Journal of Theoretical and Applied Information Technology
15th January 2023. Vol.101. No 1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ALGORITHM MODELING TO PREDICT STUDENTS LEARNING ACHIEVEMENT BASED ON BEHAVIORAL PARAMETERS AS THE IMPLEMENTATION OF LEARNING MANAGEMENT

AHMAD QURTUBI[1]

[1] Department of Islamic Education Management UIN Sultan Maulana Hasanuddin Banten, Indonesia
E-mail: [1]aqurtubijurnal@gmail.com

ABSTRACT

The rapid development of online technology-based education has driven the implementation of intelligent campuses and provides a broad platform for information, new learning style of students and smart learning system. Supporting student success in smart learning is an issue that every university needs to solve. The paper aims to design learning achievement prediction to classify students' campus behavior characteristics, and then algorithms are used to analyze the correlation between student behavior characteristics and their academic success. This was then applied algorithms of Naive Bayes, Random Forest, K-means and C4.5. The simulation results showed that the algorithm with high predictability accuracy is Naive Bayes (71%), followed by Random Forest (63%), K-means (51%), and finally C.4.5 (39%). Then the number of criteria is selected to set the total limit and the best value that chosen by calculating the ratio of internal and outer distances. K-means method is used to analyze performance on student learning style. The best K-means algorithm has predicted the student's success well, and the average score of the academic performance. These study concluded that naïve Bayes algorithm have higher accuracy than Random Forest and C4.5 algorithms. Colleges and universities can support the student learning achievement by measuring their style initiatives for different types of students, which will not only help improve the academic performance of students but also further improve the effectiveness of the learning style of the students. The proposed model in this study could be more developed to predict student success and help them learn better.

Keywords: *Naive Bayes, Random Forest, C4.5, K-Means Clustering, Students Learning Achievement*

1. INTRODUCTION

Since the Covid-19 epidemic, the learning system has turned online, using mobile internet, cloud computing, artificial intelligence and simulation result technology[1]. The students and teachers must rely on the support of science and learning style and the activities of universities is no exception need the development of academic information from digital campus 1.0 to college and university information to enter new era of 4.0 Industrial revolution[2]. In recent years, the construction of intelligent campuses in colleges and universities has been steadily progressing and gradually increasing. More and more application systems are being used by students in colleges and universities. Different application systems, such as one card systems, wireless campus networks, educational information systems and student learning style systems, generate large amounts of data every day[3]. These figures provide a solid basis for replicating mining outcomes in education.

Studies on implementation of learning style to predict students' learning achievement are still being developed to support their achievement. However, the learning style is rarely documented in research mainstream especially to analyze students' education, life, achievements and other habits in detail and to assess students' preferences[4]. The advantage of simulation model is useful for helping students learn and also simulation learning style is useful for teachers and colleges to match student development with the way students learn.

Different methods are done to find out the reason for anything that affects learning. With the rapid development of learning style and knowledge-based computer systems, teachers should start using mock learning styles to help solve problems using data mining ratings to detect student learning predictions[5]. To do so, it needs a procedure that can process student data which

collected from the university data. Based on previous research, the study was conducted by applying data mining methods to compare and detect student achievement by using Naïve Bayes, Random Forests, K-means and C4.5[6]. This study also will compare which algorithm has the highest accuracy, in terms of prediction in the student achievements that previous studies have never done.

The novelty of this study is to suggest an clustering of student behavior parameter representing their learning style to predict students' learning success from data collected in the school database[7]. The update of this study is that this study analyzes the comparison of the average student achievement of the five Naive Bayes algorithms, C4.5, Random forest and K-means. The student behavior in schools is analyzed by comparing five types of algorithms to determine the accuracy and performance levels of each algorithm.

The choice to use the C4.5, Naive Bayes, and Random Forest algorithms in this study is based on several reasons, namely: some advantages of each. C4.5 is a decision tree classification algorithm that is efficient in handling discrete and numeric type attributes [26]. The Naive Bayes algorithm, Han et al., [26] explains that this algorithm only requires one scan of the training data. Meanwhile, the Random Forest algorithm is based on a statement Tan, [27] which states that the random forest algorithm can handle very large amounts of training data efficiently and is an effective method for estimating missing data.

