Jurnal Pendidikan dan Teknologi Indonesia (JPTI)
Vol. 5, No. 9, September 2025, Hal. 2930-2939

DOI: https://doi.org/10.52436/1.jpti.1060
p-ISSN: 2775-4227
e-ISSN: 2775-4219

Deteksi Potensi Faktor Keberangkatan Jemaah Haji Menggunakan Algoritma Klasifikasi *Machine Learning*

Oxana Farah Maulida[1], Tohirin Al-Mudzakir[2], Hilda Yulia Novita[3], Tatang Rohana[4]

[1,2,3,4]Fakultas Ilmu Komputer, Universitas Buana Perjuangan, Karawang, Indonesia
Email : [1]if21.oxanamaulida@mhs.ubpkarawang.ac.id, [2]tohirin@ubpkarawang.ac.id, [2]hilda.yulia@ubpkarawang.ac.id, [4]tatang.rohana@ubpkarawang.ac.id

Abstrak

Haji merupakan salah satu rukun islam yang memiliki makna spiritual dan sosial mendalam bagi umat muslim diseluruh dunia Dengan meningkatnya jamaah haji di Indonesia setiap tahunnya, pengelolaan dan pelayanan terhadap calon jamaah haji menjadi tantangan. Faktor yang mempengaruhi seperti faktor demografis dari usia, pendidikan dan pekerjaan yang mempengaruhi keberangkatan jamaah. Penelitian ini bertujuan untuk mendeteksi faktor keberangkatan jamaah haji menggunakan algoritma *machine learning*, khususnya metode *Naïve Bayes, Random Forest* dan *Decision Tree*. Dataset yang dikumpulkan dari Kantor Kementerian Agama Karawang dan diolah menggunakan bahasa pemrograman *Phyton*. Proses penelitian meliputi pengumpulan data, *preprocessing*, split data, implementasi algoritma, dan evaluasi. *Random Forest* mencapai akurasi tertinggi sebesar 99.23%, *Decision Tree* mencatat akurasi 98.75%, dan *Naïve Bayes* memiliki akurasi 76.69%. Hasil evaluasi menunjukkan model mampu memberikan akurasi signifikan dalam mengidentifikasi kategori jamaah haji. Diharapkan penelitian ini akan memeberikan wawasan mendalam tentang klasifikasi data jamaah haji dan membantu instansi dalam perencanaan sumber daya yang kebih baik sehingga instansi dapat mengoptimalkan penggunaan anggaran dan alokasi sumber daya yang lebih efisien.

Kata kunci: *Jemaah haji, Klasifikasi, Machine Learning*

Detection of Potential Hajj Pilgrim Departure Factors Using Machine Learning Classification Algorithm

Abstract

Hajj is one of the pillars of Islam that has deep spiritual and social meaning for Muslims around the world.. With the increasing number of hajj pilgrims in Indonesia every year, the management and service of prospective hajj pilgrims becomes a challenge. Influencing factors such as demographic factors of age, education and occupation that affect the departure of pilgrims. This study aims to detect factors of hajj pilgrim departure using machine learning algorithms, especially the Naïve Bayes, Random Forest and Decision Tree methods. The dataset was collected from the Karawang Ministry of Religious Affairs Office and processed using the Phyton programming language. The research process includes data collection, preprocessing, splitting data, implementing algorithms, and evaluation. Random Forest achieved the highest accuracy of 99.23%, Decision Tree recorded an accuracy of 98.75%, and Naïve Bayes had an accuracy of 76.69%. The evaluation results showed that the model was able to provide significant accuracy in identifying categories of hajj pilgrims. It is hoped that this research will provide in-depth insight into the classification of Hajj pilgrim data and assist agencies in better resource planning so that agencies can optimize budget use and more efficient resource allocation.

Keywords: *Classification, Hajj Pilgrims, Machine Learning*

1. PENDAHULUAN

Ibadah haji termasuk salah satu rukun islam yang diwajibkan oleh Allah bagi umat muslim yang mampu melaksanakannya. Kementerian Agama (Kemenag) Karawang mencatat ada sebanyak 2.055 calon jamaah haji yang akan berangkat ke tanah suci pada tahun 2024. Pada tahun 2025 tersedia sebanyak 2.055 kuota calon jamaah haji. Faktor-faktor demografis seperti usia, pendidikan, dan pekerjaan sangat mempengaruhi kebutuhan jamaah selama menunaikan ibadah haji. Hal ini disebabkan dengan pertumbuhan jumlah jamaah yang signifikan, terdapat beberapa tantangan yang kompleks terkait pengelolaan dan pelayanan selama pelaksanaan ibadah haji mengenai calon jamaah. Merujuk pada penjabaran sebelumnya solusi pada permasalahan ini yaitu dengan menerapkan

p-ISSN: 2775-4227
e-ISSN: 2775-4219

algoritma *Machine Learning* dalam mengelola data calon jamaah haji menggunakan metode pengelompokkan data dengan sistem partisi dan pemodelan tanpa *supervisi* yang dilakukan oleh *data mining* [1].

