Logaritma : Jurnal Ilmu-ilmu Pendidikan dan Sains
Vol. 11, No. 02 Desember 2023

Predictive Data Modeling: Student's Obstacle in Mathematical Literacy Tasks Focusing on Ratio and Proportion using The K-Nearest Neighbor Algorithm

Ambarsari Kusuma Wardani[1]; Dadan Dasari*[2]; Sufyani Prabawanto[3]
[1] Universitas Islam Negeri Raden Fatah Palembang, Sumatera Selatan
[2,3] Universitas Pendidikan Indonesia, Jawa Barat
ambarsariks_uin@radenfatah.ac.id[1] ,dadan.dasari@upi.edu*[2],sufyani@upi.edu[3]

Abstract

Ratio and proportion have a fundamental role in understanding mathematics and science. However, the fact is still found that students still face difficulties in carrying out the stages of formulating, applying and interpreting the process of solving mathematical problems, especially those related to ratio and proportion of material. These three problem solving processes are processes of mathematical literacy. Situations involving students' difficulties in understanding the concepts of ratio and proportion can be considered a learning obstacle. Three factors cause students to experience learning barriers, namely ontogenic barriers (mental ability to learn), didactic barriers (the impact of teacher teaching), and epistemological barriers (student knowledge that has limited context). Therefore, the aim of this research is to predict a data model related to learning obstacles experienced by students in the process of mathematical literacy skills in ratio and proportion material. Data model predictions are carried out using a data mining algorithm, namely K-Nearest Neighbor (K-NN) in Python language via Google Colab. Evaluation of the KNN algorithm using the confusion matrix method shows that the results of calculating the average accuracy of the K-NN method can predict data with an accuracy level of 89%.

Keywords: *K-NN Model prediction; Learning Obstacle; Mathematical Literacy; Ratio and Proportion.*

Abstrak

Rasio dan proporsi memiliki peranan fundamental dalam pemahaman matematika dan sains. Namun masih ditemui fakta bahwa siswa masih menghadapi kesulitan dalam melaksanakan tahapan merumuskan, menerapkan, dan menafsirkan pada proses penyelesaian masalah matematika, terutama yang terkait dengan materi rasio dan proporsi. Ketiga proses penyelesaian masalah tersebut merupakan proses dari literasi matematis. Situasi yang melibatkan kesulitan siswa dalam memahami konsep rasio dan proporsi dapat dianggap sebagai hambatan belajar (*learning obstacle*). Tiga faktor penyebab siswa mengalami hambatan belajar, yaitu hambatan ontogenik (kemampuan mental belajar), hambatan didaktis (dampak dari pengajaran guru), dan hambatan epistemologi (pengetahuan siswa yang memiliki konteks yang terbatas). Maka dari itu, tujuan penelitian ini adalah ingin memprediksi model data terkait hambatan belajar yang dialami siswa pada proses kemampuan literasi matematis

*Correspondence:
Email: dadan.dasari@upi.edu

pada materi rasio dan proporsi. Prediksi model data dilakukan dengan algoritma data mining yaitu K-Nearest Neighbor (K-NN) dengan bahasa python melalui Google Colab. Evaluasi algoritma KNN menggunakan metode confussion matrix menunjukkan hasil dari perhitungan rata-rata akurasi dari metode K-NN ini dapat memprediksi data dengan tingkat akurasi sebesar 89%.

Kata Kunci: K-NN Prediksi Model; Hambatan Belajar; Literasi Matematis; Rasio dan Proporsi.

INTRODUCTION

Ratio and proportion have a fundamental role in understanding mathematics and science. Developing concepts and skills related to slope, constant rate of change, and algebraic concepts in mathematics rests on a solid understanding of ratios and proportions (Jitendra et al., 2009). These concepts also support an understanding of triangles, trigonometry, and more advanced algebra instruction (Dougherty et al., 2016). In general, ratios and proportions play a role in describing relationships between quantities and facilitating a deep understanding in mathematics, science, and their application in everyday life.

