

## PENDETEKSIAN POTENSI FAKTOR RISIKO TINGGI KESEHATAN JEMAAH HAJI MENGGUNAKAN ALGORITMA KLASIFIKASI DATA MINING

<sup>1</sup>ARIF BUDIARTO, <sup>2</sup>AGUS SUBEKTI, <sup>3</sup>HARIS DARMAWAN

Jurusan Ilmu Komputer, Fakultas Ilmu Komputer, Universitas  
Nusamandiri Jalan Jatiwaringin No. 2, Cipinang Melayu,  
Jakarta Timur

Email: <sup>1</sup>14002437@nusamandiri.ac.id, <sup>2</sup>agus@nusamandiri.ac.id,  
<sup>3</sup>muhammad.uhs@nusamandiri.ac.id

### ABSTRAK

Menurut Peraturan Menteri Kesehatan nomor 15 tahun 2016, Ibadah haji adalah ibadah fisik, oleh karena itu Jemaah Haji dituntut mampu secara fisik dan rohani, sehingga dapat melaksanakan ibadah haji dengan baik dan lancar. Berkaitan dengan kondisi tersebut ilmu data mining diperlukan untuk mengklasifikasi pendeteksian potensi faktor risiko tinggi kesehatan Jemaah Haji. Model klasifikasi yang digunakan bertujuan untuk memperoleh akurasi yang tepat dalam pendeteksian potensi faktor risiko tinggi kesehatan Jemaah Haji. Dataset yang digunakan adalah dataset pemeriksaan kesehatan Jemaah Haji. Berdasarkan perhitungan *confusion matrix* yaitu *accuracy*, *precision*, *recall* dan *AUC*. Hasil penelitian membuktikan bahwa algoritma klasifikasi *whitebox* mampu memberikan klasifikasi yang baik dalam upaya pengendalian faktor risiko tinggi bagi kesehatan Jemaah Haji.

**Kata kunci:** *Decision Tree*, *Datamining*, Istithaah, Risiko Tinggi Kesehatan Haji

### I. PENDAHULUAN

Menurut Peraturan Menteri Kesehatan nomor 15 tahun 2016, ibadah haji adalah ibadah fisik, oleh karena itu Jemaah Haji dituntut mampu secara fisik dan rohani, sehingga dapat melaksanakan ibadah haji dengan baik dan lancar. Kondisi ini dapat dicapai bila jemaah dalam kondisi istithaah kesehatan Jemaah Haji, yaitu kemampuan Jemaah Haji dari aspek kesehatan fisik dan mental yang terukur dengan pemeriksaan yang dapat dipertanggungjawabkan sehingga dapat menjalankan ibadah haji sesuai tuntunan agama Islam.

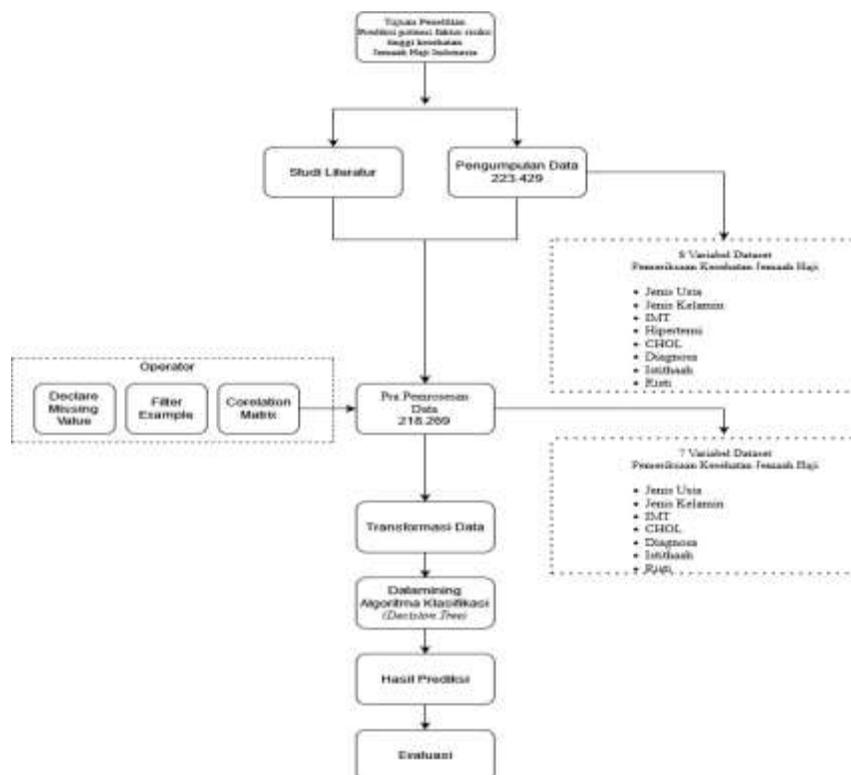
Berkaitan dengan kondisi tersebut ilmu data mining diperlukan untuk mengklasifikasi pendeteksian potensi faktor risiko tinggi kesehatan Jemaah Haji dengan menggunakan metode *Knowledge Discovery in Database* (KDD) dan model klasifikasi algoritma *whitebox* yang dibandingkan dengan algoritma *blackbox* untuk memperoleh akurasi yang tepat.

