

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Hadits merupakan salah satu panduan yang digunakan oleh umat Islam dalam melaksanakan berbagai macam aktivitas baik yang berkaitan dengan urusan dunia maupun aktivitas yang berkaitan dengan urusan akhirat, berupa sesuatu yang disandarkan kepada Nabi Muhammad ﷺ baik berupa perkataan atau perbuatan. Hadits merupakan sumber hukum agama Islam yang kedua setelah kitab suci Al – Qur'an. Jika suatu perkara tidak dijelaskan di dalam Al – Qur'an, maka umat Islam akan menggunakan sumber yang kedua yaitu Hadits[1].

Dalam hadits terdapat banyak bab yang membahas beberapa perkara, misalnya bab tentang iman, ilmu, shalat, wudlu, haji, puasa dan masih banyak lagi. Pada sumber asalnya hadits menggunakan bahasa Arab, namun seiring perkembangan jaman sudah banyak terdapat terjemahannya terutama kedalam bahasa Indonesia. Dengan banyaknya bab dalam hadits akan menjadi hal yang menarik untuk dipadukan dengan teknologi dan teknik data mining terutama text mining guna mengelompokkan hadits berdasarkan isi dari topik hadits yang bersangkutan. Pada referensi penelitian sebelumnya, masih jarang ditemukan penelitian mengenai pengelompokan hadits berdasarkan matan hadits. Adapun beberapa aplikasi mengenai hadits yang mengatur pengelompokan hadits masih secara manual melalui relasi *database*, belum ter-otomatisasi menggunakan metode atau algoritma tertentu.

Berbagai metode telah digunakan untuk melakukan pengelompokan pada suatu data tertentu salah satunya klasterisasi. Klasterisasi merupakan sebuah teknik

pengelompokan dimana data dikelompokkan berdasarkan kemiripannya dengan objek lainnya pada kluster yang sama, namun tidak akan mirip dengan objek lain pada kluster yang berbeda[2]. Tingkat kemiripan masing-masing objek antar kluster akan berbeda, ketidakmiripan ini didasarkan pada nilai atribut yang menjelaskan objek. Klusterisasi berbeda dengan klasifikasi, dimana klasifikasi menggunakan sebuah acuan data atau biasa disebut dengan sebuah data training untuk mengukur sebuah data yang baru masuk lalu dibandingkan dengan data training untuk dimasukkan ke dalam sebuah kategori yang sudah ditetapkan. Pada klusterisasi, pengelompokan data dilakukan tanpa berdasarkan acuan data training tertentu. