International Journal of Computer (IJC)

ISSN 2307-4523 (Print & Online)

https://ijcjournal.org/index.php/InternationalJournalOfComputer/index

Student Attrition Prediction Using Machine Learning Techniques

Doris Chinedu Asogwa[a*], Emmanuel Chibuogu Asogwa[b], Emmanuel Chinedu Mbonu[c], Joshua Makuochukwu Nwankpa[d], Tochukwu Sunday Belonwu[e]

[a,b,c,d,e] *Nnamdi Azikiwe University, Awka, Anambra State, 234, Nigeria*

[a]*Email: dc.asogwa@unizik.edu.ng*, [b]*Email: ec.asogwa@unizik.edu.ng*, [c]*Email: ec.mbonu@unizik.edu.ng*

[d]*Email: jm.nwankpa@unizik.edu.ng*, [e]*Email: ts.belonwu@unizik.edu.ng*

Abstract

In educational systems, students' course enrollment is fundamental performance metrics to academic and financial sustainability. In many higher institutions today, students' attrition rates are caused by a variety of circumstances, including demographic and personal factors such as age, gender, academic background, financial abilities, and academic degree of choice. In this study, machine learning approaches was used to develop prediction models that predicted students' attrition rate in pursuing computer science degree, as well as students who have a high risk of dropping out before graduation. This can help higher education institutes to develop proper intervention plans to reduce attrition rates and increase the probability of student academic success. Student's data were collected from the Federal University Lokoja (FUL), Nigeria. The data were preprocessed using existing weka machine learning libraries where the data was converted into attribute related file form (arff) and resampling techniques was used to partition the data into training set and testing set. The correlation-based feature selection was extracted and used to develop the students' attrition model and to identify the students' risk of dropping out. Random forest and random tree machine learning algorithms were used to predict students' attrition. The results showed that the random forest had an accuracy of 79.45%, while the random tree's accuracy was 78.09%. This is an improvement over previous results where 66.14% and 57.48% accuracy was recorded for random forest and random tree respectively. This improvement was as a result of the techniques used. It is therefore recommended that applying techniques to the classification model *can improve the* performance of the model.

Keywords: Machine learning; Predictive model; Random Forest; Random Tree algorithm; Student Attrition; Feature selection method; (Java Virtual Machine (JVM); Netbeans Integrated Software Development Environment (IDE); Weka Tool;Weka Plugin.

Received: 7/17/2023
Accepted: 8/22/2023
Published: 9/2/2023

----

* Corresponding author.

1. Introduction

The consequences of attrition in final year student today in our educational system are considerable, both for the individuals as well as the affected institutions. Indeed, attrition imposes costs on all parties involved, be it resources, time or money [1, 2]. Consequently, preventing educational attrition poses a major challenge to institutions of higher education [3].

The predicting of student attrition has generally been subjected to a machine leaning model as a way to make a better decision. Machine learning model is a computer program or a specific of field of computer system, used to analyze data without explicit program. This approach can help an institution to reduce student attrition, by identifying and maintaining student relationships with the assistance of predictive data mining techniques proposed by [4].

The identification of risk cases constitutes the first step to improving retention policies such as learning assistance or mentorship program. Quantifying attrition risks may prove to be helpful in allocating pedagogical, psychological or administrative resources in an efficient way. This study serves as a pilot to assess the feasibility to make attrition predictions utilizing study progression data [5].

However, understanding and tackling attrition problem at Nigeria higher institution is particularly relevant for two reasons. Firstly, with about 30% of the students not completing their studies, attrition constitutes a widespread domestic phenomenon in Nigeria higher education. Secondly, economic growth and demographic change have led to a shortage of qualified specialists in the Nigeria labour market [6].

In this paper, the objective is to apply machine learning algorithms to predict student attrition problems, as well as to gain insight and further explore the problem, and design an application based on one of these algorithms that can predict student attrition with high accuracy. The primary dataset of student's assessment from 2015-2022 was collected from the Federal University Lokoja (FUL), Kogi State Nigeria and was used to perform this analysis.

