Perspective of Artificial Intelligent Performance Prediction Model Student Analysis of Bakti Nusantara Institute of Technology and Business - IBN

Agus Suryana, Fauzi, Eka Ridhawati, Elisabet Yunaeti Anggraeni, Sucipto
Institut Bakti Nusantara, Lampung, Indonesia
E-mail: suryana.suryani64@gmail.com

Abstract :
Educational data mining has received considerable attention in an effort to improve the quality of learning. Many data mining techniques have been proposed to extract the hidden knowledge from educational data. The extracted knowledge helps institutions to improve their respective methods and learning processes of their teaching and learning activities. All these improvements lead to an increase in the performance of students and with efforts to improve the quality of better learning output. In this study, the authors propose a new student performance prediction model based on data mining techniques with new data attribute behavior features, called student's behavioral features. These types of features are related to learner interactivity with the e-learning management system. The performance of the student predictive model is avoided by the classifier set, namely; Neural Networks, Naive Bayesian and Decision Trees. In addition, the authors apply the ensemble method to improve the performance of this classifier. The author uses Bagging, Boosting and Random Forest (RF), which is a common ensemble method used in the literature. The results obtained reveal that there is a strong relationship between learner behavior and academic achievement of the heirs. The accuracy of the proposed model using behavioral features is achieved up to 22.5% A 1% improvement compared to the results when removing such features and reaches up to 25% An 8% increase in accuracy using ensemble methods. By testing the model using newcomer students, more than 80% accuracy was achieved.

Keywords: Performance Analysis, IBN Student Prediction Model, Data Mining, ensemble method
DOI Number: 10.14704/NQ.2022.20.12.NQ77359 **NeuroQuantology2022;20(12): 3510-3521**

1. INTRODUCTION

Predicting student performance is an important task in a web-based educational environment to be able to provide an overview in developing strategic plans for educational institutions. To build predictive models, there are several data mining (DM) techniques used, namely classification, regression and grouping. The most popular technique for predicting student performance is classification. There are several methods classified as Decision Tree (DT), Artificial Neural Networks (ANN) and Naive Bayes (NB) [1].

A decision tree is a set of conditions arranged in a hierarchical frame. Most researchers use this technique because of its simplicity, where it can be converted into a set of classification rules. Some well-known DT algorithms are C4.5[2] and CART. [3] Romero et al in using the DT algorithm to predict the final grades of students based on their usage data in the Moodle system. Moodle is one of the most frequently used Learning Content Management Systems (LCMS). The author has collected real data from seven Moodle courses with Cordoba University to classify students into two groups: pass and fail. The purpose of this study was to classify students with the same final grade into different groups based on the activities carried out in the web-based course. Neural networks are another popular technique that has been used in educational data mining. Neural networks are biologically inspired intelligent techniques consisting of connected elements called neurons that work together to produce an output function [4]. Arshad et al. in [5] used the ANN model to predict the academic performance of undergraduate engineering students. This study takes the Grade Point (GP) of the basic subjects assessed by students as input without considering their demographic background, while it takes the Cumulative Grade Point Average (CGPA) as output. Neural Network (NN) trains GP Engineering Degree students to get targeted output. This study shows that fundamental subjects have a strong influence on the final CGPA after graduation.

The authors at[6] used Bayesian networks to predict CGPA based on the applicant's background at the time of admission. Nowadays, educational institutions need a method to evaluate qualified applicants who graduate from various institutions. This study presents a new approach that integrates case-based components with predictive models. The case-based component takes past students who are most similar to the applicants being evaluated. The challenge is to define the similarity of the cases (applicants) in a way that is consistent with the predictive model. This technique can be applied in any institution that has a good database of student and applicant information.

2. LITERATURE REVIEW

Predicting student performance is an important task in a web-based educational environment. To build predictive models, there are several DM techniques used, namely classification, regression and grouping. The most popular technique for predicting student performance is classification. There are several methods classified such as Decision Tree (DT), Artificial Neural Networks (ANN) and Naive Bayes (NB).