The topic of ratio and proportion often appears in word problems at the elementary and middle school levels. Traditionally, this story problem is used to teach mathematical modeling concepts and train students to solve problems (Van Dooren et al., 2005). However, the fact is that students still need help solving problems related to ratios and proportions. In their study, Tiflis et al. (2019) stated that students needed help interpreting ratio and proportion questions and developing strategies to develop a solution plan. This leads to errors in selecting the operations involved in the chosen solution. In line with what has been written previously, Ambarwati & Ekawati (2022) also found that students still face difficulties formulating, applying, and interpreting the process of solving mathematical problems, especially those related to ratio and proportion material. These three problem-solving processes are processes of mathematical literacy.

Mathematical literacy ability is defined as a person's skills in formulating, employing, and interpreting mathematical concepts in solving mathematical problems involving various life contexts. (OECD, 2021). Mathematical literacy is one of the aspects tested in PISA. PISA (Program for International Student

Assessment) was held to evaluate education systems in 72 countries worldwide (OECD, 2012). The PISA results show that Indonesian students are generally still at the level of using knowledge to complete routine tasks and interpreting problems before solving them using formulas. In line with this, Meyer (2010) adds that students with low mathematical literacy skills need help solving non-routine problems related to ratios and proportions.

One of the reasons why it is difficult for students to find solutions to problems solving problems regarding ratios and proportions is that students experience various obstacles in studying this material. Çalisici (2018) stated that students find it challenging to understand that ratio compares with the concept of multiplication, not addition. Students also struggle to differentiate between fraction and ratio forms (Singh, 2000). Situations that involve students' difficulties in understanding the concepts of ratio and proportion can be considered learning obstacles.

Learning barriers can be defined as various obstacles students face during the learning process. These obstacles can make it difficult for students to understand the mathematical concepts being studied (Moru, 2009). Brousseau explains that three factors cause students to experience learning barriers, namely ontogenic barriers (mental ability to learn), didactic barriers (the impact of teacher teaching), and epistemological barriers (student knowledge that has limited context) (Suryadi, 2016).

From the description above, the researcher wants to predict a data model related to learning obstacles experienced by students in the process of mathematical literacy skills in ratio and proportion material. Data model predictions are made using a data mining algorithm, K-Nearest Neighbor (KNN). K-NN is a method for classification and prediction to place data in different categories or classes (Kuhkan, 2016). K-NN works by evaluating the similarity between new and old data. This method produces predictions based on objects with the closest parallel, measured based on the nearest distance (Jawthari & Stoffová, 2021). Therefore, it is necessary to form a prediction model by applying the K-NN algorithm to predict the learning obstacles experienced by students based on the mathematical literacy process they carry out when solving ratio and proportion problems.

RESEARCH METHODS

The researcher conducted this study to predict a data model regarding the learning obstacles experienced by students when finding solutions from ratio and proportion material using the mathematical literacy process. Prediction modeling was carried out using the K-NN algorithm with the confusion matrix method. The K-NN algorithm is implemented with the Python programming language.