The study consists of five parts. The first section consists of the formation of the problem. The second part includes a review of literature and algorithm theories. The third part includes model development process. The fourth part includes modeling initiatives; the fifth includes simulation results, discussion and conclusion.

2. LITERATURE REVIEW

2.1 K-Means Algorithms To Create Clustering Of Student Behavior Criteria

K-Means algorithm, also known as the K-mean Procedure or K-means approach, is a classical compound algorithm used for many computer functions[8]. This is the distance of the database as a task to improve the quality of the original database. The algorithm finds the end of the function, and divides the dataset into categories such as the J-Score Index, making it equal space between categories[9]. The search for the K-means computation algorithm is limited to a small part of the total potential space. If the sample size between each class is small, the main algorithm tends to achieve better results. However, if the similarities of the samples between the classes are high, then there may be more grouping. Therefore, because of the pairs of algorithms, it is possible to get the smallest solution assessment work for a region rather than the whole dataset.

The algorithm method is used to measure some points in the sample. When faced with synthesis problems, the algorithm is more effective at the end of work. Algorithms are also highly measurable and effective in the case of large databases. Since complexity depended upon N is a collection of all objects, the value of k is defined by the user, and its combination of iteration[10]. The advantage of this algorithm is that when the means clustering algorithm is used, the difference between criteria is clear and the clustering effect is better[11]. The main disadvantage of the Mains algorithm is that determining different values often leads to completely different results. You can use this algorithm to analyze data distribution, such as center, control group, and density, and select the appropriate value until the cluster center stops.

K-means algorithm is a common algorithm for dealing with mixed attribute grouping. The idea of the algorithm process is inherited, and the inequality equation between sources and mixed attribute data is added. For the population the general definition is $X = \{X_1, X_2, X_3, \dots, X_n\}$ represents a data set , where a feature of data is M. data $A_j = \{A_1, A_2, A_3, \dots, A_m\}$ where A_j Stands for attribute$j$For digital features, domain ($A_j$stands for limitation; for clear features, domain (A_j) is a set of standards. X_{ij} represent j Data Features i. Similarly Data X_{ij} can also be expressed as

$$X_{ij} = (A_i = X_i) \wedge (A_{i_2} = X_{i_2}) \wedge (A_{i_3} = X_{i_3}) \dots \wedge (A_{i_n} = X_{i_n}) \dots\dots\dots (1)$$

Total Data by M Attribute, want to set the first P Attribute for Digital Attribute R, for rating feature c Follow the following rules

$$[X_{i1}^m, x_{i2}^m, \dots, X_{in}^m] \wedge [x_{i1}^p, x_{i2}^p, \dots, X_{in}^p] \wedge [X_{i1}^r, x_{i2}^r, \dots, X_{in}^m] \dots\dots\dots (2)$$

K-means algorithm establishes an objective function, which is similar to the S a set of squares of errors, and repeats until the objective function is permanent [12]. At the same time, K means algorithm suggests a means of clustering of mixed attributes; we can understand that means is the center of a cluster of digital attributes. Mixed properties have numerical characteristics and hierarchical characteristics, whose sources are defined as the average values of all attribute values in attributes for the meaning of digital attributes, and the hierarchical attribute means attribute with

the highest frequency in the selection of attribute values. Before you group data, the data is standard and dimensionless. Then cluster analysis is done using SPSS modeler software, and a new data stream is created.

2.2. Optimized K-Means Algorithm

This paper suggests a custom algorithm by modifying conventional K-means into Optimized K-means algorithm. The purpose of the optimization is to select which method to complete the initial grades. The trick is to limit the scope of grouping according to the actual situation. The algorithm will run V Time and choose the best number of these criteria as the best number of criteria as per the following equation. The value of v is calculated in equation 3:

$$V = \frac{d_{inside\ distance}}{d_{outer\ distance}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

$$d_{inside\ distance} = \sum_{i=1}^{K}\sum_{j=1}^{n_j}(x_{ij} - c_i)^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots (4)$$

$$d_{outer\ distance} = \frac{1}{k(k-1)}\sum_{i=1}^{K}\sum_{j=1}^{K}(c_i - c_j)^2 \quad \ldots\ldots\ldots\ldots (5)$$

from k There are number of centres. This type means that when the ratio of coordination with outer distance is smallest, it means that the cluster is more harmonious and the criteria are joined together, that is, the number of criteria is the best number of criteria in the equation (4).

