

BAB III

METODE PENELITIAN

A. Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksplorasi data (*Data Mining*) berbasis kerangka kerja *Knowledge Discovery in Database* (KDD) untuk mengungkap pola hubungan antar rawi Hadis secara sistematis berdasarkan data yang tersedia. Penelitian ini bersifat deskriptif analitik, bertujuan untuk menjelaskan pola keterkaitan antar rawi guna memperoleh pemahaman yang terstruktur dan mendalam mengenai pola serta hubungan yang terjadi di antara para perawi. Pendekatan ini dirancang untuk memberikan hasil yang terukur dan dapat dipercaya.

B. Setting Penelitian

Penelitian ini dilakukan dengan pendekatan eksplorasi menggunakan *database* yang diperoleh dari aplikasi *Haditssoft*. Aplikasi ini menyediakan data Hadis yang lengkap, termasuk Sanad dan informasi perawinya. Penelitian berfokus pada pengumpulan data, proses pra-pemrosesan (*pre-processing*), dan penerapan algoritma *FP-Growth*, yang dilakukan dalam rentang waktu tertentu.

C. Data dan Sumber Data

Penelitian ini bertujuan untuk mengidentifikasi dan menganalisis pola hubungan antar perawi Hadis yang muncul dalam rangkaian Sanad dari empat kitab Hadis utama, yaitu *Shahih Al-Bukhari*, *Shahih Muslim*, *Sunan Al-Tirmidhi*, dan *Sunan Ibnu Majah*. Data penelitian diperoleh dari *database* aplikasi *Haditssoft*, yang memuat catatan Sanad Hadis dari keempat kitab tersebut. Secara keseluruhan, *dataset* mencakup 20.593 jalur Sanad, yang terdiri atas 7.008 Sanad dari *Shahih Al-Bukhari*, 5.362 Sanad dari *Shahih Muslim*, 3.891 Sanad dari *Sunan Al-Tirmidhi*, dan 4.332 Sanad dari *Sunan Ibnu Majah*. Selain itu, *dataset* ini juga mencakup 8.624 perawi unik yang tersebar dalam seluruh jalur

Sanad tersebut. Seluruh *dataset* tersebut melalui tahapan *pre-processing* sebelum dianalisis

Setiap Sanad memiliki beberapa perawi yang saling terhubung, membentuk jalur Sanad yang menggambarkan ketersambungan riwayat Hadis. *Shahih Al-Bukhari* dan *Shahih Muslim* dijadikan sebagai rujukan utama dalam penelitian ini, karena kualitas perawi yang terdapat dalam kedua kitab tersebut dinilai lebih unggul dibandingkan dengan perawi pada kitab *Sunan Al-Tirmidhi* dan kitab *Sunan Ibnu Majah*.

F. Alat dan Bahan

Penelitian ini membutuhkan berbagai alat dan bahan yang diperlukan untuk mendukung proses implementasi dan analisis data. Berikut ini adalah rincian alat dan bahan yang digunakan:

