

Pengembangan Sistem Bank Soal dengan Pengelompokan dan Klasifikasi Tingkat Kesulitan Menggunakan *K-Means* dan *Naïve Bayes* (Development of a Question Bank System with Difficulty Level Grouping and Classification Using *K-Means* and *Naïve Bayes*)

Supervised by Moh. Munih Dian Widianta, S.Kom, M.T

Alvina Sugiarti
Study Program of Informatic Engineering
Majoring of Information Technology
Program Studi Teknik Informatika
Jurusan Teknologi Informasi

ABSTRACT

School examinations at MA Manbaul Ulum still rely on a paid, limited third-party application, Madrasah CBT, which lacks an objective question difficulty analysis feature, leaving difficulty levels dependent on teachers' subjective judgment. This study aims to develop a website based question bank system capable of grouping and classifying question difficulty levels using the K-Means Clustering and Naive Bayes algorithms. The system was developed using the Waterfall model, covering requirements analysis, design, implementation with Laravel and MySQL, testing, and maintenance. K-Means Clustering groups questions into three clusters (Easy, Medium, Hard) based on the percentage of correct answers and average completion time, while the clustering results serve as training data for Naive Bayes to automatically classify the difficulty of new questions. Testing was conducted on 160 questions in Mathematics and Indonesian Language, completed by 22 grade XII students. Results show K-Means converged at the 2nd iteration for Indonesian Language and the 4th for Mathematics, yielding distributions of 30% Easy, 40% Medium, and 30% Hard for Mathematics, and 13% Easy, 58% Medium, and 27% Hard for Indonesian Language; the Naive Bayes model achieved 90.00% accuracy (F1-Score 0.8963) for Mathematics and 90.91% (F1-Score 0.8633) for Indonesian Language. The comparison between teacher and system labels showed a conformity rate of 72.5% for Mathematics and 85.0% for Indonesian Language. Black Box Testing showed that all features functioned as required, and User Acceptance Testing (UAT) confirmed the system is suitable for use. This study proves that combining K-Means Clustering and Naive Bayes effectively produces an objective, automatic question difficulty detection system based on students' performance data.

Keywords: *Question Bank, K-Means Clustering, Naive Bayes, Question Difficulty Level, Web-Based System.*