Particle Swarm Optimization for Parameter Determination and Selection Features in K-Nearest Neighbor Algorithm on the Prediction of the Mis Smpit Students' Memory Targets

¹Muhammad Evandry Dewaksara, ²Ema Utami, ³Muhammad Rudyanto Arief ¹Magister of Informatics Engineering, Amikom University Yogyakarta, Indonesia ²Magister of Informatics Engineering, Amikom University Yogyakarta, Indonesia ³Magister of Informatics Engineering, Amikom University Yogyakarta, Indonesia

Abstract:- Memorizing the Qur'an is an activity of instilling Allah's verses into the heart and mind which can later be read without looking at the text. There are many techniques in memorizing the Qur'an and the factors that influence it, for instance motivation. SMPIT Makassar Islamic School is a boarding school that provides memorizing the Our'an as one of its activities. The purpose of this study was to determine the factors which influence students who pass and do not pass in memorizing the Qur'an. The number of surahs memorized is limited to 37 suras (Juz Amma) in a period of 1 year. In this study there were 63 students consisting of 37 students who passed, and 26 students did not pass. This study uses a quantitative descriptive method with a correlational approach. In the classification process, researchers used K-Nearest Neighbor, Forward Selection and Particle Swarm Optimization (PSO). Forward Selection and Particle Swarm Optimization are used as features selection to optimize the number of features used. The Forward Selection and Particle Swarm Optimization were compared, and the results showed that the optimal use of features were three features with an accuracy of 84% and K=6 on the use of K-Nearest Neighbor and PSO, 84% with K=2 with three features on the use of K-Nearest Neighbor and Forward Selection. The results of this study revealed that the factors that influence students who pass and do not pass in memorizing the Qur'an are internal and external factor.

Keywords: - Al-Qur'an Memorization, Motivation, K-Nearest Neighbor, Particle Swarm Optimization (PSO).

I. INTRODUCTION

SMPIT Makassar Islamic School is a boarding school that focuses on teaching students language skills, information technology, sports, and other general subjects; besides that, there will be a one-year Al-Qur'an quarantine period that frees students from other school burdens. Moreover, only focus on memorizing the Qur'an (Juz Amma). Memorizing the Qur'an is not impossible because Allah SWT has guaranteed ease for the memorizers, which explain in Q.S Al-Qamar verse 22.

وَلَقَن يَسَّرْنَا الْقُرْأَنَ لِلنِّكُرِ فَهَلُ مِنُ مُّدَّكِرٍ ﴾

"And in truth We have made the Qur'an easy to remember; but is there any that remembereth?" (QS. Al-Qamar: 22).

Memorizing the Our'an is a noble activity in the sight of Allah SWT, a person who memorizes it will become Ahlullah on earth. The miracle of the Al-Our'an, even though it uses Arabic but all Muslims in the world can read and memorize it. Every Muslim can memorize the Qur'an, even if only a few short letters, such as Surah An-Nas, Al-Ikhlas, Al-Falaq and other suras, is part of juz 30 or juz Amma. (Kurniawati, 2022). In memorizing Qur'an, internal and external factors can affect the smooth process of memorization. Personality factor as a differentiator from each individual becomes the most influential internal factor in the process of memorizing the Our'an, and then the memorization method is an external factor which, in the process, becomes a uniform factor for each individual. (Fitriyah, 2008). In the learning process, everything is processed in the form of memory so that the remembered things can be firmly bound within oneself to adapt to everyone's life. Therefore, when there are students who are not so strong in memorizing but with perseverance which have great learning motivation so that they can memorize very well, this can be used as an example for others. (Kemampuan et al., 2018). Many types of research on factors in memorization have been carried out, such as studies that measure the level of memorization based on behaviour factors to increase students' motivation in memorizing. (Marlina, 2018) in this study, it was found that the provision of awareness that motivates students to memorize can increase the level of student persistence in memorizing so that students with high motivation will reach the target of memorization faster than students who have the low motivation or are not even motivated. At all. (Muliati & Oktavia, 2022).

