Determining the Law of Reading Tajweed (*Idgham Qomariyah* and *Syamsiyah*) in the Qur'an Using the Naïve Bayes Algorithm

Haikal Azhar
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
1207050044@student.uinsgd.ac.id

Ibham Bathsyi Hizbullah
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
1207050143@student.uinsgd.ac.id

Abstract— Learning tajweed to recite the Qur'an is important. Because if there is a mispronunciation, the meaning will be different. This research aims to detect two of the many tajweed, namely *idghom qomariyah* and *syamsiyah* using machine learning technology with a classification approach. This research uses the Naive Bayes algorithm to classify *idghom qomariyah* and *syamsiyah* in Al-Quran text documents. Based on experimental results using 82,173 text data, Naive Bayes was able to classify *idgham qomariyah* and *syamsiyah* with an accuracy rate of 96,80%.

Keywords- Naïve Bayes, Tajweed, Qur'an

I. INTRODUCTION

Tajweed science is a science about the rules and ways to read the Qur'an as well as possible. Protecting the reading of the Qur'an from errors and changes and protecting the oral (mouth) from reading errors is the goal of Tajweed Science. Studying Tajweed Science is *fardhu kifayah*, reading the Qur'an well (in accordance with Tajweed Science) is *fardhu 'Ain*. There are many mandatory arguments that require the practice of tajweed in every reading of the Qur'an.

The Qur'an is studied to understand the meaning or message behind the text. So to get the meaning that is in accordance with the Qur'an, you need to understand qiraat and how to read the Qur'an correctly. How to read the Qur'an properly and correctly can be learned with Tajweed Science.

In the Qur'an there are many reading laws, one of which is the *Qomariyah idgham* and *Syamsiyah idgham*, where these laws greatly influence the pronunciation of the verses read in the Qur'an. So it is necessary to know the laws of reading so

that there are no mistakes in reciting the holy verses of the Qur'an.

The rapid development of machine learning technology can be utilized to study the science of tajweed [1]–[4]. With a classification approach, every tajweed in the Qur'an can be detected and grouped. Therefore, this research aims to detect the laws for reciting *idgham syamsiyah* and *qomariyah* using the classification method in the Al-Qur'an.

II. RELATED WORKS

Many studies have been carried out regarding the classification of Arabic letters using the Naïve Bayes method which proves that this algorithm has good performance in classifying Arabic text. The following is some previous research that is relevant to this research:

1. Nurul A'ayunnisa's research regarding performance analysis of the Gaussian Naïve Bayes method for classifying handwritten images of Arabic characters [5]. The research was conducted using a classification approach using the Gaussian Naïve Bayes method, then data processing was conducted to look for patterns and the information needed. And using the moment invariant stage to determine the value of the rotation, translation, and reflection treatment of an object.

2. There is a study that employs a classification strategy based on the Naive Bayes method, with three groups or categories: Sholat (Prayer), Hajj, and Wedding. These verses were compiled using data from the book LubaabutTafsir Min Ibn Kathir, which yielded accuracy values of 75.52% [6].

Determining the Law of Reading Tajweed (Idgham Qomariyah and Syamsiyah) in the Qur'an Using the Naïve Bayes Algorithm
Haikal Azhar, Ibham Bathsyi Hizbullah
Khazanah Journal of Religion and Technology
Online ISSN: 2987-6060

3. Another research compares Naïve Bayes and Hidden Markov Model for *hijaiyah* pronunciation classification using punctuation letters [7].
4. The other researches that use machine learning algorithms, including Naïve Bayes, for tajweed of Qur'an reciter [8]–[12].

III. RESEARCH METHODS

A. Research framework

Learning to read the Qur'an is a very important topic in learning Sharia science, and the selection and use of appropriate teaching methods can have a significant impact on the success of the learning process. In this context, the use of classification methods such as Naive Bayes can be a relevant approach.

