dataset. This task particularly refers to the problem of ascertaining the category, label, or class of a new observation taking into consideration that the category of the new observation is present among the categories, labels, or classes and that this new observation has naturally occurred among the training set which defined the model and its categories, labels or classes. Because the observations as well as the categories of observations contained in the model are already known and naturally occurred as part of the model, supervised machine learning is implemented [20]. These ideas are exhibited in Figure 6.

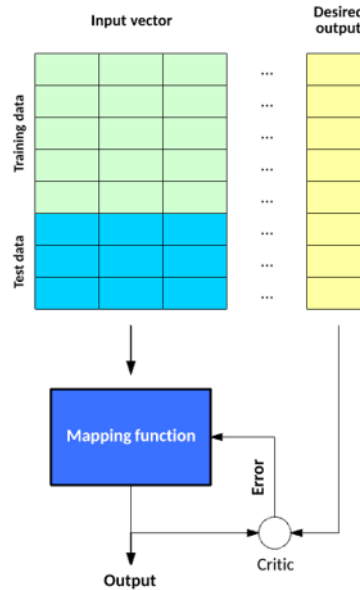


Figure 6. Supervised learning model
(image source: IBM <https://developer.ibm.com/articles/cc-models-machine-learning/>)

As shown in Figure 6, supervised learning has a critical function that is triggered before rendering an output. This critic function serves as a rater of how accurate its output is and is, therefore, a crucial part of the learning system as it provides an idea of whether where adjustments can be made or whether a new observation is not present in the list of labels.

Under supervised learning, the task is to train the machine to predict the category of a new observation by referring to the results of all the previous observations. In simple terms, it's like teaching a child to learn the red, green, and blue colors among objects around him and then showing him a new object and asking him to identify its color whether it is red, green, or blue [22].

Technically speaking, labels or categories can be identified because they can be quantifiable numerically, as in our example, color, which is a combination of certain numerical values of either Red, Green, or Blue (RGB) or cyan, magenta, yellow, black (CMYK) color combinations. Thus, classification may involve a certain process to identify or recognize labels, categories, or classes. If the method of identification involves a series of meeting feature conditions in an if-then-else fashion or finding differences with other feature's values, then it basically falls under decision trees. Meanwhile, if the recognition procedure involves connecting or linking the feature based on how frequently that feature naturally occurred in the dataset, it involves associative modeling or simply association. Finally, if the process of classification is consisted of meeting certain criteria, threshold, or weight, then it involves rule induction [23].

Currently, there are several commonly-used classifiers in the field of data mining and AI. However, for purposes of conciseness, this paper will succinctly discuss decision trees, linear regression, multilayer perceptron, and random forest which were utilized in this research thru Weka.

a. Decision trees and rules

Decision trees and rules use univariate splits to have a simple representational form of decision-making structure of an inferred model for end-users to comprehend easily [1]. The algorithm depends on likelihood-based model-evaluation methods, with varying degrees of sophistication in terms of penalizing model complexity. Technically, multiple nested if-then-else is used to model varying inputs and with similar outcomes until a decision can arrive which then becomes a rule [21]. Figure 7 on the next page illustrates the general diagram of a decision tree.

A decision tree can be formalized to become decision rules where the contents of each leaf node become the outcome while the if clause is the path being formed relating to conditions. Each rule has the following general form [1] if condition1 and condition 2 and ... condition n then outcome.

