

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang Masalah**

Ibadah haji merupakan salah satu rukun Islam yang wajib ditunaikan bagi setiap muslim yang telah mencukupi syarat-syaratnya [1]. Setiap tahun ribuan calon jemaah haji Indonesia diberangkatkan ke tanah suci untuk menunaikan ibadah haji[2]. Data pada Kementerian Agama RI dalam skala nasional melaporkan jumlah pendaftar jemaah haji mengalami tren yang meningkat. Pada tahun 2021 sebanyak 67.345 orang telah tercatat sebagai pendaftar baru Jemaah haji dan pada tahun 2022 meningkat 71% menjadi 398.622 orang pendaftar baru (satudata.kemenag.go.id) [17]. Namun, setiap tahun, sejumlah calon Jemaah haji membatalkan pendaftaran hajinya karena berbagai alasan, seperti kondisi kesehatan yang memburuk, usia lanjut, meninggal, atau kendala finansial.

Berdasarkan data dari Siskohat Kantor Kementerian Agama Kabupaten Pringsewu, alasan "lain-lain" menjadi penyebab terbesar dari pembatalan pendaftaran haji, mencakup 59,7% dari total kasus. Pembatalan akibat sakit mencapai 5,4%, sementara 34,9% disebabkan oleh wafatnya calon Jemaah haji. Secara keseluruhan, terdapat 149 kasus pembatalan pendaftaran haji dalam rentang tahun 2023-2024.

Tabel 1.1 Data Pembatalan Jemaah Haji

Sebab Batal	Tahun 2023	Tahun 2024	Jumlah	%
Lain - lain	56	33	89	59,7
Sakit	8		8	5,4
Wafat	39	13	52	34,9
Total	103	46	149	100

Membatalkan pendaftaran haji menjadi suatu kerugian bagi Jemaah tersebut, baik dari segi biaya, waktu, maupun kesempatan yang langka untuk menunaikan ibadah haji. Selain itu, pembatalan pendaftaran Jemaah haji juga berdampak pada kebijakan pemerintah dan pengelolaan kuota haji yang telah ditetapkan sebelumnya. Ahli hukum keuangan negara, Siswoyo Sujanto, dalam sidang uji materi terkait pengelolaan dana haji yang digelar oleh Mahkamah Konstitusi menyatakan, pembatalan pendaftaran Jemaah haji oleh calon Jemaah haji dapat menyebabkan kekacauan administrasi yang berdampak pada penyelenggaraan haji secara keseluruhan [18]. Oleh karena itu, prediksi akurat terkait Jemaah haji yang berpotensi membatalkan pendaftaran Jemaah haji perlu dilakukan, sehingga dapat dilakukan tindakan pencegahan atau kebijakan yang tepat sebelum calon Jemaah haji membatalkan keberangkatan haji ke Tanah Suci.

Penelitian ini menggunakan algoritma klasifikasi dalam melakukan prediksi calon Jemaah haji Indonesia yang berpotensi membatalkan keberangkatan haji. Algoritma klasifikasi adalah suatu teknik dalam *data mining* yang digunakan untuk mengklasifikasikan suatu data menjadi kelompok atau kategori tertentu berdasarkan ciri-ciri atau atribut yang dimilikinya.

Namun, dalam proses klasifikasi, pemilihan fitur yang relevan sangat penting untuk meningkatkan akurasi prediksi [3].

Beberapa studi terkait telah dilakukan mengenai algoritma klasifikasi dengan seleksi fitur. Penggunaan *Recursive Feature Elimination* dengan *Cross-Validation* (*REFCV*) dapat meningkatkan akurasi model pembelajaran mesin dengan mengidentifikasi dan memilih fitur-fitur yang paling relevan. Studi ini menunjukkan bahwa meskipun model pembelajaran mesin yang sederhana digunakan, hasil yang lebih baik dapat dicapai dengan seleksi fitur yang tepat. Dalam konteks deteksi diabetes tipe II, *REFCV* berhasil mengurangi dimensi data tanpa kehilangan akurasi prediksi yang signifikan. Dengan menerapkan *Cross Validation*, masalah *overfitting* dapat dihindari, dan hasil yang lebih akurat dicapai dibandingkan dengan menggunakan semua fitur yang tersedia [4].

Dalam penelitian yang dilakukan oleh Ardea dan Fahrurrozi pada tahun 2019, algoritma *Naive Bayes*, *K-Nearest Neighbor*, *Decision Tree*, dan *Random Forest* dibandingkan untuk mengklasifikasikan data terkait penyakit jantung koroner. Data yang digunakan adalah data penyakit jantung koroner yang berasal dari lokasi Cleveland Clinic Foundation yang telah diadopsi oleh instansi Hungarian Institute of Cardiology. Penelitian tersebut menemukan bahwa algoritma *Random Forest* memiliki nilai akurasi terbaik dibandingkan dengan algoritma lainnya untuk klasifikasi data [5].

