

Academic City University College – Accra Ghana
Society for Multidisciplinary & Advanced Research Techniques (SMART) Africa
Tony Blair Institute for Global Change
FAIR Forward – Artificial Intelligence for All - Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

Accra Bespoke Multidisciplinary Innovations Conference (ABMIC)

& The Africa AI Stakeholders' Summit

14th December, 2021

A KNN Classification Advisory System for Higher Institution Students' Performance

Royransom Nzeh, Collins Udanor, Izuchukwu Uzo, Nnamdi J. Ezeora, Nnaemeka E Ogbene, Caroline Asogwa & Assumpta O Ezugwu.

Department of Computer Science
University of Nigeria
Nsukka, Nigeria

E-mails: royransom.nzeh@unn.edu.ng, collins.udanor@unn.edu.ng, izuchukwu.uzo@unn.edu.ng, nnamdi.ezeora@unn.edu.ng, nnaemeka.ogbene@unn.edu.ng, caroline.asogwa@unn.edu.ng, assumpta.ezugwu@unn.edu.ng

Phone No: +234755901022



Proceedings Citation Format

Royransom Nzeh, Collins Udanor, Izuchukwu Uzo, Nnamdi J. Ezeora, Nnaemeka E Ogbene, Caroline Asogwa & Assumpta O Ezugwu. (2021): A KNN Classification Advisory System For Higher Institution Students' Performance. Proceedings of the Accra Bespoke Multidisciplinary Innovations Conference. University of Ghana/Academic City University College, Accra, Ghana. December 2021. Pp 1-8 www.isteam.net/ghanabespoke2021. DOI <https://doi.org/10.22624/AIMS/ABMIC2021-V2-P1>

A KNN Classification Advisory System For Higher Institution Students' Performance

Royransom Nzeh, Collins Udanor, Izuchukwu Uzo, Nnamdi J. Ezeora,
Nnaemeka E Ogbene, Caroline Asogwa & Assumpta O Ezugwu

ABSTRACT

This study focuses on enhancing students' performance through data mining techniques to reduce the challenges of low grades associated with some students in higher institutions in Nigeria. Presently, the amount of data stored in most higher institutions' databases are increasing rapidly. Nevertheless, some higher institutions still rely on paper files to store students' results. Even those institutions with databases do not know what to do with their data after graduating the students. Thus, this study proposes a method that will use data mining techniques such as the k-nearest neighbor's algorithm to extract the hidden knowledge available in the higher institution databases. The proposed method of classifying students' performance is useful in identifying the poor students by providing the framework that will guide them to acquire better grades or change to other departments where they may be better suited.

Keywords: Advisory, classification, data mining, KNN, performance.

1. INTRODUCTION

The classification of students' performance is important for educational institutions to assist the students in improving their academic performance by providing them with the opportunity to better their grades. According to [1, 2, 3, 4, 5, 6] as cited in [7], a student's academic accomplishment may be assessed using a combination of features such as course grades, semester Grade Point Average (GPA), cumulative GPA (CGPA) and final GPA. In the same vein, [8] observed that assessing students' performances is a crucial step in preserving their achievement in learning. Nevertheless, the existing studies have shown an increase in the number of students obtaining low grades when they rewrite the courses previously failed. The problem above might be ascribed to the lack of adequate methods that can be used to identify/advise the weak students in tertiary institutions in Nigeria, to assist them to score higher grades.

Background studies have shown that using data mining techniques in classifying the student's performance is a possible solution to mitigating the problem of low grades in higher institutions. Data mining is used to discover patterns in large data sets involving approaches that connect machine learning, statistics, and database systems [9]. The general goal of the data mining process is to extract information from a data set and transform it into a conceivable structure for further use. Data mining in education, known as Educational Data Mining (EDM), is a growing area of research [10]. In this sense, [11] stated that Educational Data Mining (EDM) is a developing method concerned with emerging methods for investigating the distinctive types of data that come from educational settings.

Therefore, to solve the above problem associated with low grades, existing studies used data mining classification techniques such as Naïve Bayes classifier, Decision tree (C4.5), Artificial neural network, Support vector machine, and J48 algorithm to predict students' performance in educational institutions. However, their classifiers presented an overall accuracy below 80%, which implies that the error ratio was a little bit high. Thus, there is a concern about the reliability of their prediction.