A decision tree is a set of conditions arranged in a hierarchical frame. Most researchers use this technique because of its simplicity, where it can be converted into a set of classification rules. Some well-known DT algorithms are C4.5 [2] and CART. Romero et al in [7] used the DT algorithm to predict the final grades of students based on their usage data in the Moodle system. Moodle is one of the most frequently used Learning Content Management Systems (LCMS). The author has collected real data from seven Moodle courses with Cordoba University to classify students into two groups: pass and fail. The purpose of this study was to classify students with the same final grade into different groups based on the activities carried out in the web-based course.

Neural networks are another popular technique that has been used in educational data mining. Neural network is one of the intelligent techniques inspired by biology which consists of connected elements called neurons that work together to produce output functions [8], [9]. Arshad et al. in [10] used the ANN model to predict academic performance of engineering undergraduates. This study takes the Grade Point (GP) of the basic subjects assessed by students as input without considering their demographic background, while it takes the Cumulative Grade Point Average (CGPA) as output. Neural Network (NN) trains GP Engineering Degree students to get targeted output. This study shows that fundamental subjects have a strong influence on the final CGPA after graduation.

The authors in [6] used Bayesian networks to predict CGPA based on the applicant's background at the time of admission. Nowadays, educational institutions need a method to evaluate qualified applicants who graduate from various institutions. This study presents a new approach that integrates case-based components with predictive models. The case-based component takes past students who are most similar to the applicants being evaluated. The challenge is to define the similarity of the cases (applicants) in a way that is consistent with the predictive model. This technique can be applied in any institution that has a good database of student information and applications. In summary, various studies have been investigated to solve educational problems using data mining techniques. However, there are very few studies that explain student behavior during the learning process and its impact on student academic success. This research will focus on the impact of student data interaction with the e-learning system. In addition, the extracted knowledge will help the campus to improve student academic success and assist administrators in improving the learning quality assurance system.

3. METHOD

In this study, the author uses a model of student performance analysis using the ensemble method. The ensemble method is a learning approach that combines several models to solve problems. In contrast to traditional learning approaches that train data with a single learning model, the ensemble method tries to train data using a set of models, then combine them to take a say in the results. Predictions made by an ensemble are usually more accurate than predictions made by a single model. The aim of the approach is to provide an accurate evaluation of the features that might impact a student's academic success. Figure 1 shows the main steps in the proposed methodology