Penelitian sebelumnya yang dilakukan R. Ahuja dengan menggunakan dataset pima indians diabetes terdiri dari 768 record, diterapkan kombinasi dengan 5 (lima) algoritma klasifikasi, J. R. Lambert dalam penelitian *Chronic Kidney Disease* (CKD) menerapkan fitur Teknik seleksi untuk diterapkan dalam mengekstrak atribut penting dan mengklasifikasikannya ke dalam CKD atau bukan CKD. Penelitian penyakit batu ginjal yang dilakukan oleh Y. Widiastiwi dibentuk menjadi sebuah informasi untuk mengidentifikasi faktor risiko dari sebuah kejadian menggunakan algoritma klasifikasi *Decision Tree* dengan menghasilkan akurasi terbaik sebesar 95,71%.

Dalam penelitian ini algoritma klasifikasi *whitebox Decision Tree* dibandingkan dengan algoritma *blackbox* (*Random Forest, AdaBoost, Naive Bayes*) agar mampu memberikan klasifikasi yang baik dalam upaya pengendalian potensi faktor risiko tinggi kesehatan Jemaah Haji Indonesia menggunakan atribut jenis usia, jenis kelamin, indeks masa tubuh, hipertensi, kolesterol, diagnosa, istithaah dan risti.

## II.METODE PENELITIAN

Peneliti melakukan eksperimen terhadap data Pemeriksaan Kesehatan Haji untuk prediksi akurasi potensi faktor risiko tinggi kesehatan Jemaah Haji Indonesia menggunakan metode *Knowledge Discovery in Database* (KDD) dengan 5 tahapan yang harus dilakukan dalam metode ini, yaitu: Seleksi Data, Pra Pemrosesan, Transformasi, Data Mining, dan Evaluasi Interpretasi.



Gambar 1. Tahapan Penelitian

Berdasarkan Gambar 1 diatas, 5 (lima) tahap yang dilakukan menggunakan metode KDD, menggunakan langkah berikut:

#### A. Pemilihan Data

Pada langkah ini, data dipilih dari serangkaian data riwayat Pemeriksaan Kesehatan Jemaah Haji pada tahun keberangkatan 2019 dengan total 223.429 data.

#### B. Pra Pemrosesan Data

Pada tahap ini, pembersihan data dilakukan termasuk menghapus duplikasi data, memeriksa inkonsistensi data, dan kesalahan data dengan total menjadi 218.269 data, langkah selanjutnya adalah melakukan transformasi data.

#### C. Transformasi Data

Pada tahap ini, data diubah menjadi bentuk untuk dilakukan mining.

#### D. Data Mining

Penelitian ini menggunakan algoritma data mining untuk dilakukan proses analisis data, dengan algoritma klasifikasi.

#### E. Interpretasi atau Evaluasi

Data yang dihasilkan dari algoritma menjadi informasi yang diharapkan dapat membantu prediksi akurasi potensi faktor risiko tinggi kesehatan Jemaah Haji Indonesia.

### III. HASIL DAN PEMBAHASAN

#### 3.1 Hasil

##### Deskripsi Data

*Dataset* yang digunakan terdiri dari kumpulan *dataset* pemeriksaan Kesehatan Jemaah Haji dengan 7 variabel dan 1 label, terlihat pada tabel 1.