In the memorization process, other factors that influence it, such as the influence of teacher and student interpersonal communication to improve memorization in the form of motivation (Oktarina & Yanti, 2021), teacher and student interactions through communication have been shown to

affect the level of student motivation in memorizing, based on the measurements that have been made it can be explained that 33.3% of interpersonal communication variables affect the interest in memorizing the Qur'an. The factor of student interest in memorizing has excellent potential in the smooth process of memorizing the Our'an. (Wijava et al., 2021) Then the motivational and stimulus factors (changes in the internal or external environment that can be known) become factors that affect the student's memorization process. Age, location, and time management are pretty influential factors in memorizing (Simanjuntak, 2021). Not only that, the length and shortness of the verses of the Qur'an to be memorized are factors that affect the time of memorization. Previous research shows that internal factors that affect student memorization are personality, interests and talents, and memory abilities. In contrast, external factors include the method used, motivation, fluency in reading and writing, routine practice, time, facilities and places. Based on this, the researcher will make this factor a reference variable in measuring the level of student memorization. One of the obstacles SMPIT Makassar Islamic School faces is the difficulty in knowing the factors that affect students' performance in memorizing the Qur'an. So the purpose of this study is to find out what factors influence students in memorizing the Qur'an.

II. METHODS

This research is a quantitative descriptive study with a correlational approach. Correlational research is research that aims to determine the cause-and-effect relationship between the independent variable and the dependent variable. (Saputra et al., 2019) Then this research was conducted on 63 students of SMPIT Makassar Islamic School. 18 features are used to measure the factors that influence the process of namely memorizing the Our'an, semester grades, cooperation, parental roles, discipline, evaluation implementation, communication patterns, motivation, level, read and write Quran, activity, personality, environmental changes, interests, reward, adaptation, motivation, age, and gender. In collecting data, the researcher used a value scale of 0-100 with the criteria of values 0-30 = 100, 31-60 = 100moderate, and 61-100 = high. Then the data that has been collected is classified using the k-nearest neighbour (KNN) algorithm with k = 3. The data classification results are then optimized using a forward selection approach and particle swarm optimization (PSO) to reduce the number of features used to determine the factors which are very influential in the process of memorizing the Al-Ouan. In the testing process, the author uses google colab, an application for data processing and machine learning model-making. The last step is to compare the results of forwarding selection optimization and particle swarm optimization (PSO), which one of the optimizations is the best to use in conjunction with KNN.

Table 1. Student Database

FEATURE	Student 1	Student 2	Student 3	Student 4	Student 5
SEMESTER GRADES	92	92	91	70	93
COOPERATION	90	90	90	90	30
PARENTAL ROLES	45	93	92	87	45
DISCIPLINE	90	90	85	85	88
EVALUATION IMPLEMENTATION	85	85	85	85	85
COMMUNICATION PATTERNS	70	85	85	80	70
MOTIVATION	75	85	85	60	75
LEVEL	90	90	80	80	90
READ AND WRITE QURAN	90	90	90	80	90
ACTIVITY	30	85	85	75	30
PERSONALITY	75	92	92	25	75
ENVIRONMENTAL CHANGES	30	30	30	87	20
INTERESTS	86	86	86	88	30
REWARD	86	86	86	87	87
ADAPTATION	30	95	95	75	30
MOTIVATION	95	95	85	95	85
AGE	14	14	14	14	14
GENDER	2	1	1	1	1
GRADUATION STATUS	0	1	0	1	1

Source: Google Collab Results, 2022

Google Colab Preparation

Google Collaboratory, or "Google Colab" for short, is a product of Google Research. Colab allows anyone to write and execute arbitrary python code through a browser and is perfect for machine learning, data analysis, and education. (Gunawan et al., 2020) More technically, Colab is a hosted Jupyter notebook service that requires no setup for use while providing free access to computing resources, including GPUs.

Set up a google account first, then visit https://colab.research.google.com and create a new Notebook with File Menu > New Notebook. Import the dataset using the commands shown in Figure 1, click Connect to Google Drive, select a Google Drive account, connect to, and click Allow.

from	google.c	olab	import	drive
drive	e.mount('	/cont	ent/gdr	ive')

Fig 1. Connect Google Colab to Google Drive

Once Google Colab has successfully connected to Google Drive, import the dataset to be used, as shown in Figure 2.

