Extended Abstract

SAPIENS: Self Learning Application for Elementary Students Based on Artificial Intelligence Aprroach to Improve Indonesia's Education

Ahmad Miftakhul Ilmi, Andrea Ernest, Aulia Kharis Rakhmasari, Dede Yusuf P. Kuntaritas, Zulfani Alfasanah

Advisor : Dr. Dra. Kartika Fithriasari, M.Si

INSTITUT TEKNOLOGI SEPULUH NOPEMBER

takhulilmi99@gmail.com

Abstract

At the level of reality (Das Sein), the quality of the Indonesian education system is low at 46.4%, placing the quality of Indonesian education in the 70th position out of 93 countries (CEO Magazine, 2019). Ideally (Das Sollen), various parties must consider education so that the quality of education in Indonesia is guaranteed under any circumstances. Therefore, the authors are interested in offering a solution in the form of a student learning system independently according to the interests and abilities of students. This research aims to prevent and overcome academic loss learning by utilizing artificial intelligence to improve the quality of education so that students can become candidates for superior and quality human resources and contribute to the current era of industrial revolution 4.0. The analysis knife of this research: learning loss, self-directed learning, and artificial intelligence-based education system. The approach used is a qualitative approach with library research methods with primary and secondary data sources.For this reason, data source documentation techniques with content analysis methods are used. The creation of this artificial intelligence system applies a combination of concepts between combinatorics and stochastic processes in modeling specific fields of knowledge. The first stage of this algorithm is an initial assessment to see students' potential and the classification of students' abilities. The second stage is user clustering based on the results of the initial assessment. Then, users can view learning outcomes, see progress, and repeat topics if they do not meet the target. This system is appropriate to prevent and overcome academic loss learning by utilizing artificial intelligence to improve the quality of education so that students can become candidates for superior and quality human resources and contribute to development in the current era of globalization.

Keyword : *self learning, learning loss, artificial intelligence*

INTRODUCTION

Quality, competent, and ethical Human Resources (HR) are valuable assets of the Indonesian nation. Having competitive human resources is also the hope of institutions in the

field of education. Industrial era 4.0 is characterized by the need for digitally skilled human resources to support the application of industry 4.0, which is thick with artificial intelligence (AI). Experience in many countries shows that quality human resources are more critical than abundant natural resources. However, in the last few decades, the competitiveness of the Indonesian nation among other nations has tended to be less encouraging. This is evidenced by Indonesia's GTCI (Global Talent Competitive Index), which is ranked 65 out of 132 countries. According to GTCI, the development of human resources is influenced by several factors (GTCI, 2020). Indonesia has several factors that are far behind foreign countries. Those factors are academic, soft skills, and hard skills. Quality human resources can only be realized with quality education. Therefore, efforts to improve the quality of education are non-negotiable to improve the quality of Indonesian human resources. Education is the essential capital for creating human resources and fields that are very positively affected by AI.

Data from BPS shows that the number of primary and secondary education institutions has continued to increase since the independence era. Likewise, the number of Indonesian children who have access to primary and secondary education, where the participation rate in basic education continues to increase and various programs to eradicate illiteracy continue to be promoted. The number of students in Indonesia also continues to grow, reaching 7.3 million students in 2018 (PDDikti, 2018). These data should be a breath of fresh air for the world of education in Indonesia. However, it is undeniable that the quality of the education system in Indonesia is still inferior. The survey conducted by the CEO of WORLD magazine 2019 stated that the quality of the Indonesian education system was low at 46.4%, placing the quality of Indonesian education in the 70th position out of 93 countries (CEOworld Magazine, 2018). In addition, research conducted by PISA (Program for International Student Assessment) in 2018 noted that Indonesia's reading competence was ranked 72 out of 77 countries. While the value of Mathematics is ranked 72 out of 78 countries and the value of Science is ranked 70 out of 78 countries. These show the low academic ability of Indonesian students compared to other countries.

