M A T H L I N E
JURNAL MATEMATIKA DAN PENDIDIKAN MATEMATIKA

ISSN 2502-5872 (Print)
ISSN 2622-3627 (Elektronik)

Volume 11 Number 2, May 2026, 423-436

CATEGORIZATION OF GRADE 9 STUDENTS' ACHIEVEMENT BASED ON SCHOOL EXAMINATION RESULTS AND QUR'ANIC ACHIEVEMENT USING K-MEANS AND K-NEAREST NEIGHBORS

Diki Mulyana[1], Dadan Dasari[2*], Nurjanah[3], Luthfiyati Nurafifah[4]
[1,2,3] Departement of Mathematics Education, Universitas Pendidikan Indonesia, West Java, Indonesia
[4] Departement of Mathematics Education, Universitas Wiralodra, West Java, Indonesia

*Correspondence: dadandasari@upi.edu

ABSTRACT

Current achievement assessment in Integrated Islamic Schools primarily emphasizes academic performance, with limited recognition of spiritual achievement, thereby constraining comprehensive student evaluation. Therefore, this study proposes a student achievement categorization framework that integrates School Examination Scores and Qur'anic Achievement using a combination of K-Means and K-Nearest Neighbor (kNN) algorithms. This research employed a quantitative approach, including educational data mining, to collect data from 324 ninth-grade students at SMPIT during the 2020–2025 academic years. Data were obtained from academic records and data extraction instruments. The K-Means algorithm in educational data mining served as the primary method for clustering students based on academic and spiritual characteristics, while the resulting clusters were used as target classes for the kNN classification algorithm. Model performance was evaluated using a confusion matrix based on accuracy, precision, and recall metrics. The findings demonstrate that K-Means successfully generated distinct achievement clusters reflecting the diversity of student performance, while the kNN model achieved high performance in consistently predicting the categories of Passed and Conditionally Passed. These results indicate that the integrated K-Means and kNN model is effective in establishing data-driven student achievement categories. From a broader perspective, this study contributes to helping Integrated Islamic Schools develop a student achievement evaluation model that encompasses both academic and spiritual dimensions.

Keywords: K-Means; K-Nearest Neighbors (kNN); Qur'anic Achievement; Student Achievement

How to Cite: Mulyana, D., Dasari, D., Nurjanah, N., & Nurafifah, L. (2026). Categorization of Grade 9 Students' Achievement Based on School Examination Results and Qur'anic Achievement Using K-Means and K-Nearest Neighbors. *Mathline: Jurnal Matematika dan Pendidikan Matematika, 11*(2), 409-422. http://doi.org/10.31943/mathline.v11i2.1146

PRELIMINARY

Student learning achievement evaluation at the junior secondary level can no longer be adequately viewed solely through a single dimension of academic cognitive performance. In Integrated Islamic Schools, academic achievement, represented by School Examination Scores, operates alongside character and spiritual achievement, one of which is reflected in Qur'anic Achievement (including memorization, *tahsin*, and consistency in *muraja'ah*). These two indicators represent a more comprehensive profile of student competence, aligned with the paradigm of holistic 21st-century education (Al-Qoyyim et al., 2024). During the 2019–2025 academic years, data from ninth-grade students at SMPIT revealed variations in achievement patterns: some students demonstrated consistently high performance in both dimensions, some excelled academically but showed moderate Qur'anic achievement, while others exhibited the opposite pattern. These variations require an analytical approach capable of objectively clustering and classifying students based on data. In this context, the application of K-Means as an unsupervised learning method and K-Nearest Neighbors (kNN) as a supervised learning method is both relevant and strategic. More broadly, the integration of machine learning techniques in education, particularly within educational data mining, offers opportunities for data-driven decision-making in determining student development programs, academic interventions, and strengthening character education (Ng, 2016).

Despite its significant potential, the application of K-Means and kNN algorithms in educational data contexts faces several challenges. First, the heterogeneity of educational data, which combines numerical academic scores and non-academic indicators, may introduce distance bias and cluster instability. K-Means, for instance, is highly sensitive to centroid initialization and the choice of distance metrics. (Ikotun et al., 2023; Jin & Han, 2017). Second, as a distance-based classification algorithm, kNN faces challenges related to computational efficiency and sensitivity to data distribution, particularly as the number of attributes and historical data increases (El Morr et al., 2022; Syriopoulos et al., 2025). Another challenge lies in determining the optimal value of k and ensuring feature relevance in the classification decision process (Qiu et al., 2022). In the context of Integrated Islamic education, an additional challenge arises from the limited number of studies that integrate spiritual-religious achievement as a primary analytical feature, rather than merely as a supporting variable.