A total sample of 4407 student assessment was manually extracted from the FUL database of the previous year of student assessment, which was used for predicting student's attrition. The dataset was utilized, in which each line starts with the class label of each student assessment, followed by annotating the data. After pre-processing of the data and extraction of features, machine learning techniques e.g., random forest and decision tree model was used to train in other to make prediction. Based on conventional study as progression information or data that may appear on the typical student's transcript of records, can oppose to survey student data and this may appear as a promising result as well as easily replicable at other institutions. Clearly, the main purpose of this student attrition is to predict based on the factors provided in the dataset. The classification techniques namely random forest and decision tree techniques were applied to the dataset to predict the attrition level. The cross-validation technique was used to solve the problem of over fitting for feature selection. The feature selection is one of the major goals of this study. Feature selection is performed using correlation-based feature selection technique to improve the results.

Below is the working flow of student attrition prediction.

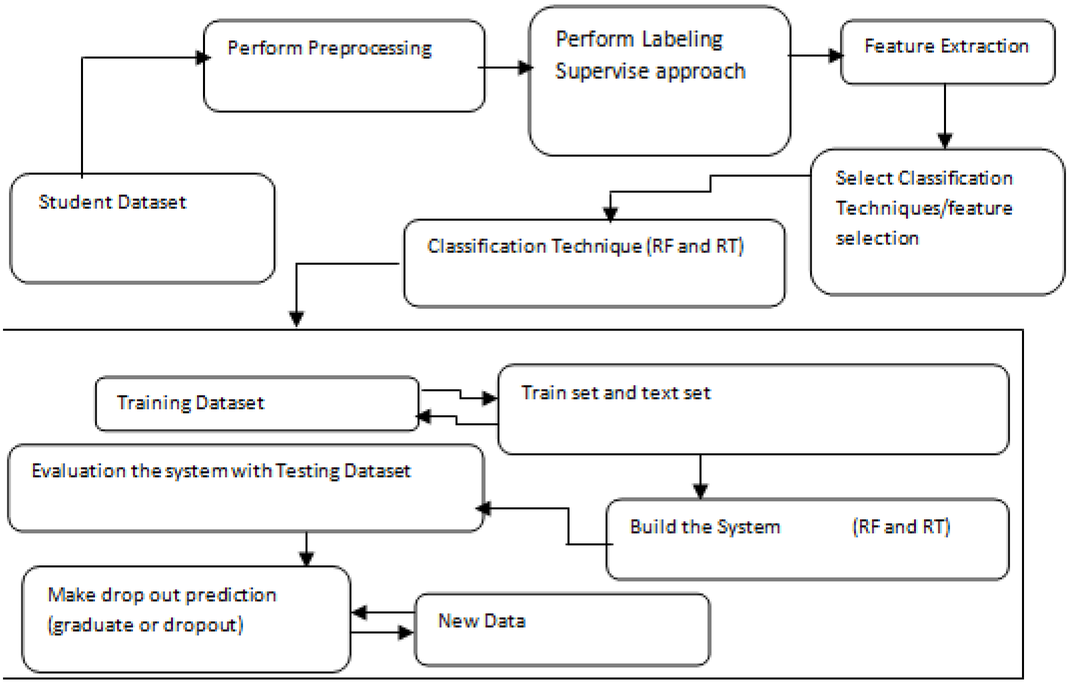

Figure 1.1: working/data flow of Student Attrition Prediction.

Finally, the performance of classifier were summarized and evaluated. Feature extraction and initial analysis of data were done with library Weka, and then applying machine learning algorithm (weka library plugin) was done in NetBeans IDE for the implementation of model.

The paper is organized as follows: section II discusses the related works on data mining, but the work is focused on random forest and random tree. In section III, analyzed the system problem, gave the details of the methodology, dataset description, feature set description and the experimental setup. Section VI presented the system implementation, evaluation then discussion of the results. Section V, presents conclusion of the report, recommendation and point out areas for future works.

