Jurnal Gantang VIII (2) (2023): 193-197

JURNAL GANTANG

e-ISSN: 2548-5547
p-ISSN: 2503-0671

http://ojs.umrah.ac.id/index.php/gantang/index

Evaluation of the implementation of machine learning algorithm K-Nearest Neighbors (KNN) using rapid miner on junior high school student learning outcomes

Lucky Heriyanti Jufri*, Dadan Dasari

Universitas Pendidikan Indonesia, Jl. Dr. Setiabudhi No. 229 Bandung, 40154, Indonesia
*Corresponding Author: luckyheriyantijufri@upi.edu

Submission: December 1st 2023; Accepted: December 24th 2023; Published: December 31st 2023
DOI: https://doi.org/10.31629/jg.v8i2.6590

Abstract

The focus of the PISA 2022 assessment is on the subjects of mathematics, language and science. Therefore, these subjects are compulsory subjects at every level of education. Because learning activities are the most important activities, the success of a learning activity is measured by the learning outcomes that have reached completeness or failed. Prediction of completeness or failure can be done by classifying data using the K-Nearest Neighbors (KNN) algorithm using the RapidMiner application. The KNN algorithm is one of the classification methods for a set of data based on learning data that has been classified before. The data used are student learning outcomes in Mathematics, Indonesian Language and Science subjects at the junior high school education level in Padang city. This research aims to predict student learning outcomes in Mathematics, Indonesian and Science subjects based on student score completeness by comparing various k values to obtain the best performance of this algorithm. The results obtained after analyzing the KNN algorithm are that classification using the KNN algorithm is most accurate when the values are k = 5 and k = 7. Where by using the value of k, the accuracy of the KNN algorithm reaches the maximum result of 94.12%. Thus, this algorithm can help teachers predict or determine how appropriate student completeness is.

Keywords: maximum K-Nearest Neighbors (KNN); rapidminer; classifier; learning outcomes

How to cite: Jufri, L. H., & Dasari, D. Evaluation of the implementation of machine learning algorithm K-Nearest Neighbors (KNN) using rapid miner on junior high school student learning outcomes. *Jurnal Gantang*, *8*(2), 193–197. https://doi.org/10.31629/jg.v8i2.6590

I. Introduction

PISA assessment focuses on the core school subjects of reading, math and science. The assessment does not just ascertain whether students can reproduce knowledge; it also tests how well students can extrapolate from what they have learned and apply that knowledge in unfamiliar environments, both inside and outside of school. This approach reflects the fact that modern economies reward individuals not for what they know but for what they can do with what they know (OECD, 2023).

Therefore, the subjects of mathematics, Indonesian language and science are also important subjects in the assessment. Mathematics and Natural Sciences (MIPA) have special characteristics that cause MIPA education to need to be handled specifically as well. One of

the characteristics is the cooperation between experiment and theory (Rahim & Rafiun, 2012). Meanwhile, the purpose of studying language is to improve the ability to understand and use language, both of which are the keys to communication activities (Solchan, 1998). So that these three subjects are an important concern for educational units, one of which is the junior high school level. In the entire educational process at school, learning activities are the most important activities, and the success of a learning activity is measured by learning outcomes that have reached completeness or failed. All students are required to complete this subject.

Predicting whether or not a student has passed can be done by classifying student completeness. One algorithm that can be used to classify is K-Nearest Neighbor (KNN). KNN is a popular and simple machine learning classifier algorithm. T. Cover and P. Hart first introduced KNN in 1967, where this algorithm classifies the sample class based on its nearest neighbour class (Fajri et al., 2020). KNN is often referred to as a lazy learner because KNN learns and classifies data without building a model. Unlike model-based classification algorithms, the KNN classifier only needs to remember all the training data in memory.

Along with its popularity, KNN is widely used to classify data in the fields of science and engineering as well as economics and business. Some recent publications that implemented KNN include the prediction of heart disease (Jabbar et al., 2013), the classification of PPA and BBM scholarship recipients (Sumarlin, 2016) as well as a selection of high school scholarships (Cholil, 2021), and so on. However, there needs to be a study of learning outcomes in the Education unit that utilizes the ability of the Machine Learning Classifier to classify and group data. Research, this time, will predict the accuracy of the students' grades on the three points of compulsory lessons on the first high school bench. (SMP). This study aims to predict the learning outcomes of students in mathematics lessons, Indonesian language and IPA based on the accuracy of students' values by comparing various values in order to obtain the best performance of these algorithms.

II. Research Method

Design Research and Data

The research uses a qualitative method using the K-Nearest Neighbor algorithm to determine the calcification result of a tourist destination recommendation. This study uses secondary data, namely data on the mid-term summative scores of the odd semester year 2023 SMP N 7 Padang. The amount of data is 255 students in class VIII in the subjects of Mathematics, Indonesian Language and Science, so the attributes in this study are the scores of Mathematics, Indonesian Language, Science and Results (pass/fail).

