International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

Utilizing J48 Algorithm in Predicting Students Dropout in Higher Education Institution

Noreen B. Fuentes
College of Computer, Information and Communications Technology
Cebu Technological University
Cebu City, Philippines
noreen.fuentes@ctu.edu.ph

Larmie S. Feliscuzo
College of Computer Studies
Cebu Institute of Technology-University
Cebu City, Philippines
larmie.feliscuzo@cit.edu

Cherry Lyn C. Sta.Romana
College of Computer Studies
Cebu Institute of Technology-University
Cebu City, Philippines
cstaromana@cit.edu

Abstract— Dropout refers to students who voluntarily withdraw from a course or program prior to completion. University dropouts continue to be a major concern for educators and represent a substantial loss of human resources for society. At Cebu Technological University, it is always a challenge of the Department Chairperson the declining student population, which resulted in the reduction of the number of sections per year level and under loading of faculty.

This study centers on the creation of a student-dropout model that predicts a student's behavior toward his studies. This model utilized the J48 decision tree algorithm, which extract data from the Student Information System (SIS) portal of the existing institution. Nine hundred sixty-one (961) demographic and academic datasets from students enrolled in the two programs under the College of Computer, Information, and Communications Technology (CCICT) of Cebu Technological University (CTU) with nineteen (19) attributes. During the testing procedure, 10-fold cross-validation was utilized. The J48 pruned tree utilized an average of 3 foliage with 4 as the measure of the tree. The Kappa statistic yields a value of 0.8617 and its Correctly Classified Instances rate of 93.7%. This algorithm helps a lot to the institution in reducing the escalation of the attrition rate and providing proactive measures to address the issue.

Keywords- academic performance; data mining; J48 algorithm; student dropout; academic analytics; higher education institution;

I. Introduction

A. *Background of the Study*

Students who discontinue their participation in a course or program prior to its completion are considered dropouts [1]. The persistence of university graduates remains a significant concern among educators and signifies a substantial depletion of valuable human resources for society. According to data provided by the National Center for Education Statistics (NCES) [2], there were 2 million withdrawals among 16- to 24-year-olds, which corresponds to an attrition rate of 5.3%. Furthermore, it is noteworthy that 24.1% of first-year college students choose to withdraw within their initial year of enrollment [3]. University graduates can be placed into one of three categories: voluntary, transient, and permanent [4].

Dropping out of higher education institutions is a global phenomenon that affects universities worldwide [5]. Higher education institutions have conducted extensive research on the various categories of graduates, the factors leading to their departure, and the consequences of their decisions for more than a century. According to data from the Philippine Statistics Authority (PSA), 9% of the estimated population of Filipinos aged six to twenty-four [6] consists of approximately 3.53 million out-of-school adolescents.

In addition, the president of the Philippine Association of State Universities and Colleges (PASUC) disclosed the findings of a survey indicating that approximately 44,000 undergraduate students may choose not to enroll for the School Year 2020-2021 owing to the economic difficulties caused by the COVID-19 pandemic [6]. This is notable in light of the government's 2016 and 2017 implementation of the Unified Student Financial Assistance System for Tertiary Education Act (UNIFAST) and the Universal Access to Quality Tertiary Education (Free Tuition) Act.

Moreover, according to [7], the attrition rate at the tertiary level in the Philippines was 83.7% in 2012 before the implementation of RA 10931. In the same study, [8] reported an increase in the number of students withdrawing from a state university, according to data from the registrar's office for the Academic Year 2012-2013, which was the primary focus of the investigation.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

42% of students who entered the College of Computer, Information, and Communications Technology (CCICT) at Cebu Technological University-Main Campus (CTU-MC) during the academic year 2018-2019 and graduated in the academic year 2021-2022 achieved academic success. As shown in Table 1, this success rate decreased to 33% for students who began their academic careers during the 2019-2020 academic year.

In addition, during the 2021-2022 academic year, the College added the Bachelor of Science in Information Systems (BSIS) to its existing programs. The attrition rates for both programs have exhibited a consistent upward trend from the 2020-2021 school year through the 2022-2023 school year, as shown in Table 1.

Table I. Students' Status for CCICT Students

School Year	No. of First Year Students	Total # of Graduates	Success Rate	Total No. of Dropouts	Dropout Rate
2018-2019	178	75	42%		
2019-2020	168	56	33%		
2020-2021	45			19	42%
2021-2022					
BSIT	140			16	11%
BSIS	115			29	25%
2022-2023					
BSIT	175			42	24%
BSIS	140			28	20%

Understanding the causes of these high attrition rates is vital for both the institution and the affected students. It sheds light on the obstacles students face in pursuit of higher education and affords CTU-MC the opportunity to evaluate and improve its support mechanisms. This study lays the groundwork for a comprehensive investigation into the causes of these attrition rates, with the goal of gaining insights that can inform interventions and policies to enhance student retention and success in CCICT at CTU-MC.