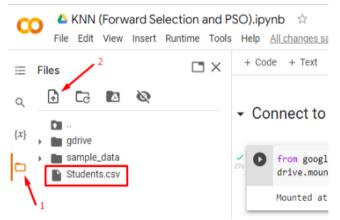


Fig 2. Import Student Dataset

> Application of KNN+Forward Selection

Forward selection is a method for selecting the most influential variable on the target variable. This method begins by selecting the variable with the most significant correlation with the target variable. Then, the variable with the most significant correlation with the remaining target variables will be selected. This process will continue until no variable has the most significant correlation with the remaining target variables. (Nanja & Purwanto, 2015) The forwarding Selection in KNN aims as a Feature Selection to select the variables that have the most influence on the target variable. Thus, there will be a decrease in the dimensions of the data so that it can speed up the KNN process.

In this stage, forward selection is used, namely choosing variables that provide information with predictive accuracy so that the performance of an algorithm can be optimized. Forward selection works on the principle that this approach builds the model, starting with no variables in the model and adding useful ones. The general flowchart or schema is as follows, Figure 3.

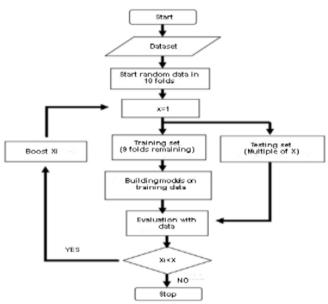


Fig 3. Flowchart Forward Selection

The training data is carried out in stages, starting from 1 variable to the level or number of variables that produces the best performance or accuracy value or the minor error. For example, testing data with two variables produces a minor error. When tested again with three variables and producing a more excellent error value compared to 2 variables, a minor error is obtained in the second variable, which means the second variable is significant; the process is stopped when all independent variables have been tested. The forward selection algorithm will be tested on each data, starting from 1-period variable data to 10-period variable data, which produces the best accuracy. Feature selection using forward can be seen in Figure 4.

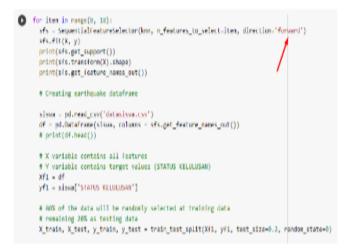


Fig 4. Feature Selection Using Forward

> Application of KNN+PSO

Particle Swarm Optimization (PSO) is a method inspired by the nature of a group of animals such as birds, termites, bees, and ants. The PSO algorithm mimics the properties of some of these organisms. This trait consists of habits carried out in daily activities and the influence of individuals on others in a population. (Okwu & Tartibu, 2021).

The word "particle" refers to an individual, for example, a bird in a bird population. Each individual or particle is interconnected with their respective intelligence and influenced by the behavior of other groups in the population. (Bharath et al., 2022) Under these conditions, if one particle finds an effective or shorter path to the food source, the other particles will follow that path even though their initial location is far from the group.

The KNN+PSO algorithm is applied to classify student achievement data sets in memorizing the Quran. The dataset was divided into the training set and the test set. The training set is used to train the KNN+PSO algorithm, and the test set is used to test the algorithm's accuracy. Python library needed for the implementation of the algorithm, as shown in Figure 5.

Import essential libraries
import os
import joblib
import numpy as np
import pandas as pd
import warnings
import matplotlib
import matplotlib.pyplot as plt
from matplotlib import ticker
import seaborn as sns
From sklearn.model_selection import train_test_split
from numpy.random import rand
import random
From random import randrange
import time
from sklearn.preprocessing import StandardScaler
From sklearn.metrics import mean_squared_error
from sklearn.model_selection import StratifiedKFold
from sklearn.linear_model import LinearRegression
import xgboost as xg
from sklearn.neighbors import KNeighborsclassifier
from sklearn import metrics
from sklearp.metrics import classification report
Fig 5. Importing KNN+PSO Libraries

After the library is imported, the dataset is loaded into the google colab environment and creates a dataframe using the pandas' library. How to import and create a dataframe can be seen in Figure 6.