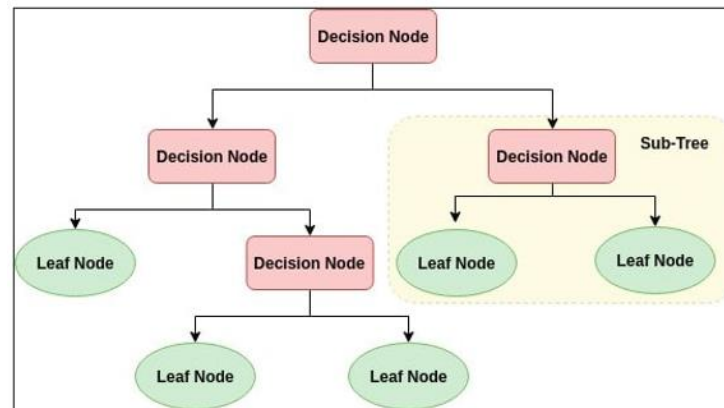


Figure 7. General diagram of a decision tree

b. Linear regression

Linear regression is a direct approach to forming the relationship through linear predictor functions such as the mean function between a dependent variable and one or more independent variables [2]. Practical uses and applications of linear regression usually fall into two categories: i) Fitting a predictive model from observed values both dependent and independent variables and ii) Quantifying the degree of strength of the relationship between the dependent and independent variables [1]. The first category is usually used during prediction or forecasting and error reduction while the second category is usually utilized in qualitatively explaining how the independent variables vary with different sets of dependent variables.

Figure 8 shows a general graph of mean values of x and y which is used in linear regression analysis and predicting the next mean values.

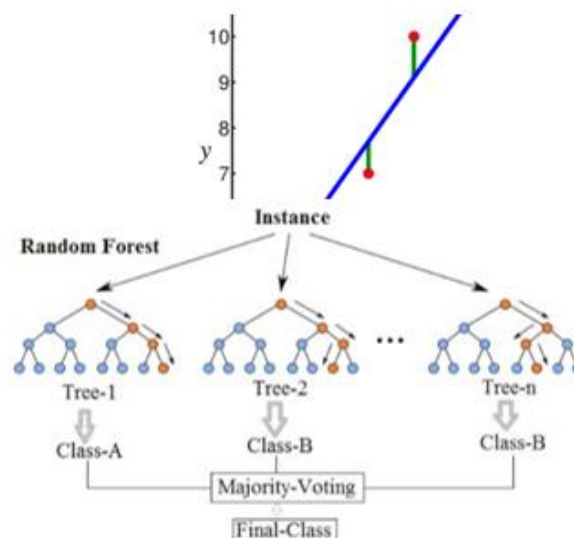


Figure 8. General graph of a linear regression thru mean values

c. Multilayer perceptron

A multilayer perceptron (MLP) is a type of feed-forward artificial neural network which is usually consisted of an input layer in one end, at least one hidden layer in the middle, and an output layer on the other end where each node is considered a neuron, except for input nodes, that uses a nonlinear activation function or a resulting value to be forwarded for use by the next layer [4]. MLP most of the times uses a form of supervised learning technique commonly known as back propagation during training or adjusting bias function values to achieve the desired output [2]. Figure 9 exhibits a general diagram of a multilayer perceptron. MLP makes a suitable classifier algorithm due to its ability to create models through regression analysis and statistically solve problems with random probability distribution[16].

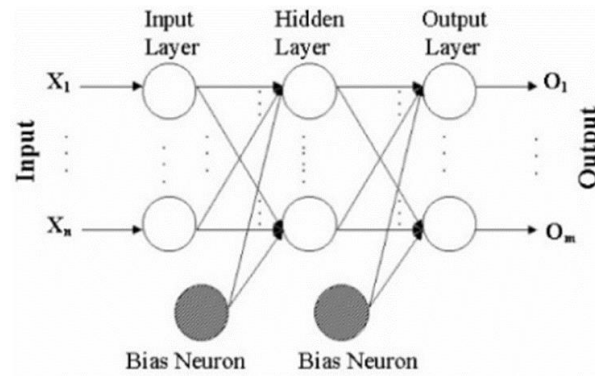


Figure 9. General diagram of a multilayer perceptron

d. Random forest

Random forest is a collaboration of learning methods for classification tasks by forming multiple decision trees during the training period and outputting the classes or mean prediction of the individual trees formed [1]. Figure 10 shows cases the general diagram of a random forest algorithm. Random forest most of the time corrects the over-fitting behavior of simple decision trees which makes it a good classifier algorithm [16].

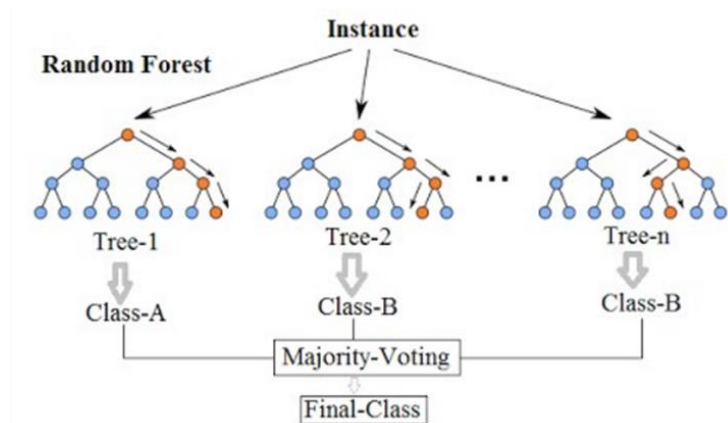


Figure 10. General diagram of a random forest

2.4.2. Clustering (unsupervised learning)

Clustering or Unsupervised learning involves the problem of classifying a new observation from a set of observations but without any categories, labels, or classes. Because it does not know what category or label it will use to identify a new observation, its aim is to find one but not really directly tagging each observation a certain label, category, or class.

To identify observations without tagging or labeling, it uses the numerical features of each observation, maps them, and group them according to their values so when a new observation arrives, it tells whether its feature values are far, near, or within the group of previous observations [22]. This process can be seen in Figure 11.

Unsupervised learning does not incorporate a critic function as clearly shown in Figure 11 and uses all or mostly all of the numerical values of features it has observed. Thus, it cannot provide an accuracy rating for each output [24]. Most unsupervised learning algorithms use the nearest neighbor concept to tell how far or near the feature of a new observation is to the previous observations [25].

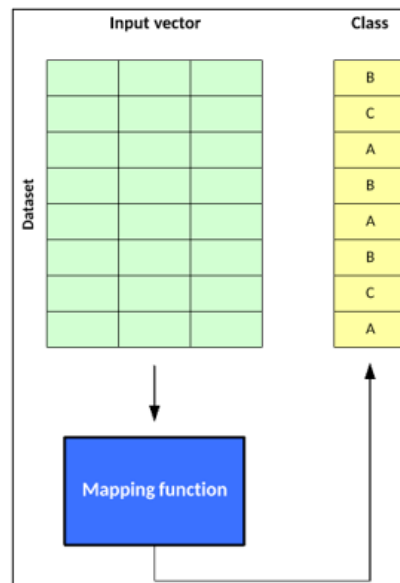


Figure 11. Unsupervised learning model

(image source: IBM <https://developer.ibm.com/articles/cc-models-machine-learning/>)

3. RESEARCH METHOD

The paper proposed a framework for an effective educational process that uses data mining techniques to uncover hidden trends and patterns and make accurate predictions based on a higher level of analytical sophistication in the admission process in the university.

The major impact of this study is that it helps the administration make better decisions by projecting the number of expected registrants for the following semester and alerting concerned colleges and departments about their enrollment readiness from their respective areas of responsibility. The number of teachers, classrooms, classroom equipment, materials and tools, pertinent supplies, and non-academic personnel can all be assessed for sufficiency using the forecasted data, and decisions, actions, and reporting can be made quickly in order to ensure a smooth enrollment and class operations.

This paper involved three major tasks: attribute selection, pattern discovery, and prediction. These three tasks were performed in Weka, an open-source data mining software with a collection of machine learning algorithms for data mining tasks. The researcher uses a feature selection technique to find the traits or features that have the greatest impact on our outcome variable. Weka offers a diverse set of attribute selection algorithms that can be categorized in a variety of ways. The way in which attributes can be evaluated is one of the most prominent categorizations of the algorithms and this way they can be classed as filters and wrappers. Wrappers employ the performance of learning algorithms to determine the desirability of an attribute subset, while filters pick and analyze features independently of learning techniques. Weka supports multiple feature selection algorithms. It contains tools for data preparation, classification, regression, clustering, association rules mining, and visualization [9].

The attribute selection task involved wrapper attribute selection using Weka's correlation attribute eval algorithm. To find which among the 11 fundamental student attributes as shown in Table 1 can be used as strong predictors of successful graduation of a student's college program. Each attribute has been assigned a model code identifier for brevity. Meanwhile, for pattern recognition and prediction tasks, a two-year record set consisting of 1,817 records of admitted students in a local university during the school year 2014 and graduated in 2018 were pre-processed and converted into a comma-separated value (CSV) format file using the attribute's model code in Table 1 as column headers. The labels chosen for the model to classify were the fourteen university program offerings as shown in Table 2.