Regarding the selection of the starting points, the method adopted in this paper is to measure the starting point should meet the condition of the distance between these central points is maximum. The points around these starting centers should be dense. Regarding the midpoint distance, we also show the average of money shown in Equation 5 so that we can get the average of all distance collections of the focal point and reflects the total size of the distance between the centers of the higher complex.

2.3. Bayesian Modeling To Calculate The Chance Of Student Success

The model algorithm spelt out in this study is based on the model algorithm. For this time, 3 models have been used in research on modeling using the Naive Bayes algorithm[13]. The model of the algorithm was processed using the fast miners 7.3 tool. Naive Bayes algorithm is used because its popularity and its good way in machine learning based on data training, using conditional probability as needed[14]. According to scholars, the Naive Bayes classification is a statistical classification that can be used to predict the possibility of class membership. In this case, Naive Bayes' theorem-based rating is named after mathematician Thomas Bayes. Naive Bayes are the rules of procedure used to calculate class possibilities. Naive Bayes algorithm provides a way to combine difficulties or development opportunities with a term that is likely to be equation (6) that can be used to calculate the difficulties of any possibility of it happening. As far as the general form of theory is concerned, the Naive Bayes rule is as follows:

$$P(H|X) = \frac{p(X|H)p(X)}{p(X)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

2.4. Random Forest Algorithm

Random Forest (RF) is an algorithm or the same decision-derivative development of the tree. RF algorithm consists of several trees or decision trees in which each tree is trained in sample data . Random Forest (RF) method is a method that can improve the accuracy of results, as every node is randomly done on children's nodes in one way or the other to wake up[15]. This method is used to create decision trees containing root nodes, internal nodes and leaf nodes by retrieving random data features and appropriate conditions. Root node is a node located above, or is commonly known as the root of the decision tree. The internal node is a branch node, where it has at least two outputs and only one input, while the life node or terminal node is the last node with only one input and no output. The following equations can be used to assess the value of entropy and the value of obtaining information:

$$Entropy\ (Y) = -\sum_{i}(c|Y)\log 2\, p(c|Y) \quad \ldots\ldots\ldots\ldots\ldots (7)$$

where Y is a collection of cases and $p(c|Y)$ and Y The ratio of classes is c And entropy to measure the power of attribute

$$(Y, a) = Entropy\ (Y) \sum_{v \in Value(a)} \frac{|Y_v|}{|Y_a|} Entropy(Y_v) \ldots. (8)$$

where $value(a)$ Represents all possible values in relation to cases. Y_V is a sub-section of Y class v Related to classes a. Y_a, all are in accordance with values a. Then choosing features as nodes based on the acquisition of information, either root or internal nodes is the most attribute in existence. We also use a gin ratio to measure comparison of estimates between multiple entropies.

$$S_D = (S \times A) = \sum_{i=1}^{c}\left(\frac{|S_i|}{|S_c|}\right) log2 \left(\frac{|S_c|}{|S_i|}\right) \quad \ldots\ldots\ldots\ldots\ldots (9)$$

where S_D There is separate information from the data. In this equation, S Input variable has an

estimated value of entropy, A Number of classes c and $\frac{|S_i|}{|S_c|}$ Probability is the status i Attribute.

$$Gain\ Ratio\ (S,A) = \frac{Information\ Gain\ (S,A)}{split\ data\ (S,A)} \dots\dots(10)$$

2.5. C4.5 Algorithm for Decision Weight

Algorithm C4.5 is so-called decision tree algorithm or an algorithm that imagines the distribution and priority approach to the classification process [16]. C4.5 is basically decided in tree rate that can be done in several stages, either the data from selected node structuring, or the selected node is the smallest value of entropy search result with the goal of reaching the final step and to make decision trees using rules.

Algorithm C4.5 There are several stages in making a decision tree among others to prepare training data, find and calculate entropy before finding each entropy class, calculate the value of the benefit and the average gain. So we will calculate entropy using the equation below:

$$H(X) = \sum_j X - p_j \times \log_{2(p_j)} \dots\dots\dots\dots\dots\dots\dots\dots(11)$$

From equation (11) we will calculate the value of profit by using the average gain equation (12) as below

$$Average\ Gain\ = H(T) - H(G) \dots\dots\dots\dots\dots\dots(12)$$

All the above algorithms are implemented using fast mining version 7.3.

