Identification of the Success of Learning Al Islam and Kemuhammadiyahan Using Machine Learning

Edi Ismanto
Dept. of Informatics Education
Universitas Muhammadiyah Riau
Pekanbaru, Riau
edi.ismanto@umri.ac.id

Eka Pandu Cynthia
Dept. of Informatics Engineering
State Islamic University of Sultan Syarif Kasim Riau
Pekanbaru, Riau
*eka.pandu.cynthia@uin-suska.ac.id

Abstract— Learning Al Islam and Kemuhammadiyahan (AIK) is one of the compulsory learning in all levels of Muhammadiyah education which has the general goal of forming human learners who are pious, noble, advanced and superior in science and technology as the embodiment of *tajdid amar makruf nahi mungkar*. In order to achieve the objectives of learning in education, the measurement and evaluation of these subjects should be carried out. The Muhammadiyah University of Riau as one of the Muhammadiyah universities in Indonesia that applies AIK education in its learning curriculum has not measured the success rate of this course. Therefore this study will identify the success of AIK learning at Universitas Muhammadiyah Riau with machine learning artificial intelligence technology. The data source is the students’ grades in this AIK course, then by using the Machine Learning method K-Nearest Neighbor classification algorithm will measure the success rate of learning. From 149 sample data that were conducted training and testing (90% of training data and 10% of test data) as well as conducting several trials of K parameter values, the best accuracy results obtained at parameter k = 2 with an accuracy value of 71.43%.

Keywords—Al Islam and Kemuhammadiyahan, K-Nearest Neighbour, Learning Success, Machine Learning.

I. INTRODUCTION

Muhammadiyah Higher Education as a higher education institution in Indonesia has a characteristic in its learning curriculum, which has a compulsory Al Islam and Kemuhammadiyahan (AIK) course. Al Islam and Kemuhammadiyahan education must be taken by students at all levels of Muhammadiyah education [4]. The general aim of AIK education is the formation of learners who are pious, noble, progressive and superior in science and technology as an embodiment of the da'wah missionary *amar makruf nahi munkar*. The AIK course is divided into four sections which are intended as a description of the general objectives above so that it is more measurable. After completing all these AIK courses, students are expected to have competencies:

(AIK 1): Knowing and understanding the nature of God, humans, and life in accordance with the guidance of the Qur’an and authentic Hadith and science, practice the correct worship procedures based on the Qur’an and Sunnah Maqbullah. (AIK 2): Having *akhlaqul karimah* attitude in *mu’amalah* that benefits themselves, society, nation, and country. (AIK 3): Able to internalize Muhammadiyah’s mission in various aspects of life. (AIK 4): Mastering and integrating Islamic values in the development and application of science.

Universitas Muhammadiyah Riau has implemented the Al Islam and Kemuhammadiyahan learning curriculum, but until now there has been no measurement of the success rate of learning this course. Measuring the success of a learning activity is quite important because it is one form of evaluation in learning. Through this measurement, it will be seen whether the objectives of learning have been achieved or not in accordance with the curriculum used, the teaching techniques carried out by lecturers supporting the course, and the anomalies that occur in the learning process.

Machine Learning is a type of artificial intelligence that provides computers with the ability to learn from data, without explicitly having to follow programmed instructions. Machine learning technology is not only applied in the industrial world but also in the world of education such as education data security, fraud detection in exams, recommendation systems and so on [2]. In this study, we will try to measure the success of AIK course learning using the K-Nearest Neighbor algorithm, which is a method of extracting data in machine learning for classification that has a high degree of accuracy [1]. Value data and supporting lecturers in AIK (1) to AIK (4) courses at the Muhammadiyah University of Riau will be conducted training and testing using the k-NN algorithm so that it will produce a classification measurement of learning outcomes for the course.


II. THEORETICAL FRAMEWORK

Machine learning has two models in the learning process, namely Supervised Learning and Unsupervised Learning. In the supervised learning model, we train the machine to give the output that we have set or expected. The algorithm applied to the machine is intended to make an input into the output that we expect. Whereas in the unsupervised learning model, we do not teach machines to produce a certain output.
We only teach them what the correct inputs are and then the machine itself will determine the output.

The work process of machine learning in this study is described as follows:

1. Data Set
   In machine learning, there are two data set models, namely public data sets and private data sets. In public data sets, the data used are data that are public and are also obtained from public repositories. On the other hand, private data sets are data specifically obtained at the place of the study. This stage is called Pre-Processing, which will process the dataset to be more specific according to research needs. There will be several sub-processes in this Pre-Processing such as Data Cleaning, Data Integration, Data Reduction, and Data Transformation.