Fig 6. Creating Student Dataframes

Divide the dataset into features and targets (X and y), then split the data into training and testing data with a ratio of 80:20, as shown in Figure 7.

```
# X variable contains all features
# Y variable contains target values (GRADUATION STATUS)
X = siswa.drop(['NAME','GRADUATION STATUS'], axis=1)
y - siswa['GRADUATION STATUS']
# 80% of the data will be randomly selected at training data
# remaining 20% as testing data
```

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)

Fig 7. Creating Feature and Target Columns

To determine the best number of K, repeat for K = 1 to 8 and calculate the accuracy values for each K. Then, select the K with the highest accuracy value. In this case, K = 6 has the highest accuracy value of 0.84. Then, classify with KNN using K = 3 and calculate the value of accuracy, precision, recall, and F1 score. as shown in Figures 8 and 9.

```
for i in range(2, 9):
 knn = KNeighborsClassifier(n neighbors=i)
 knn.fit(X_train, y_train)
 print(f"k-value = , {i}")
 print(knn.score(X_test, y_test))
 print(f"-----
                               ")
       k-value = , 2
       0.6153846153846154
            - - - - - - - - - -
       k-value = , 3
       0.6923076923076923
       k-value = .
                   4
       0.6923076923076923
           -----
       k-value = , 5
       0.6923076923076923
       k-value = , 6
       0.6923076923076923
            -----
                   7
       k-value = .
       0.6923076923076923
       k-value = , 8
       0.6923076923076923
```

Fig 8. Calculating the accuracy of KNN

target_names = ['first_value_y','second_value_y'] # target values

Print classification report after a train/test split: print(classification_report(y_test, predictions, target_names-target_names))

	precision	recall	f1-score	support	
first_value_y	0.67	0.86	0.75	7	
second_value_y	0.75	0.50	0.60	б	
accuracy			0.69	13	
macro avg	0.71	0.68	0.68	13	
weighted avg	0.71	0.69	0.68	13	

Fig 9. Calculating Precision, Recall and f1-Score

The application of PSO to this data classification is made by changing the value of k on the KNN to 1, then calculating the fitness value of each particle. The fitness value is calculated by calculating the accuracy of each particle. Afterward, the particle with the best fitness value will be used as the gbest particle. Then, the particle with the worst fitness value will be used as the lbest particle. Then, the velocity and position values of each particle will be updated. After that, the fitness value of each particle will be recalculated. This process will be carried out for 100 iterations. After 100 iterations are completed, the particle with the best fitness value will be used as the gbest particle. More details can be seen in Figures 10 and 11.

perform feature selection
start_time = time.time()
fmdl = jfs(X_train, y_train, opts)
print("Run Time --- %s seconds ----" % (time.time() - start_time))
sf = fmdl['sf']
model with selected features

num_train = np.size(xtrain, 0)
num_valid = np.size(xtest, 0)
x train = xtrain[:, sf]
Fig 10. Application of PSO to KNN (1)

```
fig, ax = plt.subplots()
ax.plot(x, curve, 'o-')
ax.set_xlabel('Number of Iterations')
ax.set ylabel('Fitness')
ax.set title('PSO')
ax.grid()
plt.show()
Iteration: 1
Best (PSO): 0.4100992952054853
Iteration: 2
Best (PSO): 0.4100992952054853
Iteration: 3
Best (PSO): 0.39938439439904794
Iteration: 4
Best (PSO): 0.39938439439904794
Iteration: 5
Best (PSO): 0.39938439439904794
Iteration: 6
Best (PSO): 0.39938439439904794
Iteration: 7
Best (PSO): 0.39938439439904794
Iteration: 8
Best (PSO): 0.39938439439904794
Iteration: 9
```

Best (PSO): 0.39938439439904794 Iteration: 10

Fig 11. Application of PSO to KNN (2)

Determine the most influential feature selection in memorizing the Qur'an using the KNN and PSO algorithms. The optimization results obtained 3 features that are most influential in memorizing the Qur'an, namely reading and writing the Qur'an, activeness in doing sunnah practices, implementation and evaluation when the memorization process is carried out. as in Figure 12.

```
feature_selection = []
items= fmdl['sf']
for index, item in enumerate(items):
    print(f"{item} = {df2.columns[item]}")
    feature_selection.append(df2.columns[item])
```

```
feature_selection
```

['BTQ', 'KEAKTIFAN', 'PELAKSANAAN EVALUASI'] Fig 12. Define PSO Result Features

III. RESULTS

The results of the KNN+Forward Selection calculation are evaluated based on the confusion matrix. The confusion matrix used to calculate classification accuracy value, (Miranda et al., 2018) confusion matrix contains true positive, true negative, false positive, and false negative values. Then the values of accuracy are calculated using the formulas in Google Colab without PSO. After optimizing PSO, the resulting accuracy, precision, recall, and f-measure values compare with the accuracy, precision, recall, and f-measure values. The results of the evaluation of data classification with KNN and Forward Selection can be seen in Table 2.

