

ABSTRAK

PERBANDINGAN PERFORMA ALGORITMA *NAIVE BAYES*, *RANDOM FOREST* DAN *K-NEAREST NEIGHBOR* PADA PREDIKSI CALON JEMAAH HAJI INDONESIA YANG BERPOTENSI MEMBATALKAN HAJI

Oleh
Feri Setiadi

Penelitian ini bertujuan untuk mengidentifikasi model klasifikasi yang paling efektif dalam memprediksi calon jemaah haji yang berpotensi membatalkan pendaftaran hajinya serta menentukan fitur-fitur yang paling berpengaruh terhadap keputusan tersebut. Penelitian ini membandingkan tiga model klasifikasi, yaitu *Naïve Bayes*, *Random Forest*, dan *K-Nearest Neighbor (k-NN)*, dengan menggunakan dataset jemaah haji dari Siskohat Kantor Kementerian Agama Kabupaten Pringsewu. Selain itu, penelitian ini juga menerapkan metode seleksi fitur *Recursive Feature Elimination Cross Validation (REFCV)* untuk mengidentifikasi fitur-fitur yang paling relevan dalam mempengaruhi pembatalan haji.

Hasil penelitian menunjukkan bahwa model *Random Forest* memberikan performa terbaik dengan akurasi, presisi, dan recall yang lebih tinggi dibandingkan model *Naïve Bayes* dan *k-NN*, baik sebelum maupun setelah seleksi fitur. Fitur-fitur seperti 'usia', 'pekerjaan', dan 'alamat' ditemukan sebagai atribut yang paling signifikan dalam mempengaruhi pembatalan haji. Penerapan metode *REFCV* terbukti meningkatkan akurasi model, khususnya pada model *Random Forest* yang mencapai akurasi 95% setelah seleksi fitur.

Penelitian ini menyimpulkan bahwa model *Random Forest* dengan seleksi fitur *REF* merupakan kombinasi yang paling efektif dalam memprediksi pembatalan pendaftaran haji, serta memberikan rekomendasi bagi pengelola haji dalam meningkatkan akurasi prediksi dan efisiensi pengelolaan data jemaah haji.

Kata kunci : *Recursive Feature Elimination Cross Validation(REF)*,
Naïve Bayes, *Random Forest*, *K-Nearest Neighbor (k-NN)*,
Pembatalan Haji