Machine Learning ialah salah satu cabang ilmu kecerdasan buatan yang berkonsentrasi pada pembuatan model dan algoritma sehingga memungkinkan *computer* untuk memahami dalam membuat prediksi atau keputusan berdasarkan data. *Machine learning* berkembang dengan cepat dan dapat menangani masalah klasifikasi, regresi, klastering sampai deteksi anomali dengan lebih efisien. [2]. Klasifikasi merupakan teknik yang digunakan untuk mengidentifikasi pola serta dapat membedakan antara satu kelas data dengan kelas lainnya, sehingga memungkinkan untuk menentukan kategori suatu objek berdasarkan atribut yang dimiliki dan kelompok yang telah ditentukan sebelumnya [3]. Proses klasifikasi meliputi tiga tahap utama, yaitu pembuatan model, penerapan model, dan evaluasi. Pembuatan model dilakukan dengan menggunakan data latih yang telah memiliki atribut dan kelas, yang selanjutnya digunakan untuk menentukan kelas objek baru. Kemudian, model dievaluasi untuk mengukur tingkat akurasi dalam pengembangan dan penerapannya pada data baru. [4].

Naïve Bayes Classifier (NBC) merupakan salah satu metode klasifikasi yang umum digunakan. Keunggulan *Naïve Bayes* yaitu kesederhanaannya yang disertai dengan akurasi tinggi. Proses klasifikasi data dilakukan dalam dua langkah. *Naïve Bayes* adalah salah satu pendekatan yang digunakan dalam data mining yang termasuk dalam kategori *supervised learning*. [5]. Langkah pertama dalam klasifikator probabilistic adalah pelatihan yang disebut sebagai model pelatihan. Pelatihan ini menggunakan teorema probabilitas yang berfokus pada gagasan bahwa keberadaan fitur tidak bergantung pada fitur lainnya, ini disebut sebagai Navie [6]. *Random Forest* (RF) merupakan metode yang menggunakan pendekatan dengan melakukan pemisahan biner rekrusif untuk mencapai node akhir dalam struktur pohon keputusan berdasarkan pohon klasifikasi dan regresi [7].

Berbagai algoritma *machine learning* telah diterapkan pada penelitian sebelumnya dalam mengklasifikasi data. Penelitian dengan algoritma *Naïve Bayes* menggunakan data rekapitulasi penjualan obat melalui proses *mining* menghasilkan tingkat akurasi sebesar 88.00% [8]. Model *Random Forest* juga telah digunakan pada klasifikasi berbasis kinerja efisiensi energi pada sistem pembangkit daya mempunyai akurasi sebesar 92,8% [9]. *Data mining* dengan menggunakan metode *Random Forest* pada prediksi penyakit jantung penerapan *Principal Component Analysis* (PCA) memiliki hasil akurasi sebesar 98% [10]. Studi lain yaitu membandingkan *Random Forest* dengan *Naïve Bayes*, dan menunjukkan bahwa algoritma *Random Forest* memiliki akurasi 86.55% dan memiliki kinerja jauh lebih baik dibandingkan *Naïve Bayes* yang hanya mencapai akurasi 36.61% dalam klasifikasi curah hujan di Indonesia [11].Penelitian selanjutnya yang menggunakan algoritma *Random Forest* memiliki akurasi 97.26% dan kinerja terbaik dalam menentukan penerima bantuan raskin, dibandingkan dengan *Support Vector Machine* (92.15%) dan *Naïve Bayes* (88.39%) [12]. Selain itu metode *Random Forest* menunjukkan akurasi sebesar 85,5% sehingga lebih unggul dari *Decision Tree* dengan akurasi 84,4% dalam memprediksi keberhasilan pengobatan *imunotherapi* untuk penyakit kutil [13]. Dalam penelitian tentang prediksi tingkat kelulusan mahasiswa, model *Random Forest* mencapai akurasi 90.92%, lebih unggul dari *Decision Tree* yang mencatat 85.37%, menunjukkan efektivitas kedua algoritma dalam mengidentifikasi faktor akademik, demografis, dan sosial ekonomi [14]. *Random Forest* dengan hyperparameter tuning terbukti menjadi algoritma terbaik untuk klasifikasi penyakit jantung dalam penelitian ini, mencapai recall 80.6% dan ROC AUC 76.3%, serta menunjukkan stabilitas performa melalui cross-validation [15]. Pada sistem pendukung berbasis *Decision Tree algorithm* hasil penelitian menunjukkan bahwa algoritma Decision Tree adalah yang paling akurat (96,35%) dan dominan untuk prediksi penyakit diabetes, serta telah berhasil diimplementasikan dalam sistem pendukung keputusan berbasis web [16]

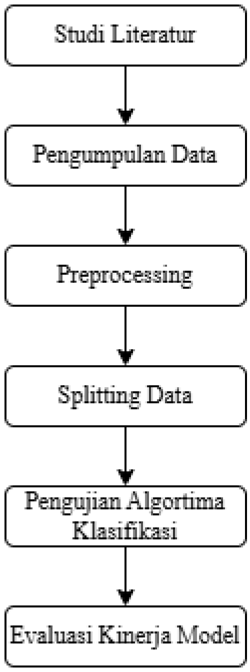
Mengacu pada paparan dan beberapa referensi tersebut, tujuan dari dilakukan penelitian ini adalah menganalisa dengan pendekatan tiga algoritma *machine learning* yaitu *Naïve Bayes, Random Forest, serta Decision Tree*, penulis berminat untuk melakukan penelitian terhadap klasifikasi data dengan jumlah yang mencapai 5000 jamaah haji, kemudian akan dilihat berdasarkan karakteristik demografis dengan mengimplementasikan algoritma *Naïve Bayes*, *Random Forest*, dan *Decision Tree*. Proses implementasi ini menggunakan bahasa pemrograman *Python*. Harapan pada penelitian ini yaitu dapat memperoleh pemahaman yang lebih mendalam terkait klasifikasi data jamaah haji sehingga memudahkan dalam menentukan keberangkatan jamaah haji.

2. METODE PENELITIAN

2.1. Tahapan Penelitian

Klasifikasi data pada penelitian ini berbasis *machine learning* yang dilakukan melalui serangkaian atau tahapan sistematis yang saling terintegrasi. Tahapan penelitian ini dirancang secara metodis agar pendekatan yang diusulkan dapat diterapkan secara terstruktur dan dapat direproduksi dalam konteks serupa. Penelitian ini akan melalui beberapa tahapan yang dapat dilihat pada Gambar 1.

p-ISSN: 2775-4227
e-ISSN: 2775-4219



Gambar 1. Alur Penelitian

Mengacu pada Gambar 1, berikut merupakan paparan dari masing-masing tahapan penelitian yang telah diimplementasikan.

a. Studi Literatur

Tahap awal penelitian ini diawali dengan melakukan *library research* (Penelitian Perpustakaan) mencari dan memahami penelitian yang dilakukan untuk data primer dengan memanfaatkan jurnal, buku, atau literatur yang tersedia, serta materi perkuliahan yang berkaitan dengan topik penelitian.

b. Pengumpulan Data

Mengumpulkan data calon jamaah haji dari Kantor Kementerian Agama Karawang bagian Pelayanan Haji dan Umrah dengan metode yang digunakan yaitu wawancara bersama narasumber. Data terdiri dari 5000 baris dan dipilih karena memiliki fitur yang representatif serta kualitas data yang memadai untuk mendukung proses klasifikasi dalam penelitian ini

c. *Preprocessing*

Tahap ini diperlukan untuk mengoptimalkan kinerja algoritma klasifikasi, di mana tujuan praproses adalah merubah data ke dalam format yang akan membuat proses selanjutnya lebih efisien dan mudah dilakukan. [17]. Memberi label pada data untuk masing-masing kelas serta pembersihan data yang mencakup mengidentifikasi dan memilih fitur yang digunakan serta menghapus data tidak konsisten, dan juga menghilangkan atribut yang tidak berkontribusi.

d. *Splitting* Data

Pendekatan *machine learning* bagian *splitting* diterapkan untuk membagi dataset menjadi dua komponen yaitu data latih (*Data Training*) dan data uji (*Data testing*) [18]. Pembagian jumlah data antara *data training* dan data testing adalah faktor penting dalam menentukan tingkat akurasi. Penelitian ini menggunakan 80% data sebagai data latih dan 20% sebagai data uji.

e. Pengujian Algoritma Klasifikasi

Model pendekatan klasifikasi yang diterapkan dalam penelitian ini adalah tiga algoritma *machine learning* yaitu *Naïve Bayes*, *Random Forest* dan *Decision Tree*. *Naïve Bayes* dipilih karena kesederhanaannya dan kemampuannya dalam menangani data dengan asumsi independensi antar fitur, serta sering digunakan dalam klasifikasi teks dan masalah dengan data besar, sedangkan *Random Forest* merupakan ensemble dari beberapa pohon keputusan yang mampu menangani *overfitting* dan memberikan hasil yang stabil. *Decision Tree* karena kemampuannya untuk memberikan interpretasi yang jelas dan visualisasi yang mudah dipahami, serta kemampuannya dalam menangani data dengan fitur kategorikal dan numerik. Algoritma ini dipilih karena untuk menentukan mana yang paling sesuai untuk dataset yang digunakan. Pada tahap ini merupakan proses implementasi ketiga algoritma tersebut dengan

p-ISSN: 2775-4227
e-ISSN: 2775-4219

menggunakan bahasa pemrograman *phyton* versi yang digunakan adalah *Python* 3.8, yang merupakan salah satu versi stabil dan banyak digunakan dalam pengembangan aplikasi *machine learning*, dan dioperasikan di *google colab* memungkinkan kolaborasi dan akses mudah ke sumber daya komputasi yang lebih besar, serta mendukung eksekusi kode *Python* secara langsung di browser.