Tabel 1. Karakteristik Dataset Pemeriksaan Kesehatan Jemaah Haji

Atribut	Description	Type
Jenis Usia	Kategori berdasarkan umur	Nominal
Jenis Kelamin	Kategori berdasarkan jenis kelamin	Nominal
IMT	Indeks Masa Tubuh	Nominal
Hipertensi	Kategori Hipertensi	Nominal
CHOL	Kategori Cholestrol	Nominal
Diagnosa	Diagnosa Penyakit	Nominal
Istithaah	Kemampuan fisik dalam kesehatan	Nominal
Risti	Risiko Tinggi	Nominal

#### 3.2 PRA PEMROSESAN DATA

Tujuan dari penelitian ini mendapatkan accuracy terbaik dari pendeteksian faktor risiko pada kesehatan Jemaah Haji Indonesia dengan menggunakan perbandingan algoritma *Decision Tree*, *AdaBoost*, *Random Forest*, dan *Naïve Bayes*. Data yang dianalisa adalah data Pemeriksaan Kesehatan Jemaah Haji 2019. Berikut hasil pengujiannya:

Tabel 2. Hasil Pengujian dengan Algoritma klasifikasi.

JENIS USIA	JENIS KELAMIN	IMT	CHOL	DIAGNOS	ISTITHAAR	RISTI	DECISION TREE	ADABOOST	RANDOM FOREST	NAÏVE BAYES
LANSIA	Pria	OBESITAS	NORMAL	YA	MEMENU	RISTI	RISTI	RISTI	TIDAK RISTI	RISTI
DEWASA	Pria	OBESITAS	NORMAL	YA	MEMENU	RISTI	RISTI	TIDAK RISTI	TIDAK RISTI	TIDAK RISTI
DEWASA	Wanita	OBESITAS	NORMAL	YA	MEMENU	RISTI	RISTI	TIDAK RISTI	RISTI	TIDAK RISTI
DEWASA	Pria	OBESITAS	NORMAL	YA	MEMENU	TIDAK RISTI	TIDAK RISTI	TIDAK RISTI	RISTI	TIDAK RISTI
LANSIA	Wanita	BB DIBAW	NORMAL	YA	MEMENU	RISTI	RISTI	RISTI	RISTI	TIDAK RISTI
LANSIA	Wanita	OBESITAS	NORMAL	YA	MEMENU	TIDAK RISTI	RISTI	TIDAK RISTI	TIDAK RISTI	TIDAK RISTI
DEWASA	Pria	BB LEBIH	NORMAL	YA	MEMENU	RISTI	RISTI	TIDAK RISTI	TIDAK RISTI	RISTI
LANSIA	Pria	BB LEBIH	NORMAL	YA	MEMENU	RISTI	RISTI	TIDAK RISTI	RISTI	TIDAK RISTI
DEWASA	Pria	OBESITAS	NORMAL	YA	MEMENU	RISTI	TIDAK RISTI	RISTI	TIDAK RISTI	RISTI
DEWASA	Pria	OBESITAS	NORMAL	YA	MEMENU	TIDAK RISTI	TIDAK RISTI	RISTI	TIDAK RISTI	RISTI
DEWASA	Wanita	OBESITAS	NORMAL	TIDAK	MEMENU	TIDAK RISTI	TIDAK RISTI	TIDAK RISTI	TIDAK RISTI	TIDAK RISTI
LANSIA	Wanita	BB LEBIH	NORMAL	TIDAK	MEMENU	RISTI	RISTI	RISTI	TIDAK RISTI	TIDAK RISTI
DEWASA	Pria	BB LEBIH	NORMAL	YA	MEMENU	TIDAK RISTI	TIDAK RISTI	RISTI	TIDAK RISTI	TIDAK RISTI
DEWASA	Wanita	OBESITAS	NORMAL	YA	MEMENU	TIDAK RISTI	TIDAK RISTI	TIDAK RISTI	RISTI	RISTI
LANSIA	Pria	OBESITAS	NORMAL	YA	MEMENU	RISTI	RISTI	RISTI	RISTI	RISTI
LANSIA	Wanita	BB DIBAW	NORMAL	YA	MEMENU	RISTI	RISTI	TIDAK RISTI	TIDAK RISTI	TIDAK RISTI
LANSIA	Wanita	BB LEBIH	NORMAL	YA	MEMENU	TIDAK RISTI	RISTI	RISTI	RISTI	TIDAK RISTI
LANSIA	Pria	BB LEBIH	NORMAL	YA	MEMENU	RISTI	RISTI	TIDAK RISTI	TIDAK RISTI	TIDAK RISTI
DEWASA	Wanita	OBESITAS	NORMAL	YA	MEMENU	TIDAK RISTI	TIDAK RISTI	RISTI	TIDAK RISTI	TIDAK RISTI
DEWASA	Pria	OBESITAS	NORMAL	YA	MEMENU	TIDAK RISTI	TIDAK RISTI	TIDAK RISTI	TIDAK RISTI	TIDAK RISTI
DEWASA	Wanita	OBESITAS	NORMAL	TIDAK	MEMENU	RISTI	TIDAK RISTI	TIDAK RISTI	TIDAK RISTI	TIDAK RISTI
LANSIA	Wanita	BB LEBIH	NORMAL	YA	MEMENU	RISTI	RISTI	RISTI	RISTI	TIDAK RISTI
LANSIA	Pria	BB LEBIH	NORMAL	YA	MEMENU	RISTI	RISTI	RISTI	RISTI	RISTI

Kemudian masukkan data yang telah diuji pada tabel 2 menggunakan algoritma klasifikasi yang ada ke dalam model *confusion matrix* maka didapatkan hasil pada tabel 3 berikut:

Tabel 3. Hasil *Confusion Matrix* menggunakan algoritma klasifikasi

	<b>Risti</b>	<b>Tidak</b>
<b>Risti</b>	95663	22156
<b>Tidak</b>	14283	86167

Kemudian masukan nilai yang ada di dalam Tabel *confusion matrix* diatas ke dalam persamaan algoritma *Decision Tree*, sehingga menghasilkan nilai seperti gambar 1 di bawah ini:

accuracy: 83.31% +/- 0.23% (micro average: 83.31%)

	true RISTI	true TIDAK RISTI	class precision
pred. RISTI	95663	22158	81.19%
pred. TIDAK RISTI	14283	86167	85.78%
class recall	87.01%	79.55%	

Gambar 1. *Confusion Matrix* menentukan accuracy dengan algoritma C4.5

Sedangkan untuk algoritma berbasis *AdaBoost* akan menghasilkan nilai seperti gambar 2 di bawah ini:

accuracy: 83.30% +/- 0.16% (micro average: 83.30%)

	true RISTI	true TIDAK RISTI	class precision
pred. RISTI	95647	22147	81.20%
pred. TIDAK RISTI	14299	86176	85.77%
class recall	86.99%	79.55%	

Gambar 2. *Confusion Matrix* menentukan accuracy dengan algoritma *AdaBosst*

Algoritma berbasis *Random Forest* akan menghasilkan nilai seperti gambar 3 di bawah ini:

accuracy: 83.30% +/- 0.23% (micro average: 83.30%)

	true RISTI	true TIDAK RISTI	class precision
pred. RISTI	95636	22141	81.20%
pred. TIDAK RISTI	14310	86182	85.75%
class recall	86.98%	79.56%	

Gambar 3. *Confusion Matrix* menentukan accuracy dengan algoritma *Random Forest*

Algoritma berbasis *Naïve Bayes* akan menghasilkan nilai seperti gambar 4 di bawah ini:

accuracy: 82.25% +/- 0.25% (micro average: 82.25%)

	true RISTI	true TIDAK RISTI	class precision
pred. RISTI	92336	21043	81.44%
pred. TIDAK RISTI	17619	87280	83.21%
class recall	83.98%	80.57%	

Gambar 4 *Confusion Matrix* menentukan accuracy dengan algoritma *Naïve Bayes*

Data uji pada gambar *Confusion Matrix* diatas akan dinilai hasil prediksi dengan menggunakan grafik ROC untuk algoritma *Decision Tree*, visualisasi dari grafik ROC yaitu:



Gambar 5 Grafik ROC dengan model algoritma *Decision Tree*

### 3.1 Hasil Pengujian

Hasil evaluasi dengan menggunakan *confusion matrix* model klasifikasi sebagai berikut:

Tabel 4 Hasil Komparasi evaluasi algoritma Klasifikasi.

Model	Accuracy	Precision	Recall	AUC
<i>Decision Tree</i>	83.31%	85.78%	79.55%	86%
AdaBoost	83.30%	85.77%	79.55%	83%
Random Forest	83.30%	85.76%	79.56%	89%
Naïve Bayes	82.29%	83.21%	80.57%	88%

Hasil pengujian *confusion matrix* tabel 4 di atas diketahui menggunakan model algoritma *Decision Tree* mempunyai akurasi 83.31%, *AdaBoost* 83.30%, *Random Forest* 83.30%, dan *Naïve Bayes* 82.29%. Berdasarkan perbandingan model klasifikasi tingkat akurasi *Decision Tree* lebih tinggi dibandingkan algoritma klasifikasi lainnya.

### 3.2 Pohon Keputusan

Algoritma Pohon Keputusan (*Decision Tree*) membagi data *training* dengan bantuan perolehan informasi. Atribut yang memiliki frekuensi tinggi dipertimbangkan untuk memisahkan data berdasarkan informasi yang tersedia dalam dataset pemeriksaan Kesehatan Jemaah Haji yang berlabel. Salah satu metode klasifikasi *Decision tree* merupakan representasi struktur pohon (tree) dimana setiap node merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas. Berdasarkan hasil penelitian algoritma *Decision Tree* gambar 6 faktor risiko tinggi (risti) dengan diawali variabel Jenis Usia pada node pertama dengan warna merah menunjukkan status Tidak Risti sedangkan Biru Risti, seperti pada gambar berikut.



Gambar 6. *White Box Decision Tree* faktor risiko kesehatan jemaah haji.

#### Tree

```
JENIS USIA = DEWASA
| ISTITHAAN = MEMENUHI SYARAT ISTITHAAN KESEHATAN HAJI: TIDAK RISTI (RISTI=1379, TIDAK RISTI=8118)
| ISTITHAAN = MEMENUHI SYARAT ISTITHAAN KESEHATAN HAJI DENGAN PENGAMFINAN
| | DIAGNOSA PENYAKIT = TIDAK : TIDAK RISTI (RISTI=257, TIDAK RISTI=765)
| | DIAGNOSA PENYAKIT = YA: RISTI (RISTI=3435, TIDAK RISTI=16497)
| ISTITHAAN = TIDAK MEMENUHI SYARAT ISTITHAAN KESEHATAN HAJI
| | CHOL = KOLESTROL TINGGI: RISTI (RISTI=2, TIDAK RISTI=0)
| | CHOL = NORMAL
| | | DIAGNOSA PENYAKIT = TIDAK : TIDAK RISTI (RISTI=3, TIDAK RISTI=4)
| | | DIAGNOSA PENYAKIT = YA: RISTI (RISTI=46, TIDAK RISTI=27)
| ISTITHAAN = TIDAK MEMENUHI SYARAT ISTITHAAN KESEHATAN HAJI UNTUK SEHENTARA
| | DIAGNOSA PENYAKIT = TIDAK : TIDAK RISTI (RISTI=4, TIDAK RISTI=35)
| | DIAGNOSA PENYAKIT = YA: RISTI (RISTI=76, TIDAK RISTI=57)
JENIS USIA = LANJUT: RISTI (RISTI=6103, TIDAK RISTI=5561)
JENIS USIA = REMAJA
| ISTITHAAN = MEMENUHI SYARAT ISTITHAAN KESEHATAN HAJI: TIDAK RISTI (RISTI=21, TIDAK RISTI=336)
| ISTITHAAN = MEMENUHI SYARAT ISTITHAAN KESEHATAN HAJI DENGAN PENGAMFINAN
| | IMT = BB DI BAWAH NORMAL: TIDAK RISTI (RISTI=1, TIDAK RISTI=4)
| | IMT = BB LEBIH: RISTI (RISTI=9, TIDAK RISTI=7)
| | IMT = OBESITAS: RISTI (RISTI=0, TIDAK RISTI=8)
```