Research on K-Means and kNN has developed rapidly across various domains. K-Means, in particular, has been extensively examined in terms of efficiency, centroid optimization, and robustness against noisy data (Beecks et al., 2022; Jiang et al., 2025; Yang & Zhu, 2019). Various variants, such as standard K-Means, grid-based K-Means, and parallel K-Means, have been proposed to improve clustering stability and computational efficiency (James Manoharan & Hari Ganesh, 2016; Krohn & Karlsson, 2016). On the other hand, kNN has continued to evolve through the optimization of distance metrics, feature selection, as well as hybrid and parallel approaches. (Lei et al., 2023; Lubis & Lubis, 2020; Miriyala et al., 2024). Even advanced approaches such as quantum kNN indicate a research direction toward higher-level computational efficiency (Gong et al., 2024; Li et al., 2022). However, most of these studies focus on text data, image data, or synthetic datasets, with very limited research examining educational data that simultaneously incorporates academic scores and religious achievement, particularly at the junior secondary level.

K-Means excels in its ability to uncover latent structures in unlabeled data, making it suitable for the initial exploration of student achievement patterns. However, its limitations lie in its assumption of convex cluster shapes and its sensitivity to outliers (Chakraborty et al., 2020; Ikotun et al., 2023). In contrast, kNN offers high flexibility in classification because it does not require an explicit training process; however, its performance depends heavily on the quality of the training data and the selection of parameters (Amorim et al., 2018; Muller, 2023). An important lesson from previous studies is that combining clustering and classification approaches has the potential to produce a more robust categorization system than relying on either method alone.

Based on this review, several significant research gaps can be identified. First, studies integrating K-Means and kNN sequentially to develop a comprehensive student achievement categorization framework remain limited. Second, there is almost no research that treats Qur'anic achievement as a primary variable in educational machine learning analysis, particularly in Integrated Islamic Schools. Third, few longitudinal studies have utilized multi-year historical data (2019–2025) to map the consistency and dynamics of ninth-grade student achievement. Therefore, this study seeks to address these gaps by developing a student achievement categorization model based on School Examination Scores and Qur'anic Achievement using K-Means and K-Nearest Neighbors, as a foundation for more equitable, objective, and student-centered educational decision-making.

This study contributes to the field of education in three main ways. First, theoretically, it develops a student achievement assessment framework by proposing an extended evaluation model that incorporates both academic and spiritual dimensions. Second, methodologically, this study presents a novel approach that adapts the use of unsupervised learning (K-Means) and supervised learning (kNN) within the context of educational data mining, particularly for student achievement categorization models, an area that remains underexplored in Integrated Islamic Schools. Third, practically, the developed model offers schools a tool to assess student achievement, design more focused and productive intervention programs, and implement more effective data-driven decision-making processes that can enhance educational quality.

METHODS

This study applies an Educational Data Mining (EDM) approach using Orange Data Mining software as the primary tool for data analysis. Orange Data Mining was selected because it provides a structured, reproducible visual environment suitable for implementing K-Means and K-Nearest Neighbors (kNN) algorithms in the context of educational data. Student achievement data from ninth-grade students at SMPIT during the 2019–2025 period, consisting of School Examination Scores and Qur'anic Achievement from 324 students, were first imported into Orange through the File widget. The data were then verified to be in numerical and consistent format before further analysis. The preprocessing stage was conducted using the Select Columns widget to define the main research attributes, followed by the Normalize widget with the Min–Max normalization method to standardize attribute scales. This step is important because both K-Means and kNN in Orange operate based on Euclidean distance calculations, meaning that differences in attribute scales may affect the analysis results.

The student clustering process was conducted using the K-Means widget in Orange Data Mining. The number of clusters (k) was conceptually determined based on the need for student achievement categorization, namely the formation of high, medium, and low achievement groups. Orange automatically performed centroid initialization and iterative cluster updates until convergence was reached. The clustering results were visualized using the Scatter Plot and Silhouette Plot widgets to evaluate cluster separation. The cluster labels generated by K-Means were then treated as student achievement categories, representing the simultaneous combination of academic achievement and Qur'anic achievement.