f. Evaluasi Kinerja Model

Evaluasi adalah tahap akhir dalam proses kinerja algoritma klasifikasi yang diterapkan dalam penelitian ini. Selanjutnya, kinerja model dievaluasi menggunakan empat metrik utama, yaitu, *accuracy, recall*, *precision*, *F1-score*, dan *confusion matrix*. *Accuracy* merupakan total keseluruhan seberapa baik suatu model klasifikasi dalam memprediksi kelas yang benar dari data. *Recall* dan *precision* mengukur ketepatan serta kelengkapan deteksi kasus positif. *F1-score* merepresentasikan keseimbangan antara keduanya. Dan *confusion matrix* memberikan rincian prediksi benar dan salah dari masing-masing kelas. Detail formula perhitungan evaluasi perhitungan untuk metrik yang digunakan :

1. Akurasi

$$\frac{\mathrm{TP+TN}}{\mathrm{TP+FP+FN+TN}}\,\mathrm{X}\,100 \qquad (1)$$

   1. *True Positive*: Jumlah data yang aktualnya positif dan diprediksi positif oleh model.
   2. *True Negative:* Jumlah data yang aktualnya negatif dan diprediksi negatif oleh model.
   3. *False Positive:* Jumlah data yang aktualnya negatif tetapi diprediksi positif oleh model.
   4. *False Negative:* Jumlah data yang aktualnya positif tetapi diprediksi negatif oleh model.

2. Presisi

$$\frac{\mathrm{TP}}{\mathrm{TP+FP}} \qquad (2)$$

3. *Recall*

$$\frac{\mathrm{TP}}{\mathrm{TP+FN}} \qquad (3)$$

4. F1-Score

$$\frac{2(\mathrm{Recall}*\mathrm{Precision})}{(\mathrm{Recall}+\mathrm{Precision})} \qquad (4)$$

2.2. Naïve Bayes

Naïve Bayes Classifier (NBC) ialah metode klasifikasi yang sering digunakan. *Naïve Bayes* merupakan salah satu metode klasifikasi dan *statistic* pengklasifikasi yang dapat memprediksi peluang untuk menjadi anggota kelas [19]. Algoritma ini memiliki keunggulan yang sederhana namun mempunyai akurasi yang tinggi [18]. *Classifier* probabilistik sebagai penggunaan teorema probabilitas untuk klasifikasi dan berfokus pada pembuatan anggapan bahwa keberadaan fitur tertentu yang independen dari keberadaan fitur [20]. Model ini didasarkan pada persamaan (5).

$$\mathrm{P(Y|X)} = \frac{\mathrm{P(X|Y)\,P(Y)}}{\mathrm{P(X)}} \qquad (5)$$

Di mana P(Y|X) adalah probabilitas dari hipotesis Y berdasarkan kondisi X, Y adalah hipotesa dari data yang berupa suatu kelas yang spesifik, P(X) adalah probabilitas dari Y, dan X adalah kelas yang datanya belum diketahui.

2.3. Random Forest

Random forest dikembangan oleh Leo Breimean. *Random forest* juga merupakan sekelompok pohon regresi atau klasifikasi yang tidak dipangkas serta dibuat dengan cara memilih sampel acak dari suatu data [21]. *Random* *forest* juga mampu menggabungkan kekuatan dari banyak pohon keputusan untuk memberi prediksi yang akurat dan stabil. Prediksi ini dibuat dengan menggabungkan hasil dari seluruh kelompok pohon regresi atau klasifikasi [20]. Model ini didasarkan pada persamaan (6).

$$Gini(D) = 1 - \sum_{i=1}^{C} p_i^2 \qquad (6)$$

Di mana (p_i) adalah proporsi kelas (i) dalam dataset (D) dan (C) adalah jumlah kelas.

2.4. Decision Tree

Decision tree merupakan metode umum yang sering diterapkan dalam pengambilan keputusan [23]. Algoritma ini memberikan solusi dari permasalahan dengan menjadikan kriteria sebagai *node* yang saling berhubungan sehingga membentuk seperti struktur pohon. Setiap pohon memiliki cabang yang mewakili atribut yang harus dipenuhi untuk menuju cabang sehingga berakhir di daun. Konsep dalam algoritma ini adalah data yang dinyatakan dalam bentuk tabel [13]. *Decision Tree* memiliki kelebihan pada tingkat kakuratannya yang tinggi, tetapi lemah jika dilakukan penambahan atau pengurangan data [24]. Model ini berdasarkan pada persamaaan (7).

$$Entropy(D) = -\sum_{i=1}^{C} p_i \log_2 p_i \qquad (7)$$

Di mana(p_i) adalah proporsi elemen dari kelas (i) dalam dataset (D)