There are several reasons for the low level of education in Indonesia. One of the problems that arise from within the students themselves, namely laziness in studying. This problem often occurs in students who are in elementary school (SD) because, at the age of 6-13 years, they are still happy with playing activities rather than learning. This feeling of laziness can arise due to several factors, including the lack of student independence in learning and the low learning discipline of students. Students who do not have independent learning are different from students who are independent in

learning. This difference can be seen from the motivation and interest of students in learning. Students who do not have the motivation and interest in learning will not study independently and experience various academic difficulties. Students who have high learning independence will be motivated to learn something with their abilities without asking for help from others. Independent learning is learning with motivation and direction that is driven by a motive to master a competency with the knowledge of competencies that students already have (Mujiman, 2011). The level of learning discipline of each student will be different. Students who are accustomed to learning discipline will make the best use of their time at home and at school so that they will show their readiness in the learning process at school, while students who are not disciplined in their studies show less readiness in learning. They will show deviant behavior in the learning process, such as not doing homework, truancy, not paying attention to teacher explanations, violating school rules. Learning discipline is very important because discipline aims to guard against deviant behavior and things that can interfere with the learning process. Discipline makes students trained and has the habit of doing good actions and can control their every action to be obedient, obedient, and orderly towards teaching and learning activities. In learning, discipline is needed so learning will run effectively and optimally. Therefore, for learning to run smoothly, all students must be disciplined in obeying school rules, discipline in doing homework, discipline in doing assignments, and discipline in studying at home.

Independent and disciplined behavior will not grow by itself but requires self-awareness, practice, habits, and encouragement from other parties, especially elementary school students (Rafika, dkk, 2017). At an age still classified as a child, elementary school students need assistance in their learning activities. In addition, the presence of other parties can also encourage the growth of student learning motivation. When the learning motivation has been formed, independence and learning discipline can grow in students. Thus, it is necessary to have a device or system that can foster independence and student learning discipline, equipped with assistance from other parties. Therefore, the authors are interested in offering a solution in the form of a student learning system independently according to the abilities of each student and equipped with volunteers who will assist students in learning activities. This is in accordance with the educational principles promoted by Ki Hajar Dewantara, Ing Ngarso Sung Tulodho (students as the center of education). In addition, this system will also help anticipate the occurrence of learning loss, especially for elementary school students.

OBJECTIVES

This research aims to prevent and overcome academic

loss learning by utilizing artificial intelligence to improve the quality of education so that students can become candidates for superior and quality human resources and contribute to the current era of industrial revolution 4.

RESEARCH METHODS AND APPROACHES

In this research, researchers used descriptive quantitative research methods to describe or describe the data that had been collected. Quantitative descriptive research focuses on solving problems that exist in the present or on actual and meaningful events (Yusuf, 2016).

There are several steps that must be taken by researchers to create a SAPIENS application:

1. Analyzing

   At this stage, the researcher conducts an analysis related to data requirements and system requirements.
2. Designing

   At this stage, the researcher developed the SAPIENS application concept using the K-Means method for mapping. This system design focuses on designing a data structure in the system, software architecture, interface representation, and procedural details (algorithm).
3. Implementation

   Researchers complete the system design that has been designed. Then, the SAPIENS application will be developed using an android device.
4. Testing

   After the application has been built, the next stage is black box testing.
5. Maintenance

   Maintenance of the system is also carried out after the application can be run. The goal is that the SAPIENS application is always up to date on additional data.

Researchers used the K-Means algorithm to determine the best cluster to support various features in the SAPIENS application. For example, this best cluster is used for mapping elementary school students based on their learning abilities so that the right learning method can be given in each cluster.

The K-Means algorithm is a clustering algorithm that groups data based on the cluster center point (centroid) closest to the data. The purpose of K-Means is to group data by maximizing the similarity of data in one cluster and minimizing the similarity of data between clusters. The measure of similarity used in the cluster is a function of distance so that the maximization of data similarity is obtained based on the shortest distance between the data and the centroid point (Darmi,2016).

The K-means clustering algorithm process is as follows:

1. Determine k as the number of clusters to be formed, using the elbow method with the following formula:

$$SSE = \sum_{k=1}^{K} \sum_{x_j \in S_k} \| X_j - C_k \|^2$$

   where K is the number of groups used in the K-Means algorithm, Xi is the number of data, and Ck is the number of clusters in the kth cluster.

2. Determines the centroid value. In determining the value of the centroid for the beginning of the iteration, the initial value of the centroid is done randomly. Meanwhile, if determining the centroid value, which is the iteration stage, the following formula is used.

$$v = \frac{\sum_{i=1}^{n} x_i}{n}, \qquad i = 1, 2, \ldots\ldots, n.$$

3. Calculates the distance between the centroid point and the point of each object. To calculate the distance can use Euclidean Distance.

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

4. Allocates each object to the nearest centroid. For allocating objects into each cluster at the time of iteration, it is generally done using hard k-means. Each object is expressly declared as a member of the cluster by measuring the proximity of its nature to the center point of the cluster.
5. Perform iterations and then determine the position of the new centroid using the equation.
6. Repeat the third step if the new centroid position is not the same.

After getting the best number of clusters with the K-means algorithm, the researchers tested the following models to group the data into clusters obtained.

A) Logistic Regression Model

Logistic regression is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. Logistic regression competes with discriminant analysis as a method for analyzing categorical response variables. However, many statisticians feel that logistic regression is more flexible and more suitable for most modeling situations than discriminant analysis. This is because logistic regression does not assume that the independent variables are normally distributed, as is the case with discriminant analysis (Hyeoun-Ae, 2013).

B) K-Nearest Neighbors

The k-NN algorithm is a method that uses a supervised algorithm. The purpose of the k-NN algorithm is to classify new objects based on attributes and training samples. The new test sample results are classified based on the majority of the categories in the k-NN. This algorithm does not use any model to match in the classification process and is only based on memory. The k-NN algorithm uses adjacency classification as the predictive value of the new test sample. The distance used is the Euclidean Distance. Euclidean distance is the most commonly used distance in numerical data (Krisandi, 2013).

C) Support Vector Classification

In classification modeling, SVC has a more mature and clearer mathematical concept than other classification

techniques. SVC can also solve classification and regression problems with linear and non-linear (Dewi, 2016).

D) Gaussian Naive Bayes

The Naive Bayes algorithm predicts future opportunities based on past experience, so it is known as the Bayes theorem. The main characteristic of this Nave Bayes Classifier is a very strong (nave) assumption of the independence of each condition/event (Kapourani, 2018).

E) Random Forest

Random forest is a combination of each good tree which is then combined into one model. Random Forest depends on a random vector value with the same distribution in all trees where each decision tree has a maximum depth. A random forest is a classifier consisting of a classifier in the form of a tree {h(x, k), k = 1, . . .} where k is a random vector that is distributed independently, and each tree in a unit will choose the most popular class on input x (Haristu, 2019).

DATA AND DISCUSSION

The K-means clustering algorithm is a data analysis method or a Data Mining method where an unsupervised modeling

process is carried out that can group objects with the same characteristics and group data with a partition system. This method seeks to group data, where one group has almost the same characteristics, while one group with another group has different characteristics.

Based on the K-Means clustering, it was found that 3 clusters were consisting of clusters with hard, medium, and low learning periods. The division of each cluster can be seen through the following visualization

*Figure **Error! Use the Home tab to apply 0 to the text that you want to appear here.**.1 The Result of KMeans Clustering Scatter Plot between Easy and Hard Assessment*

Figure 2 The Result of KMeans Clustering Scatter Plot between Medium and Hard Assessment

Based on the visualization, it is illustrated that there are three different colors: green, yellow, and purple. Thus, it can be seen that if someone who is able to carry out a difficult and medium assessment is getting higher, it will tend to be collected in green. While a medium combination is yellow, and a low combination is purple.

Table 1 The Result of KMeans Clustering in each Criteria and Variables.

Information	Cluster 1	Cluster 2	Cluster 3
Traveltime	1.548387	1.38674	1.372881
Studytime	2	2.154696	1.762712
Failures	0.548387	0.077348	0.559322
Absences	3.096774	3.337017	19.847458
First Assessment	8.116129	13.502762	10.288136
Second Assessment	7.6	13.646409	9.898305
Third Assessment	6.664516	13.79558	9.898305
Conclusion of the tightness of Study Time	Fast	Slow	Medium

From the cluster above, it can be seen that there are three different clusters. The students are selected for each cluster based on the early assessment test. If the student gets high scores, they are more likely to be placed in the faster cluster. Therefore, the treatment between each cluster is different. Students in the slow cluster will learn the materials more extended than the above cluster.

The K-Means clustering results will be the initial model that will be used as the next benchmark. Illustrated through the following flow chart:

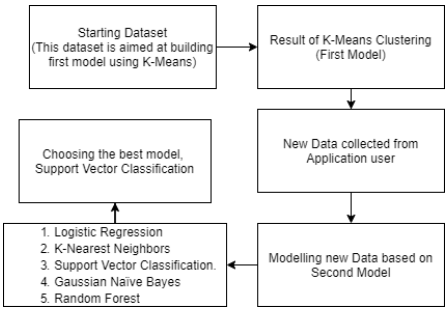

Figure 3 Flow Chart of the Machine Learning Modelling Process

The use of two Machine Learning models in this application aims to facilitate the student classification process with a high and precise level of accuracy. If only using the K-Means model, the accuracy of the clustering of each participant cannot be known with certainty, and if there is new data from participants, the model must be repeated from the beginning. However, with the second model, one of the following five models, there will be a logarithmic efficiency. Modeling does not need to be repeated from the beginning if there is new data. Researchers will compare the data by comparing five models:

1. Logistics Regression
2. K-Nearest Neighbors
3. Support Vector Classification
4. Gaussian Naive Bayes
5. Random Forest

The variables that will be used in this dataset are:

- School origin, Gender, Age, Address, Family size
- Parental status, mother's education, father's education, mother's occupation, father's occupation
- reasons for choosing this school (nominal: close to 'home', school 'reputation', preference for 'courses' or 'other')
- guardian - student guardian (nominal: 'mother', 'father' or 'other')
- travel time - commute time to school (number: 1 - 1 hour)
- study time - weekly study time (numeric: 1 - 10 hours)
- failures - number of failures of the previous class (numeric: n if 1 <= n <3, else 4)
- additional educational support (binary: yes or no)

- family education support (binary: yes or no)
- extra paid class (binary: yes or no)
- extra-curricular activities (binary: yes or no)
- want to take higher education (binary: yes or no)
- internet - Internet access at home (binary: yes or no)
- quality of family relationships (numeric: from 1 - very bad to 5 - very good)
- free time,health ,absence - number of school absences (numerical: from 0 to 93)
- First Assessment Score
- Second Assessment Score
- Third Assessment Value
- Student Cluster (Consists of 3 Clusters)

After experiencing clustering with the K-Means method, the next step is to decide which classification model will be taken as a fixed model if there will be many new data.

Table 2 The Result of Precision and Accuracy in Each model

	Cluster	precision	recall	f1-score	Accuracy
1. Logistic Regression	0	0.98	0.97	0.98	0.96
	1	1	1	1	
	2	0.87	0.93	0.9	
2. K-Nearest Neighbors	0	0.97	1	0.98	0.97
	1	1	1	1	
	2	1	0.86	0.92	
3. Support Vector Classification.	0	0.98	1	0.99	0.99
	1	1	1	1	
	2	1	0.93	0.96	
4. Gaussian Naïve Bayes	0	0.98	0.89	0.93	0.84
	1	0	0	0	
	2	0.56	0.71	0.63	
5. Random Forest	0	0.98	0.89	0.93	0.95
	1	0	0	0	
	2	0.56	0.71	0.63	

Based on the comparison of the five methods, the best accuracy value is obtained in the Support Vector Classification or SVC model. Through this SVC method, new data will be predicted whether it will enter clusters 0,1, or cluster 2.

The following is an example of a prediction that the algorithm will make. Through 20% of the initial data, predictions for each cluster will be made, and the following results will be produced.

```
[48] y_pred

     array([1, 1, 1, 0, 0, 1, 2, 0, 1,
            0, 2, 1, 0, 1, 0, 2, 0, 2,
            2, 2, 2, 1, 2, 2, 2, 0, 0,
            1, 1, 0, 1, 0, 0, 0, 1, 1,
```

Figure 4 The Result of New Data that are collected in Each Cluster

It can be seen that some students are divided into certain clusters according to their respective criteria. This will be the basis for the further classification of new data from SAPIENS users. The K-Means and SVC algorithms that have been described are two important algorithms that will determine the cluster of each student. If the cluster has been formed correctly with high accuracy, then students can enjoy the following SAPIENS features according to their respective criteria. The following are some of the features contained in SAPIENS:

A. Sign Up and Log In

On this page, users who are elementary students will register with parental assistance so that they can join and enjoy the

SAPIENS feature. Then, students need to wait for account verification from their teacher or homeroom teacher before logging in to the application homepage.

Figure 5 User Interface of SAPIENS

B. Roadmap

Once the student logs in, the application will present a roadmap as a guide for the student's activities in the application. The stages of this roadmap consist of:

- Assessment

  It is an initial test that students must complete to determine students' abilities in learning and understanding the material. The test is presented through multiple choice questions that will assess students based on reading speed, counting, and understanding a problem. The results of this assessment will be a reference for students to find learning methods, learning materials, and level of practice questions that are in accordance with students' abilities.
- Path learning

  This is a page that will display the statistics of the assessment results that contain the ability and level of understanding of students in conducting academic learning. The statistics presented will change along with the increase or decrease in students' ability to carry out a series of lessons from the SAPIENS application.
- Schedule

  This feature is used by the teacher or homeroom teacher to enter a list of student lessons for one week, including any material given at school. In addition, this feature will also encourage teachers to record the exam schedule that will last for one semester so that the application can use it to determine the time, independent learning materials, and the intensity of student learning. This feature is also equipped with a warning message every day so that students do not miss their independent study schedule.

Figure 6 User Interface of SAPIENS Roadmap

C. Let's Study

This is a student learning page that contains material and practice questions from the application and is the main feature of the SAPIENS application as the implementation stage of the available path learning features. This feature is integrated with the results of path learning, which will help students determine learning

materials that suit their abilities and must be completed within a certain period of time. Student learning outcomes will be evaluated periodically and updated on path learning statistics so that they can be used as a reference for further learning.

D.INSPIRATION

Contains inspirational articles that students can enjoy, volunteers, teachers, and parents. This page can be used to share volunteer experiences in each region, inspiration for champions and creativity from students, motivation from parents and teachers, and so on.

CONCLUSION

SAPIENS is a learning application that utilizes artificial intelligence to cluster the learning abilities of elementary school students starting from slow, medium, and fast. This application will provide material recommendations and practice questions that can be used as a source of student learning. The evaluation of each learning outcome will be used to update the student's ability level.

Researchers used the K-Means method to perform clustering followed by the Support Vector Classification method to create a fixed model for analyzing new data.

With the innovation of learning applications from SAPIENS, it is hoped that it can prevent and overcome academic loss learning in order to improve the quality of education so that students are able to become candidates for superior and quality human resources and can contribute to the current era of the industrial revolution 4.0.

RECOMMENDATION

The suggestions that the researchers convey as well as evaluations related to technology development, especially in this self-learning application, are as follows:

1. SAPIENS:
   Self-Learning Application for Elementary Students is expected to be realized directly for all elementary school students in Indonesia. This can help Indonesia in realizing a golden generation in 2045.
2. We need to cooperate with the government through the "Kampus Merdeka" program related to volunteering, which is part of the successful performance from SAPIENS application to increase student interest and contribution in advancing education in Indonesia.
3. We require help to develop methods and conduct further research, especially on the workings of artificial intelligence required by SAPIENS to produce more accurate and reliable clustering and evaluation of student learnining

REFERENCES

CEOworld Magazine. (2018). *Ranked*. Retrieved February 13, 2021, from CEOworld Magazine: https://ceoworld.biz/2020/05/10/ranked-worlds-best-countries-for-education-system-2020/.

Darmi, A. S. 2016. *Penerapan Metode Clustering K-means Dalam Pengelompokan Penjualan Produk*. Bengkulu : Jurnal Media Infotama, 148-157.

Dewi, Trysha. 2016. *Aspect Level Sentiment Classification Based On Supervised Learning*. Bandung : Universitas Telkom.

GTCI. (2020). *Countries*. Retrieved February 14, 2021, from Global Talent Competitive Index: https://gtcistudy.com/the-gtci-index/.

Haritsu, Reinardus. 2019. *Penerapan Metode Random Forest untuk Prediksi Win Ratio Pemain Player Unkown Battleground*. Yogyakarta : Universitas Sanata Dharma.

Heyoun-Ae, Park. 2013. *An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain.* Seoul : Seoul National University.

Kapourani, Andreas. 2018. *Naive Bayes & Introduction to Gaussians.*Skotlandia : University of Edinburgh.

Krisandi, Nobertus. 2013. *Algoritma K-nearest Neighbor Dalam Klasifikasi Data Hasil.*

Produksi Kelapa Sawit Pada PT. Minamas Kecamatan Parindu. Pontianak : UNTAN.

Mujiman, Haris. 2011. *Manajemen Pelatihan Berbasis Belajar Mandiri*. Yogyakarta: Pustaka Pelajar.

PDDikti. 2018. *Statistik Pendidikan Tinggi 2018.* Jakarta: Pusdatin Iptek Dikti.

Rafika, dkk. 2017. Jurnal Ilmiah Pendidikan Guru Sekolah Dasar: *Upaya Guru Dalam Menumbuhkan Kemandirian Belajar Siswa di SD Negeri 22 Banda Aceh*. Aceh: FKIP Unsyiah. Volume 2 Nomor 1.

Yusuf, Muri. 2016. Metode Penelitian Kuantitatif, Kualitatif & Penelitian Gabungan. Jakarta : Prenada Media.