

# The Investigation of Student Dropout Prediction Model in Thai Higher Education Using Educational Data Mining: A Case Study of Faculty of Science, Prince of Songkla University

Jarutas Pattanaphanchai<sup>1</sup>, Koranat Leelertpanyakul<sup>1,2</sup> Napa Theppalak<sup>1,3</sup>

<sup>1</sup> Faculty of Science, Prince of Songkla University, Thailand.

<sup>1</sup>jarutas.p@psu.ac.th; <sup>2</sup>koranat.j@outlook.com; <sup>3</sup>5810210209@psu.ac.th

## ARTICLE INFO

Submission date: 3/11/2018  
Acceptance date: 10/12/2018  
Publication date: 25/1/2019

## Abstract

The student's retention rate is one of the challenging issues that represents the quality of the university. A high dropout rate of students affects not only the reputation of the university but also the students' career in the future. Therefore, there is a need of student dropout analysis in order to improve academic plan and management to reduce students drop out from the university as well as to enhance the quality of the higher education system. Data mining technique provides powerful methods for the analysis and the prediction of the dropout. This paper proposes a model for predicting students' dropout using the dataset from the representative of the largest public university in the Southern part of Thailand. In this study, data from Faculty of Science, Prince of Songkla University was collected from academic year of 2013 to 2017. The experiment result shows that JRip rule induction is the best technique to generate a prediction model receiving the highest accuracy value of 77.30%. The results highlight the potential prediction model that can be used to detect the early state of dropping out of the student which the university can provide supporting program to improve the student retention rate.

**Keywords:** Data mining, dropout, higher-educational, prediction.

## Introduction

Education has been an important factor for developing a country as it produces skill and educated labours who in turn are the key factor for successful economic development. However, one study showed that some students could not pass through the university level [1]. This issue is not only affecting the academic field but also influences the image of the country. Especially in higher education, student retention is a challenging task that represents the efficiency and reliability of the institution. Finding hidden patterns or prediction trend in vast database helps to improve the quality of management decision-making which can allocate resources appropriately with a better understanding of student learning environment. The prediction with high accuracy in students' dropout is beneficial as it helps to identify the students at the risk stage of academic performance. Data mining has been shown the successful benefit in the

business domain and it can be a suitable tool to benefit in the educational domain for finding useful information hidden in the huge dataset.

Data mining is a technique to extract valuable data from a larger set of any raw data to discover helpful patterns and relationship. A process of processing data in data mining known as CRISP-DM (Cross Industry Standard Process for Data Mining) model[2]. The CRISP-DM involves six phases including 1) business understanding which is a phase for creating the understanding of project objectives and requirements in order to convert it to data mining problem definition; 2) data understanding that focuses on identifying the data which need to be collected; 3) data preparation as a phase for constructing the final ready-to-use dataset in the modelling tool; 4) modelling which focuses on selecting and applying various modelling techniques to find the best fit model for extracting useful information and pattern in the dataset; 5) evaluation phase for assessing the achievement of the finding model compare with the project objectives, and 6) the final phase as deployment which carries out the use of the created models [3]. The techniques that were developed and implemented in data mining consist of various methods which include classification, clustering, association rule, and regression. The classification method constructs a model based on the training set of known class labels data to classify unknown objects. Clustering is an approach for grouping records that are similar within a group. Therefore, clustering is similar to a classification method as it is an approach to organize data in classes or groups. However, the difference between classification and clustering is that in the latter class labels are unknown. The association rule assists to discover the relationship between parameters. It studies the frequency of items occurring together in the transaction [4]. Prediction approaches to create a predictive model that gives potential implications of numerical values in the future or increase/decrease trends in time [5]. The advantages of data mining are recognised in the application within educational research.

