

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1 Hasil Penelitian

4.1.1 *Business Understanding*

Tujuan utama dari penelitian ini adalah untuk membantu Kementerian Agama khususnya Kantor Kementerian Agama Kabupaten Pringsewu dalam memprediksi calon jemaah haji yang berpotensi membatalkan haji dengan memanfaatkan Teknik *data mining* dengan target kriteria keberhasilan mencapai tingkat akurasi prediksi minimal 90%.

Langkah-langkah yang dilakukan meliputi pengumpulan data historis pendaftaran calon jemaah haji, pembersihan dan persiapan data untuk analisis, analisis data untuk mengidentifikasi pola dan tren yang relevan, pengembangan model prediksi menggunakan algoritma *Naive Bayes*, *Random Forest* dan *K-Nearest Neighbor* , serta pengujian dan validasi model dengan data uji untuk memastikan akurasi dan keandalan model prediksi. Setelah itu, performa antara *Naive Bayes*, *Random Forest* dan *K-Nearest Neighbor* akan dibandingkan berdasarkan akurasi, presisi, dan recall, sebelum menerapkan model terbaik dalam lingkungan produksi dan memantau kinerjanya secara berkala untuk perbaikan lebih lanjut.

4.1.2 Data Understanding

Data yang digunakan dalam penelitian ini adalah dataset yang diperoleh dari Sistem Komputerisasi Haji Terpadu (Siskohat) di wilayah Kantor Kementerian Agama Kabupaten Pringsewu dalam rentang tahun 2019 – 2024. Jumlah data didapatkan sebanyak 1.133 record dan 22 atribut. Hasil pengumpulan data disajikan dalam Tabel 4.1 Dataset Keberangkatan dan Pembatalan Calon Jemaah Haji :

Tabel 4.1 Dataset Keberangkatan dan Pembatalan Calon Jemaah Haji

No. Porsi	Tgl Daftar	Nama	Nama Ayah	Jenis Kelamin	Pekerjaan	Pendidikan	Tempat Lahir	Tgl Lahir	Alamat	Desa	Kecamatan	Kode Pos	Kab / Kota	Bank	Cabang	Jml Setoran	S. Aktif	S. Bayar	S. Haji	No. KTP	Status
0800089740	11/01/2012	ENDY SUSILO	SUBAGIO	L	Swasta	S1	TANJUNG	20/04/1972	PRINGSEW PRINGSEW PRINGSEWU			35373	KAB. PRIN	BSM	KCP PRINGSEWU	25000000	ACTIVE	CICIL	BELUM	01200472	Berangkat
0800089742	11/01/2012	TITIN YENI HADI	SUMAR	P	Swasta	S1	GISTING	13/09/1978	PRINGSEW PRINGSEW PRINGSEWU			35373	KAB. PRIN	BSM	KCP PRINGSEWU	25000000	ACTIVE	CICIL	BELUM	01530978	Berangkat
0800091030	25/01/2012	SUHANTO	SASANASMAC	L	Pegawai Negri Sip	SLTA	SIDOHARJ	22/08/1964	SIDOHARJ SIDOHARJ PRINGSEWU			35373	KAB. PRIN	BSM	KCP PRINGSEWU	25000000	ACTIVE	CICIL	BELUM	01220864	Berangkat
0800092737	20/02/2012	MUHAMMAD NUHAIRI		L	Dagang	SLTA	PRINGSEW	15/03/1970	LINGKUNG PRINGSEW PRINGSEWU			35373	KAB. PRIN	BSM	KCP PRINGSEWU	25000000	ACTIVE	CICIL	BELUM	01150370	Berangkat
0800092738	20/02/2012	SUPRANTI MI MIYONO		P	Ibu Rumah Tangg	S1	TANJUNG	30/06/1972	LINGKUNG PRINGSEW PRINGSEWU			35373	KAB. PRIN	BSM	KCP PRINGSEWU	25000000	ACTIVE	CICIL	BELUM	01700672	Berangkat
0800096135	27/04/2012	MUNFARIDAH NASUHA		P	Swasta	SD	PONCOWI	10/06/1965	PRINGKUN PRINGSEW PRINGSEWU			35373	KAB. PRIN	BRIS	KCP LAMPUNG PF	25000000	ACTIVE	CICIL	BELUM	01500665	Berangkat
0800096147	27/04/2012	BARUDIN HASANUDIN		L	Tani / Nelayan	SD	KUTOARJ	20/06/1967	PRINGKUN PRINGSEW PRINGSEWU			35373	KAB. PRIN	BSM	KCP PRINGSEWU	25000000	ACTIVE	CICIL	BELUM	01200667	Berangkat
0800096681	10/05/2012	LILIS SUHAYA WIHATMA		P	Pegawai Negri Sip	S1	CIKONENG	11/02/1966	JL.GOTON PRINGSEW PRINGSEWU			35373	KAB. PRIN	BRIS	KCP LAMPUNG PF	25000000	ACTIVE	CICIL	BELUM	01510266	Berangkat
0800096682	10/05/2012	IMOP SUTOP SOEDARMA		L	Dagang	S1	KOTA BUN	23/03/1964	JL.GOTON PRINGSEW PRINGSEWU			35373	KAB. PRIN	BRIS	KCP LAMPUNG PF	25000000	ACTIVE	CICIL	BELUM	01230364	Berangkat
0800096865	15/05/2012	SUWARTO SC SOMO WA		L	Pensiunan	D1 / D2	MAGELAN	05/07/1947	JL.PEMUD PAGELARA PAGELARAN			35375	KAB. PRIN	BSM	KCP PRINGSEWU	25000000	ACTIVE	CICIL	BELUM	05050747	Berangkat
0800097010	21/05/2012	SUKINAH YOS YOSO SUW		P	Ibu Rumah Tangg	SD	PODOMOI	30/06/1962	PODOMOIPODOMOI PRINGSEWU			35373	KAB. PRIN	BRIS	KCP LAMPUNG PF	25000000	ACTIVE	CICIL	BELUM	01700662	Berangkat
0800097063	22/05/2012	WASIMAN ASASMA DIM		L	Pegawai Negri Sip	S1	CILACAP	10/11/1963	JL.SATRIA PRINGSEW PRINGSEWU			35373	KAB. PRIN	BSM	KCP PRINGSEWU	25000000	ACTIVE	CICIL	SUDAH	01101163	Berangkat
0800097064	22/05/2012	SITI MUFLIHA ABDUL		P	Pegawai Negri Sip	S1	PRINGSEW	20/05/1964	JL.SATRIA PRINGSEW PRINGSEWU			35373	KAB. PRIN	BSM	KCP PRINGSEWU	25000000	ACTIVE	CICIL	SUDAH	01600056	Berangkat
0800097198	24/05/2012	SUNARTI MUI MUHAMM		P	Ibu Rumah Tangg	SD	GISTING	07/12/1960	BANDUNG BANDUNGADILUWIH			35374	KAB. PRIN	BRIS	KCP LAMPUNG PF	25000000	ACTIVE	CICIL	BELUM	07471260	Berangkat
0800097205	24/05/2012	BADRUDIN BI BUKHORI		L	Dagang	SLTP	WARGO M	02/01/1960	WARGO M WARGO M PARDASUKA			35382	KAB. PRIN	BSM	KCP PRINGSEWU	25000000	ACTIVE	CICIL	BELUM	04020160	Berangkat
0800097206	24/05/2012	ALIAH MUN MUNAWIR		P	Ibu Rumah Tangg	SLTP	WARGO M	07/08/1968	WARGO M WARGO M PARDASUKA			35382	KAB. PRIN	BSM	KCP PRINGSEWU	25000000	ACTIVE	CICIL	BELUM	04470868	Berangkat
0800097357	29/05/2012	SUKINAH AM AMATDIME		P	Pegawai Negri Sip	D1 / D2	WARINGIN	29/06/1962	WARINGIN WARINGIN SUKOHARJO			35374	KAB. PRIN	BSM	KCP PRINGSEWU	25000000	ACTIVE	CICIL	BELUM	08690662	Berangkat
0800097358	29/05/2012	TUMIRANUDI KARTOREJC		L	Pegawai Negri Sip	D1 / D2	WARINGIN	17/09/1959	WARINGIN WARINGIN SUKOHARJO			35374	KAB. PRIN	BSM	KCP PRINGSEWU	25000000	ACTIVE	CICIL	BELUM	08170959	Berangkat

Hasil pengumpulan data didapatkan 1.134 record data jemaah haji yang telah berangkat dan membatalkan porsi hajinya.

4.1.3 Data Preparation

Data yang diperoleh dari pengumpulan data, kemudian diperiksa untuk mendeteksi nilai yang hilang (*missing value*) dan duplikasi data.

```
# Cek jumlah missing value per kolom
missing_values = dataset.isnull().sum()
print("Jumlah missing value per kolom:")
print(missing_values)

total_missing = missing_values.sum()
print(f"Total missing value di seluruh dataset: {total_missing}")
```

```
Jumlah missing value per kolom:
nomor_porsi      0
tanggal_daftar  0
usia_daftar     0
nama             0
nama_ayah       0
jenis_kelamin   0
pekerjaan       0
pendidikan      0
tempat_lahir    0
tanggal_lahir   0
alamat          0
desa            0
kecamatan       0
kode_pos        0
kab_kota        0
bank            0
cabang          0
jumlah_setoran  0
s_aktif         0
s_bayar         0
s_haji          0
nik             0
status          0
dtype: int64
Total missing value di seluruh dataset: 0
```

Gambar 4.1 Deteksi Nilai Hilang (*missing value*)

```
import pandas as pd

# Mengecek jumlah baris yang duplikat
jumlah_duplikasi = dataset.duplicated().sum()
print(f"Jumlah baris yang duplikat: {jumlah_duplikasi}")

# Melihat baris yang duplikat
baris_duplikat = dataset[dataset.duplicated()]
print(baris_duplikat)

Jumlah baris yang duplikat: 0
Empty DataFrame
Columns: [nomor_porsi, tanggal_daftar, usia_daftar, nama, nama_ayah, jenis_kelamin,
Index: []

[0 rows x 23 columns]
```

Gambar 4.2 Deteksi Duplikasi Data

Dari hasil deteksi tidak ditemukan duplikasi data dan nilai yang hilang (*missing value*) pada dataset.

Selanjutnya dilakukan transformasi data dengan mengubah data yang memiliki nilai kategorikal menjadi numerik. Sebagian besar algoritma *machine learning* dan statistik memerlukan data dalam bentuk numerik untuk dapat melakukan perhitungan yang relevan. Oleh karena itu, diperlukan langkah untuk mengubah data kategorikal ini menjadi format numerik. Metode yang akan digunakan pada penelitian ini menggunakan *label encoding*.

Seleksi fitur menggunakan teknik *Recursive Feature Elimination* diterapkan untuk mengidentifikasi atribut yang paling berpengaruh dan dapat meningkatkan performa model prediksi. Untuk memperoleh hasil pengujian dan performa yang lebih akurat, model prediksi akan dibandingkan antara dataset yang telah melalui

proses seleksi fitur menggunakan *Recursive Feature Elimination* dan dataset yang tidak melalui seleksi fitur.

4.1.4 Modeling

4.1.4.1 Persiapan Model

Dataset calon jemaah haji dipersiapkan untuk proses modeling dengan menerapkan teknik 10-fold *Cross-Validation*. Teknik ini membagi dataset menjadi sepuluh subset (folds), di mana setiap fold digunakan secara bergantian sebagai data uji, sementara sembilan fold lainnya digunakan sebagai data latih. Proses ini diulang sepuluh kali sehingga setiap fold berperan sebagai data uji satu kali, memastikan bahwa setiap bagian dataset diuji dan dilatih dengan merata. Pendekatan ini diterapkan baik pada dataset yang telah melalui proses seleksi fitur menggunakan teknik *Recursive Feature Elimination* maupun pada dataset yang tidak mengalami seleksi fitur.

Dengan menggunakan teknik *Cross-Validation*, evaluasi model menjadi lebih baik dan hasil performa model lebih dapat diandalkan karena meminimalkan kemungkinan *overfitting* dan memberikan gambaran yang lebih akurat mengenai kemampuan generalisasi model pada data yang belum pernah dilihat sebelumnya.

4.1.4.2 Pemodelan Tanpa Seleksi Fitur

Pada tahap ini, dilakukan pembuatan model prediksi dengan algoritma *Naïve Bayes*, *Random Forest*, dan *K-Nearest Neighbor* tanpa seleksi fitur. Semua fitur yang ada

dalam dataset akan digunakan secara langsung dalam proses pemodelan, tanpa ada pengurangan atau pemilihan fitur berdasarkan relevansi atau pentingnya terhadap target prediksi.

Pemodelan tanpa seleksi fitur memungkinkan model untuk memanfaatkan semua informasi yang tersedia dalam dataset, termasuk fitur yang mungkin dianggap kurang relevan atau memiliki korelasi rendah dengan target. Pendekatan ini memberikan keuntungan dalam beberapa kasus, terutama ketika tidak ada informasi yang jelas tentang mana fitur yang paling penting atau ketika dataset relatif kecil sehingga mengurangi risiko *overfitting*.

Namun, penggunaan semua fitur tanpa seleksi juga memiliki potensi kelemahan. Model bisa menjadi lebih kompleks dan memerlukan lebih banyak sumber daya komputasi, terutama untuk algoritma seperti *Random Forest* yang secara alami memiliki kompleksitas tinggi. Selain itu, fitur yang tidak relevan atau (noise) bisa mengurangi akurasi model dan menyebabkan *overfitting*, di mana model menjadi sangat sesuai dengan data training tetapi tidak mampu melakukan generalisasi dengan baik pada data baru.

Meskipun demikian, pendekatan tanpa seleksi fitur dapat memberikan gambaran awal tentang bagaimana model bekerja dengan menggunakan semua data yang tersedia. Hasil dari tahap ini bisa menjadi titik awal untuk kemudian melakukan analisis lebih lanjut, seperti seleksi fitur, untuk meningkatkan kinerja model dan membuatnya lebih efisien.

Pemodelan menggunakan algoritma *Naïve Bayes*, *Random Forest* dan *K-Nearest Neighbor* tanpa menggunakan seleksi fitur menghasilkan *Random Forest* sebagai model terbaik dengan rata-rata akurasi 94.97%, precision 95%, recall 94%.

Rata-rata Akurasi, Precision, dan Recall:			
	Mean Accuracy	Mean Precision	Mean Recall
Naive Bayes	0.929444	0.932869	0.929444
k-NN	0.934715	0.935113	0.934715
Random Forest	0.949721	0.950539	0.949721

Accuracy per Fold	
Naive Bayes	[0.9298245614035088, 0.8859649122807017, 0.912...
k-NN	[0.9385964912280702, 0.8947368421052632, 0.938...
Random Forest	[0.9473684210526315, 0.9210526315789473, 0.947...

Precision per Fold	
Naive Bayes	[0.947685703620754, 0.8969923292160054, 0.9122...
k-NN	[0.945793835711117, 0.9015171858216972, 0.94066...
Random Forest	[0.9513963480128893, 0.9236221867800815, 0.945...

Recall per Fold	
Naive Bayes	[0.9298245614035088, 0.8859649122807017, 0.912...
k-NN	[0.9385964912280702, 0.8947368421052632, 0.938...
Random Forest	[0.9473684210526315, 0.9210526315789473, 0.947...

Gambar 4.3 Evaluasi Model Tanpa Seleksi Fitur

Hasil pemodelan tanpa menggunakan seleksi fitur tersebut disajikan pada Tabel 4.2

Hasil Evaluasi Model Tanpa Seleksi Fitur.

Tabel 4.2 Hasil Evaluasi Model Tanpa Seleksi Fitur

Model	Accuracy (%)	Precision (%)	Recall (%)
<i>Naïve Bayes</i>	92.94%	93.28%	92.94%
<i>Random Forest</i>	94.97%	95.05%	94.97%
<i>K-Nearest Neighbor</i>	93.47%	93.51%	93.47%

Berdasarkan hasil pengujian menggunakan *Cross-Validation* dengan 10 fold, didapatkan bahwa model *Random Forest* memiliki akurasi rata-rata sebesar 94.97%, merupakan nilai tertinggi di antara ketiga model. Hal ini menunjukkan bahwa *Random Forest* memiliki performa terbaik dalam memprediksi label secara keseluruhan. Selain itu, precision rata-rata sebesar 95.05% menunjukkan bahwa model ini sangat akurat dalam mengidentifikasi positif, melebihi *Naive Bayes* dan *k-NN*. Recall rata-rata sebesar 94.97% menunjukkan bahwa *Random Forest* juga sangat efektif dalam menangkap kasus positif, sebanding dengan precision-nya.

Dengan demikian, *Random Forest* menunjukkan performa terbaik di semua metrik (akurasi, precision, dan recall), yang mencerminkan kemampuannya dalam menangani data dan membuat prediksi yang akurat.

4.1.4.2 Pemodelan Dengan Seleksi Fitur

Dalam penelitian ini, fitur-fitur yang paling relevan dalam memprediksi status keberangkatan calon jamaah haji akan dipilih menggunakan metode *Recursive Feature Elimination* dengan *Cross-Validation (REFCV)*. Metode ini bertujuan untuk mengeliminasi fitur-fitur yang tidak signifikan dan memilih fitur-fitur yang memiliki pengaruh terbesar terhadap variabel target.

Dari seluruh fitur yang ada, *REFCV* telah memilih 14 fitur terbaik, yaitu:

1. nomor_porsi
2. tanggal_daftar
3. usia_daftar
4. nama
5. nama_ayah
6. pekerjaan
7. tempat_lahir
8. tanggal_lahir
9. alamat
10. desa
11. kode_pos
12. bank
13. cabang
14. nik

Pada hasil *REFCV*, fitur seperti 'nomor_porsi', 'tanggal_daftar', 'nama', dan 'nama_ayah' terpilih sebagai fitur penting dalam pemodelan. Namun, fitur-fitur ini mungkin kurang relevan dalam konteks prediksi pembatalan haji. Hal ini bisa disebabkan oleh beberapa faktor, seperti bias dalam data atau korelasi yang tidak langsung tetapi signifikan secara statistik.

Fitur-fitur seperti 'nomor_porsi' dan 'tanggal_daftar' mungkin merefleksikan urutan pendaftaran atau kelompok tertentu yang memiliki kecenderungan berbeda dalam hal pembatalan. Sementara itu, 'nama' dan 'nama_ayah' mungkin terpilih karena variasi tertentu dalam data yang terkait dengan faktor-faktor lain, meskipun secara logika fitur-fitur ini tidak seharusnya mempengaruhi hasil prediksi.

Setelah fitur-fitur ini terpilih, model prediksi dibangun menggunakan tiga algoritma berbeda yaitu *Naive Bayes*, *K-Nearest Neighbors (k-NN)*, dan *Random Forest*. Evaluasi model dilakukan melalui metode *Cross-Validation* dengan 10 fold untuk mendapatkan estimasi akurasi yang lebih stabil dan reliabel.

Hasil evaluasi menunjukkan bahwa *Naive Bayes* memperoleh akurasi rata-rata sebesar 93% setelah menggunakan fitur-fitur yang dipilih oleh *REFCV*. Skor akurasi pada setiap fold berkisar antara 86,73% hingga 96,46%, dengan variasi yang menunjukkan konsistensi performa model dalam berbagai subset data.

k-NN juga menunjukkan performa yang sangat baik dengan akurasi rata-rata sebesar 93%. Skor per fold bervariasi antara 89,47% hingga 97,35%, yang menunjukkan bahwa *k-NN* dapat bekerja efektif dengan fitur-fitur yang dipilih.

Random Forest memberikan hasil yang paling optimal dengan akurasi rata-rata sebesar 95%. Skor per fold pada *Random Forest* berkisar antara 91,15% hingga 99,11%, menunjukkan bahwa model ini memiliki kemampuan yang kuat dalam menangani variabilitas data dan memberikan prediksi yang akurat.

Proses seleksi fitur menggunakan *REFCV* berhasil meningkatkan performa model, terutama pada *Random Forest* yang menunjukkan akurasi tertinggi di antara ketiga algoritma yang diuji. Pemilihan fitur yang tepat terbukti mampu menyederhanakan model tanpa mengorbankan akurasi, bahkan meningkatkan kemampuan model dalam memprediksi status keberangkatan calon jamaah haji.

Tabel 4.3 Hasil Evaluasi Model Dengan Seleksi Fitur

Model	Accuracy (%)	Precision (%)	Recall (%)
<i>Naïve Bayes</i>	93%	93%	90%
<i>Random Forest</i>	95%	96%	96%
<i>K-Nearest Neighbor</i>	93%	89%	90%

4.2 Pembahasan

4.2.1 Penentuan Model Terbaik

Dalam menentukan model terbaik, dilakukan perbandingan performa antara model yang menggunakan seluruh fitur dan model yang telah melalui seleksi fitur menggunakan *REFCV*.

Tabel 4.4 Perbandingan Performa dengan Seluruh Fitur dan Seleksi Fitur

Fitur/Atribut	Model	Accuracy (%)	Precision (%)	Recall (%)
Tanpa seleksi fitur (semua fitur)	<i>Naïve Bayes</i>	92.94%	93.28%	92.94%
	<i>Random Forest</i>	94.97%	95.05%	94.97%
	<i>K-Nearest Neighbor</i>	93.47%	93.51%	93.47%
Dengan seleksi fitur <i>REFCV</i>	<i>Naïve Bayes</i>	93%	93%	90%
	<i>Random Forest</i>	95%	96%	96%
	<i>K-Nearest Neighbor</i>	93%	89%	90%

Hasil perbandingan menunjukkan bahwa:

1. *Naïve Bayes*:

- Dengan seluruh fitur, model ini mencapai akurasi 92.94%, precision 93.28%, dan recall 92.94%.
- Setelah seleksi fitur dengan *REFCV*, akurasi meningkat sedikit menjadi 93%, namun precision dan recall menjadi lebih seimbang pada 93% dan 90%.

2. *Random Forest*:

- Model ini menunjukkan performa terbaik baik dengan semua fitur maupun setelah seleksi fitur. Dengan semua fitur, akurasi mencapai 94.97%, precision 95.05%, dan recall 94.97%.
- Setelah seleksi fitur, performa meningkat menjadi akurasi 95%, precision 96%, dan recall 96%. Hal ini menunjukkan bahwa seleksi fitur *REFCV* mampu meningkatkan kemampuan model dalam prediksi yang lebih tepat.

3. *K-Nearest Neighbor (k-NN)*:

- Dengan seluruh fitur, *k-NN* memiliki akurasi 93.47%, precision 93.51%, dan recall 93.47%.
- Setelah seleksi fitur, akurasi tetap pada 93%, namun precision sedikit menurun menjadi 89%, sedangkan recall menjadi 90%.

Random Forest tetap menjadi model terbaik, baik dengan seluruh fitur maupun setelah seleksi fitur. Namun, seleksi fitur dengan *REFCV* terbukti mampu meningkatkan performa secara keseluruhan, terutama untuk *Random Forest* yang mengalami peningkatan signifikan pada precision dan recall.

4.3 Keterbatasan Penelitian

Dalam penelitian ini, meskipun telah dilakukan analisis yang mendalam dan komprehensif terkait dengan performa model *Naive Bayes*, *Random Forest*, dan *K-Nearest Neighbor*, terdapat beberapa keterbatasan yang perlu diperhatikan untuk interpretasi hasil dan pelaksanaan studi di masa mendatang.

4.3.1. Keterbatasan Dataset

Ukuran dataset yang digunakan dalam penelitian ini mungkin belum mencukupi untuk sepenuhnya mewakili populasi calon jemaah haji Indonesia. Dataset yang lebih besar dan beragam dapat memberikan hasil yang lebih baik dan dapat dipercaya.

Meskipun sudah dilakukan seleksi fitur menggunakan teknik *Recursive Feature Elimination*, kualitas data tetap menjadi faktor pembatas. Data yang tidak akurat, data yang hilang, dan outlier dapat mempengaruhi akurasi prediksi dan kinerja model secara keseluruhan.

4.3.2. Pemilihan Fitur

Teknik seleksi fitur yang digunakan mungkin belum sepenuhnya optimal dalam menentukan atribut yang paling relevan untuk prediksi. Faktor lain atau metode seleksi yang berbeda bisa saja menghasilkan kumpulan fitur yang lebih baik.

4.3.3. Pemilihan Model

Parameter dan konfigurasi model (hyperparameter) yang digunakan dalam penelitian ini belum optimal. Pengaturan parameter yang lebih teliti dan penyesuaian dapat meningkatkan kinerja model. Pemilihan model pada penelitian ini terbatas pada *Naive Bayes*, *Random Forest*, dan *K-Nearest Neighbor*. Penelitian lanjutan dapat mengeksplorasi model lain seperti SVM, neural networks, atau ensemble methods untuk potensi peningkatan performa.

Penelitian ini memberikan dasar yang kuat untuk memahami faktor-faktor yang mempengaruhi keberangkatan/pembatalan calon jemaah haji dan menunjukkan bahwa pemilihan fitur yang tepat dapat meningkatkan performa model prediktif. Namun, penelitian lanjutan dengan dataset yang lebih luas dan metode seleksi fitur serta model yang lebih variatif diperlukan untuk mendapatkan hasil yang lebih akurat dan aplikatif.