To improve the accuracy of the prediction, a method is proposed in this paper that will use data mining k-nearest neighbor's (k-NN) classification method with distance measure in Minkowski distance to classify student's academic performance in a way that will help the students to improve their academic performance. The proposed approach consists of two steps. Firstly, a student's results will be obtained from the previous academic session. Then, fed into the database through a user-friendly interface and preprocessed by normalizing the results to improve the classification's exactness and efficiency. Secondly, the k-nearest neighbor (k-NN) model is utilized to classify students' anticipated performance using their CGPA. Suppose the classification outcome is negative (i.e., the student will fail in the next examination). In that case, additional efforts may be taken prior to the examination to enhance the study method for those students who may fail, thereby aiding them to pass the examination or advising them to change to another.

2. RELATED WORKS

The study [12] proposed a method based on the data mining classification technique to track student performance using the GPA, CGPA, or first-year accomplishment over a long period. The author observes that low academic performance may influence other measures, for example, graduation rates and employability after graduation. As a result, it was suggested that, by using their approach, institutions of higher learning might discover patterns and attach them to particular courses or programs to boost students' academic performance.

Also, [13] developed a system using the CART decision tree to estimate students' academic accomplishments using different features such as gender, parent education, economic background etc. They observed that results from the immediate past academic session might contribute immensely in predicting the students who are required to pass or fail in proceeding examination. Along the same line, [14] developed a meta decision tree classifier method dependent on four agent learning algorithms, such as Dagging, Adaboost, Grading, and Bagging, to create various decision trees. They looked at the four Meta-learning procedures contrastingly with respect to the training set, datasets and presumed that Adaboost is the best meta-decision classifier for analyzing students' performance. Furthermore, they inferred that predicting correctly whether a student will pass or fail in a particular course is subject to the classification approach.

In the same vein, the study [15] showed that educational data mining may be utilized to analyze students' learning behavior and boost their performance. The author applied a decision tree algorithm to identify variously hidden knowledge from students' data that may be used to envisage future low grades promptly. Similarly, [16] used a decision tree classification algorithm to predict the performance of students in subsequent examinations to enhance the quality of the higher education system.

They examined the key features such as age, student gender, student department, and study type that may influence the student's accomplishments in courses. Thus, a system was presented that allows students to predict the final grade of a course under study.

3. PROPOSED METHODOLOGY

Classifying students' performance using some defined metrics entails an initial upload of students' grade reports at the end of every semester, i.e., populating the system's database with up-to-date students' results in order to make accurate performance classification. The proposed system will contain three modules: the administrator's module: used by the system's admin to administer the system, the exam officer module: used by the system's exam officer to upload, edit students' grade reports and compute CGPA at the end of every semester and the student module: used by the student to classify subsequent academic performance.

Proceeding a successful login, the system administrator may perform any of the three functions specified for the role, i.e., training the system, creating users with the exam officer or student role (Note, users with student role can only be created after exam officer uploads students grade report) and classify students' academic performance.

Consequently, the admin officer may classify students' academic performance in 3 steps. Firstly, the admin officer will click on the "Students Performance Prediction" menu, then specify the students' year of study, and finally, click on the "Predict" button. After that, the system processes the request and returns an outcome. For example, suppose there is no student record in the database. In that case, the system displays a message "No students' grade report was uploaded to the system's database," and if a record exists for students in the database, the system displays an outcome in a table view for the concerned students with classification parameters (such as pass, fail and requires assistance or should change to another department).

On the other hand, upon successful login, a user with the exam officer role may perform any of the 3 functions allowed for the role (such as upload students' grade report, edit uploaded grade report, and compute CGPA). To upload students' grade report at the end of every semester, the exam officer, upon successful login, will be required to click on the "Upload Students' Grade Report" menu and fill out the presented form fields, thereafter, upload an Excel sheet or CSV file containing students' grade report. Also, upon successful login, a user with a student role may perform only one function, i.e., to classify their subsequent academic performance.

