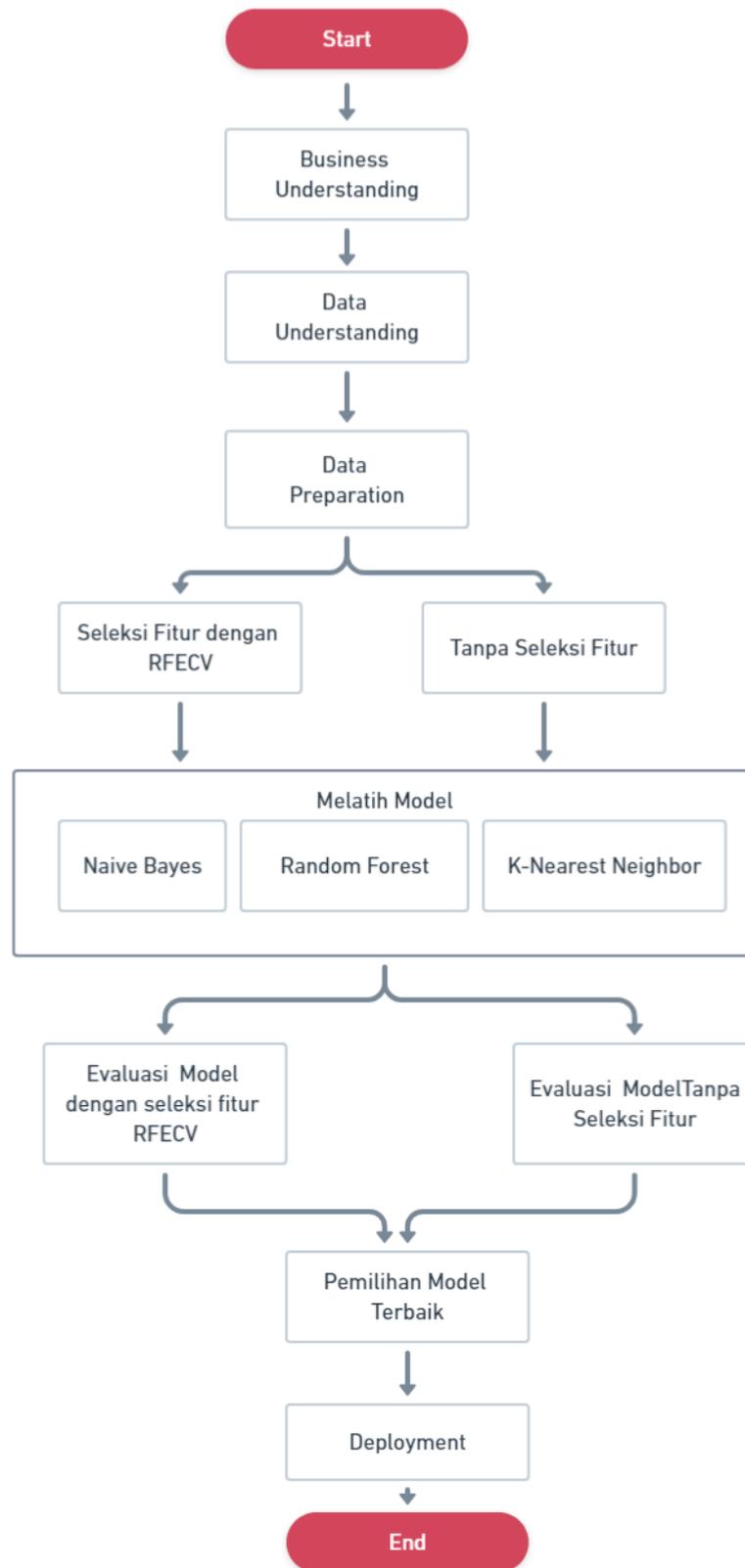


BAB III

METODOLOGI PENELITIAN

3.1 Metode Penelitian

Metode yang digunakan dalam penelitian ini menggunakan standard Cross-Industry Standard Process for *Data mining* (*CRISP-DM*) yang terdiri dari beberapa tahapan yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment* [30]. Berdasarkan tahapan *CRISP-DM* tersebut maka langkah-langkah yang akan dilakukan dalam pembuatan model prediksi pembatalan haji oleh jemaah haji dapat dilihat pada gambar 3.1 berikut :



Gambar 3.1 Langkah kerja implementasi *CRISP-DM*

Tahapan *CRISP-DM* yang akan dilakukan dalam penelitian ini adalah sebagai berikut :

1. *Business Understanding*

Penetapan tujuan bisnis dalam rangka memprediksi jemaah haji yang berpotensi membatalkan haji adalah suatu langkah krusial yang harus dilakukan untuk memastikan bahwa penelitian dapat berjalan dengan efektif dan hasilnya dapat memberikan manfaat yang signifikan bagi pemangku kepentingan terkait. Tujuan bisnis yang jelas dan spesifik akan membantu dalam menentukan parameter dan kriteria keberhasilan, serta menentukan metode dan teknik analisis yang tepat untuk mencapai tujuan tersebut.

Salah satu tujuan bisnis yang ditetapkan dalam penelitian ini adalah untuk memprediksi jemaah haji yang berpotensi membatalkan haji sehingga dapat dilakukan tindakan preventif untuk mengurangi jumlah pembatalan haji. Pembatalan haji dapat memiliki dampak yang signifikan terhadap banyak orang, termasuk jemaah haji, keluarga, dan juga bisa menjadi beban bagi penyelenggara haji. Faktor-faktor yang menyebabkan Jemaah haji membatalkan haji perlu diidentifikasi sehingga dapat dilakukan tindakan pencegahan yang tepat.

Tujuan bisnis yang jelas dan spesifik dapat membantu menentukan kriteria keberhasilan. Kriteria keberhasilan yang baik akan membantu dalam mengevaluasi keefektifan langkah-langkah yang diambil dalam pengambilan keputusan yang lebih baik. Salah satu kriteria keberhasilan yang dapat ditetapkan adalah akurasi

prediksi yang tinggi. Akurasi prediksi yang tinggi akan memungkinkan organisasi untuk mengambil tindakan preventif yang lebih efektif dan dapat membantu dalam mengurangi jumlah pembatalan haji.

Hasil prediksi dapat digunakan untuk mengambil tindakan preventif seperti memberikan edukasi dan pengawasan yang lebih intensif kepada jemaah haji yang berpotensi membatalkan haji. Dengan demikian, diharapkan jumlah pembatalan haji dapat dikurangi dan jemaah haji dapat melaksanakan ibadah haji dengan lebih aman dan nyaman.

2. Data Understanding

Pada tahap ini, dilakukan pengumpulan data yang akan digunakan. Kumpulan data tersebut selanjutnya dipilah untuk memperoleh data yang seharusnya diperlukan. Karakteristik data yang diperoleh akan mempengaruhi jenis algoritma dan tujuan yang ingin dicapai. Sumber data yang digunakan dalam penelitian ini menggunakan data jemaah haji dari Seksi Penyelenggaraan Haji dan Umroh Kantor Kementerian Agama Kabupaten Pringsewu tahun 2019-2024.

Dari data yang didapatkan akan dilakukan pemilihan kolom sebagai atribut dan label untuk proses klasifikasi dengan algoritma *Naive Bayes*, *Random Forest* dan *K-Nearest Neighbor*.

3. Data Preparation

Untuk mendapatkan model yang baik dan akurat diperlukan *Data Preparation* yang baik. Pada tahap preparation akan dilakukan beberapa tahapan sebagai berikut :

- a. Melakukan pembersihan data (data cleaning) dengan cara menghapus data yang tidak lengkap, data duplikat, dan data yang tidak relevan.
- b. Melakukan integrasi data (data integration) dengan menggabungkan data dari berbagai sumber menjadi satu dataset yang utuh.
- c. Melakukan transformasi data (data transformation) seperti melakukan pengkodean (encoding) pada variabel yang bersifat kategorikal dan normalisasi pada variabel numerik.
- d. Melakukan seleksi fitur (feature selection) dengan cara memilih variabel yang paling relevan dan berpengaruh terhadap prediksi.
- e. Memisahkan data menggunakan *Cross Validation*
- f. Mengidentifikasi kesalahan pada data
- g. Melakukan pengecekan *missing value* dan data yang tidak konsisten

4. Modeling

Pada tahap *Modeling* akan dilakukan proses klasifikasi dengan menggunakan algoritma *Naive Bayes*, *Random Forest* dan *K-Nearest Neighbor*. Langkah awal dalam tahap ini yaitu melakukan pelatihan model menggunakan data training terlebih dahulu untuk menghasilkan sebuah model yang dapat digunakan saat pengujian data. Untuk mengevaluasi performa model, hasil model, dan performa model harus ditampilkan sebagai indikator evaluasi pada tahap berikutnya.

5. Evaluation

a. Menentukan kriteria evaluasi:

Pada fase *Evaluation*, akan ditentukan kriteria evaluasi terlebih dahulu untuk membandingkan performa ketiga algoritma tersebut. Beberapa kriteria evaluasi yang digunakan yaitu Akurasi, Precision, dan Recall.

b. Mengevaluasi kinerja algoritma:

Selanjutnya, dilakukan evaluasi kinerja ketiga algoritma pada data uji dengan menggunakan kriteria evaluasi yang telah ditentukan. Hasil evaluasi kinerja nantinya akan digunakan untuk membandingkan performa ketiga algoritma.

c. Memilih algoritma terbaik:

Setelah melakukan evaluasi kinerja ketiga algoritma, langkah selanjutnya adalah memilih algoritma terbaik berdasarkan kriteria evaluasi yang telah ditentukan. Misalnya, jika kriteria evaluasi yang dipilih adalah Akurasi, maka algoritma dengan nilai Akurasi tertinggi akan dipilih sebagai algoritma terbaik.

d. Mengevaluasi model terbaik:

Setelah memilih algoritma terbaik, langkah selanjutnya adalah mengevaluasi model terbaik dengan menggunakan data uji. Hal ini bertujuan untuk memastikan bahwa model terbaik yang dipilih benar-benar memiliki kinerja yang baik pada data yang belum pernah dilihat sebelumnya.

e. Menyusun laporan hasil evaluasi:

Langkah terakhir adalah menyusun laporan hasil evaluasi yang berisi hasil evaluasi kinerja ketiga algoritma, pemilihan algoritma terbaik, evaluasi model terbaik, serta kesimpulan dan rekomendasi yang didapat dari hasil evaluasi tersebut.

6. Deployment

Hasil penelitian dapat diterapkan pada organisasi, untuk membantu menentukan Langkah-langkah strategi atau kebijakan sesuai tujuan bisnis. Hasil penelitian ini diharapkan dapat menjadi sumber referensi untuk penelitian selanjutnya, sehingga saran dan rekomendasi dari penelitian ini diperlukan agar penelitian selanjutnya dapat lebih baik lagi.

3.2 Metode Pengumpulan Data

Pada penelitian ini metode yang digunakan dalam pengumpulan data sebagai berikut :

1. Wawancara

Wawancara kepada pegawai pada Seksi Penyelenggaraan Haji dan Umroh Kantor Kementerian Agama Kabupaten Pringsewu. Wawancara dilakukan untuk mengetahui masalah dan tujuan bisnis agar relevan dengan penelitian. Dalam tahap wawancara, akan diajukan pertanyaan-pertanyaan yang tepat dan terstruktur, serta mendengarkan dengan seksama jawaban yang diberikan oleh responden. Hasil dari tahap wawancara akan menjadi bagian penting dalam analisis data dan pembuatan

model prediksi yang akan digunakan untuk mengidentifikasi jamaah haji yang berpotensi membatalkan haji.

2. Studi Pustaka

Studi pustaka dilakukan dengan mencari menggali informasi dan pengetahuan dari penelitian sebelumnya, serta publikasi dari buku, jurnal atau informasi lainnya dengan harapan pengetahuan tersebut dapat menjadi pendukung dalam penelitian ini terutama tentang teori, pendapat, serta pemikiran yang berkaitan dengan penelitian ini.

3.3 Tempat dan Waktu Penelitian

1. Tempat Penelitian

Tempat penelitian ini akan dilaksanakan di Seksi Penyelenggaraan Haji dan Umroh Kantor Kementerian Agama Kabupaten Pringsewu

2. Waktu Penelitian

Waktu penelitian dimulai pada bulan Mei 2024 sampai dengan Juli 2024.

Tabel 3.1 Jadwal Penelitian

Kegiatan	Mei				Juni				Juli			
	Minggu ke-											
	1	2	3	4	1	2	3	4	1	2	3	4
Business Understanding	■	■										
Data Understanding		■	■	■								
Data Preparation				■	■	■	■					
Modeling							■	■	■	■		
Evaluation								■	■	■	■	
Deployment												■