iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** V **Month of publication:** May 2020

DOI: http://doi.org/10.22214/ijraset.2020.5055

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

A Prediction Model for Student Attrition Using J48 Classification

Ruffa R. Ribot[1], Van M. Ribot[2], Joann G. Perez[3], Gerald T. Cayabyab[4]
[1, 3, 4]Bulacan State University, Bulacan State University, Bulacan State University, TIP, QC. Philippines

Abstract: In this paper we propose a prediction model for student attrition using J48 classification to analyse at risk student and determine what are the significant factors affect the student not to finish the program on time. Our approach for this prediction model used J48 classification algorithm. This is a kind of machine learning software. J48 is a C45 algorithm, an extension of Decision Tree machine learning. In machine learning feature selection analysis is often serve as a pre-processing stage. This tree can be used for predicting the student's enactment as pass or fail. Once the student is found at risk he/she can be provided guidance for intervention. Therefore, the model can be take advantage to identify students who are most likely at risk and assist them in succeeding better grades.
Keywords: Machine learning, Attributes Selection, Prediction Classification, Decision Tree, Data mining

I. INTRODUCTION

Various institutes have long know the need to support students who are at risk of failure in different courses. The growing usage of digital platforms such as Learning Management Systems (LMS) [1] in educational institutions has formed a massive gold mine for educators wanting to scrutinize student learning behaviour thoroughly. This kind of concern reminds universities researches to identify red flags for academic failure [3]. The development of predictive models for identifying at-risk students has increased significantly during the past decade. Early detection of such students will allow educators to employ suitable intervention procedures to guide the students to stay on track towards successful completion of the course.

The achievement of good honors in Undergraduate Degrees [6] is essential in the condition of Higher Education (HE), both for students and for the institutions that host them. An accurate assessment [7] would help in effective planning of the steps needed to adapt education toward individual needs. Traditionally, such estimates are performed subjectively. Teachers typically assess each student based on the student's recent academic performance and the teacher's knowledge about the experience with similar students in the past, such as students with similar demographics, socioeconomic background, and behavioural patterns.

Identifying at-risk students as soon as possible is a challenge in educational institutions. Decreasing the time lag between identification and real at-risk state may significantly reduce the risk of failure or disengage. It enables instructors to develop corresponding instructional interventions via course design or student-teacher communications. The classification approach has the potentials for identifying at-risk situations in a real-time manner. All the stakeholders within the educational institution can be benefited from the output of these analytical tools (i.e., administrators, teachers, and students). The system aims to correctly identify attrition students and provide early feedback based on the chances to finish the course. At-risk students receive personalized feedback to amend the possible situation of failure in their learning process and increase the probability of passing the course.

This study focuses on risk-factors for attrition (drop out) of 768, and 310 students enrolled in the Information Technology program and in BSBAE in Bulacan State University-Sarmiento Campus in the year 2012-2016 & 2013-2017. Demographic factors include socioeconomic status, gender, age, achievement attributes such as High School point Math Average, High School Type, First-year first semester general weighted average, First-year second semester general weighted average, Father Educational Attainment, Mother and Educational Attainment. Furthermore, the effects of financial incentives was considered the parent's monthly income. Consistently with the scope of this study, a review is restricted to those factors that have been previously mentioned. Moreover, transferee students are not included that have earned units from other Universities. The current study builds on identifying predictors of the attrition students at Bulacan State University, particularly in Bachelor of Science in Information Technology and Bachelor of Science in Business Administration major in Entrepreneurship.

The proposed work follows the recent research trends that focus on the objective of utilizing the prediction results in gearing up the student's academic potential during their on-going period of studies. The objective of this study is to develop a predictive modelling method to address the at-risk student who completed their studies in 2016 and 2017 with the J48 approach, an application of classification models. However, due to different learning environment might have its unique predictors, models develop for Higher Education might not apply to other school or other levels.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

II. LITERATURE REVIEWS

The development of an early feedback prediction system for at-risk students had been one of the utmost concerns of institutions. The simple Gradual At-risk (GAR) model based on students' grades has been designed [9], and its accuracy has been analysed using the data available in the whole institution. Although the system can be generalized to any type of course, the paper evaluates the system on a first-year undergraduate course within an online higher education university. These students have many difficulties since they are newcomers in higher education. Based on these needs, a few predictive algorithms and early-warning systems aimed at identifying at-risk students for early interventions have been developed [2].

To improve student academic performance, especially undergoing graduate and postgraduate courses, many experts have attempted to explore various methodologies. To date, a multitude of feature selection models has been developed that aim to extract those combinations of attributes yielding improved classification accuracy [10]. According to [11], prediction accuracy is the most important indicator for model evaluation and adopted the final grade (passed or failed) as their target variable to reflect a student's learning outcome.

[12] Utilize the following attributes; if the student enrolled the first time, Students' major, department offering major, Sports, High education for father and mother, standard test score (math) critical reading, writing, High School GPA (HSGPA) class size to determine significance of discrete attributes. They confirmed that the accuracy increased when the attributes addressing students' success in the first time semester becomes available. In addition, results indicate that the feature selection procedure based on a combination of expert knowledge and statistical test for significance can lead to useful classifiers. However, based on the study to further improve, the accuracy can be achieved by utilizing additional features (demographic, socio-economical, and academic) and/or by employing methods for automatic feature selection. In the study of [13] mention that gender is another relevant feature used in predicting. Moreover, according to [14], the most obvious measurement of individual academic performance is a student's average mark or grades.

Based on the study conducted by [15] among the several classifiers available in the literature, the Decision Tree (DT) is widely used in the health informatics area. According to [16] Decision, Tree-J48 is very simple and efficient. In addition, [17] in contribution J48 classifier combining with term weighting concept as weighted j48 classifier is used for classification. These methods increase the accuracy of the classification and feature selection process and improve system performance.

Moreover, [15] compared the performance of Decision Tree and Genetic Programming methods to distinguish healthy from post-stroke individuals. According to the result of the study Decision Tree method has shown better overall performance than GP, although the two classifiers presented the same performance during the recognition of stroke sway profiles. DT allowed a more accurate recognition of body sway, higher predictive power than Genetic Programming.

A feature selection process is also necessary to remove irrelevant, noisy or redundant features, which will improve the accuracy and interpretability of the final classifier, and also will decrease its computational complexity, allowing even the implementation of real-time systems In wrapper methods, the performance of a learning algorithm is used to assess the quality of selected feature subsets, while in filter model criterion functions evaluate feature subsets by their information content, rather than directly [21].

The J48 decision tree is an open-source Java implementation of a commonly known C4.5 supervised classification algorithm in WEKA. It is a development and extension of the ID3 (Interactive Dichotomizer 3)algorithm developed by Quinlan[28]. At this point there are numerous methods to determine statistically how accurate the output model is. There are various methods such as a percentage-split separating two thirds of the data for training and generating the model then using the final third for verification purposes, [29]. BestFirst search method has been used with default parameter settings.

III.METHODOLOGY AND DATASET OVERVIE

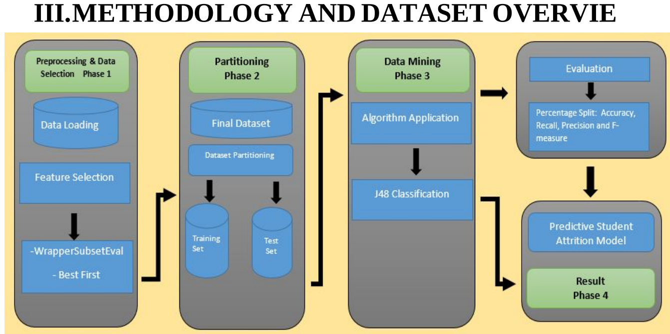


Figure 1 Model Framework

Figure 1 illustrates how the study works in which the model's formulation is composed of five great phases. The initial part is the pre-processing & data preparation in which data will be loaded in the tools then pre-processing, where the data needs to select the best features. The research uses WrapperSubsetEval for the attribute evaluator and Best First for the search method. In data mining, this involves a split factor that used (training data: testing data). This phase is about the data mining process using the J48 algorithm. The previously mentioned phases addressed matters such as data extraction and pre-processing relevant attribute selection and identification, data mining techniques for result optimization[24]. The last phase will be the result, which will take into account factors such as classification coverage, accuracy, precision, recall, and f-measure to predict the academic performance of the data mining model.

The important highlights of this work are the amount of data used. As mentioned earlier, the data used is from Information Technology and BSBAE, two of the highly regarded course in Bulacan State University –Sarmiento Campus. Bulacan State University has the largest school system in the province of Bulacan. Although we used data from this particular bachelor degree course, the school risk prediction framework is rough, that is why we decided to come up with this study.