2. Method and Algorithm
   There are five methods in extracting machine learning data, and each of which has algorithms that can help solve the problem. The following methods and algorithms are described:
   a) Estimation Method, the algorithm that can be used: Linear Regression, Neural Network, Support Vector Machine, etc.
   b) Prediction Method, the algorithm that can be used: Linear Regression, Neural Network, Support Vector Machine, etc.
   c) Classification Method, the algorithm that can be used: Naïve Bayes, K-Nearest Neighbor, C4.5, ID3, CART, Linear Discriminant Analysis, Logistic Regression, etc.
   d) Clustering Method, the algorithm that can be used: K-Means, K-Medoids, Self Organizing Map, Fuzzy C-Means, etc.
   e) Association methods, algorithms that can be used: FP Growth, Apriori, Coefficient of Correlation, Chi-Square, etc.

3. Knowledge
   At this stage, knowledge (patterns or models) will be obtained from the methods and algorithms that have been implemented to solve the problem. Five possible knowledge that will be obtained, such as Formula/function in the form of a regression formula or function, Decision Tree, Correlation Level, Rule and finally Cluster.

4. Evaluation
   At this stage, the use of machine learning algorithms in solving problems will be evaluated according to the method. The following evaluation techniques can be used:
   a) Estimation Method, the evaluation used is to calculate the error value using: Root Mean Square Error (RMSE), MSE, MAPE, etc.
   b) Prediction Method, the evaluation used is to calculate the error value using: Root Mean Square Error (RMSE), MSE, MAPE, etc.
   c) Classification Method, the evaluation conducted is Confusing Matrix to calculate the accuracy value and ROC Curve to calculate the Area Under Curve (AUC)

   d) Clustering Method, evaluations conducted are Internal Evaluation using Davies-Bouldin Index and Dunn Index, and External Evaluation using Rand Measure, F-Measure, Jaccard Index, Fowlkes-Mallows Index, and Confusing Matrix.

   e) Association method, the evaluation is done by making Lift Chart using Lift Ratio and doing the Precision and Recall (F-Measure) calculation.

III. RESEARCH METHOD

The following are the stages carried out in this study:

A. Planning Phase
   At this planning stage, literature study activities were carried out in the form of scientific journals, books and other documents that can help identify further research problems. The result of this stage is a research plan to identify the success of learning Al Islam and Kemuhammadiyahan at Universitas Muhammadiyah Riau.

B. Data Collection Phase
   At this stage, interviews and direct observations were carried out at Universitas Muhammadiyah Riau, then documented the results obtained from the values of students who had completed the learning of Al Islam and Kemuhammadiyahan (AIK 1 to AIK 4).

C. Machine Learning Process
   The process of machine learning in identifying the success of learning Al Islam and Kemuhammadiyahan in this study is described as follows:

5. Data Set
   This dataset process is intended for understanding and processing data. In this study, the dataset model
used is a private dataset of Universitas Muhammadiyah Riau.

6. Algorithm of K-Nearest Neighbour.

In this study, the K-Nearest Neighbor (k-NN) algorithm is used, which is a classification method for a set of data based on learning data that has been classified previously. This algorithm is also called a lazy learner because it stores the training provided in a record and does nothing until it is tested [5]. This algorithm is included in the supervised learning model, where the results of the new query instance will be classified based on the majority of the proximity of the k-NN category. The steps of the k-NN algorithm are as follows [6]:

a) Determine the K parameter (number of nearest neighbors).

b) Calculate the square of the Euclidean distance of the object to the given training data, using the formula:

\[ D(a, b) = \sqrt{\sum_{k=1}^{d} (a_k - b_k)^2} \]  

(1)

c) Sort the results of the second step above based on the smallest value.

d) Collecting category Y (nearest neighbor classification based on K value)

e) By using the category of the nearest majority neighbor, the object category can be predicted.


In this research, the expected knowledge is the result of identifying the success of students in learning Al Islam and Kemuhammadiyahan using the K-Nearest Neighbor algorithm.

8. Evaluation

The evaluation carried out in the research to identify the success of students in learning Al Islam and Kemuhammadiyahan using the K-Nearest Neighbor algorithm is by making Confusing Matrix to calculate the accuracy value and ROC Curve to calculate Area Under Curve (AUC).

D. Conclusion and Suggestion Phase

At this stage, conclusions are drawn from the final results of the research conducted, and suggestions are given for future research development.

IV. IMPLEMENTATION AND RESULTS

Data obtained from the academic system of the Universitas Muhammadiyah Riau for the needs of this study is in the form of a component assessment file for each Al Islam and Kemuhammadiyahan course (AIK 1 to AIK 4) from each lecturer teaching the course, the final transcript of students who have completed the course and a list of lecturers’ names teaching the course. As a sample, the data used is from the Faculty of Teacher Training and Education consisting of four study programs, with a total of 149 data.

The data were processed using the K-Nearest Neighbor algorithm learning machine as follows:

A. Data Cleaning Phase

The first process in the pre-processing stage is Data Cleaning. At this stage, we removed data duplication, examined inconsistent data and corrected data errors. Examples of data cleaning processes in this study can be seen in the image below:

![Fig. 2. Sample of Data before Cleaning](image)

In the picture above, we can see that there are some inconsistent data, that is, there is a vacancy of the name of the supporting lecturer (data with highlights). At this cleaning stage, the blank data will be corrected by filling in the name of the appropriate lecturer in the academic year, class and course that the lecturer is able to. The results can be seen in the following figure 3.

![Fig. 3. Sample of Data after Cleaning](image)
In Figure 4 above is an example of data from the results of a course lectured by one of the lecturers in a current school year in faculties, study programs, and classes at the Muhammadiyah University of Riau. In this integration process, we will combine some of the required data. In this case and in accordance with the example in Figure 4, the data that is integrated is the data from the lectures of a lecturer supporting the same course in the current school year, at the same faculty and study program. The results can be seen in Figure 5 below.

C. Data Transformation Phase

At this stage, we carried out the process of normalization and data collection to be the same. In this study, we have differences in the course code variable and the name of Al Islam and Kemuhammadiyahan courses (AIK 1 to AIK 4) due to differences in the curriculum of the student year. This would be normalized according to the latest curriculum regulations used without eliminating the objectives of each AIK course as it has been described in the Introduction. The Grade Value will also be normalized according to table 1.

Example of the transformation of the course code can be seen in the following figure.

D. Data Reduction Phase

At this stage, we broke down the data into smaller forms but still produced the same analytical results. In this study, we selected the necessary variables in the process of identifying the success of learning Al Islam and Kemuhammadiyahan at Universitas Muhammadiyah Riau. The variables needed in this study are Student Number (NIM), Student Name, AIK 1 Grade, AIK 2 Grade, AIK 3 Grade, AIK 4 Grade, and AIK Learning Success Category. The results of the reduction process can be seen in Figure 8 below.
E. K-Nearest Neighbor Algorithm Phase

Suppose a student X has completed all courses of Al Islam and Kemuhammadiyah and obtained grades as follows: AIK 1 = B-, AIK 2 = A-, AIK 3 = B + and AIK 4 = A. Based on these grades, we identify how is the final result of AIK learning using the K-Nearest Neighbor Algorithm. In this study, there are 5 categories of AIK learning identification results: Very Successful, Successful, Successful Enough, Less Successful and Fail. The following are the steps for calculating the K-Nearest Neighbor Algorithm in identifying the AIK learning outcomes of students exemplified above:

1. Parameter K = 5
2. Calculate Euclidean distance according to the formula (1)
3. Sort the data in the second step based on the smallest Euclidean value.
4. Collect category Y (classification of nearest-neighbor based on K value)

Based on Figure 11 above, it can be seen that the majority of K = 5 categorization results are "Very Good", which means that testing on a sample of new data for student X scores results in the identification of learning outcomes of Al Islam and its Kemuhammadiyah which is **Very Good**.

F. Knowledge and Evaluation

In this study using tools in the form of Rapid Miner software in training and testing data and implementing the K-NN algorithm. Rapid Miner is a machine learning data processing software with a limited license. The data sharing for the training and testing process is done directly using split data (90% for training and 10% for testing) from a total of 149 sample data held. This stage will produce a piece of knowledge, the results of the evaluation of algorithm performance and accuracy.
Fig. 14. Test Results on 10% Data and Confidence Value for Each Category of AIK Learning Success Rate

To obtain the best accuracy and performance values, a number of K parameter trials were conducted. The results of which can be seen in the following table:

<table>
<thead>
<tr>
<th>K Value</th>
<th>Accuracy Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>71.43 %</td>
</tr>
<tr>
<td>3</td>
<td>57.14 %</td>
</tr>
<tr>
<td>4</td>
<td>64.29 %</td>
</tr>
<tr>
<td>5</td>
<td>57.14 %</td>
</tr>
<tr>
<td>6</td>
<td>57.14 %</td>
</tr>
<tr>
<td>7</td>
<td>50 %</td>
</tr>
<tr>
<td>8</td>
<td>50 %</td>
</tr>
<tr>
<td>9</td>
<td>50 %</td>
</tr>
<tr>
<td>10</td>
<td>50 %</td>
</tr>
</tbody>
</table>

V. CONCLUSION

The results of the study demonstrate that machine learning can be input to the leaders and related institutions at the University of Muhammadiyah Riau in monitoring and evaluating AIK learning. Finally, based on Table II, the experimental parameter set K, obtained the best accuracy and performance values, namely the parameter set $K = 2$ with an accuracy value of 71.43%.

REFERENCES


