I. INTRODUCTION

Many people nowadays seek information on the internet, one of which is concerning the religion of Islam [1]. There are many different types of Islamic knowledge available online, such as articles, videos, and one of them is Islamic consultation in the form of questions and answers. People are welcome to ask questions, which will be answered by a trusted Ustaz. People also often read replies to other people's questions. However, due to a large number of questions and answers, the website management will find it difficult to understand each question-and-answer (Q&A) consultation and categorize it into the appropriate class because it will take a long time. Therefore, a system that can categorize each question and answer is required to make it easier and faster for the website management to group many questions and answers, and to make it possible for people to find the consultation discussion needed by class [1].

Several earlier researches used the supervised learning technique to categorize Islamic texts. In the multi-label and single label classification research of the Indonesian translation of Bukhari hadith using K-Nearest Neighbor [2] and Support Vector Machine [3], the best average Hamming loss value is 0.0886 with an ideal K value of 7 for K-Nearest Neighbor and the best average accuracy is 84.34 percent for Support Vector Machine. In the research on the classification of Islamic Q&A topics using Naive Bayes [1], 800 Q&A data were classified into eight classes, with data obtained from a single website (rumahfiqih.com) yielding an accuracy of 0.97. Because there have been few researches on the categorization of Islamic literature, a classification research using Logistic Regression was done a few years ago. The research of dangerous web classification using the Logistic Regression method [4], which used 1000 data and categorized them into 10 classes, yielding 94 percent accuracy.

The problem statement of this research are according to the main reference of this research [1], the Naive Bayes classification method is not compared with other classification systems that may result in a better evaluation (and there has been no previous research that categorizes Islamic Q&A topics using K-Nearest Neighbor, Support Vector Machine, and Multinomial Logistic Regression), the dataset's source is limited to a single website that with adding dataset sources, the actual evaluation results of the supervised learning method may appear clear and more general, and the previous research did not investigate the vocabulary words that may affect categorization results.

Supervised learning is a machine learning method that researches documents or data that are already labeled [5]. Several methods of supervised learning were used in this research. The supervised learning methods used include K-Nearest Neighbor (K-NN), Support Vector Machine (SVM), Multinomial Naive Bayes (MNB), and Multinomial Logistic Regression (MLR). Based on the primary reference of this research [1], the data were added to 920 data from two sources (rumahfiqih.com and islamqa.info), and the more generic classes became five, namely faith, worship, contemporary, marriage, and heritage.

This research aims to develop a system that can categorize questions and answers from Islamic consultations using the K-NN, SVM, MNB, and MLR methods (this research is the first to use four methods concurrently on this topic) on 920 data from two website sources (more and varied from the research on the same topic [1]), investigate the evaluation results and suitability of the four methods on the dataset, and observe the effect of each class's vocabulary words on the classification results of the system that has been developed.