Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9760

Machine Learning-Based Classification of Quranic Verses Based on Islamic Pillars

Fauzan Azima[1*],Safwandi[2], Said Fadlan Anshari[3]
[1,2,3]Department of Informatics, Faculty of Engineering, Universitas Malikussaleh, Indonesia
Jl. Cot Tengku Nie, Reuleut Timur, Muara Batu, Aceh Utara, Aceh, Indonesia
[1*]fauzan.220170022@mhs.unimal.ac.id, [2]safwandi@unimal.ac.id, [3]saidfadlan@unimal.ac.id

open access

***Corresponding Author**

Article History:
Submitted: 08-10-2026
Accepted: 17-10-2026
Published: 01-11-2026

Keywords:
FastText; Islamic Pillars; Naive Bayes; Quranic Verses; Support Vector Machine; TF-IDF

Brilliance: Research of Artificial Intelligence is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

ABSTRACT

Background: Text classification is a machine learning process used to categorize data based on specific characteristics. This study focuses on classifying the content of Islamic Pillars in Quranic verses into six categories: Shahada, Salah, Zakat, Fasting, Hajj, and Non-Islamic Pillars.
Objective: This study seeks to develop a classification application while evaluating Naive Bayes and Support Vector Machine (SVM) algorithms with TF-IDF and FastText as feature representation methods.
Methods: The dataset consisted of 800 Quranic verses selected using purposive sampling and divided into 80% training data and 20% testing data. The research stages included data preparation, text preprocessing, feature representation, model training, testing, evaluation, and implementation of the best-performing model in an application. Model performance was evaluated using accuracy, macro precision, macro recall, and macro F1-score.
Results: The results showed that the SVM model using TF-IDF achieved the best performance, with an accuracy of 81.88%, macro precision of 92.55%, macro recall of 75.63%, and macro F1-score of 81.98%. Based on these results, SVM with TF-IDF was selected as the final classification model and implemented in the application to classify Quranic verses according to the content of the Islamic Pillars.
Conclusion: The results demonstrate that the combination of SVM and TF-IDF can be applied to classify Quranic verses into the defined Islamic Pillars categories.

INTRODUCTION

The development of artificial intelligence and machine learning has enabled automatic text classification in various domains, reducing the need for manual analysis of large collections of textual data. This development has also encouraged the use of machine learning and Natural Language Processing (NLP) for Quranic text analysis and classification (Akbar et al., 2024; Bashir et al., 2023; Muslikh et al., 2024). Quranic verses related to the Islamic Pillars are distributed across different chapters and are not explicitly grouped into Shahada, Salah, Zakat, Fasting, and Hajj. Identifying these verses manually requires reading and examining their meanings, which can be time-consuming when dealing with large collections of Quranic texts.

Natural Language Processing (NLP) provides methods for automatically analyzing and classifying textual data. Machine learning algorithms such as Naive Bayes and Support Vector Machine (SVM) have been applied to Quranic text classification with different thematic categories (Fauzan et al., 2022; Ningsih et al., 2024). Feature representation can also affect classification performance. TF-IDF has been applied to Indonesian Quranic text classification, while word embedding methods such as FastText have also been used for Quranic text classification (Akbar et al., 2024; Dewantara & Lhaksmana, 2024; Muslikh et al., 2024).

Previous studies have classified Indonesian Quranic translations into general thematic categories and investigated multi-label classification using machine learning and word embeddings (Akbar et al., 2024; Muslikh et al., 2024; Ningsih et al., 2024; Shabrina & Lhaksmana, 2024). Recent research has also examined the influence of preprocessing on Quranic text classification using SVM (Solihin et al., 2026). However, classification specifically based on the Islamic Pillars into Shahada, Salah, Zakat, Fasting, Hajj, and Non-Islamic Pillars remains limited. Therefore, this study compares Naive Bayes and SVM using TF-IDF and FastText to classify 800 purposively selected Quranic verses into six classes. The best-performing model is then implemented in a Flutter-based mobile application for Quranic text classification and prediction.

LITERATURE REVIEW

Recent research has increasingly examined the use of machine learning and natural language processing for identifying thematic information in Indonesian Quranic translations. Ningsih et al. (2024) evaluated Naive Bayes and Long Short-Term Memory (LSTM) for classifying Indonesian Quranic translations. Their approach involved text preprocessing, data balancing, and feature representation using Bag-of-Words and TF-IDF. The findings indicated that

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9760

class imbalance is an important factor that can influence the performance of classification models for Quranic text. In another study, Elfianita et al. (2022) compared Naive Bayes and Support Vector Machine (SVM) for classifying Indonesian Quranic translations into five themes: Faith, Morality, Worship, History, and Law. Using Quranic translation data from the Ministry of Religious Affairs of the Republic of Indonesia, they reported that Naive Bayes produced higher classification accuracy than SVM. This work is relevant to the present study because it also examines Naive Bayes and SVM for Indonesian Quranic text, although the target categories and experimental configurations differ.

Research has also investigated word embedding techniques for Quranic text classification. Muslikh et al. (2024) combined FastText with CNN and BiLSTM for multi-label classification of Indonesian Quranic translations. Their findings indicated an improvement in F1-score when FastText was incorporated into the classification process. Akbar et al. (2024) similarly explored Word2Vec and FastText with LSTM and BiLSTM for multi-label classification involving thematic categories such as Tawheed, Worship, Akhlaq, and History. The findings from these studies suggest that word embedding techniques can provide useful representations of Quranic text, particularly when a verse may contain more than one thematic aspect.

Feature representation has also received attention in previous Quranic text classification studies. Dewantara and Lhaksmana (2024) compared TF-IDF with term-weighting techniques based on word centrality in a multi-label classification task. Their experiments involved SVM, Random Forest, and AdaBoost, demonstrating that the selection of feature-weighting techniques can affect classification performance. Solihin et al. (2026) further investigated preprocessing variations by examining the influence of different stopword sets on Quranic text classification using SVM and Backpropagation Neural Network. Their study incorporated preprocessing, stopword removal, stemming, TF-IDF, and feature selection. The results showed that SVM provided better classification performance than the Backpropagation Neural Network across the evaluated settings, while variations in stopword selection also influenced the resulting performance. These findings reinforce the importance of preprocessing and feature representation in Quranic text classification.

Overall, previous studies have addressed Quranic text classification through thematic and multi-label categories using a range of conventional machine learning and deep learning methods. However, classification specifically targeting Shahada, Salah, Zakat, Fasting, Hajj, and Non-Islamic Pillars remains relatively limited. Furthermore, the comparative evaluation of Naive Bayes and SVM combined with both TF-IDF and FastText has not been sufficiently investigated for this particular classification task. Therefore, the present study evaluates these classifier and feature representation combinations to determine the most suitable approach for classifying Quranic verses according to the Islamic Pillars.

METHOD

Research Design

A quantitative experimental design was adopted to investigate the classification of Quranic verses according to the Islamic Pillars. The experiments compared two classification algorithms, namely Naive Bayes and Support Vector Machine (SVM), using TF-IDF and FastText as alternative feature representation techniques. Four experimental combinations were evaluated: Multinomial Naive Bayes with TF-IDF, Gaussian Naive Bayes with FastText, SVM with TF-IDF, and SVM with FastText. The overall research procedure consisted of data preparation, text preprocessing, feature representation, model development and testing, performance assessment, and deployment of the selected model in a mobile application. An overview of the research methodology is shown in Figure 1.