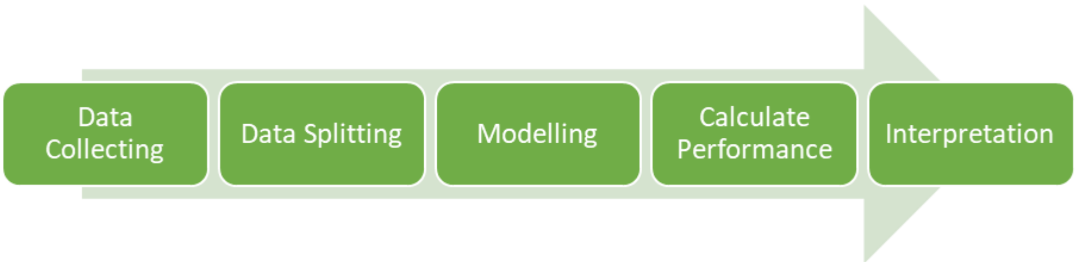

Figure 1. Research Flow

The flow of this research (Figure 1) begins with data collection. The data collected was from student test results in working on problem-solving questions on ratio and proportion material with mathematical literacy skills. The data is saved in CSV (Comma Separated Values) format, which will be processed to the data analysis stage and expressed as a dataset. The dataset was analyzed using Python programming, implemented through the Google Colab application. At this stage, researchers utilize various data analysis tools available within the Google Colab environment to gain further insight from test results and identify relevant patterns or findings (Wang et al., 2019). Then, in data processing, the dataset is divided into training data and test data. Next, activate the K-NN model to calculate the performance of the K-NN method on the dataset. The final stage is to interpret the resulting model and obtain prediction results regarding the learning obstacles experienced by students.

RESULTS AND DISCUSSION

This research begins with the data collection stage, where the researcher collects the necessary information. This study involved 92 class VIII students from five schools in South Sumatra. Each student takes a test of three questions related to the mathematical literacy process: formulating, applying, and interpreting. Researchers then analyzed students' answers to identify which learning obstacles were dominantly experienced by students in solving these problems.

After collecting data, the researcher continued to the data analysis stage using Python programming, which was run through the Google Colab application. Carneiro et al. (2018) state that applying technologies such as Google Colab and the Python programming language allows researchers to conduct data analysis efficiently and effectively. Google Colab allows researchers to utilize various Python libraries, including scikit-learn, which plays a role in building machine learning models by providing a variety of algorithms (Saabith et al., 2020). Additionally, the pandas library efficiently manages multi-dimensional arrays, while matplotlib and seaborn are used to create data visualizations through graphs or diagrams (Khandare et al., 2023). With the help of scikit-learn (Figure 2), researchers can access powerful machine-learning algorithms for classification, regression, and clustering tasks. The pandas library makes it easy to manipulate and analyze data, especially in the context of complex datasets. Meanwhile, matplotlib and seaborn provide tools for creating informative and easy-to-understand visualizations, allowing researchers to understand data patterns and present results more effectively (Naik et al., 2022). The combined use of these libraries in the Google Colab environment facilitates research by providing powerful tools and safe and efficient computing resources, speeding up the process of data analysis and model building.

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Figure 2. Importing Required Libraries

In the next step, researchers entered the data into the Python environment. As seen in Figure 3, this dataset consists of 92 data with four input variables or features and 1 column with a label called learning obstacle. It is important to note that this dataset has no missing values and extreme values (outliers). This information provides an initial picture of the characteristics of the dataset used. Ninety-two data with four features and one label "learning obstacle" shows the dimensions and structure of the dataset. In addition, the absence of missing values and outliers can increase the reliability of data analysis and the results produced from machine learning models (Seliem, 2022). Triguero et al. (2019) explain that cleaning data from missing or extreme values is essential to ensure the quality of analysis results. With a clean dataset, researchers can proceed to the next stage in the analysis and modeling process.

```python
MathematicalLiteracy = pd.read_csv('MathematicalLiteracy.csv')
print(MathematicalLiteracy)
```

```
    Formulate  Employ  Interpret Learning Obstacle
0          40      50         30       Ontogenical
1          40      40         20       Ontogenical
2          80      40         15       Ontogenical
3          40      45         20       Ontogenical
4          90      85         80        Didactical
..        ...     ...        ...               ...
87         70      85         30       Ontogenical
88         30      10          0       Ontogenical
89         90     100         80    Epistimological
90         80      80         75    Epistimological
91         20      40          0       Ontogenical

[92 rows x 4 columns]
```

Figure 2. Data Loading Process

After preparing the dataset, the next step is to divide the dataset into two main parts: training and testing data. The separation separates attributes (features) and labels in the data frame at this stage. Mahesh (2020) describes details or features as variables used as input for machine learning models, while labels are variables the model wants to predict or understand. Separating these attributes and labels is a critical step in the machine-learning process, as it allows the model to learn patterns from the details and then predict the corresponding labels (Kigo et al., 2023).

