

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Ibadah Haji**

Pelaksanaan ibadah haji yang dilakukan oleh kaum muslim pada waktu tertentu merupakan salah satu dari rukun Islam yang kelima. Semua orang muslim yang memiliki kemampuan material dan fisik untuk melakukannya hukumnya wajib. Haji secara bahasa berasal dari kata al-hajj yang berarti menyengaja atau bersungguh-sungguh. Menurut istilah, haji berarti menyengaja menuju Baitullah atau Ka'bah untuk menjalankan ibadah pada waktu yang telah ditetapkan, yaitu saat pelaksanaan ibadah haji. Puncak pelaksanaan ibadah haji terjadi pada bulan Zulhijah. Melaksanakan ibadah haji dalam Islam merupakan kewajiban bagi setiap muslim yang memiliki kemampuan dari segi materi, waktu, kesehatan, keamanan, kendaraan, dan kesiapan jiwa. Pelaksanaan ibadah haji ini, terdapat syarat-syarat yang wajib dipenuhi oleh para calon jemaah haji. Persyaratan tersebut yaitu beragama Islam, telah baligh, memiliki akal sehat, status merdeka atau bebas, memiliki kemampuan finansial, keberadaan mahram, pelaksanaan ibadah pada waktu yang telah ditentukan, dan juga menjaga kesehatan dengan memenuhi ketentuan istithaah (Nur Fadhli, 2020).

#### **2.2 Kesehatan Jemaah Haji**

Kesehatan merupakan aspek penting dalam perjalanan kegiatan ibadah haji, kondisi kesehatan yang kurang optimal dapat menghambat kelancaran pelaksanaan ibadah haji menjadi tidak maksimal. Kemampuan jasmani dan rohani merupakan salah satu syarat kelayakan untuk melaksanakan ibadah haji (istithaah). Hal ini

dapat ditentukan melalui hasil pemeriksaan kesehatan sebagai bagian dari persiapan pelaksanaan ibadah haji (Vestability, 2021). Istithaah kesehatan jemaah haji merupakan kemampuan jemaah haji untuk menjalankan ibadahnya sesuai dengan aturan agama Islam melalui pemeriksaan kesehatan fisik dan mental yang dapat dipertanggungjawabkan (Deswara, 2023).

Kriteria istithaah kesehatan jemaah haji terdapat empat, *di mana* keempatnya merupakan aspek yang penting dalam memastikan kesejahteraan jemaah haji selama menjalankan ibadah haji. Kriterianya yaitu:

1. Istithaah

Jemaah haji dengan status kesehatan haji istithaah merupakan jemaah yang memiliki kemampuan untuk menjalani seluruh rangkaian ibadah haji tanpa bergantung pada obat-obatan, peralatan kesehatan, atau orang lain.

2. Istithaah dengan pendamping

Jemaah haji dengan status kesehatan haji istithaah dengan pendamping merupakan jemaah yang membutuhkan pendamping seperti obat, alat, atau orang lain. Jemaah haji yang membutuhkan pendamping orang lain adalah jemaah yang bergantung pada bantuan orang lain dalam kegiatan sehari-hari.

3. Tidak istithaah sementara

Jemaah haji dengan status kesehatan tidak istithaah sementara merupakan jemaah yang memiliki indikasi penyakit, namun memiliki kemungkinan untuk pulih dengan penggunaan obat secara teratur dan melakukan pemeriksaan

kesehatan dengan rutin. Jika kondisi kesehatan belum terkendali, jemaah haji diberikan waktu hingga batas akhir pemeriksaan kesehatan haji. Jika kondisi kesehatan masih belum terkendali pada batas waktu tersebut, jemaah yang bersangkutan dianggap tidak memenuhi syarat istithaah kesehatan haji sementara dan penjadwalan keberangkatannya ditunda pada tahun berjalan atau ditindaklanjuti sesuai ketentuan.

#### 4. Tidak istithaah

Jemaah haji dengan status kesehatan tidak istithaah merupakan jemaah dengan hasil pemeriksaan kesehatan menunjukkan tidak memenuhi kriteria istithaah kesehatan pada satu atau lebih dari empat jenis pemeriksaan kesehatan.

Berdasarkan Keputusan Menteri Kesehatan Republik Indonesia Nomor HK.01.07/MENKES/2118/2023 tentang Standar Teknis Pemeriksaan Kesehatan Dalam Rangka Penetapan Status Istitha'ah Kesehatan Jemaah Haji, pemeriksaan kesehatan jemaah haji dilakukan oleh tim penyelenggara kesehatan haji kabupaten/kota yang dibentuk oleh Kepala Dinas Kesehatan Daerah Kabupaten/Kota. Teknis pemeriksaan kesehatan yang dilakukan pada jemaah haji termasuk pemeriksaan medis, pemeriksaan kognitif, pemeriksaan kesehatan mental, dan pemeriksaan kemampuan untuk melakukan aktivitas keseharian secara mandiri. Pemeriksaan kesehatan untuk para jemaah haji akan menjadi dasar untuk melaksanakan upaya pembinaan kesehatan, sehingga kondisi kesehatan mereka dapat ditingkatkan dan dipertahankan mulai dari di tanah air hingga di tanah suci (Menteri Kesehatan Republik Indonesia, 2023).

### 2.3 Klasifikasi

Salah satu teknik *data mining* yang paling umum digunakan adalah klasifikasi. Klasifikasi data dalam *data mining* merupakan suatu teknik yang umumnya digunakan untuk mengelola kumpulan data yang bersifat kompleks dan besar. Proses klasifikasi berfokus pada penemuan definisi kesamaan karakteristik dalam suatu kelompok atau kelas tertentu (Yuda dkk., 2022). Teknik ini biasanya menggunakan algoritma yang dapat dengan mudah disesuaikan untuk meningkatkan kualitas data. Tujuan akhir dari klasifikasi adalah untuk menghubungkan variabel yang menarik dengan variabel yang diamati dan menciptakan suatu hubungan yang relevan dalam analisis data. Terdapat berbagai jenis klasifikasi, dan setiap algoritma klasifikasi tersebut digunakan dalam mengekstrak informasi yang bermanfaat dari suatu kumpulan data. Model algoritma yang digunakan dalam klasifikasi adalah jenis algoritma yang memerlukan variabel target sebagai acuan untuk pembelajaran, sehingga dapat menghasilkan aturan atau pola yang berlaku pada data tersebut

Proses klasifikasi terdiri dari beberapa tahapan yaitu model dibangun dari data pelatihan dengan nilai label kelas diketahui. Membuat model dari dataset pelatihan, perlu menggunakan algoritma klasifikasi. Serta untuk memastikan akurasi model dari data *train*, jika hasil model memuaskan, maka model tersebut dapat diterapkan untuk mengklasifikasikan data yang belum memiliki label kelas (Marisa dkk., 2021).

## 2.4 *Data Mining*

Terdapat beragam definisi terkait *data mining*, secara umum *data mining* terbagi atas 2 kata yaitu “*data*” dan “*mining*”. *Data* adalah suatu kumpulan fakta yang terekam atau entitas yang tidak mempunyai arti dan telah terabaikan. Sedangkan *mining* adalah suatu proses penambangan. *Data mining* atau penambangan data merupakan suatu alat yang memungkinkan pengguna mengakses data dalam jumlah yang besar dengan cepat. Definisi lebih khusus dari *data mining* yaitu suatu alat dan aplikasi yang memanfaatkan analisis statistik pada data. *Data mining* menggambarkan proses pengumpulan teknik-teknik untuk menemukan pola-pola yang tidak diketahui dalam kumpulan data yang telah terkumpul. Karena banyaknya data yang tersedia dan kebutuhan mendesak untuk mengubah data tersebut menjadi informasi dan pengetahuan yang bermanfaat, maka dari itu *data mining* menjadi sangat dibutuhkan. *Data mining* juga dianggap sebagai hasil dari evolusi alami teknologi informasi. Hal ini disebabkan oleh fakta bahwa dengan semakin berkembangnya teknologi digital, manusia menghasilkan banyak data yang sangat besar di berbagai bidang, seperti pertanian, pendidikan, kesehatan dan sebagainya (Marisa dkk., 2021).

## 2.5 *Supervised Learning*

*Supervised learning* merupakan suatu metode pendekatan di mana data yang telah dilatih digunakan untuk mengelompokkan atau mengklasifikasikan data baru yang masuk. Metode ini digunakan untuk memetakan suatu inputan agar ke dalam kelompok data yang telah ada sebelumnya serta menghasilkan output yang serupa.

Metode ini sangat berpengaruh pada pola input dan output untuk mengidentifikasi informasi dari sistem (Namira Nur Az-Zahra dkk., 2021).

Metode *supervised learning* berdasar pada sekumpulan sampel data dengan label yang dimiliki. Sampel-sampel digunakan untuk menggambarkan karakteristik distribusi ukuran perilaku dalam berbagai jenis aplikasi, membentuk model perilaku dari data tersebut. *Supervised learning* dibagi lebih lanjut menjadi dua jenis masalah, yaitu klasifikasi dan regresi. Ketika variabel output berbentuk kategori seperti merah, biru, atau penyakit dan tidak ada penyakit merupakan masalah klasifikasi. Sebaliknya, ketika variabel output adalah nilai riil, seperti dolar atau berat, itu merupakan masalah regresi (Roihan dkk., 2019)

## 2.6 Seleksi Fitur

Seleksi fitur atau *feature selection* merupakan salah satu metode yang banyak digunakan dalam melakukan reduksi fitur. Metode ini banyak digunakan sebagai metode tambahan dalam melakukan klasifikasi sehingga mampu meningkatkan nilai *feature* sebelum dilatih (*train*) dengan algoritma klasifikasi (Lumbantobing dkk., 2020). *Feature Selection* sering digunakan untuk pengurangan dimensi model, dan menjadikan pengklasifikasian lebih efektif dengan mengurangi jumlah data yang dianalisa, maupun mengidentifikasi fitur yang sesuai untuk dipertimbangkan dalam proses pembelajaran. Cara kerjanya berdasar pada pengurangan ruang fitur yang besar, yaitu dengan cara mengeliminasi atribut yang kurang relevan serta dengan menggunakan penggunaan algoritma *feature selection* yang tepat sehingga dapat meningkatkan akurasi (Julianto dkk., 2022).

## 2.7 Algoritma Decision Tree

Decision Tree adalah salah satu teknik klasifikasi yang paling umum digunakan. Saat menjalani prosesnya, Decision Tree mampu mengubah data yang awalnya terwakili dalam bentuk tabel menjadi suatu struktur pohon yang memberikan gambaran visual tentang keputusan yang diambil. Pohon tersebut kemudian diubah menjadi aturan-aturan keputusan yang dapat digunakan untuk mengklasifikasikan data baru (Kurnia Sari & Firdaus Mahmudy, 2019). Algoritma Decision Tree dibuat dengan pohon keputusan dari *data training* yang terdiri dari *record-record* dalam basis data. Algoritma Decision Tree terdiri dari kumpulan *node* (simpul) yang saling terhubung melalui cabang, *di mana* cabang bergerak ke bawah dari *root* (akar) node dan berakhir di *leaf* (daun) *node* (Suntoro, 2019).

Tahapan dari algoritma Decision Tree yaitu (Akbar dkk., 2022):

1. Menyiapkan dataset pelatihan.
2. Identifikasi akar dari pohon keputusan.
3. Evaluasi *gain*, untuk memilih atribut yang akan menjadi akar, berdasarkan pada nilai *gain* tertinggi dari atribut-atribut yang tersedia. Perhitungan nilai *gain* dapat dilakukan menggunakan persamaan 1:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^N \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (1)$$

4. Lakukan langkah kedua untuk setiap cabang dengan mengikuti prosedur yang sama. Sementara itu, untuk menghitung nilai entropi menggunakan persamaan 2:

$$Entropy(S) = \sum_{i=1}^N - \pi \times \log 2 \pi \quad (2)$$

5. Ketika semua cabang node N menerima kelas yang sama, proses partisipasi pohon keputusan berakhir.

## 2.8 Algoritma Random Forest

Random forest adalah salah satu algoritma *supervised* dalam klasifikasi. Metode ini seperti namanya, membentuk sebuah hutan (*forest*) yang terdiri dari beberapa pohon (*tree*). Secara keseluruhan, semakin banyak pohon (*tree*) dalam hutan (*forest*), semakin kuat juga keseluruhan hutan tersebut. Semakin banyak pohon (*tree*), semakin tinggi tingkat akurasi yang dapat dicapai. Random Forest dimulai dengan menggunakan teknik dasar *data mining*, yaitu Decision Tree. Decision tree, input dimasukkan ke bagian atas (*root*) dan kemudian bergerak ke bagian bawah (*leaf*) untuk menentukan kelas mana data tersebut termasuk (Haristu & Rosa, 2019). Random Forest adalah salah satu algoritma pembelajaran mesin yang digunakan untuk melakukan klasifikasi. Algoritma ini terdiri dari beberapa Decision Tree yang beroperasi untuk membentuk gabungan fungsional. Keunggulan dari Random Forest dibandingkan dengan model individual lainnya adalah penggunaan Decision Tree yang tidak saling berkorelasi (Sanjaya dkk., 2020).