For the input vector, seventy percent (70%) or a total of 1,272 records were apportioned as part of the training dataset while the remaining 30% or 545 records were set aside to be part of the test dataset. By setting the 70%-30% data proportion in Weka, the application chose the respective 70% training dataset and 30% test dataset. Multiple data mining predictive techniques were utilized which included multilayer perceptron, decision trees, linear regression, and random forest.

Table 1. Student attributes and model code

#	Attribute	Description	Model code
1	General weighted average	High school general weighted average	HS_GWA
2	Admission test score (ATS)	College admission test score	ATEST_SCORE
3	Preferred program	Preferred collegiate program of the student	PREF_PROGRAM
4	Prescribed program	Prescribed college program to the student based on university policy	PRESC_PROGRAM
5	Final math grade	Final mathematics average grade of student in high school card	HS_MATH
6	Final science grade	Final science and technology average grade of student in high school report card	HS_SCIENCE
7	Final english grade	Final english average grade of student in high school report card	HS_ENGLISH
8	Admission test math score	College admission test mathematics score	ATEST_MATH
9	Admission test science score	College admission test science and technology score	ATEST_SCIENCE
10	Admission test English score	College admission test English score	ATEST_ENGLISH
11	College general weighted average (CGWA)	General weighted average of student in his chosen college	COLLEGE_GWA

Table 2. Model labels

#	College program	Assigned label
1	B.S. Biology	BS BIO
2	A.B. English	AB ENGLISH
3	A.B. Economics	AB ECONOMICS
4	A.B. Public administration	AB PA
5	B.S. Mathematics	BS MATH
6	B.S. Computer science	BSCS
7	B.S. Information and communications technology	BSICT
8	B.S. Nutrition and dietetics	BSND
9	B. S. Hospitality management	BSHM
10	Bachelor of secondary education	BSEd
11	B.S. Business administration	BSBA
12	B.S. Social work	BSSW
13	Bachelor of technology in teacher education	BTTE
14	Bachelor in industrial technology	BIT

4. RESULTS AND DISCUSSION

The major purpose of this study was to develop educational data mining and knowledge discovery system for Pangasinan State University (PSU) in order to improve its college admission services by analyzing historical data on college admissions, identifying and choosing essential features or qualities from student data entry credentials that will include a dynamic rule-based approach to data mining, knowledge discovery using various data mining approaches, assessing the resulting system's acceptability among stakeholders, and identifying how discovered patterns may be used to improve college admission services.

Each admission officer manually assesses each candidate's data against the numerous entrance requirements before making admission and placement decisions under the present admission process. This method will take far too long to complete in terms of processing time. Thus, PSU mining and knowledge discovery system was created through mining the historical databases and discovering knowledge.

Thru classification (supervised learning) and association rule mining employing Apriori algorithm to determine the most frequent attribute values in a topic dataset and then reduces the attributes to reveal distinctive patterns. In discovering patterns, which serve as the foundation for establishing the ideal course for students and admission forecasting for enrollment readiness, which is useful for enrollment capacity planning.

Figure 12 exhibits the enrollment forecasting generated from the developed customized data mining and knowledge discovery system. The researchers made use of the previous 1st semester enrollment records from school years 2013 to 2018 and projected the 1st semester 2019 enrollment. The forecast estimated 6,972 enrollees, however, only 6,216 enrolled, thus making the forecast 88.59% accurate which is still significantly close to the 85% accuracy forecast threshold. As can be gleaned from the graph.

Figure 13 shows the attribute selection task which was performed using Weka's correlation attribute eval algorithm. Based on the graph, Admission Test Score in Science shows the strongest predictor with a correlation rate of 4.94% followed by the final grade in Math during high school, admission test score, high school Science final grade, high school general weighted average, and admission test scores in Math and in English. The correlation shows also that the rest of the attributes have shown to be fewer predictors if a cut-off of 0.1% will be imposed.