After this separation, the researcher split the dataset into two parts: training data, used to train the model, and testing data, used to test the performance of the trained model. This separation is essential to measure how much the model can generalize patterns from never-before-seen data (Aldwairi et al., 2018). Sudarman & Budi (2023) explain that the separation of data into training and testing also helps in evaluating model performance objectively and identifying whether the model tends to be overfitting (too focused on training data) or underfitting (not able to generalize well).

```python
from sklearn.model_selection import train_test_split

# create feature/input data
X = MathematicalLiteracy.drop('Learning Obstacle', axis=1)

# create output data
y = MathematicalLiteracy['Learning Obstacle']

# divide training data by 80% and testing data by 20%
X_train, X_test, y_train, y_test = train_test_split(X, y,
                                   test_size=0.2, random_state=42)
```

Figure 3 Dividing Datasets

As seen in Figure 4, the data division is set with a proportion of 0.2. It means that in the machine learning process, the dataset is divided into two main parts: training data (80%), used to train the model, and testing data (20%), used to test the performance of the model that has been introduced. This method randomly divides training and testing data, so the results will differ if the researcher repeats each process. A random method was applied at the stage of dividing the dataset into training data and testing data. Therefore, the data-sharing results will differ every time this process is repeated. It can affect model performance and model evaluation carried out by researchers.

For this reason, the random_state parameter is used to keep data-sharing results identical, even if this method is carried out on different devices (Barbiero et al., 2021). To maintain consistency, researchers used a particular parameter called random_state. This parameter ensures that the data-sharing results remain identical each time the process is repeated (Mahata et al., 2021). In other words, using a specific value for random_state allows replicability of the results, even if the processing is done on different devices. It helps create reproducible experiments and facilitates the comparison of results across other conditions or environments.

Before using the K-NN model, it is essential to run a standardization process on the data to ensure that all values are on a uniform scale or magnitude (Figure 5) (Raseman et al., 2020). K-NN is a distance-based model, so distance measures have a crucial role. Differences in data scale between features can significantly influence the resulting model's performance. To overcome this potential problem, it is recommended always to scale features, especially if the difference between them is significant (Suliztia & Fauzan, 2020).

```
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()

# features scaling : standardize
scaled_X_train = scaler.fit_transform(X_train)
scaled_X_test = scaler.fit_transform(X_test)
```

Figure 4. Data Standardization

The scaling method generally applied is standardization, where the standardization process will produce each feature with an average value of 0 and a standard deviation of 1. This method will bring the data to a uniform scale, making it easier for the K-NN model to provide appropriate weights. Balance each feature and produce more consistent results (Varshney et al., 2021). The standardization process has the additional benefit of helping to prevent domination by large-scale features. It is essential because according to Prasatha et al. (2017) scale differences between components can affect distance calculations in K-NN models. With standardization, the contribution of each feature to the distance calculation becomes more balanced, so the model is more reliable and can handle situations where parts have different scale variations (Leto et al., 2023). Overall, using standardization methods helps improve the consistency and performance of K-NN models, especially in situations where the dataset has features with varying scales.

K-NN models can be built using the `KNeighborsClassifier` function in the `sklearn.neighbors` library as shown in Figure 6. This function provides various parameters that can be configured as needed. Some of the key parameters that can be adjusted include the number of neighbors (`n_neighbors`, default=4), the distance metric (`metric`, default='minkowski'), the weight of each point (`weight`, default='uniform': all k neighbors have the same weight), and other parameters.

At this stage, the K-NN model will be created using the default values for these parameters. The model will use four neighbors as default values, a Minkowski distance metric, equal neighbor weights for all points, and other values defaulted by the library sets. Using these default values, researchers can begin an initial exploration of the K-NN model and then adjust these parameters according to the specific needs and characteristics of the dataset.

```
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import confusion_matrix, classification_report


# help(KNeighborsClassifier)

knn_model = KNeighborsClassifier(n_neighbors=4)

# Create a model based on training data
knn_model.fit(scaled_X_train, y_train)

# Predict/evaluate testing data output
y_pred = knn_model.predict(scaled_X_test)

# Displays the confusion matrix
print(confusion_matrix(y_test, y_pred))

print(classification_report(y_test, y_pred))
```

Figure 5. Activating K-NN Model

After successfully building a K-NN model using training data, the next step involves evaluating the model's performance against testing data. According to Purwanto & Nugroho (2023), this evaluation process is essential to measure the extent to which the model can generalize patterns from data not used in building the model. Several standard evaluation metrics, such as accuracy, balanced accuracy, and F1 Score, are used to get a holistic picture of model performance. Accuracy provides information about the extent to which the model predicts classes correctly. Balanced accuracy becomes relevant if the dataset has a class imbalance, whereas the F1 Score delivers a balance between precision and recall, especially in the case of binary classification (Mahmudah et al., 2021).

In addition to these metrics, the confusion matrix provides a detailed view of the success and failure of the model in predicting each class. The confusion matrix includes information about true positives, false positives, and false negatives. Analysis of the confusion matrix helps researchers understand where the model can provide reasonable predictions or where improvements need to be made (Sholikhah et al., 2023). Researchers can make further decisions after getting the evaluation results, such as adjusting model parameters or considering other models. This thorough evaluation is an essential basis for ensuring the reliability and performance of the K-NN model in handling the dataset used.

```
[[4 1 0]
 [1 5 0]
 [0 0 8]]
                 precision    recall  f1-score   support

     Didactical       0.80      0.80      0.80         5
Epistimological       0.83      0.83      0.83         6
    Ontogenical       1.00      1.00      1.00         8

       accuracy                           0.89        19
      macro avg       0.88      0.88      0.88        19
   weighted avg       0.89      0.89      0.89        19
```

Figure 6. Results of The Confusion Matrix and Model Prediction

Based on the confusion matrix analysis in Figure 7, it can be concluded that the K-NN model successfully predicts students who experience ontogenic obstacles very accurately, with an accuracy level of 100%. It shows that the model can correctly recognize and classify students who face ontogenic barriers. However, the results are slightly different when we look at the model's performance against didactic and epistemological obstacles. The model could predict didactic obstacles with an accuracy rate of 80%, while for epistemological barriers, the accuracy rate was 83%. While it still performs well, there is potential to improve predictions on both types of obstacles.

This analysis provides important insights into the extent to which the model can identify different types of barriers and provides a basis for further development (Setiawan et al., 2023). If necessary, adjustments to model parameters or other methods can be explored to improve classification performance in specific categories. This evaluation plays a crucial role in understanding the strengths and limitations of the K-NN model in identifying barriers to student learning.

The confusion matrix also illustrates that the K-NN model with a value of k = 4 achieves the highest level of accuracy, namely 89%. These results show that the K-NN algorithm can effectively classify learning obstacles students face based on the mathematical literacy process and can achieve an adequate level of accuracy. These findings align with research by Zhang et al. (2017), who stated that a high level of accuracy reflects confidence in the model's effectiveness in classification tasks.

The model's success with a value of k = 4 in achieving an accuracy level of 89% indicates that this model can adequately identify student learning barriers. However, along with further exploration, research can continue to evaluate the influence of k values on model performance to ensure optimal selection of k values. This analysis provides strong support for the applicability of the K-NN algorithm in the context of learning obstacle classification, with the potential to offer valuable insights to educational advisors or policymakers.

CONCLUSION

The conclusion of this research is the creation of an application to predict the dominant learning obstacles experienced by students when solving problems regarding ratios and proportions with the mathematical literacy process using the K-NN algorithm. Evaluation of the KNN algorithm using the confusion matrix method shows that the results of calculating the average accuracy of this KNN method are 89%. The prediction with the highest level of accuracy is an ontogenic learning obstacle of 100%. For didactic and epistemological obstacles, the model can only make predictions with results accuracy of 80% and 83%. Future research could focus on developing more complex models to improve prediction accuracy. Further studies can include additional variables, such as environmental variables or student learning styles, which can enrich the analysis and understanding of the factors that influence mathematical literacy.