Tahapan dari algoritma Random Forest adalah (Amrani dkk., 2018):

1. Melakukan bootstrap untuk mengekstrak k sampel dari set pelatihan asli dengan N sampel sebanyak k kali,
2. Tetapkan k pohon keputusan.



3. Pemungutan suara menurut hasil klasifikasi seluruh pohon keputusan, hasil pemungutan suara tersebut disebut keyakinan.

## 2.9 Algoritma AdaBoost

Adaboost merupakan singkatan dari Adaptive Boosting, sebuah algoritma pembelajaran mesin yang dikembangkan oleh Robert Schapire dan Yoav Freund. Algoritma ini adalah sebuah teknik dalam *data mining* yang bertujuan untuk meningkatkan tingkat akurasi dalam metode klasifikasi. Adaboost digunakan untuk melakukan klasifikasi data dalam kelas-kelasnya yang berbeda. Adaboost menentukan kategori kelas dengan mempertimbangkan nilai bobot yang diberikan pada setiap kelas. Algoritma AdaBoost dapat membangun pengklasifikasi yang kuat dengan menggabungkan beberapa pengklasifikasi sederhana atau lemah secara linear (Febianto dkk., 2024).

Persamaan 3 merupakan persamaan dari algoritma AdaBoost (Sanjaya dkk., 2020):

$$H(x) = \sum_{t=1}^T \alpha_t h_t(x) \quad (3)$$

Di mana  $h_t(x)$  adalah klasifikasi lemah atau dasar,  $\alpha_t$  adalah tingkat pembelajaran (*learning rate*) dan  $H(x)$  adalah hasil yang berupa klasifikasi kuat atau akhir.

Tahapan dari algoritma AdaBoost adalah (Febianto dkk., 2024):

1. Persiapkan data uji (*testing*) dan data latih (*training*).
2. Memberikan nilai inisialisasi bobot pada setiap fitur/atribut.

3. Mengatur ulang bobot setiap fitur dengan nilai bobot.
4. Menghitung nilai hipotesis  $h_t$  dengan menggunakan probabilitas.
5. Menghitung kesalahan  $e_t$ , jika  $e_t > \frac{1}{2}$  maka hitung koefisien bobot, hitung bobot baru dan normalisasi. Jika  $e_t < \frac{1}{2}$  maka hitung klasifikasi data pengujian.
6. Hasil klasifikasi data uji.

### 2.10 Algoritma XGBoost

XGBoost adalah salah satu metode yang menggunakan teknik Gradient Tree Boosting yang canggih, mampu menangani masalah skala besar dengan sumber daya komputasi yang terbatas. Algoritma XGBoost memiliki keunggulan dalam hal kinerja, kompleksitas waktu, dan harga memori yang terjangkau (Septiana Rizky dkk., 2022). XGBoost merupakan penyempurnaan dari algoritma Gradient Boosting yang dapat membangun pohon klasifikasi secara efisien dan beroperasi secara paralel. XGBoost model baru memprediksi sisa model sebelumnya dan kemudian menjumlahkannya untuk mendapatkan prediksi akhir.

Tahapan dari algoritma XGBoost adalah (Guo dkk., 2020):

1. Model dimulai dengan iterasi awal, dan sub model klasifikasi dibangun pada setiap iterasi.
2. Selama proses pelatihan, model terus menghitung *loss function* untuk memilih *leaf node* dengan *gain loss* terbesar.
3. Membuat pohon keputusan baru setiap fungsi baru ( $f(x, \theta_k)$ ) agar sesuai dengan *residual* prediksi akhir.

4. Ketika K pohon diperoleh setelah pelatihan, fitur sampel prediksi akan memiliki *leaf node* yang sesuai di setiap pohon, dan setiap *leaf node* memiliki skor.
5. Skor yang sesuai dari setiap pohon dijumlahkan untuk mendapatkan nilai klasifikasi pengenalan sampel.