II. REVIEW OF RELATED LITERATURE

The growing global concern relates to the rising incidence of illiteracy among college graduates. The investigation of this phenomenon's underlying causes holds promise for the development of novel intervention strategies. Due to the high attrition rate of this cohort in the country's public universities [8], a study in Argentina assessed the perceived academic contentment of first-year university undergraduates. Retention of students is a challenge faced by universities worldwide, including Mae Fae Luang University (MFU) in Thailand and others. Considering that a significant proportion of university withdrawals occur during the first year [9], institutions with high attrition rates confront the possibility of tuition and fee revenue losses. The Philippine Statistics Authority reports that 36,238 individuals between the ages of six and twenty-four are not enrolled in school [10].

The emergence of "datafication" in tertiary education [11] has resulted in a greater emphasis on translating empirical insights into predictive models for student attrition (or success) using vast administrative data [12],[13],[14],[15],[16],[17],[18]. These applications typically seek to facilitate targeted student support and intervention programs, relying on a vast corpus of research literature on college success for engineering theory-based features. For instance, Aulck and colleagues [12] used seven categories of freshman characteristics extracted from registrar data to predict outcomes for all students at a large government-owned university in the United States. The resulting model obtained a graduation prediction accuracy of 83.2% and a retention prediction accuracy of 95.5%.

Recent research [19], [20], and [21] has investigated the fundamental causes of withdrawal intentions. Decades of research have depicted higher education as a journey with numerous turns and infrequent setbacks that have a significant impact on academic and career outcomes [22]. Multiple studies have highlighted the considerable benefits of postsecondary education in terms of income, job prestige, job matching dynamics, employment stability, and career progression. Recent comparative volumes [23], [24], [25], [26] shed further light on these dynamics. Demographics, family background, and prior academic history are prominent factors that correlate strongly with a student's academic, social, and economic resources prior to maturity, and have a significant impact on their college success [27]. Furthermore, students from multiple disadvantaged groups have an even greater likelihood of falling out of college. In contrast, research assessing individuals who began but did not complete tertiary education reveals significantly inferior labor market outcomes in terms of employment, access to professional occupations, job quality, and wages compared to graduates [28], [29], [30], [31].

A. Academic Analytics

1.6.1. Implementing Predictive Analytics on the Dropout Dataset

Previous research has established relationships between variables believed to be symptomatic of a student's intention to withdraw from their educational pursuits, such as academic history, current academic performance, and socioeconomic conditions. To classify and demonstrate the significance of these variables in predicting attrition intentions, a broader array of techniques is employed. The issue of attrition was initially highlighted in the study conducted by the author of [32], prompting further research interest in the topic. In addition, [33] authors employed logistic regression, taking categorical variables such as parental educational background, occupation, gender, and first-year academic performance into account as predictors of withdrawal intentions.

In [35], the advantages and disadvantages of numerous classification algorithms are examined in depth. Using a bank dataset, classifiers, such as Naive Bayes and J48-based algorithms, were evaluated to maximize the true positive rate while minimizing the false positive rate when identifying defaulters.

III. METHODOLOGY

A. Research Design

This study employs Exploratory Data Analysis (EDA) for Data Mining to analyze and validate data in order to generate new knowledge and build the classification model.

B. Research Methods

1) Model Development

Fig. 2 depicts the study model illustrating how data were prepared and represented.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

Figure 2.1. Model Development

B.1.1. Data Collection

The researcher acquired a total of 961 data points from the student body at the College of Computer, Information, and Communications Technology via the Student Information System (SIS) portal at Cebu Technological University. This dataset covered the Bachelor of Science in Information Systems and the Bachelor of Science in Information Technology from academic year 2018-2019 to academic year 2022-2023. The following information describes the data obtained from the SIS portal.

Table II. Features considered for the Dropout Detection

Group	Feature	Attribute	Value
Demographic Profile	Gender	Student's Gender	Male Female
	Address	Student's Address	**Nearby** - *the approximate distance is between 8 to 12 kilometers from CTU-MC both for the south and north of Cebu.* **Farther** – *roughly spanning from 15-140 kilometers both to the south and north of Cebu* **Farthest** – *Bohol, Leyte and various locations beyond Cebu*
	Age	Age	-
	Marital Status	Marital_status	- Single - Married - Separated
	Living With Parents	Living_status	- Yes - No
	Mother Occupation Status	Mothr_OccStatus	- Unemployed - Currently Working
	Mother Occupation	Mother_Occptn	- As indicated
	Father Occupation Status	Father_OccStatus	- Unemployed - Currently Working
	Father Occupation	Father_Occptn	- As indicated
	Parent's Monthly Income	Monthly_Income	
Educational Related Data	Year Entered College	Year_Entered	Year
	Degree Program	Degree_Program	BSIT BSIS
	GPA Upon Entering College	GPA	
	Computer Programming 1-Lec grade	ComProg1_LEC	1.0-1.9 = Very Satisfactory 2.0-2.9 = Satisfactory 3-10.0 = Fail 10.0 – drop or NA
	Computer Programming 1-Lab grade	ComProg1_LAB	
	Computer Programming 2-Lec grade	ComProg2_LEC	
	Computer Programming 2-Lab grade	ComProg2_LAB	
	Average Grade for the Four subjects	AveGrade	1.0-2.0 =Very Good 2.1-2.9 = Good 3.0 and up = Poor
	Total No. of Subjects Attended(based on the 4 subjects declared)	TotalNumOfSubjTaken	1.0-10.0 (based on the result)
	Student's Status	StudStatus	Non-Dropout – Dropout –

B.1.2. Data Preparation

Prior to the deployment of any data mining technique, data preprocessing is crucial. Three processes were used to prepare the data:

1. Integrate data from the SIS portal.
2. Remove entries with incomplete data from the dataset extracted from the SIS portal.
3. Categorize the final grades of the four (4) subjects to three groups: Very Satisfactory, Satisfactory, and Failed based on the University Student Manual and the student's address based on the proximity location from the university

B.1.3. Data Mining

B.1.3.1. Classification

This study aims to estimate the attrition rates of students enrolled in the two (2) programs offered by (CCICT) at CTU. The research employs a decision tree to classify unknown variables. The primary objective of this classification method is to create a decision tree with a flowchart-like structure

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

comprised of internal and leaf nodes. Leaf nodes represent class identifiers, while internal nodes represent attribute evaluations. To simplify the decision-making process and eliminate extraneous complication, pruning is utilized, which involves the removal of branches that do not contribute new information.

B.1.3.1.1. J48 Algorithm

The J48 algorithm is frequently described as an enhanced variant of C4.5. It is an extension of Ross Quinlan's earlier ID3 algorithm, known as J48 in the Weka environment, with the letter 'J' indicating its Java implementation. Using decision trees or principles derived from them, J48 facilitates classification. J48 generates a Decision Tree with nodes including the root, internal nodes, and leaf nodes as its output. As its name implies, each node in the tree represents a decision that ultimately leads to particular outcomes. At each node, the decision tree employs a dividing criterion to determine the most appropriate attribute for segmenting the trained data [36].

As the root node of the decision tree, the characteristic that best classifies the training data is selected. Entropy (amount of uncertainty in the dataset), average information, and information gain (difference in entropy before and after dividing dataset on attribute "A") are calculated using the following formula when selecting the appropriate feature of the training data.

$$Entropy = \frac{-p}{p+n} \log_2\left(\frac{p}{p+n}\right) - \frac{n}{p+n} \log_2\left(\frac{n}{p+n}\right)$$

where:

p is the number of positive examples
n is the number of negative examples

IV. RESULTS AND DISCUSSION

In this study, 961 datasets from School Year 2018-2019 to School Year 2022-2023 were obtained from the University's SIS portal. A thorough examination was conducted on the extracted dataset to identify and eliminate any duplicate entries. After the data had been cleaned, it was separated according to the specified school year. The data were then saved in CSV (comma-separated value) file format. This operation was performed using the Microsoft Excel application. In addition, the date of birth was omitted from the dataset because it was duplicated with age, which is also one of the investigated attributes. The categorical data within the dataset were organized for equivalence based on the Student Manual Grading System of the university. The dataset was subsequently divided into two subsets: training data and testing data. This division was performed to facilitate the construction of predictive models and to validate the model's accuracy.

S.Y. 2018-2019 Decision Tree Result

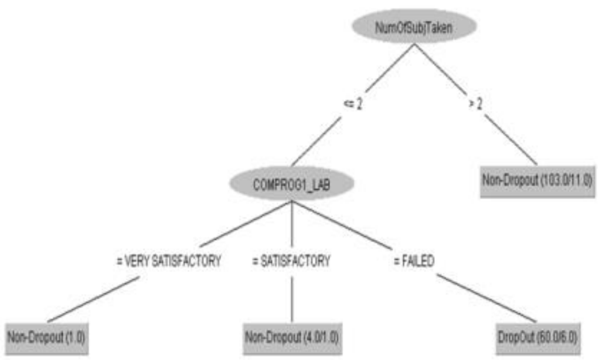


Fig.4.1. S.Y. 2018-2019 Decision Tree

Based on the results of the decision tree model for the 2018-2019 academic year, as depicted in Fig. 4.1. If a student has passed more than two subjects during the first semester, they are more likely to continue their education until graduation (4.1). In addition, the model predicts that if a student has passed at least two subjects, obtained a GPA of at least 89 in senior year of high school, and resides near or far from the university, they still have a decent chance of completing their education. The J48 Algorithm implemented tree pruning to prevent overfitting, resulting in a tree with 6 leaves and a size of 10.

In terms of efficacy, this model obtained an impressive accuracy rate of 91.01% in correctly classifying instances. The Kappa statistic, which goes beyond mere accuracy to assess the classifier's reliability, yielded a value of 0.8074. This Kappa statistic is indispensable for obtaining a deeper comprehension of the performance of the classifier. This model utilized a dataset containing 178 instances. Figure 4.2 displays the specific.

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         162               91.0112 %
Incorrectly Classified Instances        16                8.9888 %
Kappa statistic                          0.8074
Mean absolute error                      0.1368
Root mean squared error                  0.2899
Relative absolute error                 29.1189 %
Root relative squared error             59.8245 %
Total Number of Instances              178
```

Figure 4.2. S.Y. 2018-2019 Stratified Cross Validation Summary

1. ***S.Y. 2019-2020 Decision Tree Result***

As depicted in Figure 4.3, the decision tree indicates that a student is expected to continue their education until graduation if they pass more than two subjects in their first year. Moreover, the model predicts that even if a student passes at least two subjects based on the specified attributes, but receives a grade of "Very Satisfactory" or "Satisfactory" in ComProg1 Lab, there is still a substantial likelihood that they will complete their education and graduate. The pruned tree has four leaves and a total leaf count of six. This means that lesser number of leaves simplified the decision rules in the tree, making them more interpretable and simpler to comprehend and may help prevent overfitting in which the tree suits the training data and may not generalize well to new, unseen data. Figure 4.3 provides exhaustive information regarding the decision tree's outcomes.

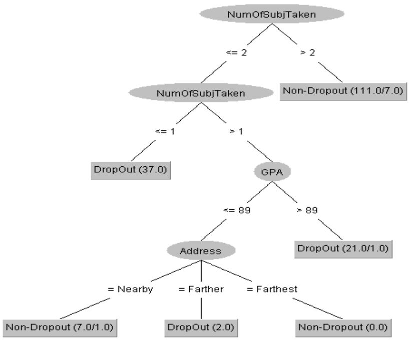
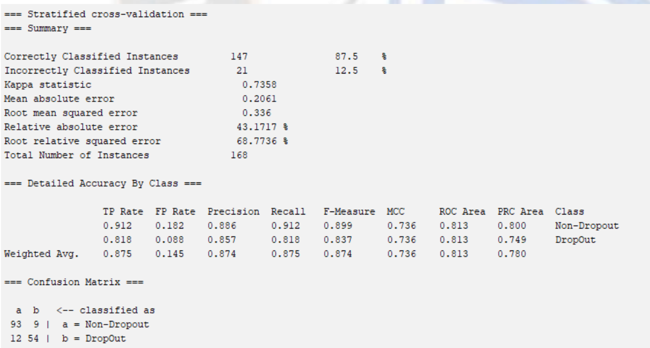


Figure 4.3. S.Y. 2019-2020 Decision Tree

The rate of Correctly Classified Instances reaches 87.5% performance-wise. This demonstrates that the model excels at predicting both university dropouts and students who continue their education. In addition, the Kappa statistic resulted in a value of 0.7358, indicating that the model's performance in predicting withdrawals is quite plausible. Mean Absolute Error and Root Mean Squared Error are currently 0.2061 and 0.336, respectively. Figure 4.4 provides a comprehensive breakdown of its performance characteristics.



```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         147               87.5    %
Incorrectly Classified Instances        21               12.5    %
Kappa statistic                          0.7358
Mean absolute error                      0.2061
Root mean squared error                  0.336
Relative absolute error                 43.1717 %
Root relative squared error             68.7736 %
Total Number of Instances              168

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.912    0.182    0.886      0.912    0.899      0.736    0.813     0.800     Non-Dropout
                 0.818    0.088    0.857      0.818    0.837      0.736    0.813     0.749     DropOut
Weighted Avg.    0.875    0.145    0.874      0.875    0.874      0.736    0.813     0.780

=== Confusion Matrix ===

  a  b   <-- classified as
 93  9 |  a = Non-Dropout
 12 54 |  b = DropOut
```

Figure 4.4. Stratified Cross Validation Summary for S.Y. 2019-2020

S.Y. 2020-2021 Decision Tree Result

The following forecast is based on the results of the decision tree model for the academic year 2020-2021:

- A student is likely to conclude their education if they have passed more than two subjects from the attribute list, regardless of how close or far away their residence is from the university.
- In contrast, it is predicted that a student will opt out if they have failed two subjects from the attribute list.
- Furthermore, even if a student has passed more than two subjects in their first year, it is possible that they will discontinue their education if they reside outside of Cebu.

To evaluate the efficacy of the model, a 10-Folds Cross Validation is employed. Figure 4.5 illustrates the structure of the association.

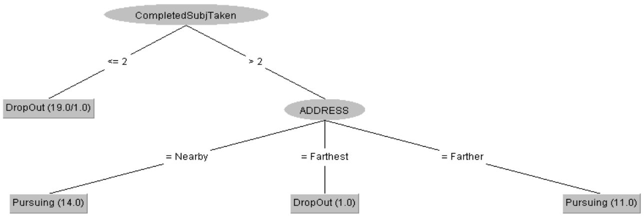
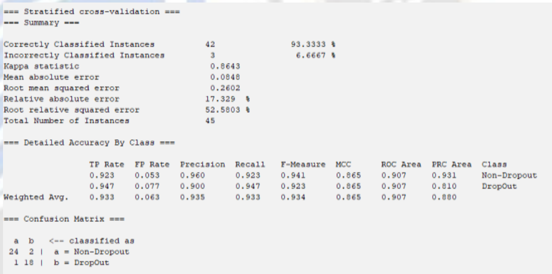


Figure 4.5. School Year 2020-2021 Decision Tree

In terms of correctly classifying instances, it obtains an impressively high accuracy rate of 93.3% in terms of its performance. In addition, the Kappa Statistics result is 0.8643, indicating that the model's ability to predict attrition rates is highly accurate and efficient. Figure 4.6 provides a more comprehensive summary of its performance characteristics.

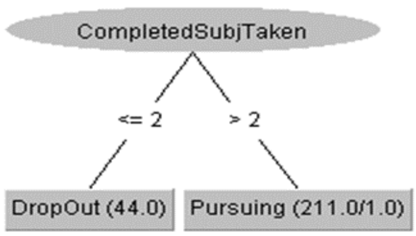


```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          42               93.3333 %
Incorrectly Classified Instances         3                6.6667 %
Kappa statistic                          0.8643
Mean absolute error                      0.0848
Root mean squared error                  0.2602
Relative absolute error                 17.329  %
Root relative squared error             52.5803 %
Total Number of Instances               45

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.923    0.053    0.960      0.923    0.941      0.865    0.907     0.931     Non-Dropout
                 0.947    0.077    0.900      0.947    0.923      0.865    0.907     0.810     DropOut
Weighted Avg.    0.933    0.063    0.935      0.933    0.934      0.865    0.907     0.880

=== Confusion Matrix ===

  a  b   <-- classified as
 24  2 |  a = Non-Dropout
  1 18 |  b = DropOut
```

Figure 4.6. Stratified Cross Validation Summary for S.Y. 2020-2021

S.Y. 2021-2022 Decision Tree Result

Compared to previous school years' decision trees, the current decision tree appears to be the simplest. As shown in Figure 4.6, it is predicted that a student will conclude their education effectively if they pass more than two computer programming courses. Examining its performance, the 99.2 percent accuracy rate approaches perfection. The Kappa Statistics result is 0.973, while the Mean Absolute Error and Root Mean Squared Error are 0.0117 and 0.0877, respectively. Given the exceedingly high rate of accuracy, these values suggest minimal error.



International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

In the Confusion Matrix, there are 209 true positives, which account for 99.5 percent, and 1 false positive, equivalent to 0.5 percent. Figure 4.7 provides details regarding the cross-validation evaluation.

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         253               99.2157 %
Incorrectly Classified Instances         2                0.7843 %
Kappa statistic                          0.973
Mean absolute error                      0.0117
Root mean squared error                  0.0877
Relative absolute error                  3.9977 %
Root relative squared error             22.9889 %
Total Number of Instances              255

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.995    0.022    0.995      0.995    0.995      0.973    0.980     0.991     Pursuing
                 0.978    0.005    0.978      0.978    0.978      0.973    0.980     0.979     DropOut
Weighted Avg.    0.992    0.019    0.992      0.992    0.992      0.973    0.980     0.989

=== Confusion Matrix ===

   a   b   <-- classified as
 209   1 |   a = Pursuing
   1  44 |   b = DropOut
```

Figure 4.7. Stratified Cross Validation Summary for S.Y. 2021-2022

A. S.Y. 2022-2023 Decision Tree Result

Regarding the decision tree model for the Academic Year 2022-2023, it has been predicted that a student is classified as a Non-Dropout if they pass more than three computer programming classes in their First Year. Alternatively, a student is more likely to complete their education if they pass at least three computer programming courses in their First Year and live with their family.