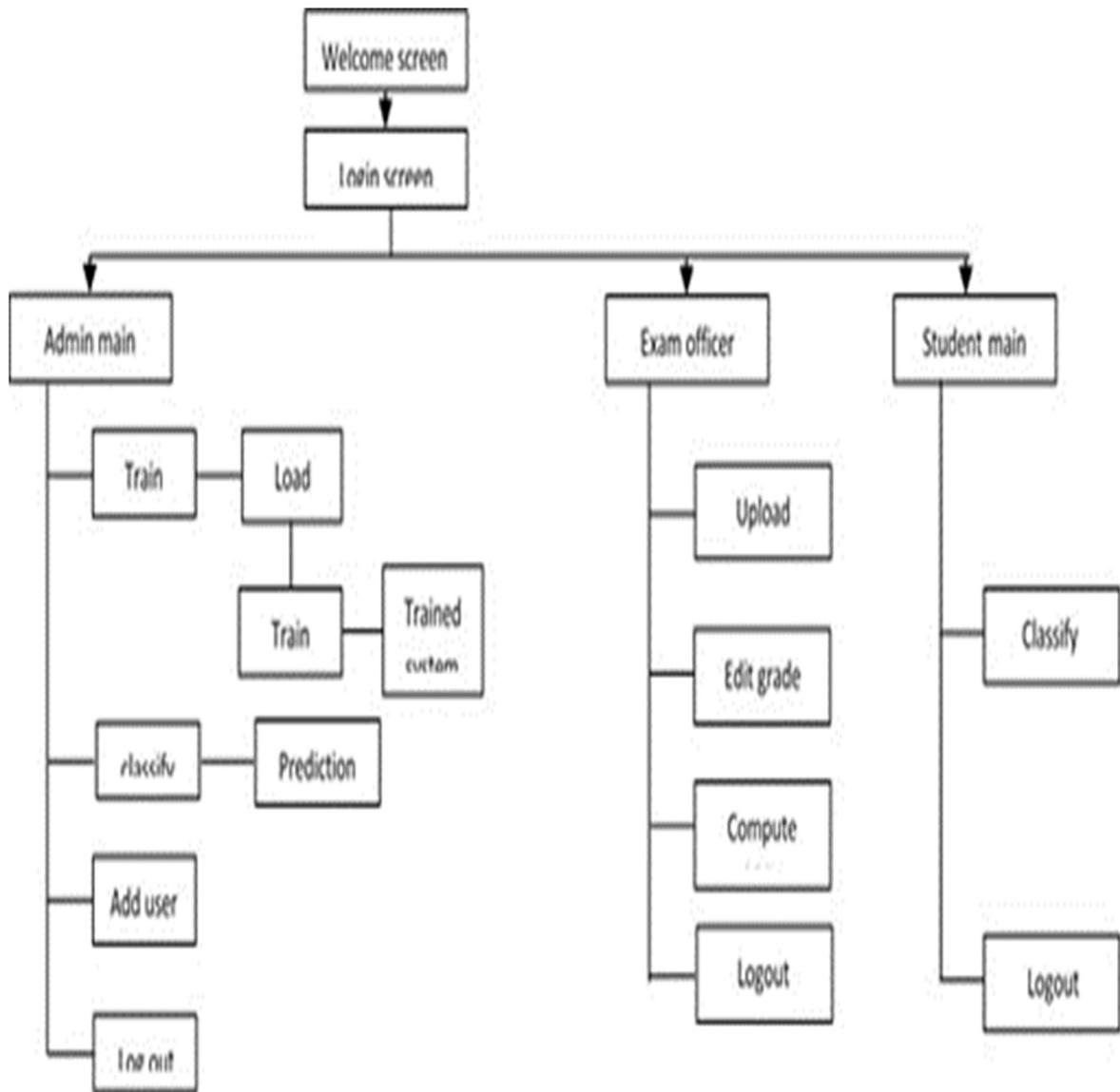


Figure 1: Proposed System Architecture

4. RESULTS

The proposed method will be evaluated using the confusion matrix, the Recall, Precision, Accuracy, and F1 score evaluation metrics [17-18]. We give a brief description of each below.

Confusion Matrix

The confusion matrix is a table that was frequently used to define the performance of a classification model on a set of test data for which the true values are known. It is a table with four different combinations of predicted and actual values. These combinations are true positive (TP), true negative (TN), false positive (FP), and false-negative (FN).

Accuracy

This is calculated as the overall number of accurate predictions divided by the overall number of the dataset.

$$\text{Accuracy} = \frac{TP + TN}{TP+TN+FP+FN}.$$

Precision

This is the proportion of correct positive classification (true positive) from cases predicted as positive.

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall

This is the proportion of correct positive classification (true positive) from truly positive cases.

$$\text{Recall} = \frac{TP}{TP+FN}$$

F1 Score

This is a metric to quantify the performance of an imbalanced dataset, and the higher the F1 score, the better the prediction power of the classification model.