Educational Data Mining (EDM) is an approach to covert raw data in education system into useful information that can help lecturers or academic staffs to perform corrective actions [6]. The works in Educational Data mining focus on developing, researching, and applying computerized methods to analyze student data. The EDM is an approach to extract useful information from educational systems similar to the extracting knowledge process in data mining. In Educational Data Mining, classification is the most popular task to build students' performance predictive model. Several algorithms in classification technique were applied to predict students' performance, namely Decision Tree, Rule Induction, Neural Network, Naïve Bayes, and Support Vector Machine (SVM) [7]. Decision Tree, a popular technique for prediction for its simplicity and comprehension, is applied to discover valuable data from a small or large data set. Rule Induction is a technique to produce rules in a set of IF-THEN statement [8,9]. Neural Network facilitates the detection of a possible interaction between predictor variable [10]. Naïve Bayes computes a set of probabilities in a given data set [11]. Support Vector Machine, a supervised machine learning algorithm, finds the hyperplane (decision boundaries that help to classify the data points) in an N-dimensional space (N, number of features) that optimally separates the data into two categories [11]. In addition, the EDM methodologically can integrate to other fields like psychology, for example, the integration of psychometric modelling with data mining techniques. Moreover, EDM

involves different groups of users which effect to the perspective of using educational information from different angles depending on the users' mission and vision [8].

This study proposes the model to predict student drop out in the university level by using Data Mining. We applied the prediction model by using classification techniques. We investigated the student records of Faculty of Science, Prince of Songkla University, Thailand such as admission method (how students enter to university), major, education status, study term, grade point average, students' high school province, and grade point average at the students' high school. In this paper, the model obtains the rule in the form of IF-THEN statements in which the IF part contains  $m$  predicting conditions of attributes, while the THEN part contains a prediction attribute or class. The classification rules were also used to identify the factors that might affect the status of students and this can help the faculty or university to offer personalised help for students in order to reduce dropout rate in university. In several experiments we compared the classification algorithms to find the best algorithm that works well with education data in Thai higher education environment. The structure of the paper is the related works, the methodology applied in this study, and the results. Finally, we present the conclusions and future works.

## Related Work

There are several studies conducted on prediction student dropout or student's performance. Márquez-Vera, Moroles, and Soto [13] proposed a method for predicting school failure and dropout of student in a middle-school applying data mining technique namely induction rules and decision tree. Moreover, Ramesh, Parkavi, and Ramar [14] investigated the factors which influence the performance of students in their final examination and found a data mining algorithm for predicting the grade of students. The results from the study showed that parents' occupation has a high impact on predicting student's performance and multilayer perceptron algorithm corresponds most closely to predict grades.

Tekin [15] proposed the model for predicting student's grade point average at graduation. Three algorithms namely Neural Networks (NN), support vector machines (SVM), and Extreme Learning Machine (ELM) were used for generating the predictive model. The results from their study showed that the SVM technique provides accurate prediction at the rate of 97.98%. In addition, Namdeo and Jayakumar [16] proposed the basic approach and concept of rough set theory to predict student's performance in course work. Moreover, Mythili and Shanavas [17] proposed the model to analyse and evaluate students' performance using classification algorithm. Datasets used for analysis consists of 6 attributes namely gender, living locality, parental education, economic status, and class attendance. The results showed that the class attendance is considerably related to student's performance. Ahmad, Ismail, and Aziz [18] proposed a prediction model of students' academic performance based on student's personal data and past academic records. The researchers used decision tree, Naïve Bayes, and rule based classification technique to predict students academic performance. The experiment results showed that the rule based algorithm was a best algorithm to create a model. The knowledge from the predicting model was used to identify and profile the students' level of success in the first semeste. Guarín, Guzmán, and González [19] presented the model which predicts the attrition rate of students at the Universidad Nacional de Colombia in the first enrollment. The classification technique used to identify academic attrition of students

were Naïve Bayes and a decision tree. The experimental result showed that the prediction of dropout status was improved when the academic data are added. In addition, Abu-Oda and El-Halees [20] proposed the prediction model of student's dropout from university using decision tree and Naïve Bayes techniques. The collected data related to students' study history, GPA, and high school grade point average. The results showed that decision tree and Naïve Bayes build the model that has the accuracy of prediction as 98.14% and 96.86% respectively. The study of Aulck, Velagapudi, Blumenstock, and West [21] proposed a model to predict the student retention in the university using a large, heterogeneous dataset of student's demographics and transcript records. The dominant predictor to indicate whether or not a student will drop out from the university were GPA in Math, English, Chemistry, and psychology courses.