If a student passes two computer programming courses and receives a Satisfactory in Comprog 1 Lab, it is predicted that they will continue their education. If the student's Comprog 1 Lab grade is Failed or Very Satisfactory, on the other hand, they are expected to discontinue their education. In addition, there is a considerable likelihood that a student will drop out of school even if they pass more than two Computer Programming courses and do not live with their families. The tree has only six leaves and is its tree size is ten. Observe Fig. 4.8 for a comprehensive illustration of the decision tree.

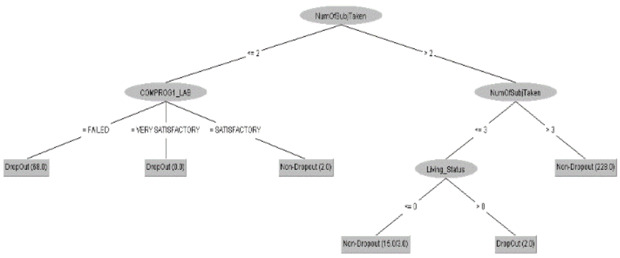


Figure 4.6. S.Y. 2022-2023 Decision Tree

In terms of its efficacy, it obtains an impressive accuracy rate of 97.46%, with only 2.5% of instances being incorrectly classified. In addition, the Kappa statistics produce a remarkable value of 0.928, indicating an exceedingly high degree of concordance between the model's predictions and the actual classifications.

The Mean Absolute Error (MAE) is determined to be 0.0327, indicating that the model's predictions are exceptionally accurate and that the typical prediction error is small. The True Positive rate is 0.988, indicating a high proportion of accurate positive predictions, whereas the False Positive rate is 0.932.

Collectively, these performance metrics demonstrate the model's robustness and dependability. Figure 4.7 contains additional performance-related information.

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         307               97.4603 %
Incorrectly Classified Instances         8                2.5397 %
Kappa statistic                          0.928
Mean absolute error                      0.0327
Root mean squared error                  0.1603
Relative absolute error                  9.1643 %
Root relative squared error             37.9798 %
Total Number of Instances              315

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.988    0.068    0.980      0.988    0.984      0.928    0.991     0.996     Non-Dropout
                 0.932    0.012    0.958      0.932    0.944      0.928    0.991     0.961     DropOut
Weighted Avg.    0.975    0.055    0.974      0.975    0.974      0.928    0.991     0.988

=== Confusion Matrix ===

   a   b   <-- classified as
 239   3 |   a = Non-Dropout
   5  68 |   b = DropOut
```

Figure 4.6. Stratified Cross Validation Summary for S.Y. 2022-2023

Each School Year, based on the result of the tree, different attributes were used to predict whether or not a student would drop out of school.

Table III. Summary of Decision Tree Performance

	2018-2019	**2019-2020**	**2020-2021**	**2021-2022**	**2022-2023**
Correctly Classified Instances	91.01%	87.5%	93.3%	99.2%	97.46%
Kappa Statistics	0.8074	0.7358	0.8643	0.973	0.928
Mean Absolute Error	0.1368	0.2061	0.0848	0.0117	0.0327
Root Mean Squared Error	0.2899	0.336	0.2602	0.877	0.1643

As depicted in Table III via the decision tree analysis, the model demonstrates a remarkable average accuracy rate of 93.7% and remarkably low error rates. This highlights the significant impact potential of implementing this model.

In addition, it is crucial to acknowledge that the model's applicability extends beyond the College of Computer, Information, and Communications Technology, thereby extending its benefits to other Colleges within the university.

Notable is the model's capacity to guide the formulation of intervention policies aimed at enhancing student retention and success in college. In addition, it has the potential to positively impact the lives of students by ensuring that they receive proactive interventions that prevent them from abandoning their educational pursuits. This unwavering support enables them to pursue their ambitions and work towards a prosperous future.

V. CONCLUSION

The study on dropout rates at Cebu Technological University-Main Campus (CTU-MC) has yielded significant findings regarding predictive models for dropout behavior and provided valuable insights into the factors influencing student attrition. In addition, predictive models for disengagement behavior made use of J48's features in Weka. According to the findings, a student must have a solid foundation in the majority of their computer programming courses in their First Year to be successful in their educational endeavor; otherwise, they will likely drop out. Most of the result were able to obtain an average accuracy rate of 92 percent of the prediction after the model was applied.

REFERENCES

[1] Spady, W. G. (1971). Dropouts from higher education: Toward an empirical model. Interchange, 2(3), 38−62. https://doi.org/10.1007/BF02282469

[2] Hussar, B., Zhang, J., Hein, S., Wang, K., Roberts, A., Cui, J., & Dilig, R. (2020). The Condition of Education 2020. NCES 2020-144. National Center for Education Statistics.

[4] Vaithiyanathan, V., K. Rajeswari, Kapil Tajane, and Rahul Pitale. "Comparison of Different Classification Techniques Using Different Datasets."Vol.6, no. 2 (2013).

[5] Fägerlind, I., & Strömqvist, G. (2004). Reforming Higher Education in Nordic Countries. UNESCO.

[6] Deiparine, C. SWS: 4.4 Million School-age Filipinos Not Enrolled as of Late 2020. Available online: https://www.philstar.com/headlines/2021/02/24/2080112/sws-44-million-school-age-filipinos-not-enrolled-late-2020 (accessed on 30 June 2022).

[7] Bodoso, E. (2018). Why students dropped?. International Journal of Current Research. 10(08). 73649-73651.

[8] Zalazar Jaime, M. F., Losano, M. C., Moretti, L. S., & Medrano, L. A. (2017). Evaluation of an academic satisfaction model for first-year university students.

[9] Meedech, P., Iam-On, N., & Boongoen, T. (2016). Prediction of student dropout using personal profile and data mining approach. In Intelligent and evolutionary systems (pp. 143-155). Springer, Cham.

[10] Out-of-School Children and Youth in the Philippines. (April 20, 2015). [Online]. Available:https://psa.gov.ph/content/out-school-children-and-youth-philippinesresults-2013-functional-literacy-education-and

[11] Neil Selwyn and Dragan Gaševic. 2020. The datafication' of higher education: discussing the promises and problems. Teaching in Higher Education 25, 4 (2020), 527–540. DOI: http://dx.doi.org/10.1080/13562517.2019.1689388

[12] Lovenoor Aulck, Dev Nambi, Nishant Velagapudi, Joshua Blumenstock, and Jevin West. 2019. Mining University Registrar Records to Predict First-Year Undergraduate Attrition. In Proceedings of the 12th International Conference on Educational Data Mining (EDM 2019). 9–18.

[13] Gerben W. Dekker, Mykola Pechenizkiy, and Jan M. Vleeshouwers. 2009. Predicting students drop out: A case study. In Proceedings of the 2nd International Conference on Educational Data Mining (EDM 2009). 41–50.

[14] Sandeep M. Jayaprakash, Erik W. Moody, Eitel J.M. Lauría, James R. Regan, and Joshua D. Baron. 2014. Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. Journal of Learning Analytics 1, 1 (2014), 6–47. DOI: http://dx.doi.org/10.18608/jla.2014.11.3

[15] Francesca Del Bonifro, Maurizio Gabbrielli, Giuseppe Lisanti, and Stefano Pio Zingaro. 2020. Student Dropout Prediction. In Proceedings of the 21st International Conference on Artificial Intelligence in Education (AIED 2020). Springer, 129–140. DOI: http://dx.doi.org/10.1007/978-3-030-52237-7_11

[16] Cédric Beaulac and Jeffrey S. Rosenthal. 2019. Predicting University Students' Academic Success and Major Using Random Forests. Research in Higher Education 60, 7 (2019), 1048–1064. DOI: http://dx.doi.org/10.1007/s11162-019-09546-y

[17] Johannes Berens, Kerstin Schneider, Simon Görtz, Simon Oster, and Julian Burghoff. 2019. Early Detection of Students at Risk - Predicting Student Dropouts Using Administrative Student Data from German Universities and Machine Learning Methods. Journal of Educational Data Mining 11, 3 (2019), 1–41. DOI:http://dx.doi.org/10.5281/ZENODO.3594771