In learning the science of tajweed, the general goal is to understand and master the rules of tajweed used in reading the Al-Quran. By using the Naive Bayes classification method, educators can develop classification models that can help identify and categorize various tajweed rules more effectively. The Naive Bayes method can be used to study tajweed patterns based on certain features or characteristics in the Al-Quran text.

This research also uses literature study methodology and practice of using the Naive Bayes algorithm. The following are the stages that can be followed in the literature study learning method and practice of using the Naive Bayes algorithm: (1) Literature study, studying literature related to the Naive Bayes algorithm, including basic concepts, probability, classification, and variants/modifications that have been developed . (2) Practice, using relevant datasets for implementing the Naive Bayes algorithm, applying the algorithm in dataset classification, using libraries or software such as scikit-learn in Python, and experimenting with variations in datasets and algorithm settings. (3) Evaluation and Correction, evaluating the performance of the Naive Bayes algorithm using evaluation metrics such as accuracy, precision, recall, or F1-score, analyzing classification results, identifying weaknesses or limitations of the Naive Bayes algorithm, and making corrections and improvements based on the evaluation results [13].

B. Ilmu Tajweed

Tajweed science is the basis for reading the Al-Qur'an properly and correctly, and tajweed science is the science of studying how to sound or pronounce the letters contained in the holy book of the Al-Qur'an [14]. So, when reading the Qur'an, you must pronounce it precisely and correctly according to valid rules, because if you read or pronounce it incorrectly it will give a different meaning.

Studying the science of tajweed is *fardhu kifayah*. This means that if in a place, region or country there are Muslims who are experts in the science of tajweed, where people can ask them questions, then that obligation has been fulfilled. However, reading the Qur'an according to the rules of

Tajweed science is *fardhu ain*. This means that everyone who reads the Qur'an must read it well and correctly in accordance with the provisions of Tajweed science.

C. Idghom Qomariyah dan Idghom Syamsiyah

"Al Qomariyah" is "Al" combined with a noun (*isim*) which begins with one of the 14 *qomariyah* letters, as summarized in the following series of letters:

ق ب غ ح ج ك و خ ف ع ي ادم

The way to read "Al" *qomariyah* must be clear (*idzhar*), that is, the reading of the lam sukun is still visible. For this reason, the law of reading "Al" *qomariyah* is often called *idzhar qomariyah*. "Al" *Syamsiyah* is "Al" or alif lam combined with a noun (*isim*) which begins with one of the 14 *Syamsiyah* letters, namely:

ط – ث – م – ر – ت – ض – ذ – ن – د – س – ظ – ز – ش - ل.

And the law is to read *idgham* (inserted) because the *makhraj* letter lam and the letter *syamsiah* are close to each other. *Syamsiyah* means sun, Alif Lam is like a star. And when it meets sunlight, it becomes invisible.

D. Supervised Learning

Supervised Learning is an approach where the system is trained first so that it can make predictions or carry out classification [15]. Supervised Learning is a method that has training data consisting of the desired input and output pairs and aims to learn the mapping between input and output spaces. Meanwhile, Supervised Learning aims to find new patterns in data by connecting existing data patterns with new data. Classification method or commonly referred to as supervised learning is a technique for collecting fully labeled data to classify unknown classes. There are many algorithms in supervised learning, some of which are C4.5 [16], [17], K-Nearest Neighbor (KNN) [18], Naive Bayes Classifier (NBC) [19], Artificial Neural Network (ANN) [20], [21], an so on.

E. Naïve Bayes

Naive Bayes is a simple probabilistic classifier that calculates a set of probabilities by adding up the frequencies and combinations of values from a given dataset. The algorithm uses Bayes' theorem and assumes that all attributes are independent or not interdependent given the values of the class variables [22]. Naive Bayes is based on the simplifying assumption that attribute values are conditionally independent of each other given output values. In other words, given an output value, the joint observing probability is the product of the individual probabilities. The advantage of using Naive Bayes is that this method only requires a small amount of training data to determine the parameter estimates needed in the classification process. Naive Bayes often performs much better in most complex real-world situations

Determining the Law of Reading Tajweed (Idgham Qomariyah and Syamsiyah) in the Qur'an Using the Naïve Bayes Algorithm
Haikal Azhar, Ibham Bathsyi Hizbullah
Khazanah Journal of Religion and Technology
Online ISSN: 2987-6060

than expected [22]. Persamaan dari teorema Bayes terdapat pada formula (1):

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \tag{1}$$

Where :

X: Data with unknown class

H: Hypothesis data is a specific class

P(H|X): Probability of hypothesis H based on condition X (posteriori probability)

P(H): Probability of hypothesis H (prior probability)

P(X|H : Probability of X based on the conditions in hypothesis H

P(X): Probability of X

IV. RESULT AND DISCUSSION

1) Training Data

Training data is a set of data used to train or build a model. The machine learning algorithm will change its parameters to adapt to the data provided during training. Just like the human brain, where synapses will make changes when humans learn. The model is trained using the training dataset, then the performance during training is tested using the validation dataset. This aims to see the ability of the model during training to see whether it can recognize patterns in general.

2) Cleaning Data

Data Cleaning is the process of preparing data for analysis by removing or modifying data that is incorrect, incomplete, irrelevant, duplicated, or incorrectly formatted. This data is usually not necessary or helpful in analyzing data because it can hinder the process or provide inaccurate results. There are several methods for cleaning data depending on how it is stored and the answers sought.

Data Cleaning is not just about deleting information to make room for new data, but rather finding ways to maximize the accuracy of data sets without having to delete information. Data Cleaning includes more actions than deleting data, such as correcting spelling and syntax errors, standardizing data sets, and correcting errors such as empty fields, missing code, and identifying duplicate data points. Data Cleaning is considered a basic element of the fundamentals of data science, as it plays a vital role in the analytical process and uncovering reliable answers.

3) Labeling Data

Data labeling is the process of identifying raw data (images, text files, videos, etc.) and adding one or more meaningful and informative labels to provide context so that machine learning models can learn from it. For example, labels might indicate whether a photo contains birds or cars, what words are spoken in an audio recording, or whether an x-ray contains a tumor. Data labeling is required for a variety of use cases including computer vision, natural language processing, and speech recognition.

4) Accuracy

In this research, accuracy is needed in data labeling, which determines whether the letter has the law of reading *idgham qomariyah* and *syamsiyah*. So as to determine a clear output from the model that has been created. Figure 1 shows an example of the results of detecting words containing *idgham qomariya*h and *syamsiyah* in the Al-Qur'an using Naïve Bayes. The accuracy of the classification model shows good results, namely 96.80% as presented in Figure 2.

	nosurah	surahnameenglish	noayah	word
1	1	Al-Fatiha	1	الله
2	1	Al-Fatiha	1	الرحمن
3	1	Al-Fatiha	1	الرحيم
4	1	Al-Fatiha	2	الحمد
7	1	Al-Fatiha	2	العالمين
...	...	...	...	...
82165	114	An-Nas	4	الوسواس
82166	114	An-Nas	4	الخناس
82167	114	An-Nas	5	الذي
82171	114	An-Nas	5	الناس
82173	114	An-Nas	6	الجنة

[10013 rows x 4 columns]

Fig 1. Result of Naïve Bayes Algorithm

Accuracy: 96.80%
Classification Report:

	precision	recall	f1-score	support
syamsiyah	0.89	1.00	0.94	495
qomariyah	1.00	0.96	0.98	1505
accuracy			0.97	2000
macro avg	0.94	0.98	0.96	2000
weighted avg	0.97	0.97	0.97	2000

Fig 2. Evaluation Result of Naïve Bayes Algorithm

This research contributes to the legal classification of reading *idgham qomariyah* and *syamsiyah* in Al-Qur'an letters using the Naïve Bayes method. In this case, the Naïve Bayes method is used because it can provide good performance in classification based on probability and feature independence assumptions.

It is hoped that the results of this research will help in the introduction and understanding of the law of reading *idgham qomariyah* and *syamsiyah* in the text of the Al-Qur'an. Accurate classification can provide benefits in the context of learning Arabic, especially in understanding the applicable reading rules and regulations.

However, this study also has several limitations. First, classification results can be influenced by the quality and representativeness of the dataset used. In future research, a larger and more representative dataset is needed to increase the reliability and generalization of the classification model. In addition, other factors such as variations in the writing of the Qur'an by different authors can also influence classification accuracy.

Furthermore, this research can be expanded by combining other methods or expanding the scope of classification to involve more other laws of Al-Qur'an reading. This can improve the understanding and application of classification models in a broader context.

In conclusion, this research contributes to the legal classification of reading *idgham qomariyah* and *syamsiyah* in Al-Qur'an letters using the Naïve Bayes method. However, this research also faces limitations in the representativeness of the dataset and variations in Arabic writing. This research can be a basis for further research that is more comprehensive and has a wider scope in classifying reading laws in the language of the Qur'an.

V. CONCLUSION

This research uses the Naïve Bayes method to classify the legal readings of *idgham qomariyah* and *syamsiyah* in Arabic letters. The methods used in this research include labeling, training data, cleaning data, and accuracy measurements. This research aims to develop a classification model using the Naïve Bayes method which can recognize the laws of reading *idgham qomariyah* and *syamsiyah* in the Al-Qur'an. Through the process of training data, cleaning data, and labeling data, as well as measuring accuracy, it is hoped that reliable results can be achieved in classifying the reading law. Future research can utilize this Naïve Bayes model for other tajweed laws.

REFERENCES

[1] M. J. Aqel and N. M. Zaitoun, "Tajweed: An Expert System for Holy Qur'an Recitation Proficiency," in *Procedia Computer Science*, 2015. doi: 10.1016/j.procs.2015.09.029.

[2] N. J. Ibrahim, M. Y. I. Idris, Z. Razak, and N. N. A. Rahman, "Automated tajweed checking rules engine for Quranic learning," *Multicultural Education and Technology Journal*, 2013, doi: 10.1108/METJ-03-2013-0012.

[3] A. P. Andriyandi, W. Darmalaksana, D. S. adillah Maylawati, F. S. Irwansyah, T. Mantoro, and M. A. Ramdhani, "Augmented reality using features accelerated segment test for learning tajweed," *TELKOMNIKA (Telecommunication, Computing, Electronics and Control)*, vol. 18, no. 1, pp. 208–216, 2020, doi: 10.12928/TELKOMNIKA.V18I1.14750.

[4] A. Alfaries, M. Albahlal, M. Almazrua, and A. Almazrua, "A Rule-Based Annotation System to Extract Tajweed Rules from Quran," in *Proceedings - 2013 Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences, NOORIC 2013*, 2015. doi: 10.1109/NOORIC.2013.63.

[5] Nurul A'ayunnisa, Y. Salim, and H. Azis, "Analisis Performa Metode Gaussian Naïve Bayes untuk Klasifikasi Citra Tulisan Tangan Karakter Arab," *Indonesian Journal of Data and Science*, vol. 3, no. 3, pp. 115–121, Dec. 2022, doi: 10.56705/ijodas.v3i3.54.

[6] S. Hidayatullah, "Classification of Al-Qur'an Arabic Verses Used Naive Bayes," *Jurnal Mantik*, vol. 6, no. 1, pp. 717–725, 2022.

[7] A. Adiwijaya, A. Riyani, and M. S. Mubarok, "A Comparison of Naïve Bayes and Bayesian Network on The Classification of Hijaiyah Pronunciation with Punctuation Letters," in *Proceedings of the 2018 International Conference on Industrial Enterprise and System Engineering (IcoIESE 2018)*, Paris, France: Atlantis Press, 2019. doi: 10.2991/icoiese-18.2019.9.

[8] M. Alatiyyah, "Quran Reciter Identification: Techniques and Challenges," *J Theor Appl Inf Technol*, vol. 101, no. 16, 2023.

[9] M. M. Al Anazi and O. R. Shahin, "A Machine Learning Model for the Identification of the Holy Quran Reciter Utilizing K-Nearest Neighbor and Artificial Neural Networks," *Information Sciences Letters*, vol. 11, no. 4, pp. 1093–1102, Jul. 2022, doi: 10.18576/isl/110410.

[10] M. Hadwan, H. A. Alsayadi, and S. AL-Hagree, "An End-to-End Transformer-Based Automatic Speech Recognition for Qur'an Reciters," *Computers, Materials & Continua*, vol. 74, no. 2, pp. 3471–3487, 2023, doi: 10.32604/cmc.2023.033457.

[11] M. Maiyurita, Z. K. Simbolon, and N. Prihatin, "Aplikasi Pembelajaran Tajwid Menggunakan Linear Congruent Method Berbasis Android," *Jurnal Infomedia*, 2018, doi: 10.30811/.v1i1.287.

[12] D. Hamdhana, F. Fadlisyah, and S. Adani, "Sistem Pendeteksi Pola Tajwid Al-Qur'an Hukum Ikhfa Syafawi Dan Idgham Mimi Pada Citra Menggunakan Metode Euclid Distance Dan Bray Curtis Distance," *TECHSI - Jurnal Teknik Informatika*, 2018, doi: 10.29103/techsi.v10i2.886.

[13] K. M. Ting, "Confusion Matrix," in *Encyclopedia of Machine Learning and Data Mining*, 2017. doi: 10.1007/978-1-4899-7687-1_50.

[14] S. A. M. Mu'abbad, *Buku Panduan Lengkap Ilmu Tajwid (Taqiya)*. Taqiya, 2015.

[15] T. Jiang, J. L. Gradus, and A. J. Rosellini, "Supervised Machine Learning: A Brief Primer," *Behav Ther*, vol. 51, no. 5, pp. 675–687, Sep. 2020, doi: 10.1016/j.beth.2020.05.002.

[16] A. Rifai, "Kajian Algortma C4.5, Naive Bayes, Neural Network, dan SVM dalam Penentuan Kelayakan Kredit," *Jurnal Sistem Informasi*, vol. 5, no. 2, pp. 176–182, 2017.

[17] E. Darmawan, "C4.5 Algorithm Application for Prediction of Self Candidate New Students in Higher Education," *Jurnal Online Informatika*, vol. 3, no. 1, p. 22, 2018, doi: 10.15575/join.v3i1.171.

[18] L. Andiani, S. Sukemi, and D. P. Rini, "Analisis Penyakit Jantung Menggunakan Metode KNN Dan Random Forest," in *Annual Research Seminar (ARS)*, 2020, pp. 165–169.

[19] B. Y. Pratama and R. Sarno, "Personality classification based on Twitter text using Naive Bayes, KNN and SVM," in *Proceedings of 2015 International Conference on Data and Software Engineering, ICODSE 2015*, 2016, pp. 170–174. doi: 10.1109/ICODSE.2015.7436992.

[20] M. H. Hassoun, "Fundamentals of Artificial Neural Networks," *Proceedings of the IEEE*, 2005, doi: 10.1109/jproc.1996.503146.

[21] A. Ahmad, "Mengenal artificial intelligence, machine learning, neural network, dan deep learning," *J. Teknol. Indones., no. October*, p. 3, 2017.

[22] M. Q. (Muhammad Q. Shihab, *Tafsir Al-Mishbah : pesan, kesan, dan keserasian Al-Qur'an volume 1-15 / pengarang, M. Quraish Shihab.*, 1st ed., vol. 1–15. Tanggerang: Lentera Hati, 2021., 2021.