2. 1 Related works

This section summarizes the concept of relevant work on predicting student attrition using a machine learning model. Here, the various stages of attrition prediction in higher institution were discussed and approaches to model and analyze student attrition in higher education reach forty years back to the famous survival model [7]. While most of the data mining perspectives on predicting individual student attrition add to our knowledge, not yet all have been taken-on in Nigeria use cases. Meanwhile, several international studies have been conducted. In [8], explores student retention probability on the basis of freshmen background data, including demographic, academic and course participation data. Methodologically, they use classification trees, neural nets and multivariate adaptive regression splines to identify transferred hours, residency and ethnicity as crucial factors for retention

Also, an overview on some of the main studies conducted on data mining approach for predicting study success solely based on enrollment data is conducted [9]. His dataset consists of socio-demographic and course data on which he applies classification/regression trees and feature selection. Ethnicity, course programs and course blocks are identified as most reliable predictors. The maximum prediction performance, however, reaches only an accuracy of 60%. Consequently, it was concluded that data from only the enrollment process is not sufficient to predict study success with satisfactory accuracy [9].

In yet another investigation to understand university student retention from a data mining perspective as executed by [10]. The paper provides not only an analysis on how to determine students at-risk of dropout, but also give practical advice how to use the gathered data to improve retention of those students afterwards. They built their analysis on a dataset, combining several resources from university institutions such as the usage of the library, online resources and the student record system. Thus, they provided a versatile look on the students' academic involvement. Analyzing the algorithms, they found, interestingly, that a naive Bayes approach outperforms support vector machines and decision trees in predicting student success.

Furthermore, in [11], they compared individual classification algorithms with those of the ensemble learning algorithms for identifying students who are having more likelihood of attrition and found that balanced dataset provided better predictions than unbalanced data set. [12] used unsupervised machine learning algorithm namely unique clustering with affinity measures (UCAM) clustering algorithms to avoid the problems of fixing the initial cluster size and seeds.

2.2 Conceptual Framework

In [13], it was proved that multi-layered neural network-based approach performed better than other machine learning algorithms in terms of accuracy without variable reduction to classify the students in to high, medium and low categories.

In [14] they compared classification algorithms for predicting student dropouts in higher educational institutes using three different student representations and they found that random forest and gradient boost ensemble algorithms performed better while Naive Bayes performed the least because of the assumption of strong interdependence. Predicted student success in English exit exam using a c4.5 algorithm (j48) showed that English placement test result was a key predictor for student success in English exit exam success[15]. Research shown, in [16], compared the classification accuracy of various decision trees algorithms such as random tree, reptree, and c4.5 decision tree in predicting whether student will graduate in time or not. There are other notable works which have compared the classifiers based on metrics other than overall classification accuracy such as in [17], proposed ICRM2 algorithm for predicting student dropout early and demonstrated that the ICRM2 algorithm outperforms other classification algorithms in terms of true negative rate and geometric means of true negatives and true positives. Notably, in [18] they compared the three classification algorithms, Naive Bayes, Bagging, and C4.5 in determining non active students in higher education setup at Indonesia Open University based on cross-validation, confusion matrix, ROC curve, recall, precision and F-measure

The Random Tree classifier proved a 57.48% accuracy which means it does not perform well in predicting student failure in university examination using machine learning algorithms. In conclusion, the accuracy of student prediction attrition was enhanced incredibly when the algorithm with feature selection was applied into the entire process and that classifiers of tree shape are more efficient in predicting student attrition from pursing computer science degree.

Another study followed the guidelines of the previous prominent research works to bridge the research gap by comparing the classification algorithms based on other metrics like accuracy, precision, recall and f score as put forward by [19]. The study conducted several many enrolments and attrition rate predictions that have utilized machine learning approaches to identify the student enrolment and attrition pattern.

Classification approaches and algorithms, such as SVM, NB, J48 and CCM, in addition to different features sets in [20], focused on the student's performance as an indicator for the student dropout rate in their freshman year. This study used naïve Bayes, decision tree, and rule-based induction machine learning algorithm to build the best model that predicted the attrition rate with 80% accuracy.

In [21], examined the dropouts rate in an online program for 189 students who registered in the online information technology certificate program. This study used four machine learning algorithms K Nearest Neighbour KNN, Decision Tree (DT), Naive Bays (NB) and Neural Network (NN), resulting in 87% accuracy for NN, and 80% accuracy for DT and finding that student demographics play a very critical role in dropout rates.

In [22], a study with the aim of understanding the student drop rate factors using machine learning algorithms was carried out. ID3 and J48 decision trees were used with weka on 220 undergraduate students in Information Technology courses. In this research; it was found out that personal factors are the most important factors that effect the student attrition and weights for 28% of dropout rate. On the same study, the institutional factors such as the university environment, and the course cost weight for 17% of dropout rate, also it was noticed that few students are most likely to drop out due to homesickness and adjustment problems [23].

In the same prospect; [24] managed to predict the student dropout rate and success factor using logistic regression and decision trees on a resample examination data at the Karlsruhe Institute of Technology (KIT) with 95% accuracy after three semesters of the student enrolments.

On the other hand, multiple authors and studies used machine learning predictive methods and data science techniques to predict student behaviour and performance in educational settings. The researchers [25] performed three different machine learning algorithms; mainly J48, naïve Bayes, and neural network to build three prediction models that predict the student's enrolment at the department level for higher education institutions in both private and public universities.

3. 1 Materials And Method

This section focuses on the concept of predicting students' attrition from pursuing computer science degree using machine learning techniques. The approaches were based on the sample of student’s data with respect to their class label. Based on this fact, the system was built with the available data set collected from Federal University

Lokoja, Nigeria with other related literature review such as journal or articles.

The analysis of the proposed system methodology is based on the concept of collecting sample of student's dataset; these samples were used to form the basis of the approach towards solving the problem definition as follow:

3.2 Machine learning Approach

Collection of the sample data

- Pre-processing (that is the data were provided with two labels, graduate or dropout), since it is a supervised learning approach, then it is a binary classification.
- Apply feature extraction with Weka library (to convert the dataset into binary classification analysis)
- Resample the dataset by applying the training set and testing set during system development analysis using Weka tools.
- Develop the model with Weka library and used java programming language with Netbeans IDE environment to implement the system with all the requirement stated above and used the proposed algorithm to perform the classification model and structured data analytics.

3.3 Random Forest

RF significant improvements in classification accuracy have resulted from growing an ensemble (group) of trees and letting them vote for the most popular class. In order to grow these ensembles, often random vectors are generated that govern the growth of each tree in the ensemble. An early example is bagging, where to grow each tree a random selection (without replacement) is made from the examples in the training set.

Another example is random split selection where at each node the split is selected at random. This generates new training sets by randomizing the outputs in the original training set. Another approach is to select the training set from a random set of weights on the examples in the training set. The method which does a random selection of a subset of features to use to grow each tree was adopted in this research work.

3.3.1 Random forest

Predicting student attrition using random forest takes the next level by combining trees with the notion of an ensemble. Thus, in ensemble terms, the trees are weak learners and the random forest is a strong learner.

Here is how such a system is trained; for some number of trees T:

Sample N cases at random with replacement to create a subset of the data. The subset should be at least about 76% of the total set.

At each node:

- For some number m, m predictor variables need to be selected at random from all the predictor variables.
- The predictor variable that provides the best split, according to some objective function, is used to do a binary split on that node
- At the next node, choose other m variables at random from all predictor variables and do the same

Depending upon the value of m, there are three slightly different systems:

1. Random splitter selection: m =1
2. Breiman's bagger: m = total number of predictor variables
3. Random forest: m<<number of predictor variables. Brieman suggests three possible values forms: ½√m, √m, and 2√m

3.3.2 Random tree

A random tree [26] is a collection of tree predictors that is called forest. It can deal with both classification and regression problems. The classification works as follows: the random trees classifier takes the input feature vector, classifies it with every tree in the forest, and outputs the class label that received the majority of votes. In case of a regression, the classifier response is the average of the responses over all the trees in the forest. All the trees are trained with the same parameters but on different training sets.

3.3.3 System Design

The method to achieve this work as follow:

Student data collection

Student data pre-processing

Student Feature extraction

Training set and Test set

Build the model

Based on this above supervised learning will be used for training of the algorithm with labeled as to which class it belongs. Using the labeled data, the algorithm learns the relationship between the feature sets and the output, and hence it then classifies the unlabeled data from the learned relationship. Hence, conceptual framework of the model.

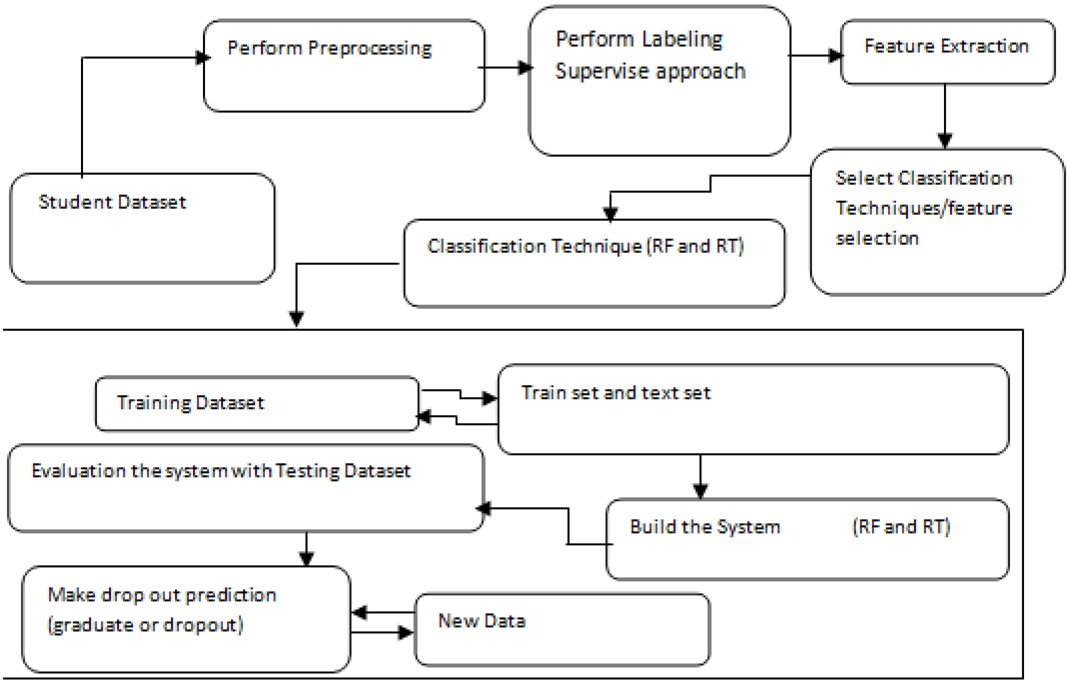

Figure 3.1: Steps for Student Attrition Prediction.

3.4 Pre-Processing

In this step complete geometric correction and filtering is done. The preprocessing uses the output of the classifier to take the required action to improve the performance.

3.4.1 Supervised Classification

Supervised classification requires the prior information which is gathered by the analyst. The analyst must have the sufficient known dataset to generate representative parameters for each class and also algorithms are used to decide decision boundaries. This process is known as training step. Once the classifier is trained it categorizes according to the trained parameters. Commonly used supervised classification approaches for maximum likelihood classification, minimum distance classification and classification techniques.

The main advantage of supervised classification is that, the operator can easily detect an error and try to fix it. The disadvantage is that it becomes costly and time consuming to set a training data and some time the selected training data may not represent the conditions all over the image. The analyst can make errors in the selection of training sets.

3.4.2 Unsupervised Classification

In unsupervised classification there is no need to have the prior knowledge of the classes. There is no interference of human as it is fully automated process. Some clustering algorithms are used to classify an image data. The basic idea is that values within a given data type should be close enough in the measurement space. The result of the unsupervised classification is the spectral classes that are based on the natural grouping of image value. It becomes more popular in the field of GIS database maintenance because this system uses clustering procedure which is fast and uses little operational parameters. The most commonly used unsupervised classifier is migrating means clustering classifier (MMC).

The main advantages of unsupervised classification are time taken is less and it minimizes the possibility of human error and there is no need of prior knowledge. The disadvantage is that sometimes the clusters in spectral region may not match to our perception of classes.

3.4.3 Dataset Description

The dataset used in this paper is student's data of 2015 till date were collected from the Federal University, Lokoja Kogi State, Nigeria was used for this analysis. This dataset consists of attributes and instances. First attribute i.e., class attribute has two possible values graduate or drop out which are nothing but class labels. The dataset has 4407 instances. Attribute one represents name of the label. Second attribute is students course enrollment whose values are nothing but course work or assessment.

3.4.4 Experimental Set Up

All the experiments were carried out in this section are computed using open-source tool Weka using java programming language with Netbeans IDE. WEKA [27] is a data mining software developed by the University of Waikato in New Zealand that apparatus data mining algorithms using the JAVA language. Weka is a milestone in the history of the data mining and machine learning research communities, because it is the only toolkit that has gained such widespread adopt ion. Weka is a bird name of New Zealand. WEKA is a modern feature for developing machine learning (ML) techniques and their application to real- world data mining problems'. Following subsection discusses more on content of dataset, pre-processing of dataset and performed classification and detection with respect to feature selection, random forest, and random tree.

3.4.5 Selection of Training Data

In this step the particular attributes are selected which best describes the pattern for predicting either the attribution is graduate or dropout.

3.4.6 Classification of Outputs

The output of the expected result is classified as to different categories accordingly namely graduate or dropout.

3.5 Implementation and Result

This section focuses on the general implementation and results obtained in this paper. The system developed student dataset with two classes namely graduate or dropout, based on this data set that was collected from Federal University Lokoja, Kogi State Nigeria which formed the basis of student attrition prediction using machine learning algorithm, were achieved with the proposed namely RF and RT. These algorithms were used to train the data collected and model was built from it. This dataset was under gone a pre-processing and feature extraction before applying classification algorithm on it. The classification used for this work was based on RF and RT which able to capture all the required training sample of data and used the test to make student attrition prediction.

This model was implemented with the set of feature set or attributes to distinguish their performance when those factors was structured into the Weka library with java language were used to implement an information table.

3.5.1 Model Evaluation

Experiment of classification model was done on two folds, which are the sample of dataset collected from which was used to perform prediction. And the training set was used to build the model and then used the test set for predicting the result with unknown class label as well as to predict a new class label with their respective class. Below is model evaluation of RT.

Correctly Classified Instances 3166 79.7883 %

Incorrectly Classified Instances 802 20.2117 %

Kappa statistic 0.5582

K&B Relative Info Score 151764.6823 %

K&B Information Score 1471.6516 bits 0.3709 bits/instance

Class complexity | order 0 3847.2872 bits 0.9696 bits/instance

Class complexity | scheme 2470.8673 bits 0.6227 bits/instance

Complexity improvement (Sf) 1376.4199 bits 0.3469 bits/instance

Mean absolute error 0.3104

Root mean squared error 0.372

Relative absolute error 64.791 %

Root relative squared error 75.9992 %

Total Number of Instances 3968

Below is the classification results with random tree:

Correctly Classified Instances 3112 78.4274 %

Incorrectly Classified Instances 856 21.5726 %

Table 4.1: Results and Analysis with Random Forest.

Class	Precision	Recall	F-Measure	ROC Area
graduate	0.783	0.920	0.846	0.876
dropout	0.834	0.613	0.707	0.876

Kappa statistic 0.5418

K&B Relative Info Score 194843.3643 %

K&B Information Score 1889.3826 bits 0.4762 bits/instance

Class complexity | order 0 3847.2872 bits 0.9696 bits/instance

Class complexity | scheme 436219.3642 bits 109.9343 bits/instance

Complexity improvement (Sf) -432372.077 bits -108.9647 bits/instance

Mean absolute error 0.2443

Root mean squared error 0.4337

Relative absolute error 50.9968 %

Root relative squared error 88.6255 %

Total Number of Instances 3968

Table 4.2 Results and Analysis with Random Tree.