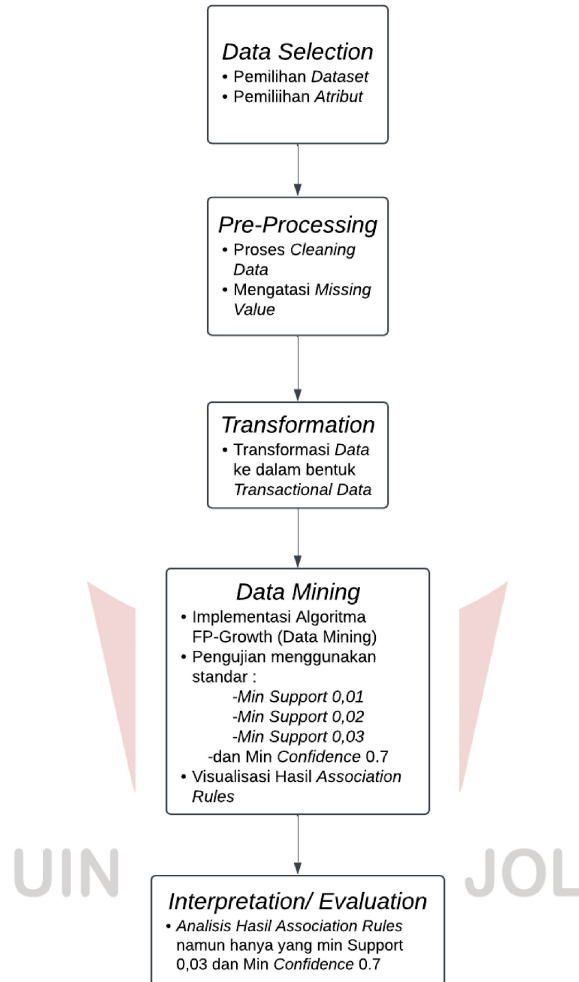
1) Alat

Komputer atau laptop dengan spesifikasi memadai digunakan sebagai perangkat utama untuk pengolahan data dan implementasi algoritma menggunakan *Python*, yang didukung oleh *library* seperti *pandas*, *numpy*, dan *mlxtend*, serta data diambil dari *Database* aplikasi *Haditssoft* yang menyediakan informasi lengkap mengenai Sanad dan perawi Hadis.

2) Bahan

Dataset Hadis yang mencakup data Sanad dan perawi dari kitab *Shahih Al-Bukhari*, *Shahih Muslim*, *Sunan Al-Tirmidhi*, dan *Sunan Ibnu Majah* menjadi bahan utama dalam analisis. Dokumentasi penelitian, seperti artikel jurnal, buku, dan sumber literatur lainnya, digunakan sebagai referensi untuk mendukung penerapan algoritma *FP-Growth*. Data tersebut telah diubah ke dalam bentuk *transactional data* sebagai notasi *Data Mining*, sehingga mempermudah proses eksplorasi pola hubungan antar perawi.

G. Kerangka Berpikir



Gambar 3. 1 Kerangka Berpikir Penelitian

Pada Gambar 3.1 ditampilkan kerangka berpikir penelitian ini yang menjelaskan tahapan *Knowledge Discovery in Database (KDD)*, yang terdiri atas proses *Data Selection*, *Data Preprocessing*, *Data Transformation*, *Data Mining*, serta *Intrepretation/Evaluation*.

a) *Data Selection*

Pada tahap *Data Selection*, peneliti menyeleksi 131 file mentah yang diperoleh dari *database* aplikasi *Haditssoft* menjadi 5 file utama yang dianggap relevan, yaitu *daftarrawi*,

PerawiBukhari, *perawiMuslim*, *perawiibnumajah*, dan *PerawiTirmidzi*. File *daftarrawi* memiliki peran penting dalam memetakan kode perawi dengan nama perawi sehingga memudahkan proses identifikasi dan analisis hubungan antar perawi. Sementara itu, empat file lainnya berisi kumpulan Sanad dari masing-masing kitab yang menjadi objek penelitian, yakni *Shahih Al-Bukhari*, *Shahih Muslim*, *Sunan Ibnu Majah*, dan *Sunan Al-Tirmidhi*. Adapun atribut yang digunakan dari data ini adalah *no_Hadis* dan *kode_perawi* pada setiap file per kitab, serta *kode_perawi* dan *nama_perawi* pada file *daftarrawi*. Dengan demikian, data yang telah terseleksi ini dapat mendukung analisis lebih lanjut dalam menemukan pola keterkaitan antar perawi Hadis secara lebih terstruktur dan valid.

b) *Pre-Processing*

Tahap *Pre-processing* berfokus pada proses pembersihan data agar siap digunakan pada tahap berikutnya. Pada tahap ini dilakukan *data cleaning* untuk memastikan bahwa tidak terdapat data yang kosong (*missing value*) atau atribut yang tidak terbaca oleh sistem. Hal ini dilakukan agar data yang sudah diseleksi benar-benar valid dan dapat diolah pada tahap analisis selanjutnya. Dengan demikian, kualitas data tetap terjaga serta mendukung kelancaran proses *Data Transformation* dan *Data Mining*.