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan algoritma *machine learning* untuk mendeteksi kemungkinan keberangkatan haji dengan menyiapkan dataset untuk proses pelatihan dan pengujian. *Dataset* dalam penelitian ini berasal dari Kantor Kementerian Agama Karawang yang memiliki jumlah kurang lebih 5000 data yang mencakup 6 fitur yaitu, umur, jenis kelamin, pekerjaan, pendidikan, kecamatan dan status haji. Penelitian ini menggunakan presentase 80% data *training* dan 20% data *testing*. Berikut contoh data yang telah melewati berbagai tahapan *preprocessing* disajikkan pada gambar 2

	umur	jenis_kelamin	pekerjaan	pendidikan	kecamatan	status_haji
0	74	1	1	7	12	SDH
1	30	1	6	8	12	SDH
2	50	0	6	8	27	SDH
3	51	1	5	3	4	SDH
4	51	0	0	3	4	BLM
...	...	...	...	...	...	...
5211	43	1	1	9	3	BLM
5212	49	1	6	8	27	SDH
5213	60	1	1	8	10	BLM
5214	81	1	1	7	9	SDH
5215	81	0	4	7	9	SDH

5216 rows × 6 columns

Gambar 2. Dataset Setelah Preprocessing

Gambar 2 memperlihatkan dataset yang telah melalui tahap *preprocessing*, dimana terdapat beberapa fitur yang digunakan dalam pendeteksian faktor keberangkatan haji yang telah diubah menjadi numerik sebelum diimplementasikan menggunakan algoritma. Pada kolom umur mencakup usia individu dengan variasi yang cukup luas, jenis kelamin yang terdapat dua kategori diwakili oleh angka, pekerjaan yang terdapat variasi mencakup berbagai status ekonomi, Pendidikan yang dicatat bisa bervariasi dari yang rendah hingga tinggi, kecamatan mencakup informasi mengenai lokasi tempat tinggal individu yang dapat mengindikasikan distribusi geografis dan status haji merupakan keterangan sudah atau belum melaksanakan haji.

3.1. Implementasi Algoritma Naïve Bayes

Fungsi Gaussian Naïve Bayes digunakan dalam pengujian dataset algoritma *Naïve Bayes*. Proses ini dilakukan dengan mengimpor library sklearn pada Python, sebagaimana ditunjukkan pada *source code* yang ditunjukkan pada Gambar 3.

```
# train model naive bayes
model = GaussianNB()
model.fit(X_train, y_train)
```

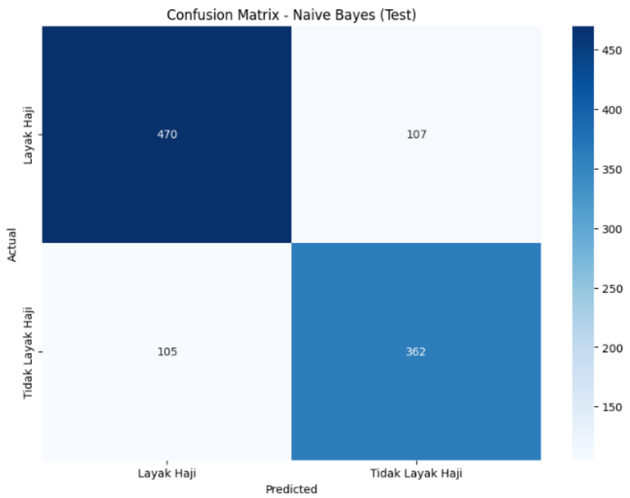
Gambar 3. Model Naïve Bayes

Pengujian tahap awal menggunakan algoritma *Naïve Bayes*, yaitu dengan melakukan evaluasi sehingga mendapatkan nilai yang akurat menggunakan *splitting data* dengan rasio 80:20. Hasil yang diperoleh menunjukkan akurasi sebesar 76,69%, presisi 79,70%, *recall* 79,69%, dan *F1-score* 79,70%. Evaluasi dilakukan dengan menggunakan Confusion Matrix untuk mendapatkan hasil klasifikasi multikelas, dengan rincian matriks dapat dilihat pada Tabel 1.

Tabel 1. Hasil Evaluasi Naïve Bayes

Kelas	**Precision**	**Recall**	**F1-Score**
Layak Haji	82%	81%	82%
Tidak Layak Haji	77%	78%	77%

Tabel 1 menyajikan hasil evaluasi menggunakan algoritma *Naïve Bayes* dengan masing-masing kelas tertentu yaitu kelas "Layak Haji" memiliki *Precision* 82%, *Recall* 81%, dan *F1-Score* 82%, sementara untuk kelas "Tidak Layak Haji" memiliki *Precision* 77%, *Recall* 78%, dan *F1-Score* 77%. Visualisasi confusion matrix pada algoritma ini dapat dilihat pada Gambar 4.