TABLE 1 Student Background

1. Age	Age admitted to the university	Numeric
2. Gender	Student gender	Nominal
3. High School Type(Public/Private)	Type of High School	Nominal
4. High School GPA	High School Grade point Ave.	Numeric
5. High School Math Average	4th yr. High School Math Ave.	Numeric
6. 1st yr. 1st Sem GPA	Grade Point Average in first-year first semester	Numeric
7. 1st yr. 2nd Sem GPA	Grade Point Average in first year second semester.	Numeric
8. Parents Monthly Income	Family Monthly Income	Numeric
9. Father Educational Attainment	Highest educational attainment of the father	Numeric
10. Mother Educational Attainment	Highest educational attainment of the father	Numeric
11. Completed	The student completed the program on-time	Nominal

Data that contain of values of all attributes describing a study in a point of time have been extracted from the system, and the others are manually collected from the Registrars' office. The data set contained 768 students, 339 females, and 429 males in total with the BSIT program and 339 students form BSEntrep 230 female and 109 males. Both datasets consist of 11 attributes and 1 class attributes, some attributes are numeric, and some are nominal. The dataset is in .arff format.

The datasets were extracted from the database in structured comma-separated-value (CSV) text files, which then were transformed into an Attribute-Relation File (ARFF) –the default file format used by the data mining suite for WEKA. Then it was loaded to WEKA. It was an extremely powerful visualization tool in analysis and attribute selection.

WEKA – means Waikato Environmental for Knowledge Analysis, which is an innovator device in the historical backdrop of the data mining and machine learning research groups. By putting endeavours since 1994, this device was created by WEKA group. WEKA contains numerous inbuilt algorithms for data mining and machine learning. It is open-source and unreservedly accessible stage free programming[30].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

To gain a good sense of data, it begins with pre-processing. Data Pre-Processing is for making the quality of the dataset better. Several processes are being done in data pre-processing such as integration, feature selection, cleaning, reduction, and transformation. In this study, there were two datasets being used. Data pre-processing has the goal of making data suitable for the purpose of analysis. It improves the quality as well and fits a specific data mining technique or tool better [27]. From the selection of 11 data columns, the extreme right column appears to be the classifier.

In this study full training set is used for attribute selection mode. WrapperSubsetEval used as an attribute evaluator. It evaluates attribute sets by using a learning scheme. Percentage split is being used to estimate the accuracy of the learning scheme for a set of attributes. Together with a non- parametrized J48 algorithm. The decision tree has used J48, is an implementation of C4.5. C4.5 is a class for generating a pruned or unpruned C4.5 decision tree. The confidence factor which we have used is 0.25. For the search method, BestFirst is being used. It searches the space of attribute subsets by greedy hill-climbing augmented with a backtracking capacity. Setting the number of consecutive non-improving nodes permits controls the level of backtracking done. Best first may begin with the empty set of attributes and search forward, or start with the full set of attributes and search backward, or begin at any point and search in both directions (by considering all possible single attribute additions and deletions at a given point). Then, instances that didn’t contribute much to the model were removed.

With the experiment, a split factor that has used (training data: testing data) is 70:30. In this study, we will test all the data converted, examine the logic of tree relation generated, and try to find the suitable variable to be the indicator factor for this prediction. The variable that shows the logic tree relation with a higher accuracy rate will be considered as an appropriate indicator or variable to predict.

The classification is based on the J48 Decision Tree algorithm, which has been proposed in this study. According to [18], decision Tree is constructed using selected optimal features and intelligent agents to ensure maximum accuracy of classification from the dataset. These are the steps to the proposed classification algorithm are as follows.

(1) For selected attributes and using records in the datasets, you now have to choose a classifier which happened that J48 algorithm is being used.

(2) Then, decide what test option you have to do; Use training set, Supplied test set, Cross-validation folds, and Percentage split if how many percent you are going to test.

(3) After choosing the test option, you are now able to see the results.

The result obtained for the application of the feature selection and classification algorithms proposed in this study will be used for the performance analysis of the attrition students of two program; Bachelor of Information Technology and Bachelor of Science in Business Administration major in Entrepreneur from Bulacan State University-Sarmiento campus graduated in 2016 and 2017.

Evaluation of the algorithms is one of the key points in any process of data mining. The most common tools used in analysing the results of classification algorithms applied. Confusion matrix, learning curves and receiver operating curves (ROC) have been measured for accessing the performance accuracy of the proposed classifier. The confusion matrix shows the number of correct and incorrect predictions made by the model compared with the actual classifications in the test data. Confusion matrix for a classifier with two classes True and False [22]. The performance was measured in terms of accuracy (the number of correctly classified samples over the number of all samples) and True Positive Rate (the number of correctly classified samples from the class of unsuccessful students). In this study, the evaluation parameters are measured in terms of precision (1), recall (2), F-measure (3), accuracy (4), time taken to build a model, correctly classified instances percent, and incorrectly classified instances percent.

$$\text{Precision} = TP/(TP+FP) \quad (1)$$

$$\text{Recall} = TP/(TP+FN) \quad (2)$$

$$\text{F-measure} = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (3)$$

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN) \quad (4)$$

Where:

TP = the number of positive cases is classified as positive

TN = the number of negative cases is classified as positive

FP = the number of negative cases is classified as negative

FN = the number of positive cases is classified as negative

Accuracy is the proportion of the total number of correct predictions and calculated as the ratio between the number of cases correctly classified and the total number of cases.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

IV. RESULTS AND DISCUSSION

As a result of this study, the experiment used a popular data mining tool called WEKA. Figure 2 & 3 shows the output generated by Weka under classification using the J48 algorithm. We used the J48 decision tree for this experiment. Color coding, which means blue was the one finished the program, and red means didn't complete the course on time. The sample training data consist of two programs; Bachelor of Information Technology (768 datasets) and Bachelor of Science in Entrepreneurship (339 datasets). In testing mode, percentage split was being used 70 % training and 30 % in testing.

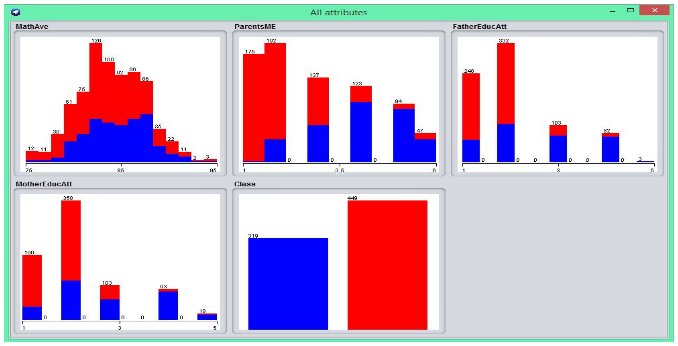


Figure 2 Attributes Visualizer (BSIT)

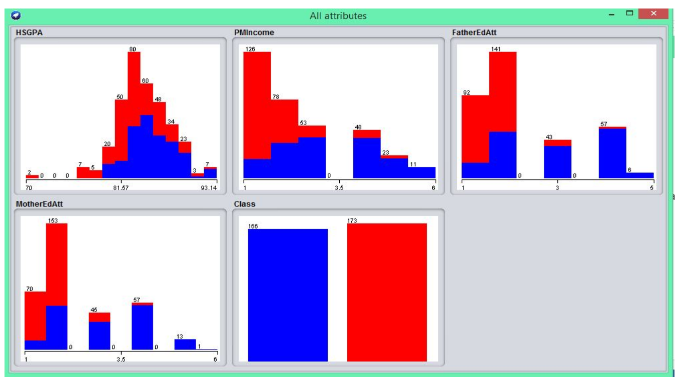


Figure 3 Attributes Visualizer (BS Entrep)

Table 2 Selected Features using Weka Tool

Attributes	
BSIT	**BS Entrepreneurship**
High School Math Ave.	High School General Point Average
Parents Monthly Income	Parents Monthly Income
Father Educational Attainment	Father Educational Attainment
Mother Education Attainment	Mother Education Attainment

Table 2 shows the experiment results features selected by Weka tool with the two programs being used in this study. Among 11 attributes with the BSIT & BS Entrep course, four features appeared on the experiment using J48 classifiers.

```
ParentsME <= 3
|   FatherEducAtt <= 2
|   |   MotherEducAtt <= 2: NO (394.0/41.0)
|   |   MotherEducAtt > 2
|   |   |   MotherEducAtt <= 3
|   |   |   |   MathAve <= 86.25: NO (26.0/3.0)
|   |   |   |   MathAve > 86.25: YES (3.0)
|   |   |   MotherEducAtt > 3: YES (20.0/4.0)
|   FatherEducAtt > 2
|   |   ParentsME <= 1: NO (13.0)
|   |   ParentsME > 1
|   |   |   MathAve <= 80: NO (4.0)
|   |   |   MathAve > 80: YES (44.0/8.0)
ParentsME > 3: YES (264.0/44.0)

Number of Leaves  :     8

Size of the tree :      15
```

Figure 4 Decision Tree Generated by Weka (BSIT)

```
FatherEdAtt <= 2
|   MotherEdAtt <= 3
|   |   PMIncome <= 2: NO (165.0/23.0)
|   |   PMIncome > 2
|   |   |   HSGPA <= 83.8: NO (26.0/7.0)
|   |   |   HSGPA > 83.8: YES (19.0/1.0)
|   MotherEdAtt > 3: YES (22.0/1.0)
FatherEdAtt > 2
|   PMIncome <= 1
|   |   FatherEdAtt <= 3: NO (7.0)
|   |   FatherEdAtt > 3: YES (4.0)
|   PMIncome > 1: YES (96.0/3.0)

Number of Leaves  :     7

Size of the tree :      13
```

Figure 5 Decision Tree Generated by Weka (BS Entrep)

In Figure 4&5, in this case, we see two figures the BSIT and BS Entrep. As an output, we got a decision tree. A tree visualization generated by Weka. In the first program, the total number of leaves were 8, and the size of the tree is 15. While on the second program, there are 7 number of leaves and 13, size of the tree. The time taken to build the model is 0.01 second. For the evaluation, the time to test the model on the test split is 0 second. To build a model, correctly classified 195 (84.7826%) instances and 35 (15.2174%) instances incorrectly. While on BS Entrep, the correctly classified instances are 94, which equivalent to 92.1569%, and the incorrectly classified instances are 8 (7.8431%). Figures 6 below shows a detailed summary of two programs.

A. Confusion Matrix

A confusion matrix is a method that has the primary function to calculate the performance of a classification model according to the calculation of testing data, which the result of predictions are in two classes, namely positive and negative class[16]. Below is the result of a confusion matrix of the two programs from Weka tool

Table 3 Confusion Matrix Bsit Program

BSIT	a	b	Classified as
	75 16	19 120	a = Yes b = No

Table 4 Confusion Matrix Bs Entrep Program

BS Entrepreneurship	a	b	Classified as
	45 1	7 48	a = Yes b = No

Table 5 Detail Accuracy By Class

Programs	Accuracy	Precision	Recall	F-Measure
BSIT	84.7826 %	0.824	0.798	0.811
BS Entrep	92.1565%	0.978	0.865	0.918

Table 5 depicts the detailed accuracy by class. To which the study uses the following parameters; accuracy, precision, recall, and f-measure to test the model.

B. Performance Measurements Results

In testing, the percentage split 70 % was being used. The Weka tool randomly chooses 30% from the dataset. As a result, in BSIT, we got 230, which is 30% of 768 and 102/339 from the BS Entrep (refer Table 3 and 4). From that table, we see all the calculated possible label values that have a maximum calculated probability. J48 algorithm classifier shows that on BSIT results, it has 91/230 labeled Yes and the probability of label No 139/230. While on BS Entrep from 102, 46 labeled as Yes and 56 labeled as No.

TABLE 6 DETAIL ACCURACY BY CLASS

Program	Accuracy: 84.78%		
BSIT	a = Yes	a = No	Classified as
	TP = 75	FN = 19	a = Yes
	FP = 16	TN = 120	b = No
	Precision	Recall	F-measure
	82.4%	79.8%	81.1%
Program	Accuracy: 91.17%		
BS Entrep	a = Yes	b = No	Classified as
	TP = 45	FN = 7	a = Yes
	FP = 1	TN = 49	b = No
	Precision	Recall	F-measure
	97.8%	86.5%	91.8%

Table 6 shows 30% of BSIT and BS Entrep records were used to test the model performance through the confusion matrix in terms of accuracy, precision, recall, and F-measure.

1) For BSIT

a) Accuracy: (TP+TN)/(TP+TN+FP+FN)
(75+120)/(75+120+16+19) = 84.78%

b) Precision: TP/(TP+FP)
75/(75+16) = 82.4%

c) Recall: TP/(TP+FN)
75/(75+19) = 79.8%

d) F-measure: (2 x Precision x Recall)/(Precision + Recall)
(2 x 824 x 798)/(824+798) = 81%

2) For BS Entrep

a) Accuracy: (TP+TN)/(TP+TN+FP+FN)
(45+48)/(45+48+2+7) = 91.17%

b) Precision: TP/(TP+FP)
45/(45+2) = 95.7 %

c) Recall: TP/(TP+FN)
45/(45+7) = 86.5%

d) F-measure: (2 x Precision x Recall)/(Precision + Recall)
(2 x 957 x 865)/(957+865) = 90.9 %

The results above shows the performance of the model using the two datasets from different programs. The result shows a positive good considerable accuracy.

V. CONCLUSIONS

This study has applied classification analytics to the students who didn't make it finish the particular program. Data use in this study is a primary source of data where the researcher collected the data directly from the subjects for five consecutive years. To classify the different applications and perform an accurate result of the classification and prediction, we applied the J48 classification algorithm.

In this paper, with the use of the J48 algorithm, we achieve extremely high classification rates given enough training samples to create the decision-tree model. Moreover, in the two programs, experiment results show accuracy for BSIT got 84.7826% and 91.1765% for BS Entrep. In addition to our focus on accuracy, this method can easily be implemented in the area of academe. Using Weka tool, we use the decision tree as an instrument to analyse what factors affect the child why a lot of them didn't finish the program.

A. The generated results of the application through WrapperSubEval in selecting attributes under classifiers trees J48 algorithm has a significant role in determining the four factors which affect the student completion of the course on time. On the results, 3 out of 4 attributes appeared the same on different times of testing with the two datasets coming from two programs.

B. This paper discusses a different approach for the prediction of the completion of the course on time by the students to a particular two programs through applying the J48 classification algorithm of data mining. J48 gives a result in the form of a decision tree. The experiment of this study dataset used came from BSIT and BS Entrep program. Dataset has come up with the information of the students both programs in five years' time. The generated model using the J48 algorithm was based on Parents Monthly income, Mother Education Attainment, Father Educational Attainment, and Math Average(BSIT) & High School Point Average (BS Entrep). The J48 prediction model for student attrition has been successfully developed by using percentage split 70% training and 30% for testing. Confusion Matrix gives the results of the algorithm. It contains information about the actual & predicted classification. The distribution model for label attribute completed for Class YES 91 and for class NO 139 for BSIT and for BS Entrep Class YES 47 and Class No is 55.

C. The model generated correct predictions or accuracy of 84.78% for BSIT and 91.17 for BS Entrep. The result indicated that the J48 classification algorithm could be applied to develop a prediction model that may provide a powerful tool that can determine and predict student attrition. In BSIT, there are 35 incorrectly classified instances, which indicate that the model is incorrect for 15.21% of the cases in the dataset. Then 195 correctly classified instances, which means that the model is correct for 84.78 % of the cases in the dataset. In the BS Entrep program, the correctly classified instances are 93 equivalent to 91.17% of the cases in the dataset. Further, it has 8 incorrectly classified instances, which corresponds to 8.82% of the cases in the dataset.

VI. RECOMMENDATIONS

There are many advances that still need to be done, more approaches, analysis of the study of data mining practices in higher education, needs further development to the current states. The dataset can also be used for testing in other classifiers such as Neural Network to compare the performance result generated using Decision Tree J48 algorithm. In the future, it also must be done by using other methods for comparison material. Or maybe a hybrid method might be used, namely the combination of the Decision Tree J48 method being used in this study combined with other methods.

VII. ACKNOWLEDGEMENT

I would like to thank first our dear Lord, who gives me the wisdom, guidance, and strength for my everyday living. To my husband, who's very supportive to me and to all my family members, thank you! To all who constantly praying for me in all I do, thank you! To the President of Bulacan State University Dr. Cecilia S. Navasero-Gascon for being supportive, considerate, and being kind always to all the scholars of the university. Thank you to Ched- FDP for giving me the chance to avail this scholarship. To Bulacan State University BulSU family thank you. To Ma'am Joan Perez , Dr. Elenita Capariňo, and the rest of my Industrial and Information Technology faculty family, thank you for the help and constantly encouraging and cheering me up. Thank you to Sir Gil Fabon who's one chat away and to my adviser in TIP Dr. Gerald T. Cayabyab. To Dr. Ruji Medina Dean of Graduate Program for always being helpful to us, thank you so much. To Ma'am Sheila Reyes, thank you for always accommodating me and helping me in gathering my data. It's never been easy and wouldn't be possible without you all, whom the Lord used to be a channel of blessings to me.

REFERENCES

[1] L. C. Ho and K. J. Shim, "Data Mining Approach to the Identification of At- Risk Students," *2018 IEEE Int. Conf. Big Data (Big Data)*, pp. 5333–5335, 2018.

[2] J. Hung, M. C. Wang, S. Wang, M. Abdelrasoul, and Y. Li, "Identifying At-Risk Students for Early Interventions – A Time-Series Clustering Approach," vol. 6750, no. c, pp. 1–11, 2015.

[3] J. H. Amirkhan and Y. B. Kofman, "Stress overload as a red fl ag for freshman failure and attrition," *Contemp. Educ. Psychol.*, vol. 54, pp. 297–308, 2018.

[4] R. Alshabandar *et al.*, "The Application of Gaussian Mixture Models for the Identification of At-Risk Learners in Massive Open Online Courses," *2018 IEEE Congr. Evol. Comput.*, pp. 1–8, 2018.