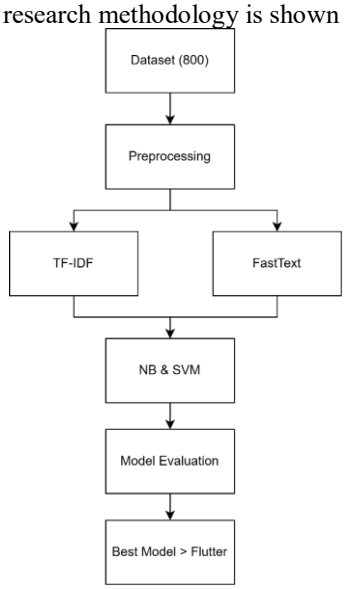


Figure 1. Research methodology

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, Number 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9760

Dataset and Data Preparation

A total of 800 Quranic verses were selected through purposive sampling for the experiments. The samples were drawn from a collection of 6,236 verses representing the official Indonesian Quran translation issued by the Ministry of Religious Affairs of the Republic of Indonesia (KEMENAG RI) (Kurnia et al., 2026). Each verse was categorized into six classes: Shahada, Salah, Zakat, Fasting, Hajj, and Non-Islamic Pillars. The dataset was divided into training and testing data using an 80:20 stratified split, resulting in 640 training instances and 160 testing instances. The split used a random state of 42.

The textual data consisted of Indonesian translations and short interpretations of the verses. The translation and interpretation texts were processed separately during feature representation. Stratified splitting was used to preserve the class distribution between the training and testing sets.

Text Preprocessing and Feature Representation

Text preprocessing was applied to the translation and interpretation texts before feature representation. The preprocessing included case folding, text cleaning, stopword removal, and stemming using the Sastrawi library. Similar preprocessing and text representation steps have also been applied in previous NLP research on Quranic text (Purwati et al., 2023; Puspasari & Rusmin, 2022). These preprocessing steps are commonly applied in text classification to reduce irrelevant textual variation and prepare the data for machine learning-based classification (Hassan et al., 2022). The same preprocessing procedure was applied to the training and testing data.

Two feature representation methods were evaluated. The first was TF-IDF, which was applied separately to the translation and interpretation texts. TF-IDF has also been applied to Quranic Tafseer text in previous text mining research (Ahmed et al., 2023). TF-IDF was employed to capture the relevance of terms by considering their occurrence within individual documents and throughout the document collection, following its use in previous Quranic text processing research (Puspasari & Rusmin, 2022). For each text source, the TF-IDF vectorizer was configured to generate unigram and bigram features, with the minimum document frequency set to two and lowercase conversion disabled. The vectorizers were trained exclusively on the training data before being used to transform the testing data. Subsequently, the TF-IDF feature matrices generated from the translation and interpretation texts were horizontally combined to produce a single feature representation for the classification models.

The second representation method was FastText. Separate FastText models were trained for the translation and interpretation texts using a vector size of 100, a context window of 5, a minimum word count of 1, skip-gram training, four workers, and 20 training epochs (Umer et al., 2022). For each verse, the available word vectors were averaged using mean pooling to produce a 100-dimensional sentence representation. This process was applied separately to the translation and interpretation texts.

Classification and Model Evaluation

The classification stage evaluated Multinomial Naive Bayes and Support Vector Machine (SVM) using the extracted text representations. Both SVM and Naive Bayes are established machine learning approaches for text classification and have been compared in previous text classification studies (Hassan et al., 2022). The experiments included Multinomial Naive Bayes with TF-IDF, Gaussian Naive Bayes with FastText, SVM with TF-IDF, and SVM with FastText.

The training data were used to train the classification models, while the testing data were used to evaluate their performance on previously unseen verses. The SVM model was implemented using LinearSVC with a random state of 42. For the TF-IDF-based SVM, sigmoid-based probability calibration was applied using five-fold cross-validation.

Model performance was assessed using accuracy, macro precision, macro recall, macro F1-score, and weighted F1-score. In addition, a confusion matrix was generated to assess the classification results for each of the six classes. Macro-averaged metrics were emphasized due to the imbalanced distribution of the dataset. These metrics provide a more balanced assessment of performance across individual classes and reduce the influence of the majority class on the overall evaluation (Opitz, 2024).

Model Selection and Application Implementation

The best-performing model was selected based on the evaluation results, with particular consideration given to the macro F1-score. The selected model was subsequently integrated into a mobile application developed using Flutter.

The application allows users to select a Quranic surah and verse as input. The selected verse is processed using the preprocessing and feature representation procedures applied during model development. The trained SVM model with TF-IDF then predicts the corresponding Islamic Pillar category, and the classification result is displayed to the user.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, Number 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9760

RESULT

This section presents the results of the Quranic verse classification experiments and the implementation of the selected model in a mobile application.

Dataset and Class Distribution

The classification experiments were conducted on 800 Quranic verses obtained through purposive sampling. The verses were assigned to six categories: Shahada, Salah, Zakat, Fasting, Hajj, and Non-Islamic Pillars. A stratified split with a ratio of 80:20 was applied, resulting in 640 instances for training and 160 instances for testing.

The dataset consisted of 301 verses in the Non-Islamic Pillars class, 301 in the Shahada class, 126 in the Salah class, 32 in the Hajj class, 31 in the Zakat class, and 9 in the Fasting class. The distribution indicates an imbalanced dataset, particularly in the Fasting class.

Feature Representation

TF-IDF was applied separately to the Indonesian translation and short interpretation texts. Each vectorizer used unigram and bigram features with a minimum document frequency of two. The resulting representations were horizontally concatenated, producing a combined feature representation with 8,047 features.

FastText was also applied separately to the translation and interpretation texts. Each model generated 100-dimensional word vectors, and the vectors within each verse were averaged to obtain a document representation.

Classification Performance

Four classification configurations were evaluated: Multinomial Naive Bayes with TF-IDF, Gaussian Naive Bayes with FastText, SVM with TF-IDF, and SVM with FastText. The evaluation was performed on 160 testing instances.

Table 1. Comparison of classification model performance

Model	Feature	Accuracy	Macro Precision	Macro Recall	Macro F1-score	Weighted F1-score
Multinomial Naive Bayes	TF-IDF	66.25%	37.99%	33.33%	33.22%	62.24%
Gaussian Naive Bayes	FastText	68.13%	49.79%	64.51%	52.18%	69.14%
SVM	FastText	80.00%	74.89%	57.53%	62.75%	79.26%
SVM	TF-IDF	81.88%	92.55%	75.63%	81.98%	82.03%

The SVM with TF-IDF achieved the highest performance, with an accuracy of 81.88% and a macro F1-score of 81.98%. It also obtained a macro precision of 92.55%, a macro recall of 75.63%, and a weighted F1-score of 82.03%. Considering the comparative results, the SVM model using TF-IDF was chosen as the classification model for implementation in the application.

Confusion Matrix

The confusion matrix of the selected SVM model with TF-IDF features is presented in Figure 2. It illustrates the distribution of correct and incorrect predictions across the six classes in the test dataset.

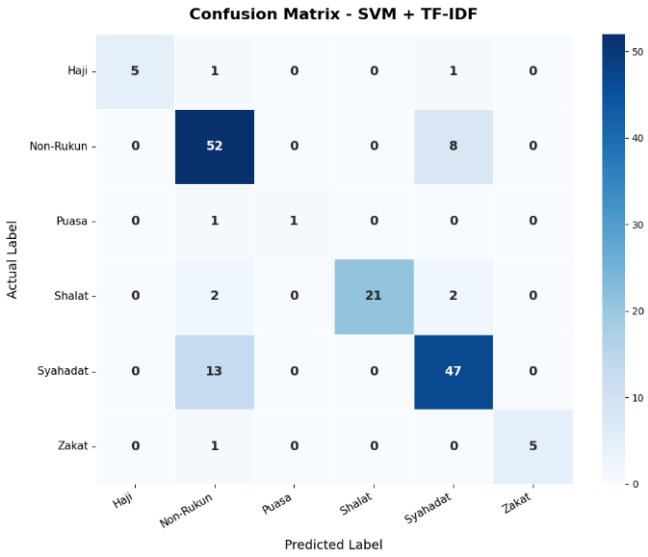


Figure 2. Confusion matrix of SVM with TF-IDF

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence
E-ISSN : 2807-9035
Volume 6, Number 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9760

The class labels in Figure 2 are presented in Indonesian. In this figure, Syahadat, Salat, Puasa, and Non-Rukun Islam correspond to Shahada, Salah, Fasting, and Non-Islamic Pillars, respectively, while Haji and Zakat correspond to Hajj and Zakat.

The model correctly classified 131 of the 160 test instances, achieving an accuracy of 81.88%. The Non-Rukun Islam class had 52 correct predictions out of 60 instances, whereas the Syahadat class had 47 correct predictions out of 60 instances. Misclassifications were particularly evident between these two classes, with eight Non-Rukun Islam instances incorrectly classified as Syahadat and 13 Syahadat instances misclassified as Non-Rukun Islam. The Salat class also showed strong classification performance, with 21 of 25 instances correctly classified. In contrast, the Puasa class had only one correct prediction out of two instances. This result should be interpreted cautiously because of the limited number of test samples. Overall, the confusion matrix indicates that the model performed well across most classes, although distinguishing between Non-Rukun Islam and Syahadat remains a challenge.

Application Implementation

The SVM model with TF-IDF features was integrated into a Flutter-based mobile application, with Flask serving as the backend for classification. Users can select a Quranic surah and verse, after which the system processes the input and displays the predicted Islamic Pillar category.



Figure 3. Mobile application interface for Quranic verse classification

Figure 3 illustrates the application's classification interface, where users select a surah and verse from the Quran. After the user submits the selection, the backend processes the input using the trained SVM model with TF-IDF features and returns the predicted category to the application. This implementation demonstrates how the trained model is integrated into the mobile application to classify Quranic verses according to the Islamic Pillars.

DISCUSSION

Among the evaluated configurations, SVM combined with TF-IDF produced the strongest classification results, achieving 81.88% accuracy and a macro F1-score of 81.98%. The model also obtained 92.55% macro precision, 75.63% macro recall, and 82.03% weighted F1-score. These findings indicate that SVM with TF-IDF was the most effective configuration for distinguishing the six Islamic Pillar categories in the dataset used in this study.

The findings are consistent with previous studies demonstrating the applicability of SVM to Quranic text classification. Fauzan et al. (2022) employed SVM to classify Indonesian Quranic translations into thematic categories and demonstrated its applicability to Quranic text. Similarly, Solihin et al. (2026) reported that SVM outperformed Backpropagation Neural Network in their Quranic text classification experiments and highlighted the influence of preprocessing on classification performance. While these studies support the use of SVM for Quranic text classification, the present study applies the algorithm to a more specific classification task based on the Islamic Pillars.

The results also provide a comparison between SVM and Naive Bayes in the context of the present classification task. SVM with TF-IDF obtained a macro F1-score of 81.98%, whereas Multinomial Naive Bayes with TF-IDF achieved 33.22%. This outcome differs from the findings of Elfianita et al. (2022), who reported higher classification accuracy for Naive Bayes than SVM when classifying Indonesian Quranic translations into thematic categories. Such

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

E-ISSN : 2807-9035
Volume 6, Number 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9760

differences may be associated with variations in classification categories, dataset distribution, feature representation, and textual characteristics across the respective studies. Consequently, the relative performance of SVM and Naive Bayes should be interpreted in relation to the specific dataset and classification task rather than generalized across different contexts.

The results further indicate that TF-IDF performed better than FastText when combined with SVM. SVM with TF-IDF achieved a macro F1-score of 81.98%, whereas SVM with FastText achieved 62.75%. This difference suggests that the term-based representation provided by TF-IDF was more suitable for distinguishing the textual characteristics of the six classes in the present dataset. In contrast, Muslikh et al. (2024) and Akbar et al. (2024) demonstrated the application of FastText to multi-label classification of Indonesian Quranic translations. The different results may be associated with differences in the classification task, dataset characteristics, and label structure. The present study uses a single-label classification scheme with an imbalanced class distribution, including only nine verses in the Fasting class.

The Naive Bayes results were lower than those obtained by SVM in the present experiment. Multinomial Naive Bayes with TF-IDF achieved a macro F1-score of 33.22%, while Gaussian Naive Bayes with FastText achieved 52.18%. These results indicate that SVM was better able to distinguish the six classes under the experimental conditions of this study. The findings also demonstrate that classifier performance is influenced by the feature representation used. Therefore, classifiers and feature representations should be evaluated as combined configurations rather than as independent components.

The relatively high macro precision and lower macro recall of the selected SVM model indicate that the model was generally precise when assigning predictions to the classes, but some relevant instances were still missed. This pattern may be associated with the unequal distribution of the classes. In particular, the Fasting class contained only nine verses, which limited the amount of training information available for learning its textual characteristics. Similar concerns regarding class imbalance have been reported in Quranic text classification research (Ningsih et al., 2024).

The main limitation of this study is the relatively small and imbalanced dataset. The classification also uses a single-label scheme, although some Quranic verses may contain more than one thematic aspect. Therefore, the results should be interpreted within the scope of the selected dataset and classification scheme. Future research could employ a larger and more balanced dataset, investigate multi-label classification, and optimize model parameters to improve classification performance.

CONCLUSION

This study investigated the classification of Quranic verses according to the Islamic Pillars using Naive Bayes and Support Vector Machine (SVM) with TF-IDF and FastText feature representations. The results showed that SVM with TF-IDF achieved the best performance, with an accuracy of 81.88%, a macro precision of 92.55%, a macro recall of 75.63%, a macro F1-score of 81.98%, and a weighted F1-score of 82.03%. The selected model was implemented in a mobile application to support automatic classification and prediction of Quranic verses.

The study demonstrates that machine learning can be applied to classify Quranic verses into Shahada, Salah, Zakat, Fasting, Hajj, and Non-Islamic Pillars. However, the findings are limited by the relatively small and imbalanced dataset, purposive sampling, and the use of a single-label classification scheme. Future research should use a larger and more balanced dataset, evaluate the model on additional data, explore multilabel classification, and apply hyperparameter optimization to improve model performance and generalizability. Further development of the application can also provide additional educational features and relevant tafsir information to support the classification results.

REFERENCES

Ahmed, M. A., Baharin, H., & Nohuddin, P. N. E. (2023). K-Means Variations Analysis For Translation of English Tafseer Al-Quran Text. *International Journal of Electrical and Computer Engineering (IJECE)*, *13*(3), 3255–3265. https://doi.org/10.11591/ijece.v13i3.pp3255-3265

Akbar, I., Faisal, M., & Chamidy, T. (2024). Multi-Label Classification of Indonesian Qur'an Translation using Long Short-Term Memory Model. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, *4*(3), 119–128. https://doi.org/10.22219/KINETIK.V9I2.1901

Bashir, M. H., Azmi, A. M., Nawaz, H., Zaghouani, W., Diab, M., Al-Fuqaha, A., & Qadir, J. (2023). Arabic Natural Language Processing For Qur'anic Research: A Systematic Review. *Artificial Intelligence Review*, *56*(7), 6801–6854. https://doi.org/10.1007/S10462-022-10313-2

Dewantara, M. P., & Lhaksmana, K. M. (2024). Multilabel Classification in Indonesian Translation of Religious Text using Word Centrality Term Weighting. *Jurnal Media Informatika Budidarma*, *8*(2), 1179–1187. https://doi.org/10.30865/MIB.V8I2.7351

Elfianita, F. D., Putra, O. V., Harmini, T., Trisnani, A., & Taufiqurrahman. (2022). Klasifikasi Teks Terjemahan Ayat Al-Qur'an dalam Bahasa Indonesia dengan Machine Learning. *Prosiding Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat Bidang Ilmu Komputer, 1,* 49-57.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.9760

Fauzan, M., Junaedi, H., & Setyati, E. (2022). Klasifikasi Al – Qur'an Terjemahan Bahasa Indonesia Dengan Menggunakan Algoritma Support Vector Machine (SVM). *KONVERGENSI*, *18*(2), 42–49. https://doi.org/10.30996/KONV.V18I1.6912

Hassan, S. U., Ahamed, J., & Ahmad, K. (2022). Analytics of Machine Learning-Based Algorithms For Text Classification. *Sustainable Operations and Computers*, *3*, 238–248. https://doi.org/10.1016/J.SUSOC.2022.03.001

Kurnia, A. R., Anggai, S., & Handayani, M. (2026). Analysis and Evaluation of Qur'an Translation Topics Using Classical, Neural, and Transformer-Based Topic Modelling. *Jurnal Ilmiah Multidisiplin Indonesia*, *5*(2), 367–383 Retrieved from https://ejournal.seaninstitute.or.id/index.php/esaprom/article/view/8099/6129

Muslikh, A. R., Akbar, I., Setiadi, D. R. I. M., & Islam, H. M. M. (2024). Multi-label Classification of Indonesian Al-Quran Translation Based CNN, BiLSTM, and FastText. *Techno.Com*, *23*(1), 37–50. https://doi.org/10.62411/TC.V23I1.9925

Ningsih, S., Safaat, N., Agustian, S., Yusra, Y., & Cynthia, E. P. (2024). Pengaruh Penyeimbangan Data Pada Klasifikasi Terjemahan Al-Quran Dengan Metode Naïve Bayes dan Long Short Term Memory. *Journal of Computer System and Informatics (JoSYC)*, *5*(3), 626–635. https://doi.org/10.47065/JOSYC.V5I3.5181

Opitz, J. (2024). A Closer Look at Classification Evaluation Metrics and a Critical Reflection of Common Evaluation Practice. *Transactions of the Association for Computational Linguistics*, *12*, 820–836. https://doi.org/10.1162/tacl_a_00675

Purwati, Y., Utomo, F. S., Trinarsih, N., & Hidayatulloh, H. (2023). Feature Selection Technique to Improve the Instances Classification Framework Performance for Quran Ontology. *JOIV : International Journal on Informatics Visualization*, *7*(2), 615–620. https://doi.org/10.30630/JOIV.7.2.1195

Puspasari, I., & Rusmin, P. H. (2022). Klasifikasi Wazan pada Kata-Kata Al Qur'an Menggunakan Natural Language Processing. *Journal of Technology and Informatics (JoTI)*, *3*(2), 41–48. https://doi.org/10.37802/JOTI.V3I2.224

Shabrina, G. A., & Lhaksmana, K. M. (2024). Multi-Label Topic Classification on the Qur'an Using the K-Nearest Neighbor and Latent Semantic Analysis Methods. *Jurnal Indonesia Sosial Teknologi*, *5*(12), 7009–7018. https://doi.org/10.59141/JIST.V5I12.1340

Solihin, A. A., Utomo, F. S., & Barkah, A. S. (2026). Impact of Stopword Variation on Qur'anic Text Classification Using Support Vector Machine and Backpropagation. *Journal of Innovation Information Technology and Application (JINITA)*, *8*(1), 173–183. https://doi.org/10.35970/JINITA.V8I1.3069

Umer, M., Imtiaz, Z., Ahmad, M., Nappi, M., Medaglia, C., Choi, G. S., & Mehmood, A. (2022). Impact of Convolutional Neural Network and FastText Embedding on Text Classification. *Multimedia Tools and Applications, 82*(4), 5569–5585. https://doi.org/10.1007/S11042-022-13459-X

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.