Gambar 4. Confusion Matrix Naive Bayes

3.2. Implementasi Algoritma Random Forest

Algoritma *Random Forest* digunakan pada proses ini lalu diselesaikan dengan menggunakan impor library sklearn pada Phyton, source code model dapat dilihat pada Gambar 5 dibawah ini.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

```
# train model random forest
model_rf = RandomForestClassifier(n_estimators=100, random_state=42)
model_rf.fit(X_train, y_train.values.ravel())
```

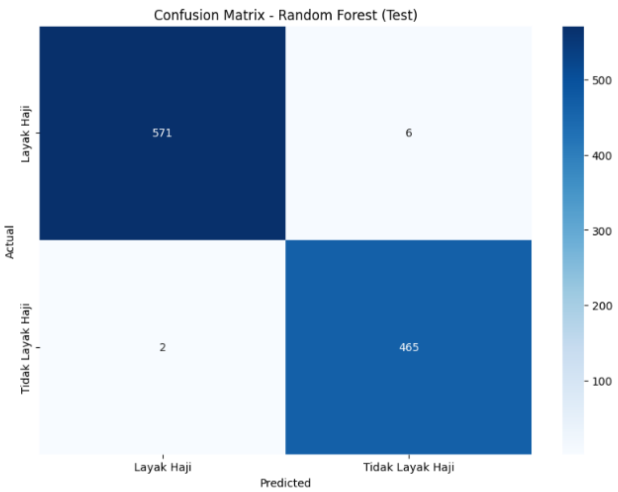
Gambar 5. Model Random Forest

Pengujian dilakukan dengan menggunakan algoritma *Random Forest* dengan rasio 80:20, yang menghasilkan akurasi sebesar 99,23%, *precision* 99,24%, *recall* 99,23%, dan *F1-Score* 99,23%. Hasil evaluasi diperoleh melalui *Confusion Matrix* untuk mendapatkan hasil klasifikasi, dan tabel *confusion matrix* yang dapat dilihat pada Tabel 2.

Tabel 2. Hasil Evaluasi Random Forest

Kelas	Precision	Recall	F1-Score
Layak Haji	100%	99%	99%
Tidak Layak Haji	99%	100%	99%

Tabel 2 menyajikkan hasil evaluasi klasifikasi data jamaah haji menggunakan algoritma *Random Forest* dengan akurasi mencapai 99%. Dimana kelas "Layak Haji" memiliki presisi yang sempurna yaitu 100%, sementara kelas "Tidak Layak Haji" juga menunjukkan kinerja yang baik dengan presisi dan F1-Score sebesar 99%. Nilai sempurna kedua kelas tersebut adalah hasil keseimbangan antara *precision* dan *recall*. Visualisasi *confusion matrix* dengan *random forest*, seperti yang ditunjukkan pada Gambar 6



Gambar 6. Confusion Matrix Random Forest

3.3. Implementasi Algoritma Decision Tree

Pada tahap ketiga, pengujian dilakukan menggunakan algoritma *Decision Tree* dengan menggunakan fungsi decision tree yang tersedia melalui impor library sklearn dalam bahasa Python. *Source code* dapat dilihat pada Gambar 6 di bawah ini.

```
# train model decision tree
model_dt = DecisionTreeClassifier(random_state=42)
model_dt.fit(X_train, y_train.values.ravel())
```

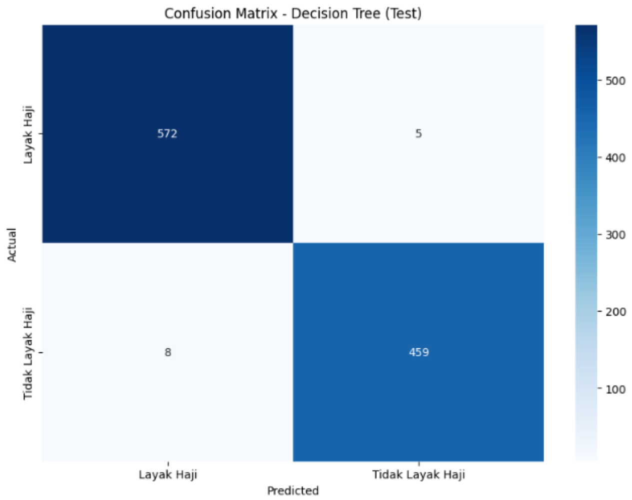
Gambar 7. Model Decision Tree

Pada tahap ini hasil dari model *decision tree* dengan rasio 80:20 memiliki hasil akurasi sebesar 98.75%, *precision* 98.76%, *recall* 98.75, dan *F1-Score* 98.75%. Hasil yang diperoleh merupakan dari proses evaluasi dengan menggunakan *Confusion Matrix* untuk mendapatkan hasil model klasifikasi. Tabel *confusion matrix* dapat dilihat pada tabel 3.

Tabel 3. Hasil Evaluasi Decision Tree

Kelas	Precision	Recall	F1-Score
Layak Haji	99%	99%	99%
Tidak Layak Haji	99%	98%	99%

Tabel 3 menunjukkan hasil evaluasi klasifikasi data Jemaah haji menggunakan algoritma *decision tree* yang menunjukkan kinerja yang sangat baik dengan hasil *precision*, *recall*, dan *F1-score* masing-masing mencapai 99% untuk kelas "Layak Haji" dan 99% *precision*, 98% *recall*, serta 99% *F1-Score* untuk kelas "Tidak Layak Haji". Visualisasi confusion matrix menggunakan algoritma decision tree dapat dilihat pada gambar 8.