[5] T. Goradia and A. Bugarcic, "SC," *Biochem. Pharmacol.*, 2018.

[6] Z. Alharbi, J. Cornford, L. Dolder, and B. D. La Iglesia, "Using Data Mining Techniques to Predict Students at Risk of Poor Performance," pp. 523–531, 2016.

[7] U. States, "On early prediction of risks in academic performance for students," vol. 59, no. 6, pp. 1–14, 2015.

[8] T. Barbé, L. P. Kimble, L. M. Bellury, and C. Rubenstein, "Journal of Professional Nursing Predicting student attrition using social determinants : Implications for a diverse nursing workforce," *J. Prof. Nurs.*, 2017.

[9] D. Baneres, M. E. Rodríguez, and M. Serra, "An Early Feedback Prediction System for Learners At-risk within a First-year Higher Education Course," vol. 1382, no. c, 2019.

[10] M. Singh and J. Singh, "Feature Extraction Model to identify At – Risk Level of Students in Academia," pp. 221–227, 2014.

[11] J. Hung, B. E. Shelton, J. Yang, and X. Du, "Improving Predictive Modeling for At-Risk Student Identification : A Multi-Stage Approach," 2019.

[12] D. D. Pokrajac, K. R. Sudler, P. Y. Edamatsu, and T. Hardee, "Prediction of Retention at Historically Black College / University using Artificial Neural Networks," 2016.

[13] R. Deuja, "Data-Driven Predictive Analysis of Student Performance In College Using Neural Networks," *2018 IEEE 3rd Int. Conf. Comput. Commun. Secur.*, pp. 77–81, 2018.

[14] R. Trakunphutthirak and V. C. S. Lee, "Detecting Student at Risk of Failure : A Case Study of Conceptualizing Mining from Internet Access Log Files," *2018 IEEE Int. Conf. Data Min. Work.*, pp. 365–371, 2018.

[15] L. H. G. Marrega, S. M. Silva, E. F. Manffra, and J. C. Nievola, "Comparison Between Decision Tree and Genetic Programming to Distinguish Healthy from Stroke Postural Sway Patterns," pp. 6820–6823, 2015.

[16] M. Adnan, R. Sarno, and K. R. Sungkono, "Algorithm," *2019 Int. Semin. Appl. Technol. Inf. Commun.*, pp. 121–126, 2019.

[17] R. Patil and V. M. Barkade, "Class-Specific Features Using J48 Classifier for Text Classification," *2018 Fourth Int. Conf. Comput. Commun. Control Autom.*, pp. 1–5.

[18] I. T.- Genetic and B. Feature, "An Intelligent Intrusion Detection System Using Genetic Based Feature Selection and Modified J48 Decision Tree Classifier," 2013.

[19] M. N. Mustafa and L. Chowdhury, "Students Dropout Prediction for Intelligent System from Tertiary Level in Developing Country," 2012.

[20] D. N. N. C. Pradeep KR, "Predictive Analysis of Diabetes using J48 Algorithm of Classification Techniques," vol. 6, pp. 347–352, 2016.

[21] J. González, J. Ortega, M. Damas, P. Martín-smith, and J. Q. Gan, "A new multi-objective wrapper method for feature selection – Accuracy and stability analysis for BCI," no. xxxx, pp. 1–12, 2019.

[22] A. Daud, U. T. Mara, U. T. Mara, U. T. Mara, U. T. Mara, and U. T. Mara, "Predictive Analytics : The Application of J48 Algorithm on Grocery Data to Predict Obesity," pp. 1–6, 2018.

[23] J. Bayer, H. Bydˇ, and G. Jan, "Predicting drop-out from social behaviour of students," no. Dm, pp. 103–109.

[24] S. Milena and M. Rubiano, "Formulation of a predictive model for academic performance based on students ' academic and demographic data Systems Engineering Program," 2015.

[25] A. Wang, N. An, J. Yang, G. Chen, L. Li, and G. Alterovitz, "Wrapper-based gene selection with Markov blanket," *Comput. Biol. Med.*, vol. 81, no. November 2016, pp. 11–23, 2017.

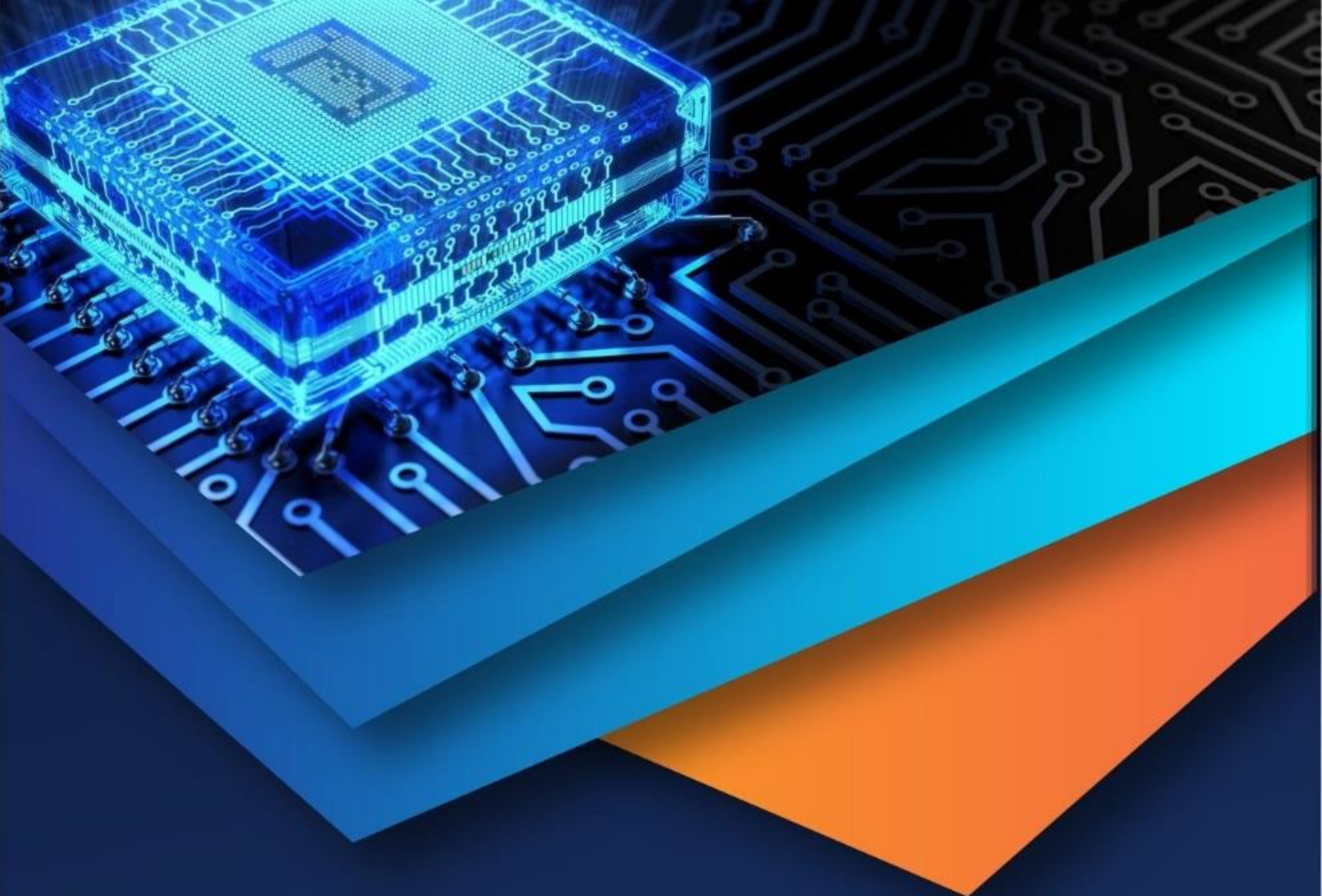
[26] P. Nrodkndm *et al.*, "A recommender system by using classification based on frequent pattern mining and J48 algorithm," pp. 1–7, 2015.

[27] R. P. Poonam Pandey, "An Analysis of Machine Learning Techniques (J48 & AdaBoost-for Classification." 2016.

[28] S. Perveen, M. Shahbaz, A. Guergachi, and K. Keshavjee, "Performance Analysis of Data Mining Classification Techniques to Predict Diabetes," *Procedia - Procedia Comput. Sci.*, vol. 82, no. March, pp. 115–121, 2016.

[29] K. Ying, A. Ameri, A. Trivedi, D. Ravindra, D. Patel, and M. Mozumdar, "Decision Tree-based Machine Learning Algorithm for In-node Vehicle Classification," pp. 71–76, 2015.

[30] N. Saravanan, "CLASSIFICATION OF DENGUE DATASET USING J48 ALGORITHM AND ANT COLONY BASED AJ48 ALGORITHM," no. Icici, pp. 1062–1067, 2017.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY