

BAB I

PENDAHULUAN

1.1 Latar Belakang

Melaksanakan ibadah haji merupakan salah satu rukun islam yang menjadi suatu kewajiban bagi setiap muslim bagi yang mukallaf dan mampu. Kewajiban ini diberlakukan ketika telah memenuhi syarat wajib haji, yaitu beragama Islam, memiliki akal sehat, merdeka, mencapai usia baligh, dan berkemampuan atau istithaah (Sarwat, 2019). Agar dapat menjalankan ibadah haji sesuai dengan ajaran islam, jemaah haji perlu memenuhi persyaratan status kesehatan istithaah. Istithaah ini melibatkan kemampuan fisik serta mental yang dapat diukur melalui pemeriksaan kesehatan (Deswara, 2023). Memeriksa status kesehatan jemaah haji sebelum melakukan ibadah haji menjadi aspek penting yang perlu diperhatikan dalam pelayanan kesehatan haji.

Berdasarkan data yang diperoleh dari Direktorat Penyelenggara Haji dan Umroh Kemenag RI, pada tahun 2023 tercatat sebanyak 824 jemaah haji meninggal dunia. Selama 10 tahun terakhir, angka ini merupakan yang tertinggi dalam penyelenggaraan haji Indonesia. Dilansir dari NU Online pada tanggal 30 Oktober 2023, Peristiwa ini disebabkan oleh sistem yang tidak memprioritaskan syarat status kesehatan istithaah (Mahbib Khoiron, 2023). Bila melihat angka kematian jemaah haji yang meningkat pada tahun 2023, maka pengolahan data sangat diperlukan yang dapat digunakan sebagai bahan evaluasi pencegahan dan perawatan kesehatan lebih lanjut.

Untuk dapat menganalisis pola yang dikelompokkan berdasarkan karakteristik yang relevan pada data pemeriksaan kesehatan jemaah haji, maka diperlukan pendekatan yang efisien untuk mengolah data tersebut melalui beberapa tahapan agar dapat mengidentifikasi informasi penting ataupun menemukan kemungkinan lain yang sulit ditemukan (Safitri & Fitriani, 2022). Dalam pengolahan data pemeriksaan kesehatan jemaah haji, *data mining* dapat menangani hal ini. Penelitian ini menggunakan pendekatan *supervised learning* yang melakukan pendekatan data latih dan data baru yang masuk akan dikelompokkan dengan data yang sudah ada (Tasia dkk., 2023). Algoritma yang digunakan pada penelitian ini yaitu algoritma Decision Tree, Random Forest, Adaboost dan XGBoost. Keempat algoritma tersebut merupakan algoritma berbasis *tree* (pohon) yang dapat membuat klasifikasi lebih mudah, sederhana dan juga memiliki tingkat akurasi yang tinggi (Halimah dkk., 2022).

Pemilihan fitur terbaik sangat penting untuk meningkatkan akurasi, penelitian ini menggabungkan beberapa algoritma klasifikasi dengan penerapan metode seleksi fitur. Seleksi fitur dapat menghilangkan fitur yang tidak relevan dengan mengidentifikasi fitur yang optimal. Salah satu metode seleksi fitur berbasis filter yang umum digunakan adalah Chi-Square. Chi-Square merupakan salah satu seleksi fitur yang mampu menyeleksi data yang akan digunakan dengan menghilangkan banyak fitur tanpa mengurangi tingkat akurasi (Rahmadeyan & Mustakim, 2023).

Banyak penelitian menggunakan metode atau algoritma klasifikasi tentang kesehatan, seperti pada penelitian yang dilakukan oleh Arif Budiarto, Agus Subekti

dan Haris Darmawan tahun 2022. Penelitian ini melakukan pendeteksian potensi faktor risiko tinggi kesehatan jemaah haji untuk mendapatkan hasil akurasi yang tepat. Hasil dari penelitian ini mendeteksi dengan 2 kelas yaitu risiko tinggi dan tidak risiko tinggi, di mana hasil algoritma klasifikasi Decision Tree mendapatkan akurasi terbaik yaitu sebesar 83.31%. Meskipun akurasi yang didapat cukup baik, penelitian ini tidak menjelaskan lebih lanjut mengenai parameter yang mempengaruhi hasil klasifikasi (Budiarto dkk., 2022).