The cluster labels generated by K-Means were subsequently used as pseudo-labels in the classification stage. This process was implemented by connecting the K-Means widget to the Data Table and Select Columns widgets in Orange Data Mining, allowing the cluster labels to be assigned as the class attribute. This approach enables the use of unsupervised learning results as the foundation for supervised learning, thereby strengthening categorization consistency. The classification stage was conducted using the kNN widget in Orange. The dataset containing cluster labels was divided into training and testing sets using the Data Sampler widget. The value of k in kNN was determined through iterative testing to obtain the best classification performance. The kNN algorithm classified each test instance based on the majority class among its k nearest neighbors using Euclidean distance. This approach enables the system to classify new students into the previously established achievement categories.

The classification model performance was evaluated using the Test & Score widget in Orange Data Mining. The evaluation metrics included accuracy, precision, recall, and F1-score. This evaluation aimed to ensure that the developed model is not only computationally accurate but also reliable for application in educational decision-making contexts..

RESULT AND DISCUSSION

1. Clustering K-Means

The clustering analysis workflow of student achievement was implemented using Orange Data Mining, beginning from data loading to the storage of clustering results. This workflow represents a systematic and reproducible analytical approach for examining achievement patterns of ninth-grade students based on School Examination Scores and Qur'anic Achievement. The initial stage begins with the File widget, which functions to import the student achievement dataset into the Orange environment. This dataset includes the main numerical attributes used in the analysis, ensuring that subsequent processes operate on structured data. The loaded data can then be reviewed through the Data Table widget, which enables researchers to conduct an initial examination of data completeness and consistency before further analysis. This stage plays an important role in maintaining the validity of the clustering results.

Next, the data were directed to the K-Means widget, which served as the core of the clustering process. At this stage, the K-Means algorithm grouped students into a number of clusters based on similarities in academic scores and Qur'anic achievement. The iterative K-Means process generated centroids representing the average characteristics of each cluster, enabling the latent structure of student achievement to be identified objectively. This stage marked a shift from individual-based achievement evaluation toward pattern-based group analysis. The clustering results were then visualized through the Scatter Plot widget, which provided a visual representation of student distribution and cluster separation. This visualization facilitated the interpretation of clustering results and served as an initial means of evaluating cluster separation quality. In the discussion context, the scatter plot helps explain that student achievement is not distributed homogeneously, but rather forms specific patterns that can be associated with profiles of academic and Qur'anic achievement.

The generated data were also presented in tabular form through the Clustering Table widget in Orange Data Mining. This data displayed information for each student along with the assigned cluster labels, allowing more detailed examination at both the individual and group levels. This stage is particularly important in educational contexts, as the clustering results can be directly linked to specific student development needs and instructional intervention strategies. The final stage of this workflow was the Save Data widget, which was used to store the clustered dataset. This data storage enables the clustering results to be reused in the classification stage using kNN, as well as to serve as a basis for educational decision-making at the school level. Thus, this workflow is not merely exploratory, but also oriented toward the practical utilization of analytical results.

Overall, the analytical workflow in Orange Data Mining illustrated in the figure confirms that the student achievement clustering process was conducted through transparent, structured, and replicable stages. From an educational perspective, this approach supports the implementation of data-driven decision-making, in which instructional policies and student development programs are based on systematically identified empirical patterns rather than solely on subjective judgment.

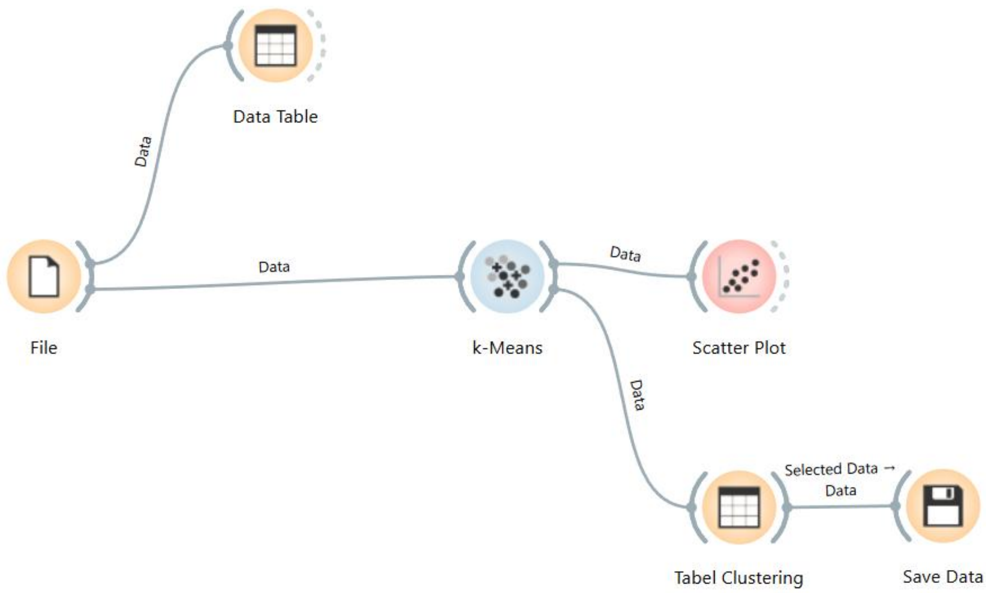

Figure 1. K-Means Workflow in Orange Data Mining

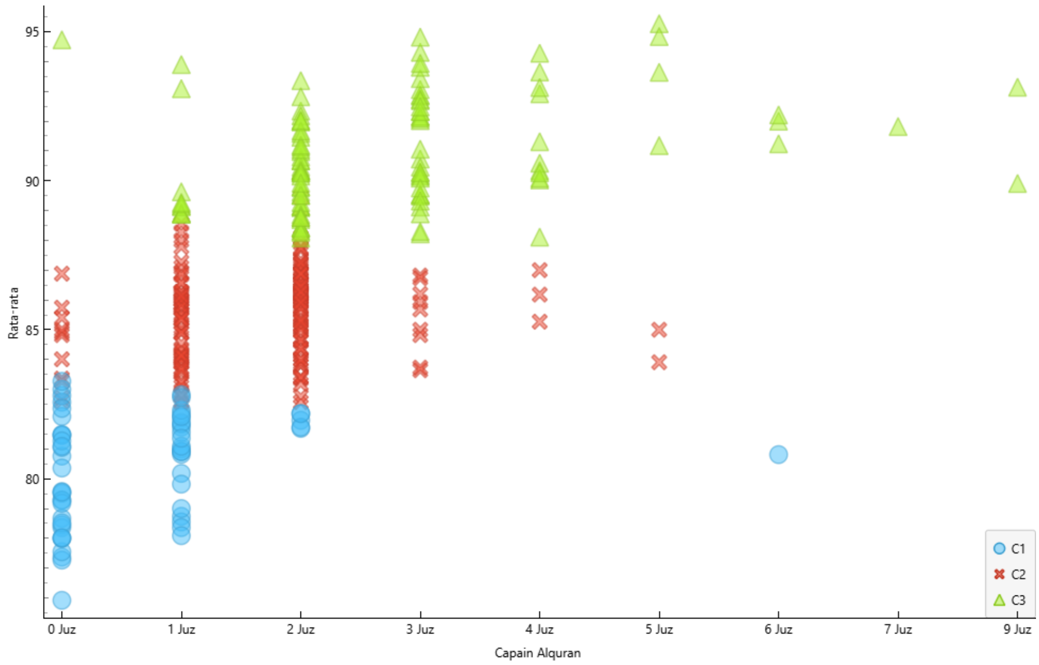

Figure 2. Visualization of Student Achievement Clusters Using a K-Means Scatter Plot

The figure presents a visualization of the clustering results for ninth-grade student achievement based on two main variables: Qur'anic Achievement on the horizontal axis and average School Examination Scores on the vertical axis. Each point represents one student, while differences in color and symbol indicate the three clusters (C1, C2, and C3) generated by the K-Means algorithm.

Cluster C1 (indicated by blue circles) is dominated by students with low Qur'anic achievement (0–2 juz) and relatively low to moderate average school examination scores. The data distribution in this cluster suggests that improvements in Qur'anic achievement have not been fully accompanied by significant increases in academic performance. This cluster may be interpreted as a group of students requiring intensive support, both in academic development and in strengthening Qur'anic learning.

Cluster C2 (indicated by red crosses) represents students with low to moderate Qur'anic achievement (0–5 juz), but with average school examination scores in the moderate to high category. This pattern indicates a group of students demonstrating relatively stable academic performance, although their Qur'anic achievement has not yet reached an optimal level. Pedagogically, this group reflects an imbalance between academic performance and spiritual achievement, thus requiring more integrated intervention strategies.

Cluster C3 (indicated by green triangles) represents students with moderate to high Qur'anic achievement (≥2 juz to 9 juz) and high average school examination scores. This cluster is relatively positioned in the upper region of the graph, indicating consistency between academic and Qur'anic achievement. It can be categorized as a high-achievement group with potential for further development through enrichment programs and advanced mentoring.

Overall, this visualization shows a positive relationship between Qur'anic achievement and academic performance, although the relationship is not perfectly linear. These findings confirm that student achievement in Integrated Islamic Schools is multidimensional, and that data-driven clustering through K-Means can reveal variations in student achievement profiles more objectively than one-dimensional assessment approaches.

2. K-Nearest Neighbors (kNN) Classification

The student achievement classification workflow was developed using Orange Data Mining as a continuation of the previous clustering stage. This workflow illustrates how structured student achievement data were systematically processed to produce a classification model performance evaluation. The process began with the File widget, which was used to load the dataset resulting from the student achievement clustering stage. This dataset contained the main attributes, namely School Examination Scores, Qur'anic Achievement, and the cluster labels obtained from the previous K-Means stage. Next, the Select Columns widget was used to define the role of each attribute, namely predictor attributes and the class attribute. This stage is crucial to ensure that the kNN model only uses relevant features in the classification process.

The selected data were then directed to the Data Sampler widget, which functioned to divide the dataset into training data and testing data. This partitioning enabled an objective evaluation of the kNN model's ability to classify previously unseen data. A portion of the training data was also displayed through the Data Table widget for inspection and verification purposes. In the modeling pathway, the training data were processed using the Preprocess widget, which performed normalization and data adjustment. This preprocessing stage is important because the kNN algorithm is highly sensitive to data scale, given that classification is based on distance calculations among data points. Normalization ensures that the contribution of each attribute to distance computation is proportional.

As shown in Figure 3, the kNN widget served as the core of the classification process. At this stage, the kNN model was built using the preprocessed training data. The kNN algorithm classified each test instance based on its proximity to a number of nearest neighbors (k) from the training set. This approach enables the system to classify students into predefined achievement categories in an adaptive and data-driven manner. The prediction results from the kNN model were then directed to the Predictions widget, which displayed the classification results for each test instance. To evaluate model performance, these predictions were compared with the actual labels through the Confusion Matrix widget. This confusion matrix presents the distribution of correct and incorrect predictions across achievement categories, providing a clear representation of classification accuracy and errors.

From a discussion perspective, this workflow demonstrates that the kNN model can utilize clustering-generated labels as a consistent basis for classification. The integration of Data Sampler, Preprocess, kNN, and Confusion Matrix confirms that the classification process was conducted in a controlled and measurable manner. This strengthens the validity of using kNN as a predictive tool for student achievement categorization in subsequent academic years.

Pedagogically, this workflow has important implications. Schools not only obtain static student achievement categories, but also acquire a predictive model that can be used to classify new students based on historical patterns. Thus, this approach supports the implementation of data-driven assessment and assists teachers and school administrators in designing more targeted instructional interventions.

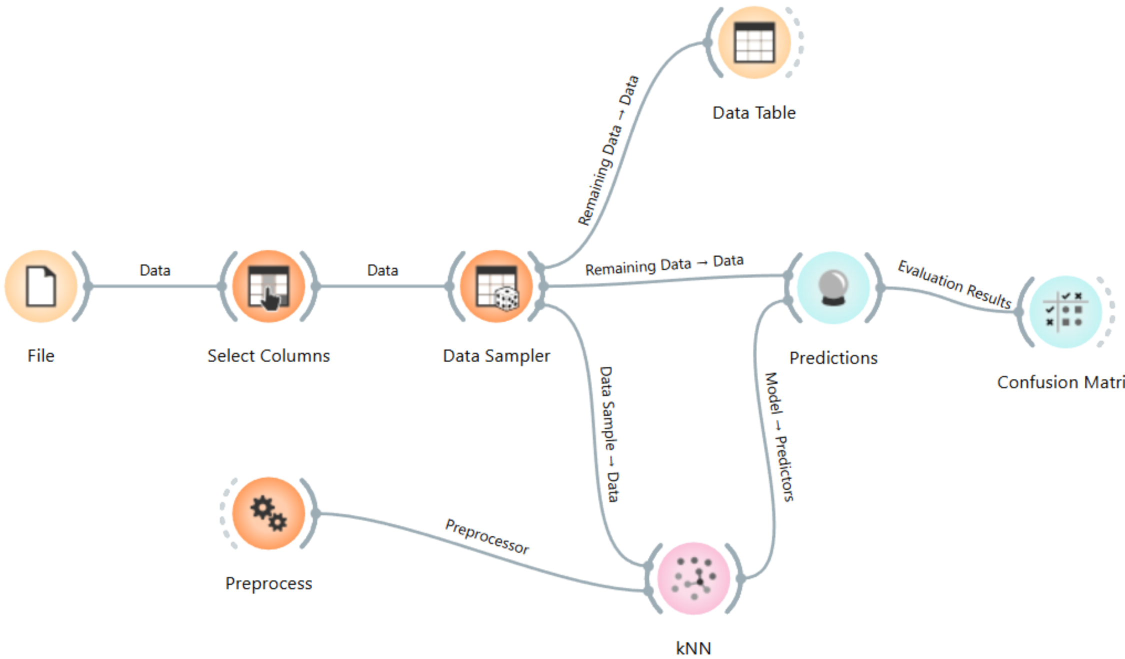

Figure 3. Workflow K-Nearest Neighbors (kNN) Orangе Data Mining

Tabel 1. Confusion Matrix K-Nearest Neighbors (kNN) Orange Data Mining

		Predicted		
		Pass	*Conditional Pass*	Σ
Actual	*Pass*	87	0	87
	Conditional Pass	0	10	10
	Σ	87	10	97

The results of the classification evaluation using a Confusion Matrix with the K-Nearest Neighbors (kNN) model are presented in Table 1, based on the output generated by Orange Data Mining. The confusion matrix compares actual classes and predicted classes across two student achievement categories, namely Pass and Conditional Pass. Based on the evaluation results, out of a total of 97 student records tested, the kNN model successfully classified all data with perfect accuracy. All 87 students who were actually in the Pass category were correctly predicted as Pass, with no classification errors. Likewise, all 10 students in the Conditional Pass category were accurately predicted by the model, with no misclassification into the Pass category.

The absence of false positives and false negatives in the confusion matrix indicates that the kNN model achieved 100% precision and 100% recall for both categories. Thus, the model not only demonstrated excellent overall accuracy, but also showed consistency in distinguishing students who genuinely met the criteria for Pass from those requiring special attention through Conditional Pass status.

The findings from the confusion matrix demonstrate that the student achievement pattern configuration obtained through K-Means clustering is sufficiently concrete and independent to be effectively learned by the kNN algorithm. These results support the similarity-based classification theory proposed by Zhang (2022) which states that kNN performs optimally when the dataset exhibits strong class separability. Accordingly, School Examination Scores and Qur'anic Achievement can be considered robust independent features for categorizing student graduation status. This study is consistent with the findings of Rahma & Ulfah (2025), which show that both clustering and classification approaches can successfully identify patterns in educational systems and support improved data-driven academic management. Furthermore, the results align with the holistic education assessment framework proposed by Benjamin Bloom, which emphasizes that student achievement evaluation should extend beyond cognitive dimensions to include affective and character-related aspects (Gable, 1986).

On the other hand, classification accuracy analysis particularly when extremely high must be carefully interpreted, as it may indicate potential overfitting in analytical models. In this study, however, such indications are justified by the well-formed clustering structure and strong class separation. From an educational perspective, the proposed model is valuable as an objective criterion for determining student graduation status, identifying Conditional Pass students, and supporting contextual validity of model assumptions, classification decisions, and data-driven management techniques.

CONCLUSION

This study developed a student achievement categorization model for ninth-grade students based on School Examination Scores and Qur'anic Achievement by integrating K-Means and K-Nearest Neighbors (kNN) using Orange Data Mining. The results show that educational data mining can effectively identify student achievement patterns in a clear and multidimensional way. The K-Means clustering successfully formed meaningful achievement groups that represent both academic performance and Qur'anic achievement, confirming that student achievement in Integrated Islamic Schools is not one-dimensional, but a combination of cognitive skills and spiritual development. The integration of clustering results into the kNN classification stage produced very strong model performance, as shown by the confusion matrix with high accuracy, precision, and recall in classifying student status (Pass and Conditional Pass), indicating that K-Means cluster labels are consistent and suitable for supervised learning. This model provides practical benefits for schools by supporting more objective, data-based decision-making in determining student achievement categories and designing targeted academic and character development programs, while also showing that educational data analysis can be applied in practice without complex programming.

However, this study has limitations in terms of the number of variables and the data scope, which is limited to one school. Future research is recommended to include additional variables such as learning motivation, attendance, behavior, and socio-emotional factors to build a more comprehensive model. It is also suggested that future studies test the model in different school contexts, compare K-Means and kNN with other algorithms such as Random Forest, Support Vector Machine, and deep learning methods, and validate the model using different datasets and longitudinal data to ensure stronger generalization and reduce potential bias. In addition, this study can serve as a foundation for developing early predictive systems that help schools identify students who need academic intervention at an earlier stage.

ACKNOWLEDGMENT

The author would like to express sincere appreciation and gratitude to SMPIT Bina Ummah for granting permission and providing support, particularly in supplying the academic achievement data and Qur'anic achievement data used in this study. The cooperation and assistance from the school management, teachers, and administrative staff greatly contributed to the smooth data collection process and the completion of this research. The commitment of SMPIT Bina Ummah to the development of data driven education is highly appreciated.

REFERENCES

Al-Qoyyim, T. M., Affan, A., Maulidiya Nabila, S., Muhaini, M., Hajariah, S., Suri, A. A., & Buntana, B. (2024). The Relationship Between Memorizing the Qur'an and Students' Academic Achievement: Case Study of Science Subjects at Tahfidzul Quran An-Nahl Mataram Integrated Islamic Elementary School. *Jurnal Pendidikan, Sains, Geologi, Dan Geofisika (GeoScienceEd Journal)*, *5*(3), 483–488. https://doi.org/10.29303/goescienceed.v5i3.357

Amorim, L. A., Freitas, M. F., Da Silva, P. H., & Martins, W. S. (2018). A Fast Similarity Search kNN for Textual Datasets. In de P. B. C., de S. P.S.L., de B. C.O.F., & F. R.A.C. (Eds.), *Proceedings - 2018 Symposium on High-Performance Computing Systems, WSCAD 2018* (pp. 229–236). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/WSCAD.2018.00043

Beecks, C., Berns, F., Hüwel, J. D., Linxen, A., Schlake, G. S., & Düsterhus, T. (2022). A Comparative Performance Analysis of Fast K-Means Clustering Algorithms. In P. E., D. H. P., K. I., & K. G. (Eds.), *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 13635 LNCS* (pp. 119–125). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-031-21047-1_11

Chakraborty, A., Faujdar, N., Punhani, A., & Saraswat, S. (2020). Comparative study of K-means clustering using iris data set for various distances. *Proceedings of the Confluence 2020 - 10th International Conference on Cloud Computing, Data Science and Engineering*, 332–335. https://doi.org/10.1109/Confluence47617.2020.9058328

El Morr, C., Jammal, M., Ali-Hassan, H., & El-Hallak, W. (2022). K-Nearest Neighbors. In *International Series in Operations Research and Management Science* (Vol. 334, pp. 301–318). Springer. https://doi.org/10.1007/978-3-031-16990-8_10

Gable, R. K. (1986). Affective Characteristics: Their Conceptual Definitions. *Instrument Development in the Affective Domain*, 1–13. https://doi.org/10.1007/978-94-015-7259-0_1

Gong, L.-H., Ding, W., Li, Z., Wang, Y.-Z., & Zhou, N.-R. (2024). Quantum K-Nearest Neighbor Classification Algorithm via a Divide-and-Conquer Strategy. *Advanced Quantum Technologies*, *7*(6). https://doi.org/10.1002/qute.202300221

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, *622*, 178–210. https://doi.org/10.1016/j.ins.2022.11.139

James Manoharan, J., & Hari Ganesh, S. (2016). Initialization of optimized K-means centroids using divide-and-conquer method. *ARPN Journal of Engineering and Applied Sciences*, *11*(2), 1086–1091. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84957888287&partnerID=40&md5=bcac5f2f8f3aeadd69e1d199ecf17632

Jiang, P., Cao, J., Yu, W., & Nie, F. (2025). A robust entropy regularized K-means clustering algorithm for processing noise in datasets. *Neural Computing and Applications*, *37*(9), 6617–6632. https://doi.org/10.1007/s00521-024-10899-4

Jin, X., & Han, J. (2017). K-Means Clustering. In *Encyclopedia of Machine Learning and Data Mining* (pp. 695–697). Springer Science+Business Media. https://doi.org/10.1007/978-1-4899-7687-1_431

Krohn, R., & Karlsson, C. (2016). Parallel K-means clustering with triangle inequality. In G. T. & H. G. (Eds.), *29th International Conference on Computer Applications in Industry and Engineering, CAINE 2016* (pp. 191–196). The International Society for Computers and Their Applications (ISCA). https://www.scopus.com/inward/record.uri?eid=2-s2.0-84991465779&partnerID=40&md5=82ccab9e9b4bf4f7d8079a1a1e58fa34

Lei, X., Yang, S., & Zhou, L. (2023). Extensions of Local-Mean K-Nearest Neighbor Classifier with Various Distance Metrics. *2023 IEEE 5th International Conference on Power, Intelligent Computing and Systems, ICPICS 2023*, 902–908. https://doi.org/10.1109/ICPICS58376.2023.10235730

Li, J., Lin, S., Yu, K., & Guo, G. (2022). Quantum K-nearest neighbor classification algorithm based on Hamming distance. *Quantum Information Processing*, *21*(1). https://doi.org/10.1007/s11128-021-03361-0

Lubis, A. R., & Lubis, M. (2020). Optimization of distance formula in k-nearest neighbor method. *Bulletin of Electrical Engineering and Informatics*, *9*(1), 326–338. https://doi.org/10.11591/eei.v9i1.1464

Miriyala, S. D., Daram, A. S., & Halavar, B. (2024). Analysis of Parallel K-Limit Selection Sort Based K Nearest Neighbors for Image Classification. In A. A., B. A., H. T., S. P., & M. K. (Eds.), *Lecture Notes in Networks and Systems: Vol. 1051 LNNS* (pp. 409–418). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-031-64850-2_39

Muller, P. (2023). Flexible K Nearest Neighbors Classifier: Derivation and Application for Ion-mobility Spectrometry-based Indoor Localization. *Proceedings of the 2023 13th International Conference on Indoor Positioning and Indoor Navigation, IPIN 2023*. https://doi.org/10.1109/IPIN57070.2023.10332541

Ng, W. S. (2016). *Web Data Mining in Education*. *7*(6), 58–77. https://doi.org/10.4018/978-1-5225-1877-8.ch005

Qiu, Q., Li, M., Shen, S., Deng, S., & Guan, S. (2022). An Attribute Contribution-Based K-Nearest Neighbor Classifier. *Proceedings of 2022 6th Asian Conference on Artificial Intelligence Technology, ACAIT 2022*. https://doi.org/10.1109/ACAIT56212.2022.10137909

Rahma, F. A., & Ulfah, S. Z. (2025). Clustering Students Based on Academic Performance and Social Factors: An Unsupervised Learning Approach to Identify Student Patterns. *International Journal for Applied Information Management*, *5*(3), 139–154. https://ijaim.net/journal/index.php/ijaim/article/view/109

Syriopoulos, P. K., Kalampalikis, N. G., Kotsiantis, S. B., & Vrahatis, M. N. (2025). kNN

Classification: a review. *Annals of Mathematics and Artificial Intelligence*, *93*(1), 43–75. https://doi.org/10.1007/s10472-023-09882-x

Yang, Y., & Zhu, Z. (2019). A fast and efficient grid-based K-means++ clustering algorithm for large-scale datasets. In K. P., L. Y., P. J., & Z. H. (Eds.), *Advances in Intelligent Systems and Computing* (Vol. 891, pp. 508–515). Springer Verlag. https://doi.org/10.1007/978-3-030-03766-6_57

Zhang, S. (2022). Challenges in KNN Classification. *IEEE Transactions on Knowledge and Data Engineering*, *34*(10), 4663–4675. https://doi.org/10.1109/TKDE.2021.3049250