3. MODEL DEVELOPMENT PROCEDURE

3.1. Data mining scheme

The study standardized cross-industry for data mining procedures that consisted of business processing, data production, data storage, evaluation and deployment.

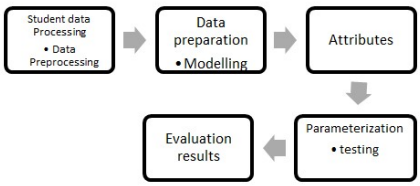

Figure 1. Data Mining Scheme To Get Data Sets Of Students Learning Results

The focus of the research is the students where their previous semester data is collected for this study. The value of the learning result is then analyzed in Rapidminer tool on this occasion. The researchers will therefore use the dataset to find out the criteria of learning styles, behavior characteristics, and learning achievement. To get high quality data, then it performed perform preprocessing techniques. The techniques used in preprocessing are as follows: Data cleaning, data integration and data reduction. The pre-processed data is considered finished at this stage. Then the preprocessed data is created for modeling and attributes for parameter development. After the preprocessing phase is completed, the selected attribute is obtained for data training and use as testing data.

3.2- Student Learning Styles

The basic function of learning style of students is to carry out cluster analysis on student campus activity data in the system, divide students into several special categories and show important features of each category. After the duration of the data mining operation, the results of the classification will be shown in the form of a histogram. After analysis, the previous analysis interface will provide student success information, category information (category information is no longer -1) and average results in various ways.

In the way colleges and universities learn, we must scientifically and reasonably classify the behavior of students in schools, develop learning patterns and service methods for students suitable for different types of students, and provide individual and accurate support measures, improve learning patterns and can make teaching services work well, quality levels for talented people can promote training[17]. In the age of simulation results on the Internet, student success data is a card, class attendance data, book loan data, internet data, and learning attendance discipline data. All this data reflects their study and the laws of life in schools, which can dynamically and accurately map the behavioral characteristics of students and behavioral habits hidden in this data.

3.2. Parameter Of Student Learning Behavior

Based on the behavioral data of a four-year undergraduate at a university, the thesis determines the characteristics of student behavior and research variables of student success, as shown in Table 1. from these criteria result, success behavior data comes mainly from the criteria items, the remaining behavioral data comes mainly from the learning style and learning behavior data comes primarily from educational information datasets. Due to additional variation and inconsistent data structure, a sample of students with seriously missing information and some out-liars was eliminated through data processing. The collected data of

students are then simplified in twelve parameter criteria (Table 1).

Table 1. Research Variables On Student Learning Behavior And Student Success

Parameters	Parameter code	Explanation
Politeness level	c_1	The ability to behave friendly to others especially to seniors.
Listening ability	c_2	The ability to sit quietly listening to concepts and meanings as the lesson progresses.
Level of creativity	c_3	Ability to bring up ideas and ideas to learning colleagues
Craft	c_3	Ability to manage study time and routine activities
Study tasks	c_3	Do college assignments correctly
presence	c_4	Be present on time and discipline
Exam	c_5	Doing the exam right
class leadership	c_6	Ability to manage learning teams
Communication skills	c_7	Speak softly and gently without interfering with other feelings.
Verbal memory	c_8	Ability to remember verbatim from the teacher's explanation.
Remember	c_9	The ability to carry out the synthesis of knowledge from previous experience.
Social interaction	c_{10}	The ability to adapt to any class situation.
Class participation	c_{11}	Collaborative and participatory in classroom activities
Team discussions	c_{12}	Connecting thoughts and ideas with other students.

4. EXPERIMENTAL RESULTS

4.1. Cluster Analysis Of Student Success

K-means algorithms are used to group student success data. According to the quality of the clustering algorithm's diagnosis, the best grouping effect is achieved when the number of criteria is set at 12. According to the actual student success situation, the average score of each cluster and the average score of the student's overall index are compared, and H is higher than the average score of the overall index of the student, which is lower than the average of the student population. The results of cluster analysis of the performance behavior of learners in different criteria are shown in Table 2.