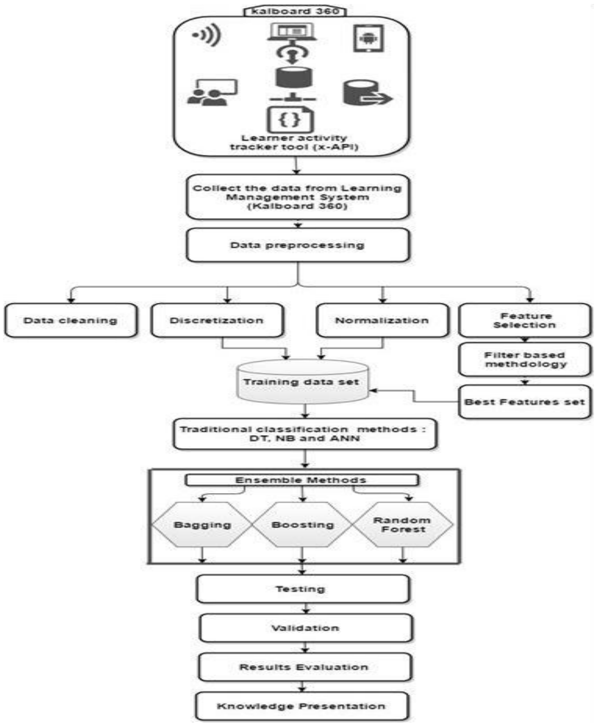

Figure 1. The proposed methodology

This methodology begins by collecting data from the Bakti Nusantara Institute of Technology and Business (LMS) system using the experience API (xAPI). This step is followed by the data preprocessing step, which deals with the transformation of the collected data into a suitable format. After that, the writer uses discretization mechanism to change the student's performance from numerical value to face value, which represents the class label of the classification problem. To achieve this step, the writer divides the data set into three nominal intervals (High Level, Intermediate Level and Low Level) based on the student's total grades/grades such as: Low Level Interval covers grades from 0 to 69, Intermediate Level interval covers grades from 70 to 89 and the High Level interval covers values from 90-100. The data set after discretization consisted of 127 students with Low Level, 211 students with Intermediate Level and 142 students with High Level. Then, the writer uses normalization to scale the attribute values into a small range [0.0 0.0 to 1.0]. This process can speed up the learning process by preventing attributes with large ranges from exceeding attributes with smaller ranges. After that, a feature selection process is applied to select the best feature set with a higher rating. As shown in Figure 1, the authors apply a filter-based technique for feature selection.

In this paper, the ensemble method is applied to provide an accurate evaluation for features that might impact student performance/grade level, and to improve student prediction model performance. The ensemble method is categorized into dependent and independent methods. In the dependent method, the output of the learner is used in the creation of the next learner. Boost is an example of a dependent method. In the independent method, each learner performs independently and their outputs are combined through a voting process. Bagging and random forest are examples of independent methods. These methods re-sampling the original data into data samples, then each sample will be trained by a different classifier. The classifiers used in the student prediction model are Decision Trees (DT), Neural Networks (NN) and Naive Bayesian (NB). The results of individual classifiers are then combined through a voting process, the class chosen by most classifiers is an ensemble decision.

Boosting belongs to a family of algorithms that are able to turn weak learners into strong learners. The general improvement procedure is simple, trains a set of learners sequentially and combines them for prediction, then focuses more on the mistakes of previous learners by editing the weights of the weak learners. A special enhancement constraint that is only used to solve binary classification problems. This limitation is eliminated by the AdaBoost algorithm. AdaBoost is an example of a boosting algorithm, which stands for adaptive boost. The main idea behind this algorithm is to pay more attention to patterns that are difficult to classify. The amount of attention was measured by the weight assigned to each subset in the training set. All subsets are given the same weight. In each iteration, the weight of the misclassified instances increases while the weights of the truly classified instances decrease. Then the AdaBoost ensemble combines learners to produce strong learners from weaker classifiers through a voting process [11].

Bagging is an independent ensemble based method. The purpose of this method is to improve the accuracy of unstable classifiers by creating a composite classifier, then combining the outputs of the studied classifiers into a single prediction. The Bagging algorithm is summarized in Figure 8, starting with resampling the original data into different training data sets (D1-Dn) called bootstrap, each bootstrap sample size is equal to the original training set size. All bootstrap samples will be trained using different classifiers (C1-Cm). The results of individual classifiers are then combined through a majority vote process, the class chosen is by the number of classifiers with the highest number of ensemble decisions [12]. In upgrading, as opposed to bagging, each classifier is affected by the performance of the previous classifier.

In pocketing, each data sample is selected with equal probability , whereas in upgrading, instances are selected with a probability proportional to their weight. In addition, bagging works best with high variance models that produce generalized variance behavior with small changes to the training data. Decision trees and neural networks are examples of high variance models

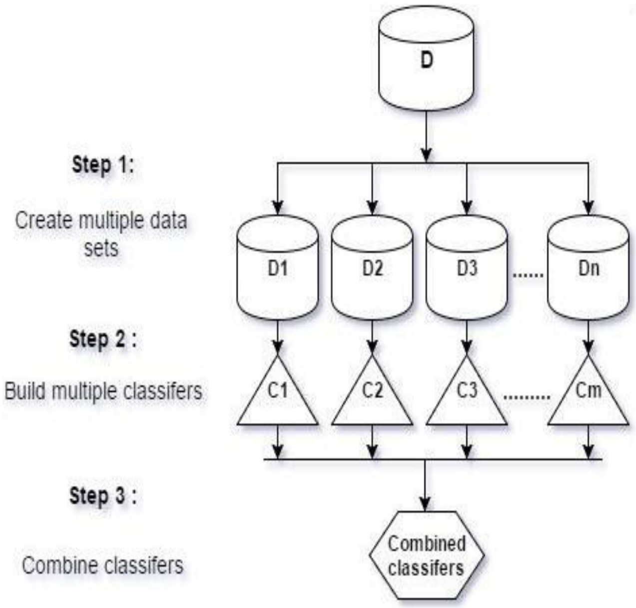

Figure 2. Decision trees and neural networks

Random Forest (RF) is a special modification of bagging where the main difference with bagging is the integration of random feature selection. Through the decision tree construction process, RF uses a random decision tree to select a random subset of features. Note that randomness is only done in the feature selection process, but the selection of the dividing point on the selected features is done by bagging. The combination of decision trees and bootstrap makes RF strong enough to overcome the problem of overfitting, and to reduce the correlation between trees that provide accurate predictions [13]. All of the above classification methods were trained using 10-fold cross-validation. This technique divides the data set into 10 subsets of the same size , nine subsets are used for training, while one is left and used for testing. This process was repeated ten times, the final result estimated as the average error rate of the test sample. After the classification model is trained, the validation process begins. The validation process is the last phase to build a predictive model, it is used to evaluate the performance of a predictive model by running the model through real data.

4. RESULTS

There are many features that directly or indirectly affect the effectiveness of the student performance model. In this section, the authors will evaluate the impact of behavioral features on students' academic performance using different classification techniques such as (DT, ANN and NB). After applying classification techniques to the data set, the results differ based on different data mining measurements. The results of the analysis show the classification results using several classification algorithms (ANN, NB and DT). Each classifier introduces two classification results: (1) classification results with student behavioral features (BF) and (2) classification results without behavioral features (WBF).

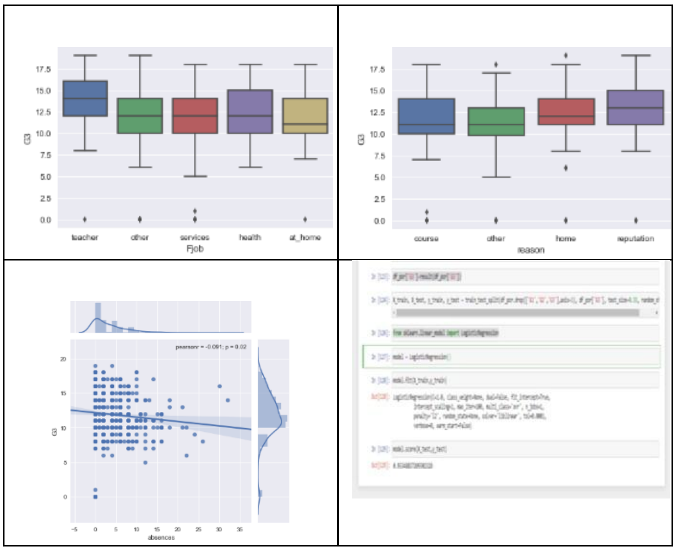

Figure 3. Visualization of Dataset Classification Algorithm

4.1 Prediction Analysis of Dataset Models for IBN Student Data Prediction Modeling

4.1.1 Predictive comparative model for IBN student performance method

Model	Decision Tree	Random Forest	Logistic Regression	SVC	Ada Boost	Stochastic Gradient Descent
Model Score	0.869041	0.962192	0.887671	0.932877	0.867123	0.856164
Cross Validation	0.891720	0.853057	0.904459	0.853503	0.856688	0.821656

Figure 4. Comparison of prediction algorithm datamining methods

Based on the prediction accuracy for the classification algorithm, the accuracy level of the algorithm can be obtained which can be used to extract information based on the data description that can be processed into defined models.