Gambar 7. Algoritma *Decision Tree* faktor risiko kesehatan jemaah haji

Penelitian menggunakan metode algoritma *Decision Tree* pada gambar 7 dilakukan dengan menggunakan metode *White Box* dengan aplikasi pendukung pengujian data Rapid Miner untuk mengklasifikasikan pengendalian faktor risiko

Kesehatan Jemaah Haji yang dapat diukur menurut beberapa faktor antara lain jenis usia, jenis kelamin, Indeks Masa Tubuh (IMT), kolesterol (CHOL), diagnosa penyakit, dan istithaah.

#### IV.KESIMPULAN

Penelitian ini dibuat sebagai pengendalian potensi faktor risiko tinggi kesehatan Jemaah Haji Indonesia dengan tingkat akurasi yang baik. Untuk memprediksi permasalahan tersebut digunakan metode *white box* pada algoritma klasifikasi *Decision Tree* dengan hasil akurasi terbaik pada *Decision Tree* dibandingkan dengan algoritma *blackbox* (*Random Forest, AdaBoost, Naive Bayes*).

Hasil penelitian model klasifikasi algoritma *Decision Tree* pada node pertama dalam memprediksi potensi faktor risiko tinggi (risti) kesehatan Jemaah Haji terdapat pada jenis usia, sehingga pengendalian faktor risiko tinggi bisa difokuskan pada jenis usia dewasa yang memiliki comorbid untuk dilakukan upaya pembinaan kesehatan beserta rangkaian tindakan medis dalam tahapan pemeriksaan kesehatan secara rutin sebelum masa keberangkatan ibadah haji agar Jemaah Haji mampu (Istithaah) secara fisik dalam melaksanakan aktifitas ibadah haji.

Penelitian dengan model *Decision Tree* tingkat accuracy sebesar 83.31%. Hal ini membuktikan bahwa penelitian algoritma *Decision Tree* mampu memberikan klasifikasi yang baik dalam upaya pengendalian faktor risiko tinggi bagi Kesehatan Jemaah Haji.

### DAFTAR PUSTAKA

- Kementerian Kesehatan, “Peraturan Menteri Kesehatan Republik Indonesia Nomor 15 Tahun 2016 Tentang Istithaah Kesehatan Haji,” no. May, pp. 31–48, 2016.
- Kementerian Kesehatan RI, “Peraturan Menteri Kesehatan Republik Indonesia Nomor 62 Tahun 2016 Tentang Penyelenggaraan Kesehatan Haji,” 2016.
- K. F. Irnanda and A. P. Windarto, “Penerapan Klasifikasi C4.5 Dalam Meningkatkan Kecakapan Berbahasa Inggris dalam Masyarakat,” *Semin. Nas. Teknol. Komput. Sains*, pp. 304–308, 2020.
- S. P. I. P. D, “Effective Use of the Kdd Process and Data Mining for,” no. January 2001, 2002.
- Ainurrohmah, “Akurasi Algoritma Klasifikasi pada Software Rapidminer dan Weka,” *Prisma*, vol. 4, pp. 493–499, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>.
- A. P. Permana, K. Ainiyah, and K. F. H. Holle, “Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 6, no. 3, pp. 178–188, 2021, doi: 10.14421/jiska.2021.6.3.178-188.
- R. Ahuja, S. C. Sharma, and M. Ali, “A diabetic disease prediction model based on classification algorithms,” *Ann. Emerg. Technol. Comput.*, vol. 3, no. 3, pp. 44–52, 2019, doi: 10.33166/AETiC.2019.03.005.
- J. R. Lambert, P. Arulanthu, and E. Perumal, “Identification of Nominal Attributes for Intelligent Classification of Chronic Kidney Disease using Optimization Algorithm,” *Proc. 2020 IEEE Int. Conf. Commun. Signal Process. ICCSP 2020*, pp. 119–125, 2020, doi: 10.1109/ICCSP48568.2020.9182206.
- K. M. Almustafa, “Prediction of chronic kidney disease using different classification algorithms,” *Informatics Med. Unlocked*, vol. 24, p. 100631, 2021, doi: 10.1016/j.imu.2021.100631.