Tuble 2. In the forward beleen in recuracy			
к	KNN accuracy	KNN + Forward Selection Accuracy	
2	0.6153	0.8461	
3	0.6923	0.7692	
4	0.6923	0.7692	
5	0.6923	0.7692	
6	0.6923	0.7692	
7	0.6923	0.7692	
8	0.6923	0.7692	

Table 2 KNN+Forward Selection Accuracy

Source: Google Collab Results, 2022

The results of the calculation of KNN+PSO are the same as being evaluated for KNN+Forward Selection based on the confusion matrix. The results of the evaluation of data classification with KNN and PSO can be seen in Table 3.

Table	3.	KNN+PSO	Accuracy
-------	----	----------------	----------

к	KNN accuracy	KNN + PSO accuracy
2	0.6153	0.7692
3	0.6923	0.6923
4	0.6923	0.7692
5	0.6923	0.7692
6	0.6923	0.8461
7	0.6923	0.8461
8	0.6923	0.8461

Source: Google Collab Results, 2022

IV. DISCUSSION

From the test results, the best accuracy using KNN+PSO is 0.84 or 84% at K = 6 and the best using KNN+Forward Selection is 0.84 or 84% at K = 2, meaning that the model can predict well whether students can memorize the Our'an. The test results also found in KNN+ PSO that the most influential features in the process of memorizing the Our'an, namely reading and writing the Qur'an, activeness in doing sunnah practices; implementation and evaluation when the memorization process was carried out. While on KNN+Forward Selection activeness in doing sunnah practices; implementation and evaluation; motivation become an influential features However, the results of this test still cannot be used directly to measure the level of memorization of the Qur'an because there are still several factors that have not been measured, such as internal factors such as personality, interests, talents, and memory abilities. External factors include the method, motivation, fluency in reading and writing, routine practice, time, facilities, and place. Thus, this study still requires further research to better measure the level of memorizing the Qur'an. However fourth features, namely reading and writing the Qur'an, activeness in doing sunnah practices; implementation & evaluation; and motivation at least, results of this research can be used as a consideration in measuring the level of memorization of the Qur'an.

V. CONCLUSIONS AND SUGGESTIONS

From the results of this study, it can be concluded that the factors that influence the process of memorizing the Qur'an are reading and writing the Qur'an, activeness in doing sunnah practices, and implementation and evaluation when the memorization process is carried out. Thus, this research can be used to measure the level of memorizing the Qur'an. The authors want to acknowledge this research as supported by the University of Amikom Yogyakarta and declare no conflict of interest.

REFERENCES

- [1]. Bharath, P., Ramachandran, M., Ramu, K., & Chinnasamy, S. (2022). A Study on Various Particle Swarm Optimization Techniques used in Current Scenario. Design, Modelling and Fabrication of Advanced Robots, 1(1). https://doi.org/10.46632/dmfar/1/1/3.
- [2]. Dewi Kurniawati. (2022). IMPLEMENTASI PROGRAM MENGHAFAL AL-QUR'AN JUZ 30 DI SMP WAHID HASYIM MALANG. 1(April), 201–212.
- [3]. Fitriyah, D. (2008). Faktor yang mempengaruhi kecepatan menghafal Al-Qur'an antara santri mukim dan nonmukim di Pesantren Zaidatul Ma'arif Kauffman Parakan Temanggung. Skripsi. Semarang: IAIN Walisongo. http://eprints.walisongo.ac.id/11330/1/3101100_Darli
- m_Skripsi_Lengkap.pdf.
 [4]. Gunawan, T. S., Ashraf, A., Riza, B. S., Haryanto, E. V., Rosnelly, R., Kartiwi, M., & Janin, Z. (2020). Development of video-based emotion recognition using deep learning with Google Colab. Telkomnika (Telecommunication Computing Electronics and Control), 18(5), 2463–2471.
- https://doi.org/10.12928/TELKOMNIKA.v18i5.16717.
 [5]. Hasanah, M. (2021). The Role of Parents in Children Memorizing the Qur'an in Middle School Based on the Amanatul Ummah Islamic Boarding School. *Tafkir: Interdisciplinary Journal of Islamic Education*, 2(2), 139–156. https://doi.org/10.31538/tijie.v2i2.43.
- [6]. Kemampuan, H., Al Qur'an, M., Motivasi, D., Dengan, B., Belajar, H., Siswa, P., Aliyah, M., Fathimiyah Banjarwati, A. L., & Lamongan, P. (2018). Hubungan Kemampuan Menghafal Al Qur'an dan Motivasi Belajar Dengan Hasil Belajar PAI Siswa Madrasah Aliyah Al Fathimiyah Banjarwati Paciran Lamongan. Darajat: Jurnal Pendidikan Agama Islam, 1(1), 78 94–78 94. http://ejournal.iaitabah.ac.id/index.php/Darajat/article/view/284.
- [7]. Marlina, M. (2018). Teori Behavior dalam Meningkatkan Motivasi Hafalan Surat Pendek Al-Quran. Irsyad: Jurnal Bimbingan, Penyuluhan, Konseling, Dan Psikoterapi Islam, 6(4), 407– 426.<u>https://doi.org/10.15575/irsyad.v6i4.907.</u>
- [8]. Miranda, E., Mutiara, A. B., Ernastuti, & Wibowo, W. C. (2018). Classification of Land Cover from Sentinel-2 Imagery Using Supervised Classification Technique (Preliminary Study). Proceedings of 2018 International

Conference on Information Management and Technology, ICIMTech 2018, September, 69–74. https://doi.org/10.110 9/ICIMTech.2018.8528122.

- [9]. Muliati, I., & Oktavia, G. (2022). Ten Students' Motivation in Memorizing Quran A Case Study at Rumah Quran in Padang Indonesia. 1(1), 1–13.
- [10]. Nanja, M., & Purwanto, P. (2015). Metode K-Nearest Neighbor Berbasis Forward Selection Untuk Prediksi Harga Komoditi Lada. Pseudocode, 2(1), 53–64. https://doi.org/10.33369/pseudocode.2.1.53-64.
- [11]. Oktarina, D., & Yanti, Y. C. M. (2021). PENGARUH KOMUNIKASI INTERPERSONAL GURU DAN SANTRI TERHADAP MINAT HAFALAN AL-QUR'AN (Studi Pada Santri Putri Takhasus Pondok Pesantren Darul Huffaz Kecamatan Gedong Tataan Kabupaten Pesawaran Tahun 2020). INTERCODE – Jurnal Ilmu Komunikasi Vol. 1 No. 1 (Maret, 2021), Hal. 49-62 ISSN, 1(1), 35–48.
- [12]. Okwu, M. O., & Tartibu, L. K. (2021). Particle Swarm Optimisation. Studies in Computational Intelligence, 927, 5–13. https://doi.org/10.1007/978-3-030-61111-8_2.
- [13]. Saputra, W. R., Hendri, M., & Aminoto, T. (2019). Korelasi Motivasi Dan Hasil Belajar Ipa Siswa Kelas Viii Di Smp Negeri Se-Kecamatan Jambi Selatan. EduFisika, 4(01), 36–45. https://doi.org/10.22437/edufisika.v4i01. 3996.
- [14]. Simanjuntak, D. (2021). Faktor-faktor pendukung dan penghambat dalam menghafal Al-qur'an. 2, 92–101.
- [15]. Wijaya, C., Siregar, M. F. S., Ruslan, M., Holid, S., & Roslaeni, R. (2021). Manajemen Pembelajaran Tahfiz dalam Peningkatan Minat Menghafal Alquran Siswa di Yayasan Tahfidzul Quran Al-Fawwaz Medan. In Edukasi Islami: Jurnal Pendidikan Islam (Vol. 10, Issue 01). https://doi.org/10.30868/ei.v10i01.1214.