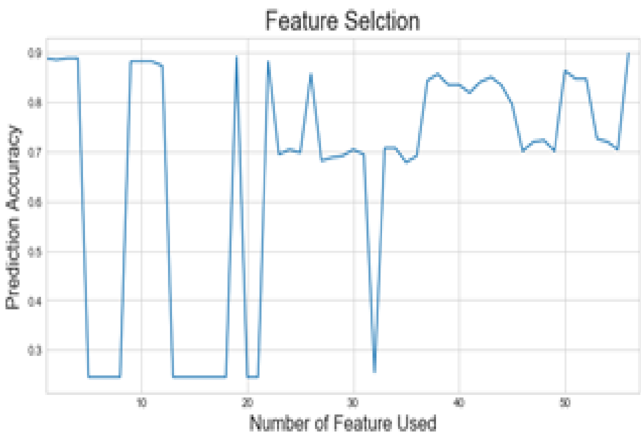

Figure 5. Prediction Accuracy Prediction algorithm classification

Table 1. Results of Classification Method Using Ensemble Method

Evaluation Measure	Traditional classification methods		Bagging			Boosting			Random Forest
Classifiers type	DT	ANN	NB	DT	ANN	NB	DT	ANN	NB	DT
Accuracy	75.8	79.1	67.7	75.6	78.9	67.2	77.7	79.1	72.2	75.6
Recall	75.8	79.2	67.7	75.6	79.0	67.3	77.7	79.2	72.3	75.6
Precision	76.0	79.1	67.5	75.7	78.9	67.1	77.8	79.1	72.4	75.6
F-Measure	75.9	79.1	67.1	75.6	78.9	66.7	77.7	79.1	71.8	75.5

4.1.2 IBN Student Analysis Matric Performance Correlation Model

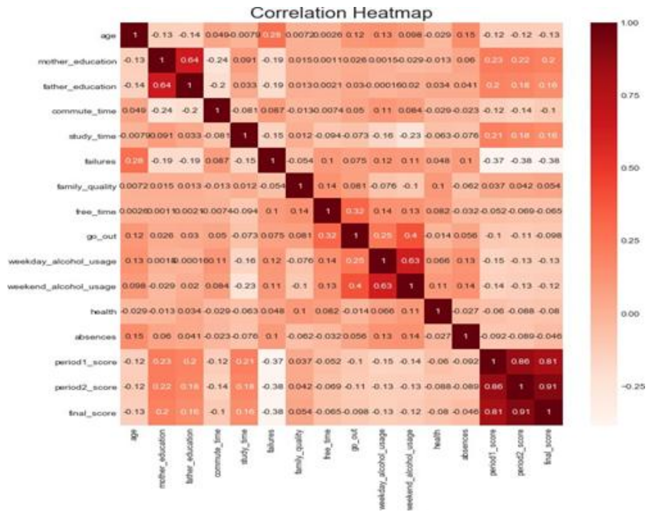

Figure 6. Prediction Model of Matric Performance Correlation Analysis of IBN Students

4.1.3 Prediction Model of Final Grade Competency Level of IBN Student Learning Outcomes

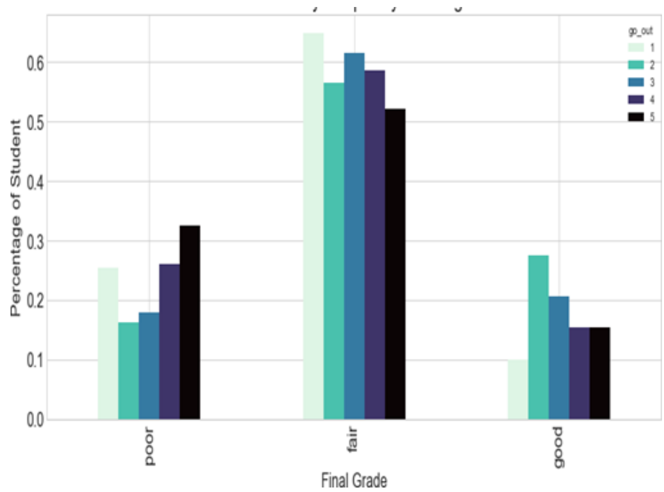

Figure 7. Model of IBN Student Competency Success Prediction Approach

This data is used to model the approach to predicting student achievement at IBN in achieving the competence of the department's curriculum. Attribute data included student grades, demographics, social and school-related features and were collected using study program evaluation reports and questionnaires. The author has classified these students into three categories, "good", "fair", and "poor", according to their final examination performance. Then the authors analyzed several features that have a significant influence on the final performance of students, including Romantic Status, Cigarette Consumption, Parents' Education Level, Exit Frequency, Desire for Higher Education and Living Room. Finally, taking advantage of the available features, the authors have created various machine learning models to predict the final classification of student performance and have compared the performance of the models based on one-time sample accuracy scores.

4.1.4 IBN Student Study Period Accuracy Prediction Model

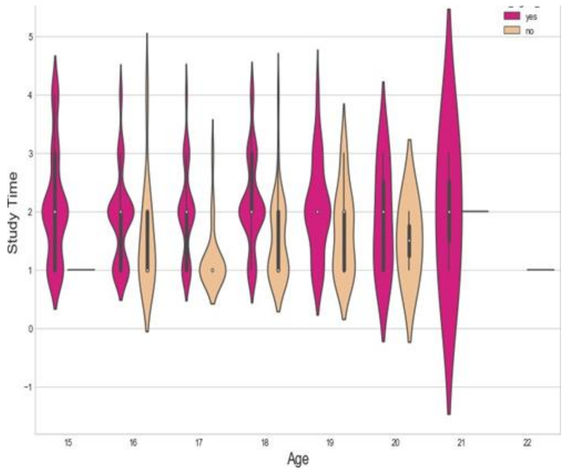

Figure 8. Prediction Model of IBN Student Study Period Accuracy

4.1.5 Prediction Model of Student Competence Success in 4 Study Programs, Faculty of Computer Science and Faculty of Economics

Prediction of student competencies that will be achieved from 4 student study programs of information systems, information management, management and digital business study programs using data intervals using intelligent learning machines that apply 5 prediction algorithms with 3 categories of “poor”, “fair” and “Good”.

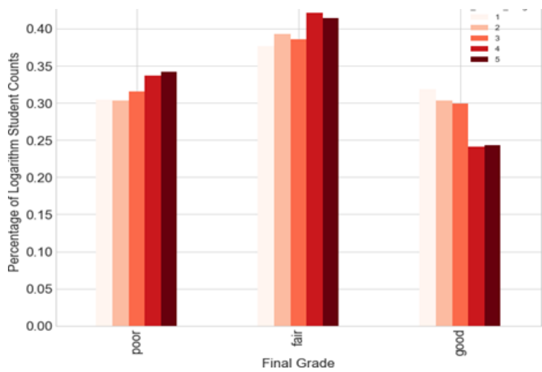

Figure 9. Prediction Model of Study Program Competency Success

4.2 Evaluation Results Using Traditional DM Techniques

Table 2. Classification Method Results with Behavioral Features (BF) and Results without Behavioral Features (WBF)

Evaluation Measure	DT (J48)		ANN		NB	
Behavioral features existence	BF	WBF	BF	WBF	BF	WBF
Accuracy Recall	75.8	55.6	79.1	57.0	67.7	46.4
Precision	75.8	55.6	79.2	57.1	67.7	46.5
F-Measure	76.0	56.0	79.1	57.2	67.5	46.8
	75.9	55.7	79.1	57.1	67.1	46.4

As shown in Table 2, we can see that the ANN model outperforms other data mining techniques. The ANN model achieved an accuracy of 79.1 with BF and 57.0 without behavioral features . An accuracy of 79.1 means that 380 out of 480 students were correctly classified to the appropriate class label (High, Medium and Low) and 100 students were misclassified. For the recall measure, the results were 79.2 with BF and 57.1 without behavioral features. A draw of 79.2 means that 380 students were classified correctly to the total number of cases that were not classified and classified correctly. For precision measures, the result is 79.1 with BF and 57.2 without behavioral features. A precision of 79.1 means that 380 out of 480 students were classified correctly and 100 students were misclassified.

