

ABSTRACT

Health is an important thing that needs to be considered and is the main factor in the smooth running of the Hajj pilgrimage. Hajj pilgrims need to have prime physical and mental conditions which are an important part of fulfilling the requirements of istithaah . The death rate of hajj pilgrims every year always occurs so that more optimal prevention and health care efforts are needed. Data mining processing is important to determine the accuracy of the classification of hajj pilgrim health data which can be used as material for evaluating the health conditions of the pilgrims. This study aims to compare the algorithms in the supervised learning approach with the Chi-Square feature selection in determining the classification of the health status of hajj pilgrims in the Tasikmalaya area in 2024. The algorithms compared are the Decision Tree, Random Forest, AdaBoost, and XGBoost algorithms. The four algorithms were chosen because they have strong capabilities in handling complex data, as well as being efficient in handling various types of data. Based on the results of the study, the application of the Chi-Square feature selection was able to increase the accuracy of each algorithm used in the supervised learning approach. The evaluation results with Confusion Matrix show that the XGBoost algorithm after using Chi-Square feature selection has the highest accuracy of 97.21% when compared to the Decision Tree algorithm which only reached 95.61%, Random Forest 96.81%, and AdaBoost 89.94%.

Keywords: *Chi-Square, Classification, Data Mining, Hajj Pilgrim Health*

ABSTRAK

Kesehatan merupakan suatu hal penting yang perlu diperhatikan dan menjadi faktor utama dalam kelancaran ibadah haji. Jemaah haji perlu memiliki kondisi fisik serta mental prima yang merupakan bagian penting dalam memenuhi syarat istithaah . Angka kematian jemaah haji tiap tahunnya selalu terjadi sehingga diperlukan upaya pencegahan dan penanganan kesehatan yang lebih optimal. Pengolahan *data mining* penting untuk menentukan ketepatan klasifikasi data kesehatan jemaah haji yang dapat digunakan sebagai bahan evaluasi kondisi kesehatan jemaah. Penelitian ini bertujuan untuk membandingkan algoritma pada pendekatan *supervised learning* dengan seleksi fitur Chi-Square dalam menentukan ketepatan klasifikasi status kesehatan jemaah haji di wilayah Tasikmalaya tahun 2024. Algoritma yang dibandingkan yaitu algoritma Decision Tree, Random Forest, AdaBoost, dan XGBoost. Keempat algoritma tersebut dipilih dikarenakan memiliki kemampuan yang kuat dalam menangani data yang kompleks, serta efisien dalam menangani berbagai jenis data. Berdasarkan hasil penelitian, penerapan seleksi fitur Chi-Square mampu meningkatkan akurasi setiap algoritma yang digunakan dalam pendekatan *supervised learning*. Hasil evaluasi dengan *confusion matrix* menunjukkan algoritma XGBoost setelah menggunakan seleksi fitur Chi-Square, memiliki akurasi tertinggi sebesar 97,21% jika dibandingkan dengan algoritma Decision Tree yang hanya mencapai 95,61%, Random Forest 96,81%, dan AdaBoost 89,94%.

Kata Kunci: Chi-Square, *Data Mining*, Kesehatan Jemaah Haji, *Supervised Learning*