Class	Precision	Recall	F-Measure	ROC Area
graduate	0.783	0.920	0.846	0.876
dropout	0.834	0.613	0.707	0.876

3.5.2 Algorithm Comparison

Table 4. 3 Detail Performance Evaluation on test set by class.

Algorithms	Accuracy %	Train Set	Test Set	Binary Class
Random Forest	97.00%	3968	439	*graduate*
Random Tree	54.98%	3968	439	*dropout*

4. Discussion of Results

The result of this paper was achieved using a set performance evaluation by class, notably, random forest achieved a good performance in these results compare to random tree in term of accuracy by algorithms with 54.98% in random forest, 97.00% in random tree respectfully, in both cases. Shown in above table 3.0

5. Conclusion and Future Work

Recent study shown that, researchers has proposed various techniques to predict student attrition, enrollment, failure etc. Basically, major works has been done on enrollment and retention prediction which are biased towards limited feature space. Therefore, there is a need to search for new features for predicting student attrition. The research work started by first providing an overview of the student attrition prediction in their degree course career phenomenon and discussing its impact on businesses, institutions, and individuals. We also explored the behavior of student attrition in a high risk of dropout. Then also summarized and discussed related works done on enrollment, retention, and student attrition prediction.

To develop student attrition prediction model, the paper started by surveying existent system analysis reports of student attrition prediction from the literature and institution and then identified the attrition process that can be converted to a feature set. We identified important of features using machine learning techniques.

As our main contribution, we also introduced correlation based feature i.e. the system didn’t directly utilized the two sets of algorithm proposed as a default, it was customized by update the parameters with the choice of java programming language, those two sets of algorithm were undergone fine-tune, and this show that the results obtained from the set of feature were better than the default algorithm, therefore with the help of feature sets as well as to predict and classify the unknown student attrition with machine learning model.

References

[1]. Gansemer-Topf, A. M. and Schuh, J. H. (2006). Institutional selectivity and institutional expenditures: Examining organizational factors that contribute to retention and graduation. *Research in Higher Education, 47(6):613–642.*

[2]. Yu, C. H., DiGangi, S., Jannasch-Pennell, A., and Kaprolet, C. (2010). A data mining approach for identifying predictors of student retention from sophomore to junior year. *Journal of Data Science,*

8(2):307–325.

[3]. Zhang, Y., Oussena, S., Clark, T., and Kim, H. (2010). Using data mining to improve student retention in higher education: a case study. *International Conference on Enterprise Information Systems*.

[4]. Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems, 49(4):498–506.*

[5]. Lorenz, k., (2018) . Predicting Student Dropout: A Machine Learning https://www.researchgate.net/publication/322919234

[6]. Office, F. L. (2017). Fachkrfteengpassanalyse. Technical report, Arbeitsmarktberichterstattung.

[7]. Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research, 45(1):89–125*

[8]. Yu, C. H., DiGangi, S., Jannasch-Pennell, A., and Kaprolet, C. (2010). A data mining approach for identifying predictors of student retention from sophomore to junior year. *Journal of Data Science, 8(2):307–325.*

[9]. Kovacic, Z. (2010). Early prediction of student success: Mining students' enrollment data. *Proceedings of Informing Science and IT Education Conference, pages 647–665*

[10]. Zhang, Y., Oussena, S., Clark, T., and Kim, H. (2010). Using data mining to improve student retention in higher education: *a case study. International Conference on Enterprise Information Systems*.

[11]. Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. Decision Support Systems, 49(4), 498–506. https://doi.org/10.1016/j.dss.2010.06.003.