For F-Measures, the results are 79.1 with BF and 57.1 without behavioral features. The experimental results prove the strong influence of student behavior on student academic achievement. We can get more accurate results by training the data set with the ensemble

In this section, the authors apply the ensemble method to improve the evaluation results of the traditional DM method. Table 2, presents the results of traditional classifiers and results of traditional classifiers using ensemble methods (Bagging, Boosting and RF). As shown in Table 2, we can see good results using the ensemble method with traditional classifiers (ANN, NB and DT). Each ensemble trains three classifiers, then aggregates the results through a majority voting process to achieve the best prediction of student model performance. The boosting method outperformed other ensemble methods, where the accuracy of the DT using boosting was increased from 75.8 to 77.7, which means that the number of correctly classified students increased from 363 to 373 from 480. Recall results increased from 75.8 to 77.7, which means that 373 students were correctly classified to the total number of cases that were not classified and correctly classified. The precision of the results also increased from 76.0 to 77.8, which means that 373 out of 480 students were classified correctly and 107 students were misclassified.

Table 3. Results of Classification Methods Using Ensemble Methods

Evaluation						
Measure Traditional						
classification methods			Baggage Boost			Random
Forest						
Classifiers						
type	DT	ANN	NB	DT	ANN	NB
DT		ANN	NB	DT		
Accuracy		75.8	79.1	67.7	75.6	78.9
67.2		77.7	79.1	72.2	75.6	
Recall	75.8	79.2	67.7	75.6	79.0	67.3
77.7		79.2	72.3	75.6		
Precision		76.0	79.1	67.5	75.7	78.9
67.1		77.8	79.1	72.4	75.6	
F-Measure		75.9	79.1	67.1	75.6	78.9
66.7		77.7	79.1	71.8	75.5	

Boosting also achieved a marked improvement with the NB model, where the accuracy of NB using boosting increased from 67.7 to 72.2, which means the number of correctly classified students increased from 324 to 346 out of 480 students. The recall results increased from 67.7 to 72.3 , which means that 347 students were classified correctly to the total number of cases that were not classified and classified correctly. The precision result also increased from 67.5 to 72.4 , which means 347 out of 480 students were classified correctly. The performance of the ANN model using the boosting method is not much different from the results of the ANN model without boosting. After the classification model is trained using 10-fold cross-validation, the validation process begins. Validation is an important phase in building a predictive model, it determines how realistic the predictive model is. In this study, the model was trained using 500 students and the model was validated using 25 newcomer students. In validation, the data set contains unknown labels to evaluate the reliability of the trained model. Table 4 shows the results of the evaluation using several classification methods (ANN, NB and DT) through the testing process and the validation process.

Table 4. Classification Methods Results through Testing and Validation

Precision	79.1	67.5	85.0	84.7	83.8
76.0					
F-Measure	75.9	79.1	67.1	81.8	79.2
80.2					

As shown in Table 4, we can see that the evaluation measurement results improved for all three predictive models through the validation process. All three prediction models achieved an accuracy of more than 80%, which means that 20 of the 25 new students were correctly classified to the correct class label (High, Medium and Low) and 5 students were misclassified. The results of the validation process prove the reliability of the proposed model.

5. CONCLUSION

Academic achievement is of great concern to academic institutions around the world. The widespread use of LMS produces a large amount of data on teaching and learning interactions. This data contains hidden knowledge that can be used to improve student academic achievement. In this study, the author proposes a new student performance prediction model based on data mining techniques with new data attributes/features, called student data behavior features. This type of feature is related to student interactivity with the learning management system. The performance of the student predictive model is evaluated by a set of classifiers, namely; Neural Networks, Naive Bayesian and Decision Trees. In addition, the authors apply the ensemble method to improve the performance of this classifier. The author uses Bagging, Boosting and Random Forest (RF), which is a common ensemble method used in the literature. The results obtained reveal that there is a strong relationship between student behavior and their academic achievement. The accuracy of the student predictive model using behavioral features achieved an increase of up to 22.1 % compared to the results when removing the feature, and achieved an increase in accuracy of up to 25.8% using the ensemble method. The resource visited feature was the most effective behavioral feature in the student performance model. In future research, the author will focus more on analyzing this kind of feature. After completing the training process, the predictive model was tested using unlabeled newcomers, achieving more than 80% accuracy. These results prove how realistic the predictive model is. Finally, this model can help lecturers as educators to understand students, identify weak students, to improve the learning process and reduce the rate of academic failure. It can also help administrators to improve learning system outcomes in the academic quality assurance process.