Figure 14 presents the prediction accuracy of the four subjects supervised learning classifiers (simple linear regression, decision Table, multi-layer perceptron, and random forest), each of which was trained with 70% of the datasets and the remaining 30% were utilized as test dataset using Weka. The graph shows that the Multilayer perceptron rendered a significant average accuracy of 92.12% on its predictions, followed by the decision Table approach (91.33%), simple linear regression (89.78%), and random forest approach (89.15%), respectively.

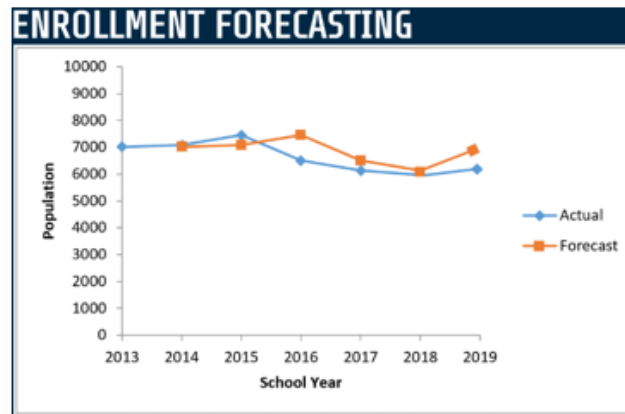


Figure 12. Graph of enrollment forecasting generated

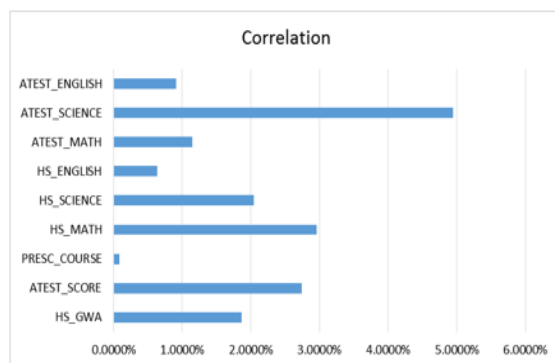


Figure 13. Graph of attribute selection task performed using Weka's correlation attribute eval algorithm

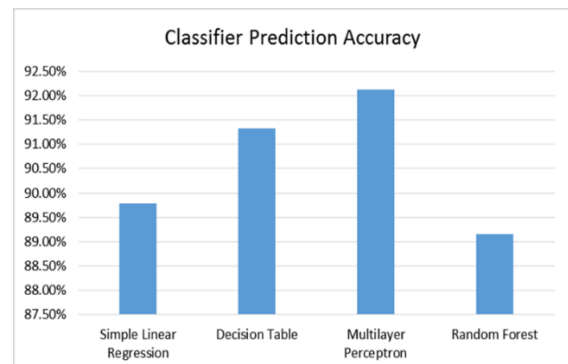


Figure 14. Graph of classifier prediction accuracy among the 4 subjects supervised learning classifiers

5. CONCLUSION

The careful selection of attributes that embody the data mining model is a crucial task in order to gain substantial accurate results. The result found that the university program for college freshmen using students' grades from high school report cards and college admission test scores as EDM model attributes can be predicted with a significant level of accuracy by using multiple supervised learning classifiers such as multilayer perceptron, decision Table, linear regression and random forest.




REFERENCES

- [1] M. J. Zaki, W. Meira Jr, and W. Meira, *Data mining and analysis: fundamental concepts and algorithms*. Cambridge University Press, 2014.
- [2] D. L. Poole and A. K. Mackworth, *Artificial Intelligence: foundations of computational agents*, 2nd ed. Cambridge University Press, 2010.
- [3] R. T. Watson, *Data Management: Databases and Organizations*. Prospect Press: Prospect Press; 6th edition, 2015.
- [4] C. Romero, M. Ventura, Sebastian Pechenizkiy, and R. Sj. Baker, *Handbook of educational data mining*. CRC press, 2010.
- [5] J. Grus, *Data science from scratch: first principles with python*. O'Reilly Media, 2019.
- [6] F. Gorunescu, *Data mining: concepts, models and techniques*. Berlin: Springer-Verlag Berlin and Heidelberg GmbH and Co. K, 2013.