REFERENCES

Aldwairi, T., Perera, D., & Novotny, M. A. (2018). An evaluation of the performance of Restricted Boltzmann Machines as a model for anomaly network intrusion detection. *Computer Networks*, *144*, 111–119.

Ambarwati, B. T., & Ekawati, R. (2022). ANALISIS LITERASI MATEMATIKA SISWA DALAM MENYELESAIKAN SOAL HIGHER ORDER THINKING SKILLS (HOTS) PROPORSI. *Jurnal Ilmiah Pendidikan Matematika Volume*, *11*(2).

Barbiero, P., Ciravegna, G., Georgiev, D., & Giannini, F. (2021). Pytorch, explain! a python library for logic explained networks. *ArXiv Preprint ArXiv:2105.11697*.

Çalisici, H. (2018). Middle School Students' Learning Difficulties in the Ratio-Proportion Topic and a Suggested Solution:" Envelope Technique". *Universal Journal of Educational Research*, *6*(8), 1848–1855.

Carneiro, T., Da Nóbrega, R. V. M., Nepomuceno, T., Bian, G.-B., De Albuquerque, V. H. C., & Reboucas Filho, P. P. (2018). Performance analysis of google colaboratory as a tool for accelerating deep learning applications. *IEEE Access*, *6*, 61677–61685.

Dougherty, B., Bryant, D. P., Bryant, B. R., & Shin, M. (2016). Helping students with mathematics difficulties understand ratios and proportions. *Teaching Exceptional Children*, *49*(2), 96–105.

Jawthari, M., & Stoffová, V. (2021). Predicting students' academic performance using a modified kNN algorithm. *Pollack Periodica*, *16*(3), 20–26. https://doi.org/10.1556/606.2021.00374

Jitendra, A. K., Star, J. R., Starosta, K., Leh, J. M., Sood, S., Caskie, G., Hughes, C. L., & Mack, T. R. (2009). Improving seventh grade students' learning of ratio and proportion: The role of schema-based instruction. *Contemporary Educational Psychology*, *34*(3), 250–264.

Khandare, A., Agarwal, N., Bodhankar, A., Kulkarni, A., & Mane, I. (2023). Analysis of Python Libraries for Artificial Intelligence. In *Intelligent Computing and Networking: Proceedings of IC-ICN 2022* (pp. 157–177). Springer.

Kigo, S. N., Omondi, E. O., & Omolo, B. O. (2023). Assessing predictive performance of supervised machine learning algorithms for a diamond pricing model. *Scientific Reports*, *13*(1), 17315.

Kuhkan, M. (2016). A method to improve the accuracy of k-nearest neighbor algorithm. *International Journal of Computer Engineering and Information Technology*, *8*(6), 90.

Leto, C., Sujana, D., Windyasari, V. S., & Muhammad, R. (2023). *KONSEP DATA MINING DAN PENERAPAN*.

Mahata, S., Shakya, P., & Babu, N. R. (2021). A robust condition monitoring methodology for grinding wheel wear identification using Hilbert Huang transform. *Precision Engineering*, *70*, 77–91.

Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, *9*(1), 381–386.

Mahmudah, K. R., Indriani, F., Takemori-Sakai, Y., Iwata, Y., Wada, T., & Satou, K. (2021). Classification of Imbalanced Data Represented as Binary Features. *Applied Sciences*, *11*(17), 7825.

Meyer, E. R. (2010). *Developing proportional reasoning in Mathematical Literacy students*. Stellenbosch: Stellenbosch University.

Moru, E. K. (2009). Epistemological obstacles in coming to understand the limit of a function at undergraduate level: A case from the national university of lesotho. *International Journal of Science and Mathematics Education*, *7*(3), 431–454. https://doi.org/10.1007/s10763-008-9143-x

Naik, P., Naik, G., & Patil, M. (2022). Conceptualizing Python in Google COLAB. *India: Shashwat Publication*.

OECD. (2021). Pisa 2021 Mathematics Framework (Draft). *Angewandte Chemie International Edition, 6(11), 951–952.*, 5–24. http://www.oecd.org/pisa/pisaproducts/pisa-2021-mathematics-framework-draft.pdf

OECD, P. (2012). Assessment and Analytical Framework-Mathematics, Reading, Science, Problem Solving and Financial Literacy, 2012b. *France: OECD Publishing*.

Prasatha, V. S., Alfeilate, H. A. A., Hassanate, A. B., Lasassmehe, O., Tarawnehf, A. S., Alhasanatg, M. B., & Salmane, H. S. E. (2017). Effects of distance measure choice on knn classifier performance-a review. *ArXiv Preprint ArXiv:1708.04321*, 56.

Purwanto, A., & Nugroho, H. W. (2023). Analisa Perbandingan Kinerja Algoritma C4. 5 Dan Algoritma K-Nearest Neighbors Untuk Klasifikasi Penerima Beasiswa. *Jurnal Teknoinfo*, *17*(1), 236–243.

Raseman, W. J., Rajagopalan, B., Kasprzyk, J. R., & Kleiber, W. (2020). Nearest neighbor time series bootstrap for generating influent water quality scenarios. *Stochastic Environmental Research and Risk Assessment*, *34*(1), 23–31.

Saabith, A. L. S., Vinothraj, T., & Fareez, M. (2020). Popular python libraries and their application domains. *International Journal of Advance Engineering and Research Development*, *7*(11).

Seliem, M. M. (2022). HandlingOutlier data as missing values by imputation methods: application of machine learning algorithms. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *13*(1), 273–

286.

Setiawan, Z., Fajar, M., Priyatno, A. M., Putri, A. Y. P., Aryuni, M., Yuliyanti, S., Widiputra, H., Meilani, B. D., Ibrahim, R. N., & Azdy, R. A. (2023). *BUKU AJAR DATA MINING*. PT. Sonpedia Publishing Indonesia.

Sholikhah, I. I., Harjanta, A. T. J., & Latifah, K. (2023). Machine Learning Untuk Deteksi Berita Hoax Menggunakan BERT. *Prosiding Seminar Nasional Informatika*, *1*(1), 524–531.

Singh, P. (2000). Understanding the concepts of proportion and ratio among grade nine students in Malaysia. *International Journal of Mathematical Education in Science and Technology*, *31*(4), 579–599.

Sudarman, E. J., & Budi, S. (2023). Pengembangan Model Kecerdasan Mesin Extreme Gradient Boosting Untuk Prediksi Keberhasilan Studi Mahasiswa. *Jurnal STRATEGI-Jurnal Maranatha*, *5*(2), 297–314.

Suliztia, M. L., & Fauzan, A. (2020). Comparing Naive Bayes, K-Nearest Neighbor, And Neural Network Classification Methods Of Seat Load Factor In Lombok Outbound Flights. *Jurnal Matematika, Statistika Dan Komputasi*, *16*(2), 187–198.

Suryadi, D. (2016). Didactical design research (ddr): upaya membangun kemandirian berpikir melalui penelitian pembelajaran. *Makalah Pada Seminar Nasional Matematika Dan Pendidikan Matematika UNSWAGATI*, *6*.

Tiflis, O., Ineson, G., & Watts, M. (2019). Ratio and Proportion: An analysis of GCSE resit students' errors. *Proceedings of the British Society for Research into Learning Mathematics*, *39*(1).

Triguero, I., García-Gil, D., Maillo, J., Luengo, J., García, S., & Herrera, F. (2019). Transforming big data into smart data: An insight on the use of the k-nearest neighbors algorithm to obtain quality data. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(2), e1289.

Van Dooren, W., De Bock, D., Hessels, A., Janssens, D., & Verschaffel, L. (2005). Not everything is proportional: Effects of age and problem type on propensities for overgeneralization. *Cognition and Instruction*, *23*(1), 57–86.

Varshney, R. K., Chauhan, S. P. S., & Sharma, V. (2021). A K-NN based Data Reduction Technique in String Space via Space Separation. *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, 223–227.

Wang, A. Y., Mittal, A., Brooks, C., & Oney, S. (2019). How data scientists use computational notebooks for real-time collaboration. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–30.