Selanjutnya pada penelitian oleh Okta Jaya Harmaja, Irvan Prasetia, Yosi Victor Hutagalung, dan Hendra Ardanis Sirait pada tahun 2023 yang melakukan penelitian mengenai perbandingan algoritma dalam klasifikasi diagnosa dini diabetes. Algoritma yang digunakan pada penelitian ini adalah algoritma Random Forest, Adaboost dan XGBoost. Hasil penelitian tersebut menunjukkan bahwa penggunaan Random Forest menunjukkan kinerja paling baik dengan nilai akurasi sebesar 86% dibandingkan ketiga algoritma tersebut (Jaya Harmaja dkk., 2023). Penelitian yang dilakukan oleh Feri Setiadi dan Handoyo Widi Nugroho pada tahun 2024 yang membandingkan algoritma Naive Bayes, Random Forest dan K-Nearest Neighbor pada prediksi calon jemaah haji Indonesia yang berpotensi membatalkan haji dengan menggunakan Recursive Feature Elimination. Hasil penelitian menggunakan algoritma Random Forest dengan penggunaan seleksi fitur menunjukkan kinerja terbaik dengan nilai akurasi sebesar 95% dibandingkan algoritma lainnya (Setiadi & Widi Nugroho, 2024).

Meskipun penelitian sebelumnya menunjukkan hasil yang baik, namun masih terdapat potensi untuk meningkatkan akurasi dan efektivitas algoritma tersebut.

Penelitian ini melakukan perbandingan algoritma Decision Tree, Random Forest, Adaboost dan XGBoost dengan seleksi fitur Chi-square pada klasifikasi status kesehatan jemaah haji di wilayah Tasikmalaya. Perbandingan terhadap algoritma dengan melakukan seleksi fitur diperlukan untuk menghasilkan hasil klasifikasi yang akurat dan meningkatkan hasil akurasi. Penggunaan seleksi fitur Chi-Square digunakan untuk meningkatkan nilai akurasi pada keempat algoritma dengan memilih fitur terbaik. Hasil penelitian ini berfokus terhadap hasil perbandingan akurasi klasifikasi dari masing-masing algoritma menggunakan proses seleksi fitur Chi-Square dan tanpa melalui proses tersebut.

1.2 Rumusan Masalah

Berikut merupakan perumusan masalah pada penelitian berdasarkan latar belakang yang telah diuraikan:

- a. Bagaimana pengaruh penggunaan seleksi fitur Chi-square dalam menentukan ketepatan klasifikasi status kesehatan jemaah haji?
- b. Bagaimana akurasi algoritma pada pendekatan *supervised learning* untuk menentukan ketepatan klasifikasi status kesehatan jemaah haji dengan dan tanpa menggunakan seleksi fitur Chi-square?

1.3 Tujuan Penelitian

Berikut merupakan tujuan penelitian berdasarkan latar belakang yang telah diuraikan:

- a. Menganalisis pengaruh seleksi fitur Chi-square untuk menentukan ketepatan klasifikasi status kesehatan jemaah haji.

- b. Mengevaluasi hasil perbandingan akurasi algoritma pada pendekatan *supervised learning* untuk menentukan ketepatan klasifikasi status kesehatan jemaah haji dengan dan tanpa menggunakan seleksi fitur Chi-Square.

1.4 Manfaat Penelitian

Berikut merupakan manfaat pada penelitian berdasarkan latar belakang yang telah diuraikan:

- a. Memberikan kontribusi secara teoritis mengenai akurasi algoritma pada pendekatan *supervised learning* dengan melakukan klasifikasi status kesehatan jemaah haji di wilayah Tasikmalaya menggunakan seleksi fitur Chi-Square.
- b. Klasifikasi status kesehatan jemaah haji pada algoritma pendekatan *supervised learning* dikembangkan dengan penambahan seleksi fitur Chi-Square sehingga model dapat memberikan fitur terbaik untuk menghasilkan algoritma dengan akurasi yang tinggi.
- c. Dapat dijadikan sebagai referensi penelitian selanjutnya mengenai algoritma pada pendekatan *supervised learning* dalam bidang kesehatan.

1.5 Batasan Masalah

Berikut merupakan batasan masalah pada penelitian berdasarkan latar belakang yang telah diuraikan:

- a. Algoritma pada pendekatan *supervised learning* yang digunakan adalah Algoritma Decision Tree, Random Forest, AdaBoost dan XGBoost.
- b. Seleksi fitur pada klasifikasi status kesehatan jemaah haji menggunakan Chi-Square.

- c. Data yang digunakan pada penelitian ini merupakan data pemeriksaan kesehatan jemaah haji di wilayah Kabupaten dan Kota Tasikmalaya tahun 2024.
- d. Status kesehatan jemaah haji diklasifikasikan menjadi empat kelas, yaitu istithaah, istithaah dengan pendamping, tidak istithaah sementara dan tidak istithaah.