- [7] J. Dean, *Big data, data mining, and machine learning: value creation for business leaders and practitioners*. John Wiley and Sons, 2014.
- [8] S. Huang and N. Fang, "Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models," *Computers & Education*, vol. 61, pp. 133–145, Feb. 2013, doi: 10.1016/j.compedu.2012.08.015.
- [9] E. Muratov, M. Lewis, D. Fourches, A. Tropsha, and W. C. Cox, "Computer-assisted decision support for student admissions based on their predicted academic performance," *American Journal of Pharmaceutical Education*, vol. 81, no. 3, p. 46, Apr. 2017, doi: 10.5688/ajpe81346.
- [10] C. C. Aggarwal and Others, *Data mining: The textbook*. Springer, 2015.
- [11] V. Kotu and B. Deshpande, *Predictive Analytics and Data Mining*. Elsevier, 2015.
- [12] D. T. Larose and C. D. Larose, *Discovering knowledge in data: an introduction to data mining*. John Wiley and Sons, 2014.
- [13] R. S. J. d. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," *Journal of educational data mining*, vol. 1, no. 1, pp. 3–17, 2009, doi: 10.5281/zenodo.3554657.
- [14] M. Heumann, Christian Schomaker and Others, *Introduction to statistics and data analysis*. Springer, 2016.
- [15] B. Christian and T. Griffiths, *Algorithms to live by: The computer science of human decisions*. Henry Holt and Co.; 1st edition, 2016.
- [16] P. Frasconi, N. Landwehr, G. Manco, and J. Vreeken, *Machine learning and knowledge discovery in databases: European Conference, ECML PKDD 2016, Riva Del Garda, Italy, September 19-23, 2016, Proceedings, Part II*. Springer, 2016.
- [17] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge University Press, 2013.
- [18] U. Waikato, "Weka - Machine Learning Software in Java," ., 2014. <http://www.cs.waikato.ac.nz/ml/weka/>.
- [19] Y. Kodratoff and R. S. Michalski, *Machine learning: An artificial intelligence approach, volume 3*. Elsevier, 2024.
- [20] J. Bibodi, A. Vadodaria, A. Rawat, and J. Patel, "Admission prediction system using machine learning," pp. 1–8, 2017.
- [21] T. H. H. Cormen, C. E. E. Leiserson, R. L. L. Rivest, and C. Stein, *Introduction to algorithms*, 3rd ed. MIT Press, 2014.
- [22] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach - 3rd Edition*. Publisher: Prentice Hall pa. Pearson; 3rd edition, 2009.
- [23] L. Rusdiana and Marfuah, "The Application of Determining Students' Graduation Status of STMIK Palangkaraya Using K-Nearest Neighbors Method," *IOP Conference Series: Earth and Environmental Science*, vol. 97, p. 012040, Dec. 2017, doi: 10.1088/1755-1315/97/1/012040.
- [24] R. E. Neapolitan and X. Jiang, *Artificial Intelligence With an Introduction to Machine Learning*, vol. 53, no. 9. Chapman and Hall/CRC, 2019.
- [25] S. Asadianfam and S. Shamsi, Mahboubeh Asadianfam, "Predicting academic major of students using bayesian networks to the case of iran," *International Journal of Computer-Aided Technologies (IJCAx)*, vol. 2, no. 3, p. 7, 2015, doi: 10.5121/ijcax.2015.2304.

BIOGRAPHIES OF AUTHORS



Julius Cesar O. Mamaril    is an Associate Professor at the College of Computing Sciences, Pangasinan State University, Philippines. He possesses a degree in Computer Science (BSCS), Master in Public Administration (MPA), Doctorate degree in Business Administration (DBA) and currently completing his Doctorate Degree in Information Technology (DIT) at the Technological Institute of the Philippines, Manila. He was a former Dean of the College of Computing Sciences, Deputy Executive Director of Open University Systems and currently the Director of ICT Management Office, formerly MIS Office. He can be contacted at email: jmamaril@psu.edu.ph.



Melvin A. Ballera    is a professor at Technological Institute of the Philippines, Manila with the degree in PhD in Computer Science (FIT/SGS), De La Salle University. He holds positions including academic coordinator, director for linkages and currently the Assistant Research Director. He has published numerous papers in the area of Artificial Intelligence, E-Learning, and Instructional Technology. He can be contacted at email: melvin.ballera@tip.edu.ph.