Gambar 8. Confusion Matrix Decision Tree

3.4. Pembahasan

Secara keseluruhan dengan berdasarkan hasil evaluasi dari ketiga algoritma yang digunakan dengan metoode klasifikasi jamaah haji menggunakan *machine learning* terbukti efektif terutama pada algoritma *Random Forest* lebih unggul dibandingkan dengan Decision Tree dan Naïve Bayes dalam mendeteksi faktor keberangkatan jemaah haji, dengan akurasi tertinggi mencapai 99,23%. Sifat dataset yang digunakan, yang terdiri dari 5000 data demografis jemaah haji, berkontribusi pada performa model yang tinggi, di mana kualitas data dan relevansi fitur-fitur demografis seperti usia dan pekerjaan sangat mempengaruhi hasil klasifikasi. Keunggulan ini disebabkan oleh sifat Random Forest sebagai metode ensemble learning yang mampu mengurangi risiko overfitting dan meningkatkan generalisasi model, terutama pada dataset yang memiliki fitur beragam.

Dibandingkan dengan penelitian sebelumnya, akurasi Random Forest dalam penelitian ini jauh lebih tinggi daripada penelitian lain yang menggunakan algoritma serupa, seperti dalam klasifikasi penyakit jantung dan prediksi kelulusan mahasiswa. Namun, akurasi yang sangat tinggi ini juga perlu diwaspadai terhadap potensi overfitting dan bias data yang tidak seimbang.Penelitian ini mengkaji dalam beberapa aspek ilmu komputer dan aplikasi teknologi, di mana hasil penelitian menunjukan bahwa penerapan pada metode analisis data yang inovatif dan dapat meningkatkan efisiensi pada klasifikasi data yang digunakan serta memberikan solusi yang dihadapi dalam penelitian ini. Meskipun terdapat beberapa keterbatasan dalam metodologi yang digunakan dan perlu diperhatikan untuk penelitian selanjutnya, sehingga hasil ini tidak hanya memberikan kontribusi ilmiah dari penelitian ini terletak pada penerapan algoritma *machine learning* dalam konteks pengelolaan haji, yang dapat membantu Kementerian Agama dalam perencanaan sumber daya dan pelayanan yang lebih efisien, serta mendeteksi potensi masalah keberangkatan lebih awal. Penelitian ini mendorong penggunaan analisis data sebagai alat bantu dalam pengambilan keputusan strategis di sektor keagamaan, khususnya dalam pengelolaan ibadah haji di Indonesia.

4. KESIMPULAN

Tujuan dari penelitian ini adalah untuk mempelajari *variable* demografis yang berkontribusi pada keberangkatan jamaah haji dengan menggunakan tiga algoritma *machine laearning*, yaitu *Naïve Bayes*, *Random*

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Forest, dan *Decision Tree* sebagai solusi terhadap permasalahan data yang tidak seimbang. Keunggulan ini disebabkan oleh kemampuan Random Forest dalam mengurangi risiko *overfitting* dan meningkatkan generalisasi model. Penelitian ini juga menyoroti pentingnya kualitas dan relevansi data demografis dalam mempengaruhi hasil klasifikasi. Selain itu, penelitian ini memberikan kontribusi praktis bagi Kementerian Agama dalam perencanaan dan pelayanan haji yang lebih efisien, serta mendorong penggunaan analisis data sebagai alat bantu dalam pengambilan keputusan strategis di sektor keagamaanNamun terdapat keterbatasan dalam metodologi, seperti potensi keseimbangan data atau bias dalam fitur yang digunakan. Oleh karena itu, saran untuk penelitian selanjutnya yaitu dengan menambahkan fitur non-demografis atau eksplorasi teknik *ensemble learning* untuk meningkatkan generalisasi model.

DAFTAR PUSTAKA

[1] N. Hidayati *et al.*, "Penerapan Algoritma Klasterisasi dan Klasifikasi pada Tingkat Kepentingan Sistem Pembelajaran di Universitas Terbuka," 2020.

[2] Y. Heryadi and T. Wahyono, "Machine Learning: Konsep dan Implementasi," 2020. [Online]. Available: https://www.researchgate.net/publication/344419764

[3] I. Romli and A. T. Zy, "Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5," 2020.

[4] A. Pebdika, R. Herdiana, and D. Solihudin, "Klasifikasi Menggunakan Metode Naive Bayes untuk Menentukan Calon Penerima PIP," Feb. 2023.

[5] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine," *TEKNOSAINS : Jurnal Sains, Teknologi dan Informatika*, vol. 10, no. 2, pp. 176–184, Jul. 2023, doi: 10.37373/tekno.v10i2.419.

[6] A. M. Siregar, "Klasifikasi Untuk Prediksi Cuaca Menggunakan Esemble Learning," *PETIR*, vol. 13, no. 2, pp. 138–147, Sep. 2020, doi: 10.33322/petir.v13i2.998.