Data Analysis Technique

The data analysis techniques used in this study are as follows:

a. Data collection
   The data collection was carried out by obtaining mathematics, IPA and Indonesian language grades from teachers in the field of study at the school.
b. Analyzing data using the KNN method assisted by RapidMiner software, with the following steps
   Data is divided into 2 groups, namely training data and testing data.

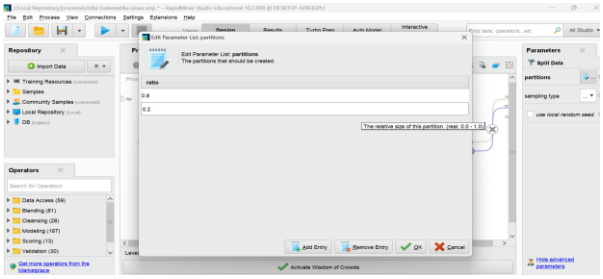

Figure 1. Ratio Determination

Testing is done by running the Dengan Analysis process using RapidMiner software and k values that vary among others for k=3, k=5, k=7, and k=9 values and will be seen on the data sum of 100 and 255. with Euclidean approach.

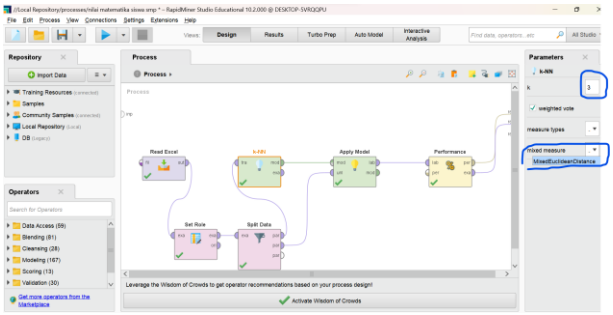

Figure 2. Determining K-value variation in the KNN algorithm

Determines the most accurate k values so that this algorithm provides the best performance to predict the accuracy of students' learning outcomes in mathematics lessons, Indonesian language and IPA.

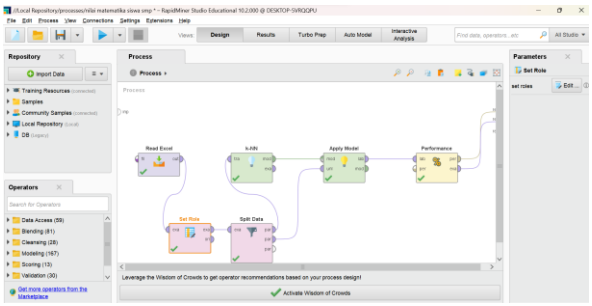

Figure 3. KNN algorithm model view on rapid miner

III. Results and Discussion

The data obtained is processed using Rapidmaner software. Tools used, among others, are:

1. Read Excel is a function that calls out the Excel data to be processed.
2. Role set is to specify the attribute to be used as a role
3. Split data is divided into two: data training and data testing.
4. KNNs that function to process data with k values can be modified according to desired needs.
5. Performance to determine the result of the KNN process framework performed.

After performing the entire step of the research procedure, then obtained the following results:

For k = 3 obtained the following result

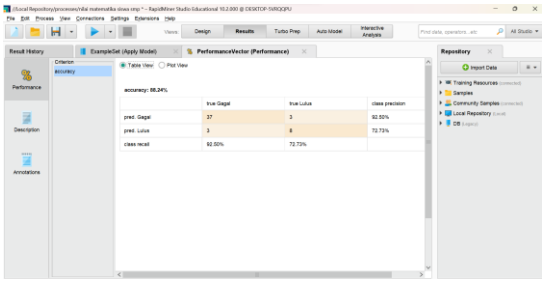

Figure 4. KNN output using RapidMiner

For k = 5, the following result

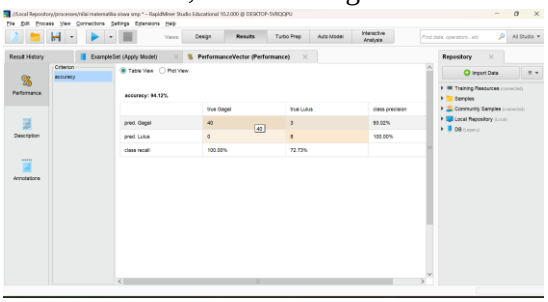

Figure 5. KNN output using RapidMiner

For k = 7, obtained the following result.

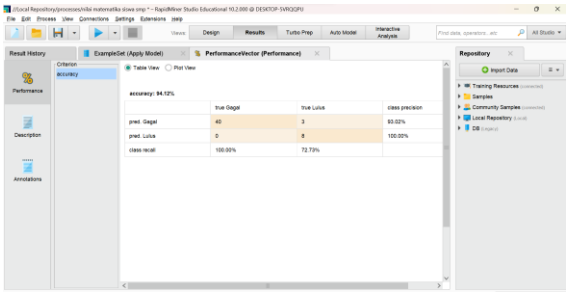

Figure 6. KNN output using RapidMiner

For k = 9, obtained the following result.

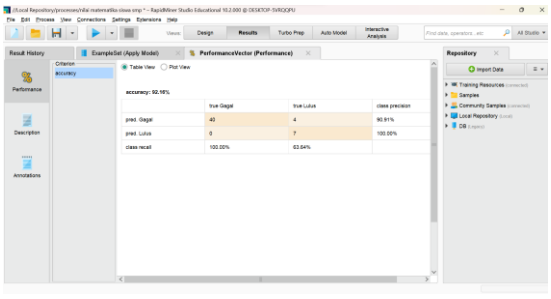

Figure 7. KNN output using RapidMiner

The data processed using RapidMiner is a total of 255 data, where 80% of the data is used for data training, and 20% is used for data testing to predict whether the KNN algorithm is well

used. Based on Table 1, it can be concluded that there are 4 k values tested to try to obtain the best k value used. The results obtained from the four k value performed are as follows:

Table 1.
KNN Algorithm Accuracy of Test Results

Jumlah Data	k=3	k=5	K=7	K=9
255	88.24	94.12	94.12	92.16

It appears in Table 1 that the highest accuracy is at k = 5, and k = 7 is 94.12%. The output of the KNN result means that out of 51 test data used (in this case, 20% of the total data 255), 11 data do not match the prediction, while 40 other data match the forecast. As for the lowest accuracy found at k=3 of 88.24%, which means out of 51 test data used, there are 14 data that do not match the prediction, while the other 37 data match the forecast. The accuracy obtained increases with the increasing K determination (Farokhah, 2020).

IV. Conclusion

Classification of students' learning outcomes using the KNN algorithm can help teachers predict or determine how the students' accuracy corresponds based on mathematics, Indonesian language, and IPA values. Classifications using this KNN algoritm give the most accurate result when k = 5 and k = 7. By using the value k, the accuracy of the KNN algorithm reaches the maximum result of 94.12%. As for the meaning of this Accuracy Level of 51 test data used (in this case, 20% of Total data 255), there are 11 data that do not match the prediction, while 40 other data match the forecast. Based on all the values obtained and the results compared to previous studies, the performance values achieved in this study are better compared with the previous study's performance, as well as show that the performance value can be better due to various simulations of testing of the entire K value on KNN.

References

Cholil, S.R., Handayani, T., Prathivi, R., Ardianita, T. (2021). Implementasi algoritma klasifikasi k-nearest neighbor (KNN) untuk klasifikasi seleksi penerima beasiswa. *IJCIT (Indonesian Journal on Computer and Information Technology)* Vol. 6 No.2, November 2021 Hal. 118-127.

Fajri, M. S., Septian, N., & Sanjaya, E. (2020). Evaluasi implementasi algoritma machine learning k-nearest neighbors (KNN) pada data spektroskopi gamma resolusi rendah. *Al-Fiziya: Journal of Materials Science, Geophysics, Instrumentation and Theoretical Physics*, *3*(1), 9–14. https://doi.org/10.15408/fiziya.v3i1.16180

Farokhah, L. (2020). Implementasi k-nearest neighbor untuk klasifikasi bunga dengan ekstraksi fitur warna RGB. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *7*(6),1129–1136. https://doi.org/10.25126/jtiik.2020722608

Jabbar, M. A., Deekshatulu, B. L., & Chandra, P. (2013). Classification of heart disease using k- nearest neighbor and genetic Algorithm. *Procedia Technology*, *10*, 85–94. https://doi.org/10.1016/j.protcy.2013.12.340

OECD, T. O. for E. C. and D. (2023). *Program for international student (PISA) 2022 assessment and analytical framework*. https://www.oecdlibrary.org/education/pisa-2022-assessment-and-analytical framework_dfe0bf9c-en

Rahim, U., & Rafiun, A. (2012). Peran matematika dalam menunjang mata pelajaran IPA, bahasa inggeris dan bahasa indonesia. *Jurnal Pendidikan Matematika*, *3*. https://www.neliti.com/publications/317669/peran-matematika-dalam-menunjang-mata-pelajaran-ipa-bahasa-inggeris-dan-bahasa-i

Panjaitan., CHP., Pangaribuan., L.J., Cahyad., C.I. (2022). Analisis metode k-nearest neighbor menggunakan rapid miner untuk sistem rekomendasi tempat wisata labuan bajo. *Remik: Riset dan E-Jurnal Manajemen Informatika Komputer*, Volume 6, Nomor 3, Agustus 2022 http://doi.org/10.33395/remik.v6i3.11701

Sichao, Z., (2020). Cost-sensitive KNN classification. *Neurocomputing*, vol. 391, pp. 234–242

Sumarlin, S. (2016). Implementasi algoritma knearest neighbor sebagai pendukung keputusan klasifikasi penerima beasiswa PPA dan BBM. Jurnal Sistem Informasi Bisnis, 5(1), 52–62. https://doi.org/10.21456/vol5iss1pp52-62

Solchan, H. 1998. Interaksi belajar mengajar bahasa indonesia. (Jakarta: Depdikbud)