### 2.11 Chi Square

Chi Square atau yang biasa disimbolkan sebagai  $X^2$  merupakan metode *feature selection* atau pemilihan fitur yang banyak digunakan saat ini untuk mengevaluasi korelasi dengan menggunakan pendekatan statistika. Algoritma Chi-Square bekerja dengan menguji independensi sebuah *term* dan kategorinya sehingga algoritma ini dapat menghilangkan fitur-fitur pengganggu dan tidak relevan (Lumbantobing dkk., 2020). Persamaan 4 merupakan persamaan dari seleksi fitur Chi-Square.

$$X^2(t, c) = \frac{N(A \times D - b \times C)^2}{(A + B) \times (C + D) \times (A + C) \times (B + D)} \quad (4)$$

Variabel  $t$  merupakan kata yang sedang diujikan terhadap suatu kelas  $c$ ,  $N$  merupakan jumlah dokumen latih,  $A$  merupakan banyaknya dokumen pada kelas  $c$  yang memuat kata  $t$ ,  $B$  merupakan banyaknya dokumen yang tidak berada di  $c$  namun memuat kata  $t$ ,  $C$  merupakan banyaknya dokumen yang berada di kelas  $c$  namun tidak memiliki kata  $t$  di dalamnya, serta  $D$  merupakan banyaknya dokumen yang bukan merupakan dokumen kelas  $c$  dan tidak memuat kata kata  $t$  (Listiowarni & Rahayu Setyaningsih, 2019)

## 2.12 Penelitian Terkait

Penelitian Terkait merupakan proses membaca dan memahami penelitian sebelumnya yang relevan dengan penelitian yang dilakukan. Para peneliti sebelumnya telah menggunakan beberapa penelitian yang terkait dengan haji, kesehatan serta penerapan *data mining* pada klasifikasi. Beberapa algoritma klasifikasi yang telah digunakan melibatkan Decision Tree, Random Forest serta metode klasifikasi lainnya.

### 2.10.1 State of the art

Tabel 2.1 merupakan penjelasan mengenai hasil penelitian disajikan dalam bentuk *state of the art*.

Tabel 2. 1 *State of The Art*

No	Nama & Tahun Penelitian	Judul Penelitian	Algoritma	Hasil/Temuan
1	Feri Setiadi dan Handoyo Widi Nugroho (2024)	Perbandingan Performa Algoritma Naive Bayes, Random Forest dan K-Nearest Neighbor pada Prediksi Calon Jemaah Haji Indonesia yang Berpotensi Membatalkan Haji	Naive Bayes, Random Forest dan K-Nearest Neighbor	Hasil penelitian menunjukkan Random Forest menjadi model terbaik, baik dengan seluruh fitur maupun setelah seleksi fitur dengan nilai akurasi 95% dibandingkan dengan Naïve bayes 93% dan K-Nearest Neighbor 93%. Seleksi fitur dengan REF terbukti mampu meningkatkan performa secara keseluruhan.
2	Hendro Nugroho, Gusti Eka Yulastuti, Andrean Firman P (2023)	Klasifikasi Diagnosis Diabetes Melitus Menggunakan Metode Naive	Naïve Bayes	Penelitian klasifikasi menggunakan dataset Diabetes melitus menunjukkan algoritma Naïve Bayes menghasilkan kinerja yang baik