[18] Stephen Hutt, Margo Gardner, Angela L. Duckworth, and Sidney K. D'Mello. 2019. Evaluating Fairness and Generalizability in Models Predicting On-Time Graduation from College Applications. In Proceedings of the 12th International Conference on Educational Data Mining (EDM 2019).

[19] Rodrı´guez-Go´mez D, Feixas M, Gairı´n J, Muñoz JL. Understanding Catalan university dropout from a cross-national approach. Studies in Higher Education. 2015; 40(4):690–703. doi: 10.1080/03075079. 2013.842966

[20] Tinto V. Research and Practice of Student Retention: What Next? Journal of College Student Retention. 2007; 8(1):1–19. doi: 10.2190/4YNU-4TMB-22DJ-AN4W

[21] Montmarquette C, Mahseredjian S, Houle R. The determinants of university dropouts: a bivariate probability model with sample selection. Economics of Education Review. 2001; 20:475–484. doi: 10.1016/ S0272-7757(00)00029-7

[22] George D. Kuh, Jillian Kinzie, Jennifer A. Buckley, Brian K. Bridges, and John C. Hayek. 2007. Piecing Together the Student success puzzle: Research, Propositions, and Recommendations. ASHE Higher Education Report 32, 5 (2007), 1–182. DOI: http://dx.doi.org/10.1002/aehe.3205

[23] Blossfeld, H. P., Schneider, T., & Doll, J. (2009). Methodological advantages of panel studies. Designing the new National Educational Panel Study (NEPS) in Germany. Journal for Educational Research Online, 1(1), 10-32.

[24] Kogan, I. (2012). Tertiary education landscape and labour market chances of the highly educated in Central and Eastern Europe. European Sociological Review, 28(6), 701-703.

[25] Gangl, M., Müller, W., & Raffe, D. (2003). 10 Conclusions: explaining cross-national differences in school-to-work transitions. Transitions from Education to Work in Europe: The Integration of Youth into EU Labour Markets: The Integration of Youth into EU Labour Markets, 277.

[26] Shavit, Y. (Ed.). (2007). Stratification in higher education: A comparative study. Stanford University Press.

[27] James S. Coleman. 1988. Social capital in the creation of human capital. Amer. J. Sociology 94 (1988), S95–S120

[28] Cristobal de Brey, Lauren Musu, Joel McFarland, Sidney Wilkinson-Flicker, Melissa Diliberti, Anlan Zhang, Claire Branstetter, and Xiaolei Wang. 2019. Status and Trends in the Education of Racial and Ethnic Groups 2018 (NCES 2019-038). Technical Report. National Center for Education Statistics.

[29] Emily Forrest Cataldi, Christopher T. Bennett, and Xianglei Chen. 2018. First-Generation Students: College Access, Persistence, and Postbachelor's Outcomes (NCES 2018421). Technical Report. National Center for Education Statistics.

[30] Davies, R., & Elias, P. (2003). Dropping out: A study of early leavers from higher education. DfES Publications.

[31] Grubb, W. N. (2002). Learning and earning in the middle, part I: National studies of pre-baccalaureate education. Economics of education review, 21(4), 299-321.

[32] Johnes, J., & Taylor, J. (1991). Non-completion of a degree course and its effect on the subsequent experience of non-

completers in the labour market. Studies in Higher Education, 16(1), 73-81.
[33] Tinto V. Research and Practice of Student Retention: What Next? Journal of College Student Retention. 2007; 8(1):1–19. doi: 10.2190/4YNU-4TMB-22DJ-AN4W
[34] Goldenhersh H, Coria A, Saino M. Desercio´n estudiantil: desafı´os de la universidad pu´blica en un horizonte de inclusio´n Desercio´n. RAES Revista Argentina de Educacio´n Superior. 2011; 3:96–120.
[35] Da'vila N, Garcı´a-Artiles MD, Pe'rez-Sa'nchez JM, Go'mez-De'niz E. An Asymmetric Logit Model to explain the likelihood of success in academic results. Revista de Investigacio´n Educativa. 2015; 33:27–45.
[34] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research. 2002; 16:321–357.
[35] S.Archana1; Dr. K.Elangovan.(2014): Survey of Classification Techniques in Data Mining, International Journal of Computer Science and Mobile Applications, 2(2), pp. 65-71.
[36] Tina R. Patil; Mrs. S. S. Sherekar.(2013):Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification, International Journal Of Computer Science And Applications,6(2).
[37] I. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, San Francisco: Morgan Kaufmann Publishers, 2000.
[37] Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. Knowl. Data Eng. IEEE Trans. 17(6), 734–749 (2005)

AUTHOR'S PROFILE

Noreen B. Fuentes is a faculty of the College of Computer, Information, and Communications Technology at Cebu Technological University-Main Campus in Cebu City, Philippines. She completed her Master of Science in Teaching with a concentration in Computer Science at the University of Cebu-Main Campus in Cebu City, Philippines. She is presently in the Dissertation Phase of her Doctor of Information Technology program at Cebu Institute of Technology-University in Cebu City, Philippines. Her teaching areas are in programming, data mining and data structures.

Larmie S. Feliscuzo holds the position of Management Information Systems Director at Cebu Institute of Technology – University. She graduated with a Bachelor of Science in Information Technology (Magna Cum Laude) and a Master's in Computer Science from the same university. Additionally, she earned her Ph.D. in Technology Management from Cebu Technological University.

Cherry Lyn C. Sta.Romana is the Dean of the College of Computer Science of Cebu Institute of Technology-University. She finished her Bachelor of Science in Computer Science and Master's in Computer Science from the University of the Philippines – Los Baños and Doctor in Information Technology from Cebu Institute of Technology-University.