

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini telah membandingkan tiga model pembelajaran, yaitu *Naive Bayes*, *Random Forest*, dan *K-Nearest Neighbor*, dalam memprediksi potensi pembatalan pendaftaran haji Indonesia oleh calon jemaah haji. Berdasarkan hasil analisis performa model menggunakan seluruh fitur dan seleksi fitur dengan *REFCV*, dapat disimpulkan bahwa pemodelan dengan seleksi fitur mampu memberikan hasil yang lebih optimal. Seleksi fitur *REFCV* terbukti meningkatkan akurasi, precision, dan recall terutama pada model *Random Forest*, yang menunjukkan peningkatan paling signifikan dibandingkan model lain.

Secara keseluruhan, *Random Forest* konsisten menjadi model dengan performa terbaik, baik sebelum maupun setelah seleksi fitur. Meskipun *Naive Bayes* dan *k-NN* juga menunjukkan performa yang baik, *Random Forest* unggul dalam hal akurasi dan keseimbangan antara precision dan recall, terutama setelah dilakukan seleksi fitur.

Dalam konteks dataset ini, *Random Forest* dengan seleksi fitur *REFCV* direkomendasikan sebagai model terbaik untuk digunakan, karena mampu memberikan prediksi yang lebih tepat dan efisien dengan mengurangi fitur yang kurang relevan tanpa mengorbankan akurasi.

Berdasarkan hasil perangkingan atribut dan analisis performa model, berikut adalah fitur-fitur yang paling mempengaruhi pembatalan haji:

1. Usia Daftar

Usia saat mendaftar dapat mempengaruhi peluang pembatalan, karena mungkin mencerminkan kesiapan fisik dan finansial calon haji.

2. Pekerjaan:

Jenis pekerjaan mungkin berkaitan dengan stabilitas ekonomi atau fleksibilitas waktu yang mempengaruhi keputusan untuk melanjutkan atau membatalkan keberangkatan.

3. Tempat Lahir dan Tempat Tinggal

Lokasi geografis dapat mempengaruhi aksesibilitas dan kesiapan logistik, sehingga memengaruhi keputusan pembatalan.

4. Tanggal Lahir

Tanggal lahir bisa menjadi indikator usia yang juga berpengaruh dalam pengambilan keputusan.

5. Alamat

Lokasi tempat tinggal juga bisa menunjukkan faktor ekonomi atau sosial yang berhubungan dengan pembatalan haji.

6. Bank dan Cabang

Bank dan cabang yang digunakan untuk setoran haji bisa saja terkait dengan akses atau kebijakan finansial yang berbeda, mempengaruhi keputusan akhir calon haji.

5.3 Saran

Berdasarkan kesimpulan di atas, beberapa saran untuk penelitian selanjutnya adalah:

1. Mempertimbangkan dataset yang lebih besar dan beragam akan sangat bermanfaat untuk generalisasi model.
2. Selain *Naive Bayes*, *Random Forest*, dan *K-Nearest Neighbor*, eksplorasi model lain seperti *Gradient Boosting*, *Support Vector Machine (SVM)*, atau algoritma ensemble lainnya dapat memberikan wawasan lebih lanjut dan potensi peningkatan akurasi prediksi.
3. Penelitian selanjutnya sebaiknya memfokuskan pada optimalisasi *hyperparameter* untuk setiap model. Teknik seperti *Grid Search* atau *Random Search* dapat digunakan untuk menemukan konfigurasi parameter yang optimal.