[12]. Banumathi, A., & Pethalakshmi, A. (2012). A novel approach for upgrading Indian education by using data mining techniques. 2012 IEEE International Conference on Technology Enhanced Education (ICTEE), 1–5. https://doi.org/10.1109/ICTEE.2012.6208603

[13]. Alam, M. M., Mohiuddin, K., Das, A. K., Islam, Md. K., Kaonain, Md. S., & Ali, Md. H. (2018). A Reduced feature based neural network approach to classify the category of students. *Proceedings of the 2nd International Conference on Innovation in Artificial Intelligence - ICIAI '18, 28–32. https://doi.org/10.1145/3194206.3194218*

[14]. Manrique, R., Nunes, B. P., Marino, O., Casanova, M. A., & Nurmikko-Fuller, T. (2019). An Analysis of Student Representation, Representative Features and Classification Algorithms to Predict Degree Dropout. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge - LAK19, 401–410. https://doi.org/10.1145/3303772.3303800*.

[15]. Puarungroj, W., Boonsirisumpun, N., Pongpatrakant, P., & Phromkhot, S. (2018). Application of Data Mining Techniques for Predicting Student Success in English Exit Exam. *Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication - IMCOM '18, 1–6. https://doi.org/10.1145/3164541.3164638.*

[16]. Supianto, A. A., Julisar Dwitama, A., & Hafis, M. (2018). Decision Tree Usage for Student Graduation Classification: A Comparative Case Study in Faculty of Computer Science Brawijaya University. *2018 International Conference on Sustainable Information Engineering and Technology (SIET), 308–311. https://doi.org/10.1109/SIET.2018.8693158*

[17]. Márquez-Vera, C., Cano, A., Romero, C., Noaman, A. Y. M., Mousa Fardoun, H., & Ventura, S. (2016). Early dropout prediction using data mining: A case study with high school students. *Expert*

Systems, 33(1), 107–124. https://doi.org/10.1111/exsy.12135

[18]. Ratnaningsih, D. J., & Sitanggang, I. S. (2016). Comparative analysis of classification methods in determining non-active student characteristics in Indonesia Open University. *Journal of Applied Statistics, 43(1), 87–97*

[19]. Hossin, M., Sulaiman, M.N. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process, 5(2), 01–11*.

[20]. Petkovski A., Stojkoska B., Trivodaliev K., and Kalajdziski S., (2016) "Analysis of Churn Prediction: A Case Study on Telecommunication services in Macedonia," *in Proceedings of 24th Telecommunications Forum, Belgrade, pp. 1-4, 2016*.

[21]. Yukselturk E., Ozekes S., and Türel Y.,(2014) "Predicting Dropout Student: An Application of Data Mining Methods in an Online Education Program," *European Journal of Open, Distance and e-learning, vol. 17, no. 1, pp.118-133, 2014*

[22]. Rai S. and Jain A., (2013) "Students' Dropout Risk Assessment in Undergraduate Courses of ICT at Residential University-A Case Study," *International Journal of Computer Applications, vol. 84, no. 14, pp. 31-36, 2013*.

[23]. Rai S. and Jain A., (2013) "Students' Dropout Risk Assessment in Undergraduate Courses of ICT at Residential University-A Case Study," *International Journal of Computer Applications, vol. 84, no. 14, pp. 31-36, 2013*

[24]. Kemper L., Vorhoff G., and Wigger B., (2020), "Predicting Student Dropout: A Machine Learning Approach," *European Journal of Higher Education, vol. 10, no. 1, pp. 28-47, 2020.*

[25]. Mulugeta M. and Borena B.,(2013) "Higher Education Students' Enrolment Forecasting System Using Data Mining Application in Ethiopia" *HiLCoE Journal of Computer Science and Technology, vol. 2, no. 2, pp. 37-43, 2013*

[26]. Wikipedia contributors, "Random_tree", Wikipedia, The Free Encyclopedia. *Wikimedia Foundation, 13 -Jul-2014.*

[27]. E. Frank, M. Hall, G. Holmes, R. Kirkby, B. Pfahringer, I. H. itten, and L. Trigg,(2005) Weka, in Data Mining and Knowledge Discovery Handbook, *Springer, 2005, pp. 1305 – 1314*