[7] F. Yulian Pamuji, V. Puspaning Ramadhan, and R. Artikel, "Jurnal Teknologi dan Manajemen Informatika Komparasi Algoritma Random Forest Dan Decision Tree Untuk Memprediksi Keberhasilan Immunotheraphy Info Artikel Abstrak," vol. 7, pp. 46–50, 2021, [Online]. Available: http://http://jurnal.unmer.ac.id/index.php/jtmi

[8] H. Derajad Wijaya and S. Dwiasnati, "Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat," *Jurnal Informatika*, vol. 7, no. 1, 2020, [Online]. Available: http://ejournal.bsi.ac.id/ejurnal/index.php/ji

[9] G. Awliya Muhammad Ashfania *et al.*, "Penggunaan Algoritma Random Forest untuk Klasifikasi berbasis Kinerja Efisiensi Energi pada Sistem Pembangkit Daya," 2023.

[10] N. H. Alfajr and S. Defiyanti, "Prediksi Penyakit Jantung Menggunakan Metode Random Forest dan Penerapan Principal Component Analysis (PCA)," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3S1, Oct. 2024, doi: 10.23960/jitet.v12i3S1.5055.

[11] N. A. Prakoso Indaryono, "Analisa Perbandingan Algoritma Random Forest dan Naïve Bayes untuk Klasifikasi Curah Hujan Berdasarkan Iklim di Indonesia," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 1, pp. 158–167, Feb. 2024, doi: 10.29100/jipi.v9i1.4421.

[12] I. Kurniawan, D. Cahya Putri Buani, W. Apriliah, R. Amegia Saputra, and P. Korespondensi, ""Implementasi Algoritma Random Forest Untuk Menentukan Penerima Bantuan Raskin Implementation of Random Forest Algorithm For Determining Recipients Of Raskin," vol. 10, no. 2, pp. 421–428, 2023, doi: 10.25126/jtiik.202396225.

[13] F. Yulian Pamuji, V. Puspaning Ramadhan, and R. Artikel, "Jurnal Teknologi dan Manajemen Informatika Komparasi Algoritma Random Forest Dan Decision Tree Untuk Memprediksi Keberhasilan Immunotheraphy Info Artikel ABSTRAK," vol. 7, pp. 46–50, 2021, [Online]. Available: http://http://jurnal.unmer.ac.id/index.php/jtmi

[14] Y. E. Yuspita, R. Okra, and M. Rezeki, "Penerapan Algoritma Klasifikasi Untuk Prediksi Tingkat Kelulusan Mahasiswa Menggunakan RapidMiner," *Jurnal Teknologi Informasi*, vol. 6, no. 1, 2025, doi: 10.46576/djtechno.

[15] D. Haganta Depari *et al.*, "Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung," *JURNAL INFORMATIK Edisi ke*, vol. 18, p. 2022, 2022.

[16] Erfan Karyadiputra and Agus Setiawan, "Sistem pendukung keputusan berbasis decision tree algorithm untuk prediksi penyakit diabetes," *Teknosains*, Dec. 2023, doi: https://doi.org/10.24252/teknosains.v17i3.38383.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

[17] T. Gori, A. Sunyoto, and H. Al Fatta, "Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 1, pp. 215–224, Feb. 2024, doi: 10.25126/jtiik.20241118074.

[18] M. Nurhariza, A. Ratna Juwita, and D. Sulistya Kusumaningrum, "Implementasi Algoritma Naive Bayes Untuk Klasifikasi Menentukan Prestasi Siswa Berdasarkan Nilai Rata-Rata," no. 1, 2024.

[19] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine," *TEKNOSAINS : Jurnal Sains, Teknologi dan Informatika*, vol. 10, no. 2, pp. 176–184, Jul. 2023, doi: 10.37373/tekno.v10i2.419.

[20] A. M. Siregar, "Accounting Information System Perbandingan Algoritme Klasifikasi Untuk Prediksi Cuaca," 2020.

[21] S. Wahyuni Kalumbang, "Perbandingan Regresi Logistik, Klasifikasi Naive Bayes, dan Random Forest (Comparison The Logistic Regression, Naive Bayes Classification, and Random Forest)," vol. 03, no. 02, p. 2021, 2021.

[22] A. Salam, L. Azhari, R. S. Septarini, and N. Heriyani, "BULLETIN OF COMPUTER SCIENCE RESEARCH Pendekatan Hybrid K-Means SMOTE dan Logistic Regression Untuk Deteksi Dini Diabetes Mellitus Pada Imbalanced Data," *Media Online)*, vol. 5, no. 3, pp. 222–230, 2025, doi: 10.47065/bulletincsr.v5i3.502.

[23] D. Sayhidin, G. Haris, and C. Juliane, "Implementasi Data Mining Tingkat Kepemimpinan Siswa dengan K-Nearest Neighbor, Decision Tree, dan Naïve Bayes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, no. 1, p. 199, Jan. 2023, doi: 10.30865/mib.v7i1.5351.

[24] Ahmad Taufiq Ramadhan, Faishal Hilmy F. G, Nadya Rafaela Puteri, and Alifya Meirza, "Penerapan Algoritma Decision Tree Dalam Melakukan Analisis Klasifikasi Harga Handphone," *Jurnal Sistem Informasi dan Ilmu Komputer*, vol. 1, no. 4, pp. 195–206, Nov. 2023, doi: 10.59581/jusiik-widyakarya.v1i4.1861.