As can be seen from Table 2, the characteristics of students' success behavior can be analyzed based on the Naïve Bayes, C4.5, Random Forest, and K-means. However, on K-means have achievement the average number of student achievement. According to cluster average standard, most of the cluster number has value grade under 4, whereas only K-means has the average student achievement with the ratio of students and average index of each cluster is shown in table 2.

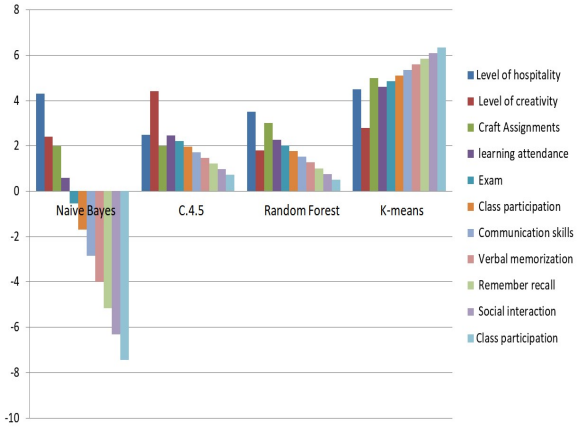

Figure 2. Average Student Achievement From Copying Five Algorithms

Negative signs indicate that the direction of effect is reversed i.e. criterion should be influenced by the attainment of learning, not the other way around. In Table 1, algorithms are compared based on the quality of learning style of students. Algorithms are used to analyze the relationship between student behavior and academic success [18]. The above cluster analysis has divided the conduct of students into three categories: success behavior, work-breaking behavior and learning behavior. To learn more about the relationship between the student's behavioral characteristics and his academic success, and to find out if there is a particular link between the characteristics of the student's behavior in school and his academic success, there is a certain link between analyzing interplay to algorithms, mining hidden correlations and rules of the consequences of simulation results[19].

The algorithms also will be used for mining association rules with one of the most influential algorithms for mining frequent item sets is the association rules[20]. The main idea is to mine item sets that often go through two stages: generation of candidate seats and detection of closed plots. The point is an algorithm based on the idea of a two-

stage frequency set. The rules of association are in terms of one-dimensional, single layers and the association rule in classification. Here, all support is more than the minimum support of a set of items known as frequent set items which so-called the frequency sets[25]. These algorithms have been widely used in businesses, network security and other sectors[21]. The main idea of using the algorithms is to find all sets of frequency firstly, this item set appears at least as often as the minimum support is determined[22]. Then strong association rules arise from established frequency, which must be met with minimal support and minimal confidence. Then use the frequency set found in step 1 to create the required rule, creating all the rules containing only items from the set, with only one item on the right side of each rule. After this rule is created, only rules remain above the minimum level of trust given by the user. A refracted method is used to produce all sets of frequency.

A large number of candidate seats can be created, and scanning databases may be required to repeat, algorithms are two main flaws. In setting the parameters of the algorithm model, the success attitudes of five types of students, the work and relaxation attitudes of three types of students, four types of student learning attitudes and three types of academic success of the students are established, set to variables before and after the rules of association[23]. Set the level of support confidence level to analyze the association rules, all of which obtained 12 parameters of association rules. The study's objectives were to choose the rules of commitment to a rate of increase of more than 1, where the latter had rules of association of students' academic achievement, as shown in Table 2.

Table 2.Results Predicting Academic Success With Proposed Algorithms

Parameter	Naive Bayes	C.4.5	Random Forest	meaning of
Level of hospitality	4.3	2.5	3.5	4.5
Level of creativity	2.4	4.4	1.8	2.8
Craft Assignments	2	2	3	5
Learning attendance	0.6	2.467	2.267	4.6
examination	-0.55	2.217	2.017	4.85
Class Participation	-1.7	1.967	1.767	5.1
Communication skills	-2.85	1.717	1.517	5.35
Verbal recall	-4	1.467	1.267	5.6
Remember	-5.15	1.217	1.017	5.85
social interaction	-6.3	0.967	0.767	6.1
Class Participation	-7.45	0.717	0.517	6.35

Based on the behavioral data of undergraduates at a university, The K-mean algorithm was used for grouping analysis and algorithmic association. Using a mock learning style to change the student's learning style from a combined learning style to a personal learning style[24]. The colleges and universities can master the behavioral habits of students to support their learning success, thus sampling the way learning is done by learning a personal student from a coherent student learning style changes in the way.

The results of the learning style have made changes in the learning style concept of the students. As a cradle of talent development, universities need to keep pace with social development. All types of application systems, such as one card systems, digital campus systems, student teaching information systems and book landing systems, provide abundant data resources for the learning style of school students. Through mining the results of these rapid simulations, we can learn from all school students to study, work and relax, so that we can change the way we learn from problem solving when they have appeared before we find problems at our own initiative, the possibility of correcting and correcting learning style mistakes.

4.2. Accuracy Level Of Each Algorithm

This section will look at the accuracy of each algorithm. We compare the accuracy of each algorithm (such as the Naive Bayes method, the C4.5 method, Random Forests and the accuracy of sources). The outcome of the decision tree may be used as an activity assessment as a decision making in the future. From Figure 3, each algorithm is compared to determine its basic difference. Random Forest Algorithm has all features as tree nodes of judgment, while C4.5 algorithms has not all features for all the nodes. This is the difference between Random Forest algorithm and C4.5. The spread of this study is by correcting prediction from selected algorithms, i.e. Naive Bayes algorithms. The deployment process takes place through several stages. After data testing was quickly done using Miner 7.3, there are results in the form of a

distribution table that will be used for algorithm prediction accuracy. Also the results of comparison of distribution table later return with previous training figures. We also found that when the comparison is complete, there is a new comparison value.

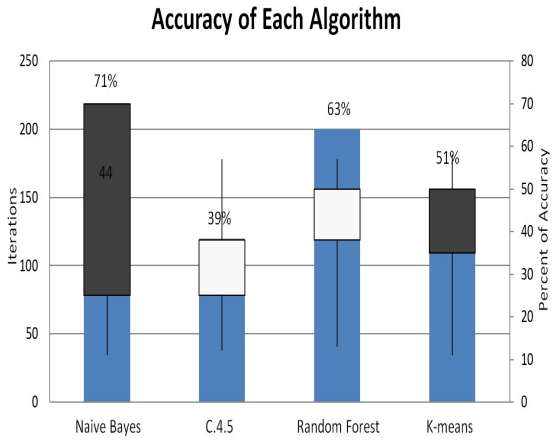

Figure 3. Accuracy level of each algorithm

For the accuracy of the spread of Naive Bayes algorithm, it found that grade prediction accuracy in early semester students reached 71 percent, after which Random Forest accuracy reached 63 percent. From Figure 3, it seems that the difference in the accuracy value of the previous grade compared to the value of accuracy is 51% using the K-means among the college students in the early semester. Finally, the accuracy limit value when we use C4.5 is 39%. Therefore, the accuracy of the Naïve Bayes algorithm is highest valued in the deployment process.

Sigitta et al. [28] found that the concentration of majors in Ta'alumul Huda Bumiayu Islamic High School students can be predicted using data mining techniques using the K-Means algorithm and the Naïve Bayes algorithm with an accuracy of 80% with a good classification predicate. The results of Tutupoly's (2020) show that overall the C4.5 classification algorithm has the greatest accuracy when compared to the Naive Bayes or Random Forest algorithms with an accuracy rate of 85.34% in the first experiment and 89.06% in the third experiment. While measurements using the ROC curve, the Naive Bayes algorithm is an algorithm that has the highest level of accuracy compared to the C4.5 and Random Forest algorithms with an AUC value of 0.925. Yusuf [30] shows the highest correlation value for the initial IP variable of r=0.783 and the leave variable has a very weak correlation level of r=0.054. The accuracy value of the naïve Bayes algorithm variable after cleaning is 78.0% and the Random Forest algorithm variable is 76.7%. Sitepu [31] states that besides the Naive Bayes Algorithm, there is also a Random Forest algorithm. The Random Forest algorithm is an algorithm that is accurate, even though the data is large, it still produces a good level of accuracy in its classification. In Adnyana [32], which predicts students' long predictions, says that the random forest algorithm has an accuracy rate of 83, 54%, which means the level of accuracy is good.

5- CONCLUSIONS

The study compares algorithms based on replicating learning results to predict students' success. We have successfully created a model to predict the learning achievement of students based on simulation using four algorithms. While research into learning systems is a difficult subject with ideological and practical importance, we were able to make it.

This research contributes to the process of implementing the student learning style system, which has special application value. Learning style systems can provide a database to match principals. They work through data processing. The modular design of the system is closely linked to the professional life and social practices of students at school. In addition, it can reflect basic data on students' daily performance in time, providing data support and a basis for decision-making for students in schools.

We also used cluster analysis method to classify the behavioral characteristics of four-year undergraduates at a university, as well as through analysis of the correlation between the behavioral characteristics of students at school and their academic success, we found that success behavior characteristics, work interval symbols and learning of different groups of students There is a close correlation between behavior and their academic achievement, it provides schools with a basis for different types of students to take learning style measures of different students, and on this basis, it provides some advice on how to use mock learning style to improve learning style in the background of building smart campuses. Colleges and universities should make full use of the achievements of building educational information and follow many data information obtained through the way they learn copy and the way students learn, to provide full data support for decision making and school development.

The study has achieved expected results, namely knowing the accuracy prediction algorithm model for student's success. These findings conclude that Naive Bayes algorithms have more

accuracy than Random Forest and C4.5 algorithms, so it can be seen that the difference in accuracy between Naive Bayes and Random Forest semantics which means Naive Bayes algorithm can thus better predict the student learning success.

Finally, from these findings, it can be concluded that Naive Bayes algorithm is more accurate than Random Forest and C4.5 algorithms. We found that the difference in accuracy between Naïve Bayes and Random Forest to better predict students' learning achievement. In addition, the Naïve Bayes algorithm can be used to predict the relationship of the students' learning success. Finally, it is recommended for the implementation of the research model in university that can provide benefit to the teachers to improve the student success.

This study has limitations so that future research needs to add experimental data. Research experiments can use even more data and try it with other student graduation datasets so that the model that has been obtained will be even more tested.

REFERENCES :

[1] Ye, Qing, Jin Zhou, and Hong Wu. "Using information technology to manage the COVID-19 pandemic: development of a technical framework based on practical experience in China." *JMIR medical informatics* 8, no. 6 (2020): e19515.

[2] Butt, Rameen, et al. "Integration of Industrial Revolution 4.0 and IOTs in academia: a state-of-the-art review on the concept of Education 4.0 in Pakistan." *Interactive Technology and Smart Education* 17.4 (2020): 337-354.

[3] Yu, Haibin. "Application analysis of new internet multimedia technology in optimizing the ideological and political education system of college students." *Wireless Communications and Mobile Computing* 2021 (2021).

[4] Howard, Tyrone C. Why race and culture matter in schools: Closing the achievement gap in America's classrooms. Teachers College Press, 2019.

[5] Abu Amuna, Youssef M., Mazen J. Al Shobaki, and Samy S. Abu-Naser. "The role of knowledge-based computerized management information systems in the administrative decision-making process." (2017).

[6] Aung, Yi Yi, and Myat Myat Min. "An analysis of random forest algorithm based network intrusion detection system." *2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*. IEEE, 2017.

[7] Livieris, Ioannis E., et al. "Predicting secondary school students' performance utilizing a semi-supervised learning approach." *Journal of educational computing research* 57.2 (2019): 448-470.

[8] Zhao, Yawei, et al. "Large-scale k-means clustering via variance reduction." *Neurocomputing* 307 (2018): 184-194.

[9] Student, Sebastian, and Krzysztof Fujarewicz. "Stable feature selection and classification algorithms for multiclass microarray data." *Biology direct* 7.1 (2012): 1-20.

[10] Yuan, Guan, et al. "A review of moving object trajectory clustering algorithms." *Artificial Intelligence Review* 47.1 (2017): 123-144.

[11] Yuan, Chunhui, and Haitao Yang. "Research on K-value selection method of K-means clustering algorithm." *J* 2.2 (2019): 226-235.

[12] Fränti, Pasi, and Sami Sieranoja. "K-means properties on six clustering benchmark datasets." Applied intelligence 48, no. 12 (2018): 4743-4759.

[13] He, Qingfeng, et al. "Landslide spatial modelling using novel bivariate statistical based Naïve Bayes, RBF Classifier, and RBF Network machine learning algorithms." *Science of the total environment* 663 (2019): 1-15.

[14] Obulesu, O., M. Mahendra, and M. ThrilokReddy. "Machine learning techniques and tools: A survey." *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 2018.

[15] Probst, Philipp, Marvin N. Wright, and Anne-Laure Boulesteix. "Hyperparameters and tuning strategies for random forest." Wiley Interdisciplinary Reviews: data mining and knowledge discovery 9, no. 3 (2019): e1301.

[16] Hassani, Hossein, Xu Huang, Emmanuel S. Silva, and Mansi Ghodsi. "A review of data mining applications in crime." Statistical Analysis and Data Mining: The ASA Data Science Journal 9, no. 3 (2016): 139-154.

[17] Bao, Wei. "COVID-19 and online teaching in higher education: A case study of Peking University." Human behavior and emerging technologies 2, no. 2 (2020): 113-115.

[18] Xu, Xing, Jianzhong Wang, Hao Peng, and Ruilin Wu. "Prediction of academic performance associated with internet usage

behaviors using machine learning algorithms." Computers in Human Behavior 98 (2019): 166-173.

[19] Bhutto, Engr Sana, Isma Farah Siddiqui, Qasim Ali Arain, and Maleeha Anwar. "Predicting students' academic performance through supervised machine learning." In 2020 International Conference on Information Science and Communication Technology (ICISCT), pp. 1-6. IEEE, 2020.

[20] Yuan, Xiuli. "An improved Apriori algorithm for mining association rules." *AIP conference proceedings*. Vol. 1820. No. 1. AIP Publishing LLC, 2017.

[21] Gupta, Brij B., ed. Computer and cyber security: principles, algorithm, applications, and perspectives. CRC Press, 2018.

[22] Fournier-Viger, Philippe, Zhitian Li, Jerry Chun-Wei Lin, Rage Uday Kiran, and Hamido Fujita. "Efficient algorithms to identify periodic patterns in multiple sequences." Information Sciences 489 (2019): 205-226.

[23] Casey, Beth M., and Colleen M. Ganley. "An examination of gender differences in spatial skills and math attitudes in relation to mathematics success: A bio-psycho-social model." *Developmental Review* 60 (2021): 100963.

[24] Zou, Fanna, and Rui Li. "Construction of Student Innovation and Entrepreneurship Experience System Integrating K-Means Clustering Algorithm." *Mathematical Problems in Engineering* 2022 (2022).

[25] Zhang, Lili, Wenjie Wang, and Yuqing Zhang. "Privacy preserving association rule mining: Taxonomy, techniques, and metrics." *IEEE Access* 7 (2019): 45032-45047.

[26] Patel, Harsh H., and Purvi Prajapati. "Study and analysis of decision tree based classification algorithms." *International Journal of Computer Sciences and Engineering* 6, no. 10 (2018): 74-78.

[27] Tan, Kun, Weibo Ma, Fuyu Wu, and Qian Du. "Random forest–based estimation of heavy metal concentration in agricultural soils with hyperspectral sensor data." *Environmental monitoring and assessment* 191, no. 7 (2019): 1-14.

[28] Sigitta, Rito Cipta, M. Ghozi Ghozali, and Nurul Mega Saraswati. " Application of the K-Means Algorithm and the Naïve Bayes Algorithm in the Selection of Student Major Concentrations at Ta'alumul Huda Bumiayu Islamic High School." *Indonesian Journal of Informatics and Research* 2, no. 2 (2021): 19-25.

[29] Tutupoly, Taransa Agasya. " C4 5 Algorithm Comparison., Naive Bayes, and Random Forest for Jakarta Student Graduation Data Classification." (2020): 5-12.

[30] Yusuf, Bustami, Muthmainna Qalbi, Basrul Basrul, Ima Dwitawati, Malahayati Malahayati, and Mega Ellyadi. " Implementation of Naive Bayes and Random Forest Algorithms in Predicting Academic Achievement of Ar-Raniry State Islamic University Banda Aceh Students." *Cyberspace: Jurnal Pendidikan Teknologi Informasi* 4, no. 1 (2020): 50-58.

[31] Br Sitepu, Natalina. " Analysis of the Decision Tree Algorithm with the Random Forest Algorithm on Discretize By Frequency." (2019).

[32] Adnyana, I. Made Budi. " Prediction of Student Study Period Using the Random Forest Method (Case Study: STIKOM Bali)." *CSRID (Computer Science Research and Its Development Journal)* 8, no. 3 (2016): 201-208.