Penelitian yang dilakukan oleh Agung Purwanto, Handoyo Widi Nugroho membandingkan *Algoritma C.45* dan *Algoritma K-Nearest Neighbors* untuk klasifikasi penerima beasiswa. Dari penelitian tersebut, didapatkan bahwa *K-Nearest Neighbors* memberikan nilai akurasi yang lebih baik [6].

Berdasarkan hasil akurasi yang relatif baik pada model studi literature tersebut, penelitian ini akan membandingkan antara model klasifikasi *data mining Naive Bayes*, *Random Forest* dan *K-Nearest Neighbor* dengan menambahkan tahap *Recursive Feature Elimination Cross-Validation*. Data yang digunakan dalam penelitian ini adalah data jemaah haji dari Kantor Kementerian Agama Kabupaten Pringsewu tahun 2019-2024. Penelitian ini dilakukan untuk mendapatkan model terbaik dalam memprediksi pembatalan haji sehingga dapat memberikan informasi yang berharga dalam upaya meningkatkan efektivitas pengelolaan kuota haji dan mencegah atau mengurangi jumlah pembatalan haji oleh masyarakat.

## **1.2 Identifikasi Masalah**

Berdasarkan latar belakang masalah yang telah dijelaskan, adalah beberapa masalah yang telah diidentifikasi:

1. Tren peningkatan jumlah pendaftar jemaah haji setiap tahun berdampak pada pengelolaan kuota haji
2. Tingginya angka pembatalan pendaftaran Jemaah haji oleh calon Jemaah haji
3. Perlunya prediksi yang akurat untuk mengidentifikasi jemaah yang berpotensi membatalkan pendaftaran haji.

### **1.3 Batasan Masalah**

Untuk menjaga fokus dan relevansi hasil yang diperoleh, dalam penelitian ini ditetapkan Batasan-batasan masalah sebagai berikut :

1. Membangun model prediksi pembatalan haji oleh Jemaah haji menggunakan algoritma *Naive Bayes*, *Random Forest* dan *K-Nearest Neighbor*.
2. Dataset yang digunakan adalah data jemaah haji pada Kantor Kementerian Agama Kabupaten Pringsewu pada tahun 2019-2024.
3. *Recursive Feature Elimination* diterapkan sebagai seleksi fitur untuk menentukan fitur paling berpengaruh pada data.

### **1.4 Rumusan Masalah**

Berdasarkan latar belakang masalah dan studi literatur yang telah dilakukan, maka rumusan masalah dalam penelitian ini dapat dirangkum sebagai berikut :

1. Model klasifikasi mana yang memberikan performa terbaik dalam memprediksi calon jemaah haji yang berpotensi membatalkan pendaftaran haji?
2. Fitur/atribut apa saja yang paling mempengaruhi keputusan pembatalan haji oleh calon jemaah?
3. Bagaimana pengaruh seleksi fitur terhadap performa model klasifikasi?

## **1.5 Tujuan Penelitian**

Berdasarkan rumusan masalah diatas, maka tujuan penelitian ini adalah sebagai berikut :

1. Mengidentifikasi model klasifikasi yang paling efektif dalam memprediksi calon jemaah haji yang berpotensi membatalkan keberangkatannya
2. Menentukan fitur-fitur atau atribut-atribut yang paling berpengaruh terhadap keputusan pembatalan haji.
3. Menganalisis dampak seleksi fitur terhadap performa model klasifikasi.

## **1.6 Kegunaan Penelitian**

Berikut adalah beberapa kegunaan dari penelitian ini :

1. Menambah wawasan tentang efektivitas algoritma klasifikasi dalam konteks prediksi pembatalan pendaftaran jemaah haji, serta memberikan kontribusi pada pengembangan ilmu pengetahuan di bidang *data mining* dalam topik serupa.
2. Memberikan wawasan bagi pengelola haji dalam mengidentifikasi calon jemaah yang berpotensi membatalkan pendaftaran Jemaah haji, meningkatkan efisiensi administrasi haji, dan mengurangi kerugian finansial bagi jemaah dan pemerintah.
3. Penelitian ini dapat menjadi dasar bagi pengambilan kebijakan publik yang lebih efektif dalam pengelolaan haji, serta membuka peluang untuk penerapan teknologi dalam manajemen kegiatan keagamaan lainnya.