

### BAB III

#### METODE PENELITIAN

##### 3.1 Objek Penelitian

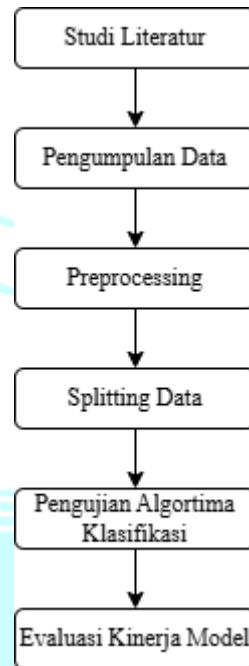
Objek penelitian ini adalah klasifikasi data calon jamaah haji tahun 2025 dengan jumlah data sebanyak 5000 jamaah haji Kabupaten Karawang yang dilihat berdasarkan faktor demografis yang menggunakan algoritma *Machine Learning* untuk meningkatkan efisiensi pengelolaan data. Penelitian ini bertujuan untuk mengklasifikasikan data berdasarkan fitur-fitur tertentu yang diambil dari beberapa kolom di dalamnya. Penelitian ini meningkatkan akurasi dan efisiensi dalam klasifikasi data calon jamaah haji dibandingkan dengan metode manual. Pengumpulan data pada penelitian ini dilakukan dengan observasi dan wawancara bersama narasumber yaitu kepala seksi bagian Haji dan Umrah Bapak H Iwan Suryawan dan bagian komputerisasi jamaah haji yaitu Ibu Hj Popon. Lokasi penelitian dilakukan di Kantor Kementerian Agama Karawang Jalan Husni Hamid No. 1, Karawang bagian Pelayanan Haji dan Umrah Terpadu. Waktu penelitian berlangsung mulai dari bulan November hingga April 2025 yang mencakup tahapan-tahapan penelitian. Waktu penelitian dapat dilihat pada tabel 3.1

Tabel 3. 1 Waktu Penelitian

Tahun: 2025					
No	Tahap penelitian	Bulan 1	Bulan 2	Bulan 3	Bulan 4
1	Pengumpulan Data	■			
2	<i>Preprocessing</i>				
3	<i>Split data</i>		■		
4	Implementasi Algoritma			■	
5	Hasil Evaluasi				■

### 3.2 Prosedur Penelitian

Untuk memperoleh hasil yang diharapkan dalam penelitian, diperlukan langkah-langkah penelitian yang terstruktur dan sistematis. Berikut prosedur yang di lakukan dapat dilihat pada gambar 3.1.



Gambar 3. 1 Prosedur Penelitian

Mengacu pada gambar, berikut adalah penjelasan dari masing-masing tahapan penelitian yang telah diimplementasikan.

#### 1. Studi Literatur

Tahapan awal pada penelitian ini dimulai 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 relevan dan berkaitan dengan topik yang dibahas.

#### 2. 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. Data yang digunakan yaitu menggunakan data dari tiga tahun sebelumnya. Contoh dataset dapat dilihat pada gambar 3.2

no	estimasi	no_porsi	tanggal_daftar	nama	nama_ayah	jenis_kelamin	0	pekerjaan	pendidikan	...	no_rekening	status_haji	no_telepon
1	2022	1000543675	2012-07-18 00:00:00	ASEP SAEPUDIN	RAMAN, H.	L	1	SWASTA	SLTA	...	7039757798	BLM	0858 85642922
2	2022	1000539819	2012-07-03 00:00:00	MEMED CARMEDI	SOFWAN SIDIK, H	L	2	PETANI	S.1	...	7039036465	BLM	0856 8600529
3	2022	1000531204	2012-05-31 00:00:00	OZI	KADAN	L	3	SWASTA	SLTA	...	7037470273	BLM	0858 90117618
4	2022	1000539231	2012-07-02 00:00:00	FAUZAN	FARHAN	L	4	SWASTA	SLTP	...	7038967953	BLM	0857 21343376
5	2022	1000536663	2012-06-21 00:00:00	AEP SAEFULLOH	ZAINUDIN, H	L	5	SWASTA	S.1	...	7038475888	BLM	0857 21343376
...	...	...	...	...	...	...	...	...	...	...	...	...	...
979	2022	1000524066	2012-05-07 00:00:00	SULASTRI	IMAN TARUJI	P	13	IBU RUMAH TANGGA	SLTA	...	7036493253	BLM	0813 11306915
980	2022	1000523022	2012-05-03 00:00:00	NENI MULYANI	SUHAEMI	P	14	SWASTA	SLTA	...	7036264076	BLM	0813 10719302

Gambar 3. 2 Contoh Dataset

### 3. Preprocessing

Pada tahap ini diperlukan untuk memaksimalkan algoritma klasifikasi, tujuan *preprocessing* yaitu mentransformasikan data ke suatu format yang prosesnya menjadi efektif. Sebelum di evaluasi menggunakan algoritma, yaitu memberi label pada data untuk masing-masing kelas serta pembersihan data yang mencakup mengidentifikasi dan memilih fitur yang akan digunakan serta menghapus data yang tidak konsisten, dan menghilangkan atribut yang tidak berkontribusi terhadap analisis yang ingin dicapai.

**Cek List Group dalam status\_haji - Before Cleaning**

```
[33] df['status_haji'].unique()
array(['BLM', 'SDH', 'BELUM', 'Belum', 'PASPOR ASLI'], dtype=object)
```

```
[34] df['status_haji'] = df['status_haji'].replace({
    'BELUM': 'BLM',
    'Belum': 'BLM',
    'PASPOR ASLI': 'SDH',
})
```

**Cek List Group dalam status\_haji - After Cleaning**

```
[35] df['status_haji'].unique()
array(['BLM', 'SDH'], dtype=object)
```

Gambar 3. 3 Contoh Preprocessing

#### 4. *Splitting Data*

Dalam *machine learning*, *splitting* data digunakan untuk membagi dataset menjadi dua komponen utama yaitu data latih (*Data Training*) dan data uji (*Data testing*). Pembagian jumlah data antara *data training* dan data testing adalah faktor penting dalam menentukan tingkat akurasi. Dalam penelitian ini, presentase data yang digunakan yaitu 80% *data training* dan 20% *data testing*

#### 5. Pengujian Algoritma Klasifikasi

Model klasifikasi yang diterapkan dalam penelitian ini adalah 3 algoritma *machine learning* yaitu *Naïve Bayes*, *Random Forest* dan *Decision Tree*. *Naïve Bayes* yang berbasis pada teorema bayes, efektif untuk data dengan asumsi indenpedensi antar fitur sedangkan *Random Forest* merupakan ensemble dari beberapa pohon keputusan yang mampu menangani *overfitting* dan memberikan hasil yang stabil. *Decision Tree* mengusulkan interpretabilitas yang tinggi dengan visualisasi yang jelas. Algoritma ini dipilih karena untuk menentukan mana yang paling sesuai untuk dataset yang digunakan. Pada tahap ini merupakan proses implementasi ketiga algoritma tersebut dengan menggunakan bahasa pemrograman *python* dan dioperasikan di *google colab*.

#### 6. Evaluasi Kinerja Model

Langkah terakhir dalam menilai kinerja algoritma klasifikasi yang digunakan dalam penelitian ini adalah evaluasi. Ada empat metrik yang digunakan untuk mengukur kinerja model 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.