Prediction Of College Academic Performance Of Senior High School Graduates Using Classification Techniques

Avigail Magbag, Rodolfo Raga Jr.

Abstract: The amount and variety of student data collected by the higher education institution, together with the potential of Educational Data Mining and analytics, provide a potential for the discovery of factors that contribute to the prediction of academic performance. The study focused on the development of a model for the prediction of academic performance of first year students in higher education. The aim is to allow early intervention that will help students stay on track and alleviate non-continuance. Predictors include pre-enrollment data such as demographics, High School performance and admission scores, combined with course taken in college. The Senior High School track or strand was introduced as a new predictor. These were collected from 4,762 first year students in the AY 2018-2019 from three Higher Education Institutions in Central Luzon. Classification techniques were used in the prediction of the performance level of the student based from the computed first year grade point average. Multiple experiments were conducted on individual and combined datasets, evaluated for accuracy, precision, recall and AUC to measure performance. Results revealed different model performances for each of the dataset, the one with the most instances got the highest accuracy at 68%. The accuracy when all data were combined was lower than the accuracy for one of the datasets but was higher than the other two. Analysis of the actual predictions was conducted to further understand the results and identify weaknesses that can be addressed in the next experiments.

Index Terms: academic performance prediction, classification techniques, senior high school

1 INTRODUCTION

The goal of elevating the quality of education in the Philippines led to the extension of the basic education to twelve years to make it at par with other countries. This initiative is known as the K-12 Program, signed into law as the Enhanced Basic Education Act of 2013 [1]. A defining feature of the K-12 program is the provision for career tracks in the Senior High School (SHS). Students are allowed to choose a specialization based on aptitude, interests, and school capacity for their upper secondary education. The choice of career track will define the content of the subjects a student will take in Grades 11 and 12 [2]. Each student in SHS can choose among three tracks: Academic; Technical-Vocational-Livelihood (TVL); and Sports and Arts. The Academic track is provided to students who wish to pursue a baccalaureate degree after SHS graduation. Students can choose among four strands compatible with the degree they wish to pursue. The strands are Accountancy, Business Management (ABM); Humanities and Social Sciences (HumSS); General Academic Strand (GAS) and Science, Technology, Engineering, Mathematics (STEM). According to the data from the Department of Education (DepEd), 763,593 students in Grade 12 were enrolled in the Academic track in Academic Year (AY) 2017-2018, comprising 61.70% of total students in the private and public SHS in the country. Those who took the TVL track make up 37.76% of total number of enrollees with only .37% and .17% taking Arts and Sports respectively [3]. These numbers are possible indication of the preference of students to pursue a college degree which may lead to better job opportunities [4],[5].

- *Avigail Magbag is a faculty member of the School of Computing of Holy Angel University. She is currently pursuing doctor of information technology degree in the Graduate School of Angeles University Foundation, Angeles City, Philippines. E-mail: apmagbag@gmail.com*
- *Rodolfo Raga Jr. is a PhD in Computer Science from the De La Salle University, Philippines. He is currently a full-time faculty member of the Computer Studies and Engineering Department of Jose Rizal University. E-mail: rodolfo.raga@jru.edu*

The preference of pursuing a college degree is also evident in the recent college enrollment statistics of the Commission on Higher Education (CHED) in the AY 2018-2019 which totaled to 3,212,542 with 982,049 new enrollments [6]. Although, the total enrollment does not represent SHS graduates entirely, it is an indication of the importance of college education to the students. But even with increasing registration, the overall completion rate is still very low, with only an average of 49% of first year students reaching their senior year and only 30% completion rate. An alarming 83.7% dropout rate was reported for enrollees from 2001-2012 [7]. With a fresh batch of SHS graduates who entered Higher Education Institutions (HEIs) in 2018 and with one of the thrusts of the program to prepare students for further education [2], the SHS track or strand is considered a new attribute for consideration as a predictor for academic performance. The relationship that may be established with the SHS track or strand, and the student performance in college may provide insights on how the students are prepared for higher education courses. The results may also be used as inputs in the review of the SHS curriculum in relation to the success of SHS graduates in college. The study focused on the development of a model for the prediction of academic performance of first year students in higher education using SHS background, entrance exam performance and the courses taken in college. The motivation behind the chosen group is to allow early intervention and appropriate support services that can make a difference in the students' academic success. Predicting students' performance on their first term in college allows for the institution to outline programs for students to help them stay on track and eventually graduate [8]. It is important to point out that based on CHED's admission policy, all Grade 12 graduates are eligible to enter college regardless of the track or strand in SHS [9]. Given this policy, it becomes even more important to predict ahead of time the type of support that SHS graduates would need once they move on to higher education especially those enrolled in a program that is not aligned with the track or the strand taken in SHS.

2 RELATED STUDIES

2.1 Importance of Early Prediction

The first year in college is a transition period for students coming from high school and the new learning environment introduces challenges they need to adjust to. Intervention programs in the universities can help them get through the stage of adjustment to college life when given in a timely manner. Prediction of the performance of college freshmen prove to be beneficial as it allows for the earliest possible intervention for students who are adjusting from high school to the University [10]. It has been found out that academic performance in the first year of college can be used as a predictor in determining student attrition and retention [11],[12]. The bulk of attrition occurs in the first two years at the university, with the rate of academic failures as one of the major reasons for high attrition level [13]. A means to identify low performing students at an early stage will prevent the student to drop-out through directed intervention. In return, the rate for non-continuance of freshmen may be alleviated since academic performance in the first year of college can be used as a predictor in determining student attrition and retention [14]. Graduation rate may also be improved through the retention of first year college students. Several studies conducted for prediction of graduation performance also showed first year academic performance as a strong predictor of graduating GPA [15],[16]. Raju and Schumacker [17] indicated high school GPA and first semester GPA in freshman year as two of the most important variables associated with retention leading to graduation.

2.2 Predictors of Performance

There are several attributes considered for the prediction of performance and determining which relevant attributes will have the most effect on performance led to continuous research in performance prediction. There are no standard set of indicators for performance prediction due to many factors pertaining to what is being predicted and what data is available in the academic institution. Several studies have been conducted to determine the variables that affect the performance of college freshmen. Pre-enrollment data collected include demographics, high school background and performance in college entrance exams. High school grades and admission scores which are a reflection of their previous performance are used as basis of predicting future performance due to the lack of current performance data of incoming freshmen [18],[19],[20],[21]. Admission score in particular proved to have a major impact on student performance as well as a contributing factor for student drop-out [22],[23]. In another study, only enrolment data were utilized for early prediction of student success, which combines demographics and the course enrolled by the students [24]. The challenges in performance prediction include the lack of standard set of predictors leading to limited portability amongst developed models [25]. It may be argued that the differences in findings are brought about by the differences on used variables, but even with similar predictors the results are not always strongly similar. For demographics, in some prediction gender and place of origin are found to be insignificant determinants [20] while in another research both gender and province were selected as predictors after applying filter model [26]. In the Philippines, lack of student data is still a concern for some HEIs while for some the concern is the management of multiple data in different formats in separate repositories. These are opportunities explored in the study. The data utilized were those commonly available across all HEIs pertaining to incoming freshmen.

3 METHODOLOGY

3.1 Dataset

The data were obtained from 3 HEIs in Central Luzon, specifically in the cities of Angeles, San Fernando and Olongapo. The subject of the study were first year students from AY 2018-2019 from 8 academic departments; Arts and Sciences, Business and Accountancy, Computer Studies, Criminology, Education, Engineering and Architecture, Hospitality and Tourism, and Nursing and Allied Medical Sciences. A total of 4,762 instances were derived from the consolidated records of the 3 HEIs. Table 1 presents the distribution of student records collected per HEI.

TABLE 1

LIST OF ATTRIBUTES

HEI	Freshmen
HEI A	3,588
HEI B	596
HEI C	415
Total	4,762

Data collected include pre-enrollment data available such as registration records, admission exam results and SHS background. Grades from the first semester of their freshman year were also collected and used the basis of the predicted academic performance.

3.2 Pre-processing

Several steps were taken to come up with the final dataset due to the diversity in the records of each HEI as well as the different formats of the data coming from multiple offices. Only those who have completed SHS with a strand or track were included. Records were matched using a unique identifier for each student to consolidate the registration records with those of the admission exam results and SHS background. Records with missing values were removed from the dataset. A total of 3,466 usable records remained from the collected data. The number of available features from each HEI varies, as such only those that were found similar across the HEIs were selected for modelling. Table 2 lists these attributes:

TABLE 2

LIST OF ATTRIBUTES

Attribute	Description
ID Number	Unique Identifier
age	Age
gender	Gender [Male/Female]
municipality	Municipality
province	Within or Outside the Province
SHS name	Name of SHS
type	Type [Public/Private]
strand	SHS Track / Strand
GWA	Grade Weighted Average of Grade 12
english	Admission score in English Exam
math	Admission score in Math Exam
verbal	Admission score in Verbal Exam
non-verbal	Admission score in Non-Verbal Exam
college	College or Department
course	College Course / Program
units	Total Units Enrolled
GPA	First Year GPA

Labels were replaced for uniformity across all records, specifically for college and department names. Due to the differences in the range of scores for the admission exams for each HEI, the values were normalized by span for each set within the range of 0 to 1. The grade weighted average from SHS was also normalized within the same range. The GPA was derived from the grades in each of the courses enrolled in the semester computed according to the specific HEI's policy. The GPA was the basis for the academic performance which was categorized as either high, medium or low. The basis for the high category corresponds to the required grade for an academic award at the end of the semester while low performing includes not only the failing students but those with grades that also need intervention to perform better in the semester. Prior to model development, feature selection was conducted using Correlation-based Feature Selection (CFS), Gain Ratio and Information Gain were used for feature ranking. These methods were utilized to rank the attributes of most importance for predicting the performance [26][27]. Feature selection was conducted for each dataset before merging, and features were ranked again after merging. Features selected on individual datasets vary in rank and number. Table 3 lists the corresponding features for each dataset selected using feature ranking. Municipality, SHS name and type were not selected in all of the datasets in the process, while GWA, college, course, units and math were included in all. Gender and strand were also ranked in all datasets except for HEI C.

TABLE 3

FEATURES SELECTED PER DATASET

Dataset	List of Features
HEI A	GWA, college, course, gender, strand, units, english, verbal, math and non-verbal
HEI B	gender, course, college, strand, units, province, GWA, age, math and non-verbal
HEI C	GWA, course, units, college, math, english, general and science
Merged	course, college, GWA, units, strand, gender, english and math

3.3 Modelling and Evaluation

After pre-processing, the dataset was divided into training and testing data with a 70-30 percent distribution. The training data was fitted to Logistic Regression (LR) and Neural Network (NN) for modelling. These were selected based from the nature of data available for prediction [28] as well as from the results from an initial study conducted using pre-enrollment data [29]. Experiments were conducted to compare the performance score of both algorithms. Performance metrics include AUC, F-measure, accuracy rate, precision and recall. Stratified 10-fold cross validation was utilized to further divide the training set for training and validation. The data was shuffled randomly before splitting into folds and fitted to the algorithms. The performance metrics were measured in each fold and the mean was computed after all iterations. Standard deviation was also computed to measure how close the accuracy results for each validation were to the mean. The dataset collected presented an imbalanced distribution of the instances. To resolve this issue, a technique known as Synthetic Minority Over-sampling Technique (SMOTE) was applied. It works by creating synthetic observations based on existing minority observations by calculating the k-nearest neighbor for each observation. To avoid overfitting, SMOTE was only applied to the minority class during model training on the validation set for each fold in the cross validation [30]. The models were also evaluated using the 30 percent of the data as testing set to measure how they will perform on unseen data and how they compare with the training results.

4 RESULTS AND DISCUSSION

4.1 Experiments

Several experiments were conducted for the purpose of comparing the performance of the classification algorithms on the sets of data. The values are the means of the performance measures after the 10-fold cross validation. Table 4 presents the performance of Logistic Regression for each HEI's dataset during training. It must be noted that the HEI with the greatest number of instances recorded the highest values on all metrics while the one with the least instances recorded the lowest scores.

TABLE 4

TRAINING RESULTS USING LOGISTIC REGRESSION

HEI	Accuracy	Precision	Recall	F-score	AUC
HEI A	60.97	66.53	60.97	62.13	68.97
HEI B	50.33	57.65	50.33	48.59	62.03
HEI C	44.49	51.43	44.49	45.25	55.79

Table 5 presents the performance of Neural Network for each HEIs dataset during training. Similar with Logistic Regression, the HEI with the most number of instances recorded the highest values on all metrics although the lowest scores were mostly on the second HEI.

TABLE 5

TRAINING RESULTS USING NEURAL NETWORK

HEI	Accuracy	Precision	Recall	F-score	AUC
HEI A	64.02	67.94	64.02	64.37	70.09
HEI B	46	60	46	46	60.7
HEI C	46.48	50.01	46.48	46.07	55.85

Neural Network performed better for the first and last HEIs while Logistic Regression was a better fit for the second HEI except in precision. The number of features in the prediction models for each dataset varies according to the ranked features during selection. It was important to note that in HEI C, the SHS strand was not included as one of the selected features. Table 6 and 7 show the results of both algorithms for the combined datasets, where Neural Network outperformed Logistic Regression on all metrics. In Table 5, all the attributes similar across all datasets were included, while Table 6 only includes the 8 features selected after applying feature ranking on the combined datasets which included the SHS strand. It must be observed that the scores on the metrics decreased when some features were eliminated based from feature ranking.

TABLE 6
TRAINING RESULTS ON COMBINED DATASETS

HEI	Accuracy	Precision	Recall	F-score	AUC
LR	48.43	58	48.43	49.84	61.97
NN	56.98	64.66	56.98	58.58	66.93

TABLE 7
TRAINING RESULTS ON COMBINED DATASETS WITH FEATURE RANKING

HEI	Accuracy	Precision	Recall	F-score	AUC
LR	48.06	57.86	48.06	49.25	61.67
NN	55	63.35	55	57.2	65.93

The standard deviation of the accuracy scores of both classifications were very low at .02 to .04 indicating that the values computed in each fold for cross validation are close to the mean accuracy score.

4.2 Classification Performance
Testing was also conducted on unseen data to validate the consistency of the performance of the models. Tables 8 and 9 present the testing results for Logistic Regression and Neural Network on each dataset respectively. The results were consistent with the HEI having the most instances getting the highest scores in all metrics. It is important to note though that the metrics in all the datasets improved for both classifications.

TABLE 8
TESTING RESULTS USING LOGISTIC REGRESSION

HEI	Accuracy	Precision	Recall	F-score	AUC
HEI A	61.18	66	61	62	71.63
HEI B	51.84	59.53	51.84	52.35	63.27
HEI C	51.96	63	52	55	62.92

TABLE 9
TESTING RESULTS USING NEURAL NETWORK

HEI	Accuracy	Precision	Recall	F-score	AUC
HEI A	68	70	68	69	72.65
HEI B	56	64	56	59	64.96
HEI C	54	63	54	56	65.83

Tables 10 and 11 on the other hand, show the testing results of both algorithms for the combined datasets, where Neural Network outperformed Logistic Regression on all metrics. The results are consistent in terms of the metric scores between the datasets, with the first dataset having all similar features included getting higher scores. Similar with the results from the testing with each dataset, all metrics improved for both classifications on unseen data.

TABLE 10
TESTING RESULTS ON COMBINED DATASETS

HEI	Accuracy	Precision	Recall	F-score	AUC
LR	52.55	62	53	54	66.5
NN	62.66	66	63	64	69.54

TABLE 11
TESTING RESULTS ON COMBINED DATASETS WITH FEATURE RANKING

HEI	Accuracy	Precision	Recall	F-score	AUC
LR	51	59	51	52	64.97
NN	59	65	59	60	70.42

In the analysis of the confusion matrix, it was seen that Logistic Regression was able to correctly classify high performing students at a rate of 75% and low performing students at 72% compared to Neural Network at 76% and 54% respectively on HEI A's dataset. On the combined datasets with selected features, Logistic Regression was consistent at better predicting low performing students at 69%, while Neural Network predicts high performing students better at 78% when all features were included. Although, Neural Network recorded a lower value at predicting low performing students as high, which is consistent with the initial experiments on similar data [29].

5 CONCLUSION AND FUTURE WORKS
The main goal of the study was to predict the academic performance of freshmen students in college using data that are available prior the start of the first term. The attributes that were included are gender, SHS grade and strand, and entrance exam performance as pre-enrollment data and freshmen details such as college, course and the number of units enrolled. The SHS strand as a new attribute proved to be a predictor for academic performance being included in the ranking using feature selection algorithms on two individual datasets and the combined datasets. The results showed that end of semester performance can be predicted at a rate of almost 63% accuracy at the start of the semester with only data collected before the start of classes. Although it is important to note that when evaluated separately, the HEI with the most number of instances reached an accuracy of 68%. This is consistent with the initial results conducted on a single set of data [29]. The results are also comparable with those of [24] where a 60.5% accuracy rate was achieved using CART for Information System students, while [31] recorded a 50.39-58.63% accuracy rate on training and testing respectively using Naïve Bayes for Engineering students. Reference [32] also recorded a range of 52-67% prediction rates on six different classifiers using data prior the start of the term. Considering that these studies only include students from the same programs, the current study included students from courses across three different institutions and achieved a slightly higher accuracy rate. Due to factors including uniqueness of data collected by each HEI, results may not be generalizable to all institutions. It might be considered by these institutions to include in their collection similar data that were ranked as predictors for academic performance. Also,

the models considered all incoming freshman students across three HEIs in all programs. Investigating smaller units of students, such as those in similar fields or specializations may allow a more accurate prediction of academic performance [33]. It is also important to state that Neural Network provided a higher accuracy at predicting high performing students with very low percentage of predicting low performing students to belong to high performing students. These results can be used as a starting point for planning intervention services for students who are at risk of low performance at the earliest time possible. Neural Network will be further investigated for the model development, with consideration of optimizing parameters for deep learning to improve prediction performance. Further study will be conducted with an increased number of data samples from other Universities and Colleges to reveal other data available prior the start of classes that may be used for prediction. Also, to support or negate the observation in the current study that an increase in the number of instances does not necessarily equate to an improved model performance. A web-based application will also be developed to allow improved collection and storage of data that may help enhance data collection and preparation prior to modelling. The application will also utilize the prediction model to help identify students who might be at risk of failing.

REFERENCES

[1]. Enhanced Basic Education Act of 2013, Retrieved from https://www.officialgazette.gov.ph/2013/05/15/republic-act-no-10533/

[2]. The K-12 Basic Education Program, Retrieved from https://www.officialgazette.gov.ph/k-12/#section-7

[3]. J. Nelz, “First Batch of Students of K-12 Program Graduate Next Month,” Retrieved from Phil News: https://philnews.ph/2018/03/27/first-batch-students-k-12-program-graduate-next-month/, March 27, 2018.

[4]. Tymon and S. Batistic, “Improved Academic Performance and Enhanced Employability? The Potential Double Benefit of Proactivity for Business Graduates," Teaching in Higher Education, 21:8, pp. 915-932, 2016, doi: 10.1080/13562517.2016.1198761.

[5]. P. Carnevale, S. J. Rose and B. Chea, “The College Payoff: Education, Occupations, Lifetime Earnings,” Retrieved from https://repository.library.georgetown.edu/handle/10822/559300, 2013.

[6]. CHED, Retrieved from https://ched.gov.ph/statistics/, 2018.

[7]. T. Tanhueco-Tumapon, “Creating a Culture of Student Engagement,” The Manila Times, September 8, 2017.

[8]. V.L. Miguéis, A. Freitas, P.J.V. Garcia and A. Silva, “Early Segmentation of Students According to their Academic Performance: A Predictive Modeling Approach,” Decision Support System, 2018, doi:10.1016/j.dss.2018.09.001.

[9]. CHED, Retrieved from https://ched.gov.ph/cmo-105-s-2017/, 2017.

[10]. G. Gray, C. McGuinness, P. Owende and M. Hofmann, “Learning Factor Models of Students at Risk of Failing in the Early Stage of Tertiary Education,” Journal of Learning Analytics, 3(2), pp. 330–372, 2016, doi:10.18608/jla.2016.32.20.

[11]. J. R. Casanova, A. Cervero, J. C. Núñez, L. S. Almeida and A. Bernardo, “Factors that Determine the Persistence and Dropout of University Students,” Psicothema, 30(4), pp. 408–414, 2018, doi:10.7334/psicothema2018.155.

[12]. V. Hegde and P. P. Prageeth, “Higher Education Student Dropout Prediction and Analysis Through Educational Data Mining,” Proc. 2nd International Conference on Inventive Systems and Control, IEEE, pp. 694–699, 2018, doi: 10.1109/ICISC.2018.8398887.

[13]. P. Sander, “Using Learning Analytics to Predict Academic Outcomes of First-year Students in Higher Education,” Capstone Report, University of Oregon Applied Information Management Program, 2016.

[14]. M.E. Orozco and J. Castro-Niguidula, “Predicting Student Attrition Using Data Mining Predictive Models,” 143rd The IIER International Conference, pp. 20-26, 2016, unpublished.

[15]. R. Asif, S. Hina and S. I. Haque, “Predicting Student Academic Performance Using Data Mining Methods,” IJCSNS International Journal of Computer Science and Network Security, Vol. 17, 2017.

[16]. S. Gershenfeld, D. Hood and M. Zhan, “The Role of First Semester GPA in Predicting Graduation Rates of Underrepresented Students,” Journal of College Student Retention: Research, Theory & Practice, pp. 469-488, 2016.

[17]. D. Raju and R. Schumacker, “Exploring Student Characteristics of Retention that Lead to Graduation in Higher Education Using Data Mining Models,” Journal of College Student Retention: Research, Theory & Practice, 16(4), pp. 563–591, 2018, doi: 10.2190/cs.16.4.e.

[18]. H. Chanlekha and J. Niramitranon, “Student Performance Prediction Model for Early-Identification of At-Risk Students in Traditional Classroom Settings,” Proc. 10th International Conference on Management of Digital EcoSystems (MEDES ’18), September 25-28, 2018, Tokyo, Japan, ACM, New York, NY, USA, pp. 239–245, doi: 10.1145/3281375.3281403.

[19]. S. Hoffait and M. Schyns, “Early Detection of University Students with Potential Difficulties,” Decision Support Systems, 2017, doi:10.1016/j.dss.2017.05.003.

[20]. R. Kanagi, C. Tan, K. Sarimila, K. Lim, K. Haslina and D. Khezrimotlagh, “Factors Affecting First Year Undergraduate Students Academic Performance, “ Scholars Journal of Economics, Business and Management, pp. 54-60, 2015.

[21]. P. A. Westrick, H. Le, S. B Robbins, J.M.R. Radunzel and F.L. Schmidt, “College Performance and Retention: A Meta-Analysis of the Predictive Validities of Act Scores, High School Grades, and

SES," Educational Assessment, 20(1), pp. 23–45, 2015, doi:10.1080/10627197.2015.997614
[22]. Montalbo, Y. Evangelista and M. Bernal, "Admission Test as Predictor of Student Performance in Political Science and Psychology Students Of Rizal Technological University," Asia Pacific Journal of Multidisciplinary Research, 6(3), pp. 68–73, 2018.
[23]. D.V. Silfverberg and A.C. Orbeta, "The Role of Entrance Exams in Academic Performance of Students with Low Socioeconomic Background: Evidence from the SGP-PA," Proc. 13th National Convention on Statistics, p.30, 2016.
[24]. J. Z. Kovacic, "Early Prediction Of Student Success: Mining Students Enrolment Data," Proc. 2010 InSITE Conference, pp. 647–665, 2010, doi: 10.28945/1281.
[25]. R. Conijn, C. Snijders, A. Kleingeld, and U. Matzat, "Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS," IEEE Trans. on Learning Technologies, 10(1), pp. 17–29, 2017, doi:10.1109/TLT.2016.2616312.
[26]. J.D. Febro and J. Barbosa, "Mining Student at Risk in Higher Education Using Predictive Models," Journal of Advances in Technology and Engineering Research, 3(4), pp. 117–132, 2017, doi:10.20474/jater-3.4.2.
[27]. E.B. Costa, B. Fonseca, M.A. Santana, F.F. de Araújo and J. Rego, "Evaluating the Effectiveness of Educational Data Mining Techniques for Early Prediction of Students' Academic Failure in Introductory Programming Courses," Computers in Human Behavior, 73, pp. 247–256, 2017, doi:10.1016/j.chb.2017.01.047.
[28]. Shahiri, H. Wahidah & N. Rashid, "A Review on Predicting Student's Performance Using Data Mining Techniques," Procedia Computer Science, pp. 414-422, 2015.
[29]. Magbag and R. Raga, "Prediction of College Academic Performance of Senior High School Graduates Using Machine Learning," Proc. International Conference on Emerging Technologies 2019, Daejeon, South Korea, 2019, unpublished.
[30]. M. Santos, J. Soares, P. Henriques Abreu, H. Araujo and J. Santos, "Cross-Validation for Imbalanced Datasets: Avoiding Overoptimistic and Overfitting Approaches," IEEE Computational Intelligence Magazine, 13, pp. 59-76, 2018, doi:10.1109/MCI.2018.2866730.
[31]. E. P. I. García and P. M. Mora, "Model Prediction of Academic Performance for First Year Students," 10th Mexican International Conference on Artificial Intelligence, IEEE, pp. 169-174, 2011, doi: 10.1109/MICAI.2011.28.
[32]. K. Verma, A. Singh and P. Verma, "A Review on Predicting Student Peformance Using Data Mining Method," International Conference on Futuristic Trends in Engineering, Science, Humanities, and Technology (FTESHT-16), Gwalior, pp. 124-129, 2016.
[33]. A.M. Gansemer-Topf, J.I. Compton, D.R. Wohlgemuth, G.R. Forbes, and K. S. Ralston, "Modeling Success: Using Pre-enrollment Data to Identify Academically At-Risk Students," Education Publications, 37, 2015.