No	Nama & Tahun Penelitian	Judul Penelitian	Algoritma	Hasil/Temuan
		Bayes dengan Seleksi Fitur Backward Elimination		dengan nilai akurasi 90% dibandingkan dengan metode Naïve bayes dikombinasikan dengan seleksi fitur Backward Elimination sebesar 86%.
3	Hertati, Elin Haerani, Novriyanti, Fadhilah Syafria (2023)	Analisis Sentimen Masyarakat Terhadap Kenaikan Biaya Haji Tahun 2023 Menggunakan Metode Naïve Bayes Classifier	Naïve Bayes	Penelitian ini terdapat lebih banyak tanggapan positif dari masyarakat, menunjukkan bahwa kenaikan biaya haji pada tahun 2023 dapat diterima oleh masyarakat. Sementara itu, nilai akurasi tertinggi yang tercatat dalam penelitian ini adalah sebesar 81,46%.
4	Doni Andriansyah, Eka Wulansari Fridayanthie (2023)	<i>Optimization of Support Vector Machine and XGBoost Methods Using Feature Selection to Improve Classification Performance</i>	Support Vector Machine dan XGBoost	Hasil klasifikasi dengan seleksi fitur menunjukkan algoritma XGBoost memiliki performa yang lebih baik sebesar 91,4% dibandingkan dengan algoritma SVM sebesar 89,8% sehingga dapat digunakan untuk klasifikasi dan membantu dalam mendiagnosis penyakit kanker payudara.
5	Okta Jaya Harmaja, Irvan Prasetia, Yosi Victor Hutagalung, Hendra Ardanis Sirait (2023)	<i>Comprasion of Ensemble Learning Algorithm in Classifying Early Diagnostic of Diabetes</i>	Adaboost, XGBoost, dan Random Forest	Hasil penelitian menyimpulkan bahwa model dengan akurasi paling tinggi adalah Random Forest dengan nilai 0.86, diikuti oleh XGBoost dengan nilai 0.85, dan AdaBoost dengan nilai 0.82.
6	Muhammad Salabil, Nuril Lutvi Azizah (2023)	Implementasi <i>Data Mining</i> Dalam Melakukan Prediksi Penyakit Diabetes	Random Forest, dan XGBoost	Evaluasi model menunjukkan kinerja yang memuaskan dalam penelitian ini, dengan akurasi keseluruhan mencapai 74% untuk

No	Nama & Tahun Penelitian	Judul Penelitian	Algoritma	Hasil/Temuan
		Menggunakan Metode Random Forest Dan Xgboost		penggunaan Random Forest dan 76% untuk penggunaan XGBoost.
7	Rosdiana, Vera Novalia, Ilham Saputra, Mutammimul Ula, Muhammad Danil (2022)	<i>Application of Artificial Intelligence Chi-Square Model and Classification Of KNN in Heart Disease Detection</i>	K-Nearest Neighbor	Kesimpulan dari penelitian pengujian klasifikasi model KNN untuk mendeteksi penyakit jantung dari data akurasi dengan pemilihan data terbaik pada k=20 dengan akurasi tertinggi 89,4%. Data yang sebelum direduksi sebesar 83,38%.
8	Arif Budiarto, Agus Subekti, Haris Darmawan (2022)	Pendeteksian Potensi Faktor Risiko Tinggi Kesehatan Jemaah Haji Menggunakan Algoritma Klasifikasi <i>Data Mining</i>	Decision Tree, AdaBoost, Random Forest, dan Naïve Bayes	Hasil penelitian model klasifikasi algoritma Decision Tree memiliki akurasi yang tinggi sebesar 83.31% dalam memprediksi potensi faktor risiko tinggi (risti) kesehatan Jemaah Haji. Dilanjut dengan algoritma AdaBoost 83.30%, Random Forest 83.30%, dan Naïve Bayes 82.29%.
9	Deo Haganta Depari, Yuni Widiastiwi, Mayanda Mega Santoni (2022)	Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung	Naïve Bayes, Decision Tree, dan Random Forest	Akurasi model Naive Bayes mencapai 71%, model Decision Tree mencapai 72% dan model Random Forest memberikan hasil klasifikasi tertinggi, yaitu 75%, di mana metode ini dianggap sebagai yang terbaik untuk digunakan.
10	Ariansyah Arifin, Marshall Anugrah Najmi, Jovanka Samudra, Sultan Farel	Klasifikasi Dalam Mendeteksi Penyakit Hipoglikemia Dengan Menggunakan Metode Random Forest dan Adaboost	Random Forest, dan AdaBoost	Hasil penelitian menunjukkan penggunaan algoritma Random Forest menghasilkan tingkat akurasi sebesar 92%, Dengan menggunakan algoritma AdaBoost, diperoleh akurasi yang lebih tinggi mencapai 93.75%.

No	Nama & Tahun Penelitian	Judul Penelitian	Algoritma	Hasil/Temuan
	Syah Reza, Desta Sandya Prasvita (2022)			
11	Muhammad Kaddafi Nasution, Rd. Rohmat Saedudin, Vandha Pradwiyasma Widartha (2021)	Perbandingan Akurasi Algoritma Naïve Bayes dan Algoritma XGBoost pada Klasifikasi Penyakit Diabetes	Naïve Bayes, dan XGBoost	Hal tersebut dapat dilihat dari hasil akurasi terbaik dicapai oleh Algoritma XGBoost, yakni sebesar 90.10%. Sementara itu, Algoritma Naïve Bayes mencapai nilai akurasi terbaik sebesar 79.68%.
12	Syafitri Hidayatul Annur Aini, Yuita Arum Sari, Achmad Arwan (2018)	Seleksi Fitur Information Gain untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode K-Nearest Neighbor dan Naïve Bayes	K-Nearest Neighbor dan Naïve Bayes	Klasifikasi penyakit jantung menggunakan seleksi fitur Information Gain dengan kombinasi KNN dan Naïve Bayes, menghasilkan akurasi tertinggi saat menggunakan data latih dengan label kelas seimbang, yaitu sebesar 92,31% saat menggunakan enam fitur.

### 2.10.2 Matriks Penelitian

Tabel 2.2 merupakan matriks penelitian yang berisi penelitian yang berfokus untuk menyelesaikan masalah kesehatan dengan menggunakan algoritma klasifikasi.

Tabel 2.2 Matriks Penelitian

No	Penulis	Algoritma							Seleksi Fitur					
		Decision Tree	Naïve Bayes	Support Vector Machine	Ada Boost	XGBoost	Random Forest	K-Nearest Neighbor	Tanpa Seleksi Fitur	Backward Elimination	Pearson Correlation	Chi-Square	Recursive Feature Elimination	Information Gain
1	Feri Setiadi, dkk (2024)		✓				✓	✓					✓	
2	Hendro Nugroho, dkk (2023)		✓							✓				
3	Hertati, dkk (2023)		✓						✓					
4	Doni Andriansyah, dkk (2023)			✓		✓					✓			
5	Okta Jaya, dkk (2023)				✓	✓	✓		✓					
6	Muhammad Salabil, dkk (2023)					✓	✓		✓					
7	Rosdiana, dkk (2022)							✓			✓			



No	Penulis	Algoritma							Seleksi Fitur					
		Decision Tree	Naïve Bayes	Support Vector Machine	Ada Boost	XGBoost	Random Forest	K-Nearest Neighbor	Tanpa Seleksi Fitur	Backward Elimination	Pearson Correlation	Chi-Square	Recursive Feature Elimination	Information Gain
8	Arif Budiarto, dkk (2022)	✓	✓		✓		✓		✓					
9	Deo Haganta, dkk (2022)	✓	✓				✓		✓					
10	Ariansyah Arifin, dkk (2022)				✓		✓		✓					
11	Muhammad Kaddafi Nasution, dkk (2021)		✓			✓			✓					
12	Syafitri Hidayatul, dkk (2018)		✓					✓						✓
13	Penelitian sekarang, (2024)	✓			✓	✓	✓				✓			

Penelitian yang dilakukan mempunyai keterkaitan dengan penelitian sebelumnya yaitu mengenai haji dan kesehatan. Penelitian ini memiliki keterbaruan seperti pada data dan platform model klasifikasi yang digunakan. Data yang digunakan pada penelitian ini yaitu data pemeriksaan kesehatan jemaah haji di Kota dan Kabupaten Tasikmalaya tahun 2024, *platform* model yang digunakan adalah Google Collaboratory dengan menggunakan bahasa pemrograman Python.

Berdasarkan studi literatur yang telah dilakukan pada penelitian sebelumnya, masih sangat minim ditemukan penelitian data kesehatan yang melakukan seleksi fitur. Penelitian ini melakukan perbandingan akurasi algoritma pada pendekatan *supervised learning* dengan seleksi fitur Chi-square, sebagai keterbaruan dari penelitian yang juga menjadi perbedaan antara penelitian ini dan penelitian yang telah dilakukan. Algoritma pada pendekatan *supervised learning* yang digunakan yaitu algoritma Decision Tree, Random Forest, Adaboost dan XGBoost. Keempat algoritma digunakan pada penelitian ini karena algoritma tersebut termasuk algoritma *tree* yang memiliki nilai akurasi tinggi dibandingkan algoritma lainnya. Penggunaan seleksi fitur dilakukan untuk meningkatkan nilai akurasi pada keempat algoritma dengan memilih fitur terbaik. Penelitian dilakukan untuk meningkatkan pemahaman mengenai penggunaan algoritma ini lebih luas, terutama dalam penelitian mengenai status kesehatan jemaah haji.