ACKNOWLEDGMENT

This paper has been presented in National Seminar on Technology, Business and Multidisciplinary Research in Yogyakarta, Indonesia, 23 – 24 August 2022. This work is supported by Institut Bakti Nusantara, Lampung, Indonesia. We gratefully appreciate this support.

6. REFERENCES

[1] A. alfiani, F. W.-A. and A. S. Procedia, and undefined 2015, “Mapping student's performance based on data mining approach (a case study),” *Elsevier* .

[2] L. Yulianto, A. Triayudi, IS-J. Mantik, and undefined 2020, “with Comparison of K-Nearest Neighbor Data Mining Method and Decision Tree C4. 5: Implementation of Educational Data Mining For Analysis of Student Performance …,” *iocscience.org* ,vol. 4, no. 1, pp. 441–451, 2020.

[3] A. Varsi, S. Maskell, and PG Spirakis, “An o(Log2 n) fully-balanced resampling algorithm for particle filters on distributed memory architectures," *Algorithms* , vol. 14, no. 12, Dec. 2021.

[4] N. Papernot, I. Goodfellow, M. Abadi, K. Talwar, and . Erlingsson, “Semi-

supervised knowledge transfer for deep learning from private training data," *5th Int. conf. Learn. Representative. ICLR 2017 - Conf. Track Proc.* , 2017.

[5] " algorithm and data mining and performance and analysis... - Google Scholar." [On line]. Available: https://scholar.google.co.id/scholar?hl=id&as_sdt=0% 2C5&q=algorithm+and+datamining+and+performance e+and+analysis+and+students&oq=algorithm+and+da tamining+and+performance+and+analysis+and+stude n. [Accessed: 15-Aug-2022].

[6] LL di Langosco, V. Fortuin, and H. Strathmann,"Neural Variational Gradient Descent," 2021.

[7] R. Agarwal, N. Frost, C. Xuezhou Zhang, R. Caruana, and GE Hinton, "Neural Additive Models: Interpretable Machine Learning with Neural Nets."

[8] K. Kwon, D. Kim, and H. Park, "A parallel MR imaging method using multilayer perceptron :, " *Med. Phys.* , vol. 44, no. 12, pp. 6209–6224, Dec. 2017.

[9] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.* , vol. 42, no. 2, pp. 386–397, Feb. 2020.

[10] T. Vande Vyvere, G. Wilms, L. Claes, F. Martin Leon, D. Nieboer, J. Verheyden, L. Van Den Hauwe, P. Pullens, AIR Maas, and PM Parizel, "Central versus local radiological reading of acute computed tomography characteristics in multi-center traumatic brain injury research," *J. Neurotrauma* , vol. 36, no. 7, pp. 1080–1092, Apr. 2019.

[11] D. Wu, X. Wang, J. Bai, X. Xu, B. Ouyang, Y. Li, H.Zhang, Q. Song, K. Cao, and Y. Yin, "Automated anatomical labeling of coronary arteries via bidirectional tree LSTMs," *Int. J. Comput. Assists. radiol. Heaven.* , vol. 14, no. 2, pp. 271–280, Feb. 2019.

[12] M. Windarti, PP-IC on, and undefined 2020, "Prediction Analysis Student Graduate Using Multilayer Perceptrons," *atlantis-press.com* , 2020.

[13] H. Nori, S. Jenkins, P. Koch, and R. Caruana, "InterpretML: A Unified Framework for Machines Learning Interpretability."