Tran, Dang, Dinh, and Phan [22] presented a study on Predicting Student Performance (PSP) in academic systems using regression and classification methods in order to predict student performance on their courses and also help to predict which courses suited their abilities or preferences. Similarly, Pereira and Zambrano [23] used the classification technique based on decision tree to build a model for profiling student dropout pattern. Watkins, NaPhayap, and Jirasukprasert [24] presented intelligent recommendation system that can predict potential dropout of students, provide suggested programs or activities matching students' preference, and predict the expected GPA and grade point. Likewise, Iam-On and Boongoen [25] presented the results of an investigation on the effectiveness of education data mining technique for detecting the first-year students bearing a risk to fail or drop out, and factors influencing the student performance at Mae Fah Luang University.

Ramanathan, Parthasarathy, Vijayakumar, Lakshmanan, and Ramani [26] introduced the predicting student performance model based on cluster of distributed architecture. The researcher predict the student's performance using the combination of Lion-Wolf algorithm and DBN. Additionally, Burgos, Campanario, Pena, Lara, Lizcano, and Martínez [27] investigated historical student course grade data from E-learning system using data mining technique to predict student drop out from an E-learning course. The prediction model was also used for designing a tutoring action plan in which it led to a reduced dropout rate by 14% compared with previous academic years. In addition, Bhanushali, Khan, Madhia, and Majumdar [28] presented model to identify the student who might dropout based on certain attributes such as semester attendance, test marks, and aggregate CGPA of the student in the previous semesters using decision tree and Naive Bayes algorithm.

In summary, there is evidence that the classification method suits well to predict student's performance or discover factors that affect to students' dropout rate.

## Methodology

The proposed method in this study for predicting the student dropout pattern is based on the process of Knowledge Discovery and Data Mining (KDD) [29]. The processing step are data collection, data selection, data preprocessing, data mining and experimentation as shown in Figure 1.

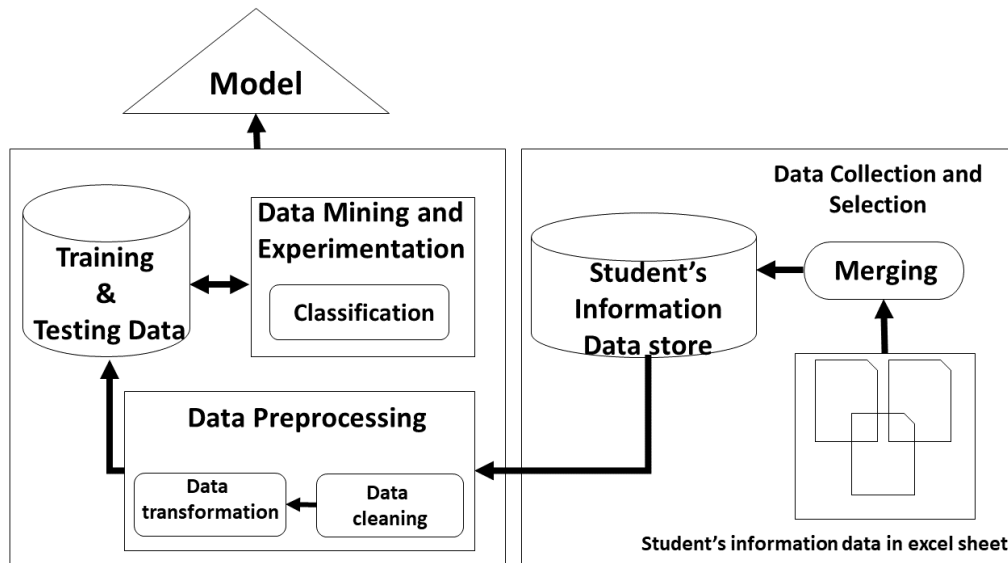


Figure 1. The Prediction of Student Dropout Process

### Data Collection and Selection

In this study, we gathered student related data from the Faculty of Science, Prince of Songkla University of 5 years from 2013 to 2017. The data were collected from three different data sources which were stored in an MS Excel file. We concentrated on the historical data related to student academic background, carry path of study, and social factors. A detailed description of the three datasets is shown in

Table 1.

**Table 1. Description of Educational Dataset**

Dataset	Description
SelectingMajor	Dataset contains 2,009 records of students and 16 attributes to consider academic factors
AdmissionScore	Dataset contains 1,101 records of students admitted in 2014-2018 with 12 attributes
StudentInformation	Dataset contains 4,228 records of students and 11 attributes to consider social factors

In this step, data from different data sources were combined into a single dataset using a python script to integrate data from different files. In total, we have 4,238 records with 7 attributes namely, admission method, major, education status, term of enrollment, grade point average of university, province of high school, and grade point average of high school.

### Data preprocessing

This step focuses on preparing data suitable for the data mining process. It consists of data cleaning and data transformation process. Data cleaning is a step to remove missing values, noise or redundant data from the dataset. We deleted the attributes that has a number of records with null data more than 80% using the filter feature of Weka. In addition, we ignored dropout type as “Graduation” because we focus on relation about dropout or continue their study. Data transformation is a step to modify

the type of attributes to match the requirement of data mining techniques. In order to facilitate the classification model and patterns extraction, the continued variables were transformed to discrete variables using discretize filter in Weka. Then, we formatted the data to ARFF format for building a predictive model. Finally, we obtained the data that will be analysed as shown in Table 2. Then, the whole dataset was divided randomly into training and testing data files.

**Table 2. The Dataset in ARFF Format**

ARFF FORMAT	
@attribute	Entrance {First, Admission}
@attribute	Major {Biology, Biotechnology, Chemistry, Chemistry-Biology, ComputerScience, ICT, Materials, Mathematics, Microbiology, Physics, PolymerScience, Statistics, PhysicalScience, LifeScience, Science, IT}
@attribute	EducationStatus {DROPOUT, Studying}
@attribute	Term {1,2,3}
@attribute	GPX nominal
@attribute	Province {Krabi, Chumphon, Trang, NakhonSiThammarat, NONTHABURI, Narathiwat, PrachuapKhiriKhan, Pattani, Phangnga, Phatthalung, Phetchaburi, Phuket, Yala, RoiEt, Ranong, SiSaKet, Songkhla, Satun, SamutPrakan, SuratThani, Ubon-Ratchathani, Bangkok, Kanchanaburi, NakhonNayok, NakhonPathom, Phayao, Ratchaburi, Phitsanulok, ChiangMai, UdonThani, Chachoengsao, BuriRam, KhonKaen, Yasothon, Phichit, Lamphun, MahaSarakhm, NongBuaLamphu, CO, Chanthaburi, PrachinBuri, Phetchabun, UthaiThani, NakhonPhanom, Saraburi, Loei, Unknown}
@attribute	HighschoolGPX nominal

The attribute ‘Entrance’ represents the method how students enter to Faculty of Science in which ‘First’ means the student apply or enter to the faculty through the direct application for a special program offered from the faculty, and ‘Admission’ means the students did the central examination tests before applying to the faculty with their examination scores. The attribute ‘Major’ represents the major field that students chose to study; ‘EducationStatus’ represents students’ status at the faculty. The attribute ‘Term’ represents the current semester term time and ‘GPX’ represents the current grade point average of students. The attribute ‘Province’ represents the province of high school where students graduated from. ‘HighschoolGPX’ represents the grade point average of student from their high school.

## Experimental Design

This step describes the experiment of data mining techniques which will be selected to use for building a prediction model of student’s dropout. We decided to test with the classification Tree Model and Rule-Induction because these classification techniques support to process a substantial number of predictor variables. In addition, this technique supports non-parametric data and capture nonlinear relationships and complex interactions between predictors and the dependent variables [30]. We performed experiments in order to attain an algorithm that provides the highest

classification accuracy. We used Weka Experiment Environment tool to run controlled experiments on our prepared datasets with machine learning algorithm. We compared three decision tree algorithms, namely C4.5 (J48), RandomTree, and REPTree and three rule-induction algorithms, namely OneR, ZeroR, and a rule-based learner (JRip). Each algorithm was evaluated using the default algorithm configuration in WEKA.

C4.5 (J48) is one of the most popular algorithms for the decision tree technique [31]. A decision tree induction of C4.5 algorithm consists of internal nodes, branches, and leaf nodes. The internal node represents a test on attribute and leaf nodes represent class label, where as a branch represents an testing outcome [11]; RandomTree, which produces a random set of data for constructing a decision tree; and REPTree, which uses the regression tree logic to creates multiple trees in different iterations [32]. JRip, where classes are examined in increasing size of data [33]; OneR, which uses the minimum-error attributes for class prediction [34]; and ZeroR, which predicts the mean for numeric values and mode for nominal class, and is considered to be a baseline classifier [35]. We considered the ZeroR classifier as a baseline and as an indicator of the predictive power compared with the other algorithms.

All classification algorithms were run on the dataset containing academic information of student e.g. grade point average, major, and high school grade point average using the 10-fold cross-validation for estimating the generalisation performance. The 10-fold cross-validation operation divides the dataset into 10 roughly equal sets. For each set, it trains the model using nine remaining sets and calculates the test error by classifying the given set. Then, the results from the 10 test sets are averaged. The statistical significance of differences in the performance of OneR and other learners is tested with the two-sided paired t-tester in Weka's Experimenter, using a significance level of 5%.

We first tested the results to identify the best algorithm for classifying the class.

Table 3 ranks the schemes according to the total number of significant wins (>) and losses (<). The first column (>-<) is the number of wins against the number of losses. In

Table 3 is also illustrated that JRip performed the best classified data against the other algorithms.

**Table 3. Ranking Algorithms in The Experimental Results**

>-<	>	<	Resultset
4	4	0	rules.JRip
3	3	0	trees.J48
2	3	1	trees.REPTree
-1	2	3	rules.OneR
-3	1	4	trees.RandomTree
-5	0	5	rules.ZeroR

Then, we tested the accuracy of each algorithm to perform the prediction. The results showed that the JRip algorithm achieved a classification of 38.97% (+/-0.53%) which is statistically significant better than ZeroR at 30.14% (+/-0.53%) as shown in

**Table 4 Algorithm Accuracy Compared To Zeror**

Dataset	(1) rules.ZeroR	(2) rules.OneR	(3) rules.JRip	(4) trees.J48	(5) trees.REPTr	(6) trees.Rando
StudentScience	30.14 (0.11)	37.43 (0.76) v	38.97 (0.53) v	38.82 (0.58) v	38.26 (0.81) v	34.78 (1.48) v
	(v/ /*)	(1/0/0)	(1/0/0)	(1/0/0)	(1/0/0)	(1/0/0)

**Table 4 Algorithm Accuracy Compared To Zeror**

Dataset	(1) rules.ZeroR	(2) rules.OneR	(3) rules.JRip	(4) trees.J48	(5) trees.REPTr	(6) trees.Rando
StudentScience	30.14 (0.11)	37.43 (0.76) v	38.97 (0.53) v	38.82 (0.58) v	38.26 (0.81) v	34.78 (1.48) v
	(v/ /*)	(1/0/0)	(1/0/0)	(1/0/0)	(1/0/0)	(1/0/0)

### Data mining model

The results from the experiment showed that JRip algorithm is the best algorithm to classify the student drop out status from our datasets. Therefore, we selected JRip to build the prediction model.

### Data mining model evaluation

In this section, we evaluate the purposed model. The classification metrics used for testing the performance of our purposed classifier is the statistical measurement from the confusion matrix: true positive rate (TP) and false positive rate (FP). The true positive rate is the pridction to be true is actually true. The false positive rate is the proportion of records which were classified as true but actually false. The true positive rate is equivalent to recall. The precision is the proportion of the records which truly have class x with all records which were defined as class x. The F-measure is a combined measure of precision and recall.

### Results and Discussion

In this section, we present prediction model rules generated by the algorithm JRip with the dataset of student record of Faculty of Science. The model is in a form of decision rules as shown below:

<p>Rule1: If Term = 2 then EducationStatus = Dropout</p> <p>Rule2: If GPX = range of 0.00-1.33 and EntranceMethod = UniversityDirectAdmission then EducationStatus = Dropout</p>
--



<p>Rule 3: If GPX = range of 0.00-1.33 and Major = GeneralScience then EducationStatus = Dropout</p> <p>Rule 4: If Term = 2 then EducationStatus = Dropout</p> <p>Rule 5: If EntranceMethod = CentralAdmission and Major = ComputerScience and GPX = range of 0.00 – 1.33 then EducationStatus = Dropout</p> <p>Rule 6: If EntranceMethod = CentralAdmission and Major = Physics and GPX = range of 0.00 – 1.33 then EducationStatus = Dropout</p> <p>Rule 7: If EntranceMethod = CentralAdmission and Province = Pattani and GPX = range of 0.00 – 1.33 then EducationStatus = Dropout</p>
---

Table 5 shows the performance of the JRip algorithm using statistical measurement which calculated from the confusion matrix, and the rates of true positive class.

**Table 5. Classifier Accuracy For Jrip**

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.871	0.227	0.891	0.871	0.881	0.636	0.826	0.869	Studying
0.773	0.129	0.773	0.773	0.754	0.636	0.826	0.742	Dropout

From our investigation, we found that the main attributes influencing student dropout are mostly concerning the academic aspect, specifically the term they are studying in and a low average grade. In addition to a low average grade, the method of entering the university and the major of the students also affect the student dropout. These results are aligned with previous results of Pereira and Zambrano [23] in which the dominant factors of student dropout from the university are related to academic factors such as a low average in grades, the semester of the program, and the faculty which the student belongs. Moreover, the location of the high school that students graduated from also related to the students’ dropout. Similar was stated by Rahman and Dash [36] who found that the location of student’s residence either rural or urban is related with the discipline they chose to study.

### Conclusion and Future Work

In this study, we investigated the student record of Faculty of Science, Prince of Songkla University. We analysed data from 2013 to 2017. In total, we have 4,238 records with 7 attributes such as admission method, major, education status, term of enrollment, grade point average of university, province of high school, and grade point average of high school in order to find a model that can predict student dropout. Among all data mining classifiers JRip showed the best accuracy, with 77.30%, when

compared all 6 classifiers. This model can help the administrative staff to plan the strategy to support the student in order to increase their retention in the faculty. As future works, we intend to expand the experiment with additional data such as the demographical information of the student, learning behavior with e-learning system, psychological factors. In addition, we will apply different data mining techniques such as association rules for investigating the relationships between socioeconomic and academic factors.

## Acknowledgment

My sincere thanks to the staff at the academic unit of faculty of science who have provided academic information for this research.

## Conflict of Interests.

There are non-conflicts of interest .

## References

- [1] J. Gairín, X. M. Triado, M. Feixas, P. Figuera, P. Aparicio-Chueca, and M. Torrado, "Student dropout rates in Catalan universities: profile and motives for disengagement," *Qual. High. Educ.*, vol. 20, no. 2, pp. 165–182, 2014.
- [2] R. Wirth and J. Hipp, "CRISP-DM: Towards a standard process model for data mining," in *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, 2000, pp. 29–39.
- [3] B. Leventhal, "An introduction to data mining and other techniques for advanced analytics," In *Journal of Direct, Data and Digital Marketing Practice*, vol. 12, no. 2, pp.137-153, 2010.
- [4] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.
- [5] B. Daniel, "Big Data and analytics in higher education: Opportunities and challenges," In *British journal of educational technology*, vol. 46, no. 5 pp. 904-920, 2015.
- [6] R. Llorente and M. Morant, "Data mining in higher education," in *New fundamental technologies in data mining*, IntechOpen, 2011.
- [7] A. M. Shahiri and W. Husain, "A review on predicting student's performance using data mining techniques," *Procedia Comput. Sci.*, vol. 72, pp. 414–422, 2015.
- [8] C. Romero and S. Ventura, "Educational data mining: a review of the state of the art," *IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.)*, vol. 40, no. 6, pp. 601–618, 2010.
- [9] M. M. N. Quadri and N. V Kalyankar, "Drop out feature of student data for academic performance using decision tree techniques," *Glob. J. Comput. Sci. Technol.*, 2010.

- [10] P. M. Arsad and N. Buniyamin, "A neural network students' performance prediction model (NNSPPM)," in *2013 IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA)*, 2013, pp. 1–5.
- [11] T. Patil, and S. Sherekar, "Performance Analysis of Naive Bayes and J48," *International Journal Of Computer Science And Applications*, vol. 6, no. 2, pp. 256-261, 2013.
- [12] S. Suthaharan, "Support vector machine," in *Machine learning models and algorithms for big data classification*, Springer, 2016, pp. 207–235.
- [13] C. Márquez-Vera, C. R. Morales, and S. V. Soto, "Predicting school failure and dropout by using data mining techniques," In *IEEE Journal of Latin-American learning technologies*, vol. 8, no. 1, pp. 7-14, 2013.
- [14] V. Ramesh, P. Parkavi, and K. Ramar, "Predicting student performance: a statistical and data mining approach," *International Journal of computer applications*, vol. 6, no. 8, pp. 35-39, 2013.
- [15] A. Tekin, "Early Prediction of Students' Grade Point Averages at Graduation: A Data Mining Approach," *Eurasian Journal of Educational Research*, vol. 54, pp. 207-226, 2014.
- [16] J. Namdeo, and N. Jayakumar, "Predicting Students Performance Using Data Mining Technique with Rough Set Theory Concepts," *International Journal of Advance Research in Computer Science and Management Studies*, vol. 2, no. 2, pp. 367-373, 2014.
- [17] M. Mythili and A. Shanavas, "An Analysis of students' performance using classification algorithms," *Journal of Computer Engineering*, vol. 16, no. 1, pp. 63-69, 2014.
- [18] F. Ahmad, N. Ismail, and A. A. Aziz, "The prediction of students' academic performance using classification data mining techniques," *Applied Mathematical Sciences*, vol. 9, no. 129, pp. 6415-6426, 2015.
- [19] C. E. Guarín, E. León Guzmán, and F. A. González, "A model to predict low academic performance at a specific enrollment using data mining," *IEEE Revista Iberoamericana de tecnologías del Aprendizaje*, vol. 10, no. 3, pp. 119-125, 2015.
- [20] G. Abu-Oda, and A. El-Halees, "Data Mining in Higher Education: University student dropout case study," *International Journal of Data Mining & Knowledge Management Process*, vol. 5, no. 1, pp. 15-27, 2015.
- [21] L. Aulck, N. Velagapudi, J. Blumenstock, and J. West, "Predicting student dropout in higher education" *2016 ICML Workshop on #Data4Good: Machine learning in Social Good Applications*, pp. 16-20, 2016. New York.
- [22] T. -O. Tran, H. -T. Dang, V. -T. Dinh, and X. -H. Phan, "Performance Prediction for Students: A Multi-Strategy Approach," *Cybernetics and Information Technologies*, vol. 1, no. 2, pp. 164-182. 2017.

- [23] R. T. Pereira, and J. C. Zambrano, J. C. "Application of Decision Trees for Detection of Student Dropout Profiles," In 16th IEEE International Conference on Machine Learning and Applications, 2017, pp. 528-531. IEEE.
- [24] K. Watkins, N. Na Phayap, and T. Jirasukprasert, "Intelligent Recommendation Systems using Data Mining techniques for Education Analytics for university," In The 2nd International STEM Education Conference, 2017, pp. 56. Chiang Mai.
- [25] N. Iam-On, and T. Boongoen, "Improved student dropout prediction in Thai University using ensemble of mixed-type data clusterings," International Journal of Machine Learning and Cybernetics, vol. 8, no.2, pp. 497-508, 2017.
- [26] L. Ramanathan, G. Parthasarathy, K. Vijayakumar, L. Lakshmanan, and S. Ramani, "Cluster-based distributed architecture for prediction of student's performance in higher education," Cluster Computing The Journal of Networks, Software Tools and Applications, 2018, pp. 1-16.
- [27] C. Burgos, M. L. Campanario, D. Pena, J. A. Lara, D. Lizcano, and M. A. Martínez, "Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout," In Computers & Electrical Engineering, vol. 66, pp. 541-556, 2018.
- [28] D. Bhanushali, S. Khan, M. Madhia, and S. Majumdar, "Student Performance Prediction And Analysis," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 7, no. 3, pp. 255-258, 2018.
- [29] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "Knowledge Discovery and Data Mining: Towards a Unifying Framework," In KDD, vol. 96, pp. 82-88, 1996.
- [30] V. García, A. Marqués, and J. Sánchez, "Non-parametric statistical analysis of machine learning methods for credit scoring," In Advances in Intelligent Systems and Computing, 2012, pp. 263-272. Berlin: Springer.
- [31] M. Mayilvaganan, and D. Kalpanadevi, "Comparison of Classification Techniques for predicting the performance of Students Academic Environment," In International Conference on Communication and Network Technologies (ICCNT), 2014, pp. 113-118. IEEE.
- [32] S. Kalmegh, "Analysis of WEKA data mining algorithm REPTree, Simple CART and RandomTree for classification of Indian news," *International Journal of Innovative Science, Engineering & Technology*, vol. 2, no. 2, pp.438-446, 2015.
- [33] A. Rajput, R. Prasad Aharwal, M. Dubey, S. Saxena, and M. Raghuvanshi, "J48 and JRIP rules for e-governance data," *International Journal of Computer Science and Security (IJCSS)*, vol. 5, no. 2, pp. 201-207, 2011.
- [34] Z. Muda, W. Yassin, M. N. Sulaiman, and N. I. Udzir, "Intrusion detection based on k-means clustering and OneR classification," in *2011 7th International Conference on Information Assurance and Security (IAS)*, 2011, pp. 192–197.
- [35] S. B. Aher and L. Lobo, "Comparative study of classification algorithms," *Int. J. Inf. Technol.*, vol. 5, no. 2, pp. 239–243, 2012.

- [36] A. Rahman, and S. Dash, "Data Mining for Student's Trends Analysis Using Apriori Algorithm," *International Journal of Control Theory and Applications*, vol. 10, no. 18, pp. 107-115, 2017.