5. CONCLUSION

The academic accomplishment of students in institutions of higher learning has become a major concern for administrators. An initial inquiry of students at the instance of poor performance may help the administration take appropriate action to enhance their performance through additional tutoring and counseling. Thus, the aim of the authors is to implement the proposed method described in section 3.

REFERENCES

- [1] Durairaj, M., & Vijitha, C. (2014). Educational data mining for prediction of student performance using clustering algorithms. *International Journal of Computer Science and Information Technologies (IJCSIT)*, 5(4), 5987–5991
- [2] Sweeney, M., Lester, J., & Rangwala, H. (2015). Next-term student grade prediction. In *IEEE International Conference on Big Data, IEEE Big Data 2015* (pp. 970–975). Santa Clara: Institute of Electrical and Electronics Engineers Inc. <http://doi.org/10.1109/BigData.2015.7363847>
- [3] Jacob, J., Jha, K., Kotak, P., & Puthran, S. (2015). Educational data mining techniques and their applications. In *2015 International Conference on Green Computing and Internet of Things, ICGCloT*, (pp. 1344–1348). Delhi: Institute of Electrical and Electronics Engineers Inc. <http://doi.org/10.1109/ICGCloT.2015.7380675>
- [4] Kolo, K. D., Adepoju, S. A., & Alhassan, J. K. (2015) A decision tree approach for predicting students academic performance. *International Journal of Education and Management Engineering*, 5, 12–19. <http://doi.org/10.5815/ijeme.2015.05.02>
- [5] Arsad, P. M., Buniyamin, N., & Manan, J. A. (2014) Neural network and linear regression methods for prediction of students' academic achievement. In *IEEE Global Engineering Education Conference (EDUCON)* (pp. 916–921). Istanbul: Institute of Electrical and Electronics Engineers Inc. <http://doi.org/10.1109/EDUCON.2014.6826206>
- [6] Tekin, A. (2014). Early prediction of students' grade point averages at graduation: A data mining approach. *Eurasian Journal of Educational Research*, (54), 207–226.
- [7] César A. Del Río & Julio A. Pineda Insuasti. (2014). Predicting academic performance in traditional environments at higher-education institutions using data mining: A review (2016)
- [8] Ibrahim, Z., Rusli, D. (2007). Predicting students' academic performance: comparing artificial neural network, decision tree and linear regression, in: *21st Annual SAS Malaysia Forum*, 5th
- [9] "Data Mining Curriculum", Available a <http://www.kdd.org/curriculum/index.html>
- [10] Shahiri, A. M., Husain, W., & Rashid, N. A. (2015). A Review of Predicting Student's Performance Using Data Mining Techniques. In *3rd Information Systems International Conference, 2015* (Vol. 72, pp. 414–422). Shenzhen: Elsevier. <http://doi.org/10.1016/j.procs.2015.12.157>
- [11] "What is Educational Data Mining (EDM)? ", Available at <http://edtechreview.in/dictionary/394-what-is-educational-data-mining>
- [12] "The 5 Most Commonly Found Metrics for Student Success", Available at <https://evisions.com/5-commonly-found-metrics-student-success/>
- [13] Zlatko J. Kovacic & John Steven Green. (2010). Predictive working tool for early identification of 'at risk' students. New Zealand.
- [14] Shanthini.A. Vinodhini.G and Chandrasekara. R.M. (2018). Predicting Students' Academic Performance in the University Using Meta Decision Tree Classifiers, *Journal of Computer Science*.
- [15] El-Halees, A. (2008). Mining Students Data to Analyze Learning Behavior: A Case Study. The 2008 international Arab Conference of Information Technology (ACIT2008) – Conference Proceedings, University of Sfax, Tunisia.
- [16] Al-Radaideh, Q., Al-Shawakfa, E. and Al-Najjar, M. (2006). Mining Student Data Using Decision Trees. The 2006 International Arab Conference on Information Technology (ACIT'2006) – Conference Proceedings.

- [17] Abhadiomhen, S. E., Wang, Z., & Shen, X. (2021). Coupled low rank representation and subspace clustering. *Applied Intelligence*, 1-17.
- [18] Abhadiomhen, S. E., Wang, Z., Shen, X., & Fan, J. (2021). Multiview Common Subspace Clustering via Coupled Low Rank Representation. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 12(4), 1-25.