c) *Transformation*

Tahap *Data Transformation* dilakukan dengan mengubah data yang sudah melewati proses sebelumnya menjadi bentuk *transactional data*. Pada tahap ini, setiap Hadis direpresentasikan sebagai sekumpulan item berupa perawi yang membentuk satu rangkaian Sanad. Format transaksional ini penting karena menyesuaikan struktur data agar dapat diproses oleh algoritma *FP-Growth*. Dengan adanya transformasi ini, pola keterhubungan

antar perawi dalam sebuah Sanad dapat dikenali lebih sistematis sehingga mempermudah proses pencarian *frequent itemsets* pada tahap *Data Mining*.

d) *Data Mining*

Tahap *Data Mining* pada penelitian ini dilakukan dengan menerapkan algoritma *FP-Growth* menggunakan bahasa pemrograman Python. Proses ini terdiri dari tiga langkah utama. Pertama, pembentukan *Conditional Pattern Base*, yaitu sub-database yang berisi lintasan awalan (*prefix path*) dan pola akhiran (*suffix pattern*) yang diekstraksi dari FP-Tree utama. Kedua, pembangunan *Conditional FP-Tree*, di mana setiap item dihitung nilai *support count*-nya berdasarkan *conditional pattern base* yang terbentuk. Hanya item yang memiliki nilai *support count* lebih besar atau sama dengan ambang batas *minimum support* yang akan dipertahankan untuk membentuk pohon baru. Ketiga, penemuan *Frequent Pattern*, yang diperoleh dengan mengombinasikan item dalam lintasan tunggal (*single path*) atau dengan melakukan proses rekursif jika lintasan bercabang, sehingga semua pola yang sering muncul dapat ditemukan secara menyeluruh.

Dalam penelitian ini, pengujian dilakukan dengan tiga variasi *minimum support* yaitu 0,01, 0,02, dan 0,03 serta *minimum confidence* sebesar 70%. Pemilihan tiga tingkat *minimum support* ini bertujuan untuk membandingkan pengaruh perubahan ambang batas terhadap jumlah pola dan aturan asosiasi yang dihasilkan. Nilai *minimum support* 0,03 dipilih sebagai batas tertinggi untuk memastikan bahwa pola yang muncul benar-benar signifikan, karena hanya perawi yang sering berulang dalam Sanad yang akan membentuk asosiasi. Sementara itu, *confidence* sebesar 70% dipilih untuk menjamin aturan yang terbentuk memiliki tingkat

kepercayaan yang tinggi sehingga layak dijadikan dasar dalam analisis hubungan antar perawi.

Implementasi FP-Growth ini dilakukan dengan Python karena sintaksisnya sederhana, memiliki pustaka yang mendukung seperti *mlxtend* untuk penambahan aturan asosiasi, serta didukung komunitas pengguna yang luas. Dengan menggunakan Python, keseluruhan proses dapat dilakukan secara lebih efisien dan terstruktur, mulai dari pembangunan FP-Tree hingga penemuan *frequent itemsets*. Hasil dari proses ini divisualisasikan dalam bentuk aturan asosiasi untuk masing-masing nilai *minimum support*, yang memberikan gambaran jelas mengenai keterkaitan antar perawi dalam Sanad pada keempat kitab Hadis yang dianalisis.

e) *Interpretation/Evaluation*

Tahap interpretasi dan evaluasi dilakukan untuk menilai kesesuaian pola keterkaitan antar perawi yang dihasilkan dari algoritma FP-Growth. Pada penelitian ini, evaluasi difokuskan pada hasil dengan *minimum support* 0,03 dan *minimum confidence* 70%, karena pada tingkat inilah pola yang muncul dapat dianggap paling signifikan dan memiliki reliabilitas tinggi. Dengan ambang batas *support* sebesar 0,03, aturan yang terbentuk hanya berasal dari perawi yang benar-benar sering muncul bersama dalam Sanad, sehingga mengurangi kemungkinan adanya pola yang bersifat kebetulan atau jarang terjadi. Sementara itu, nilai *confidence* sebesar 70% memastikan bahwa hubungan antar perawi yang terbentuk memiliki tingkat kepastian yang kuat, sehingga dapat dijadikan dasar analisis yang valid.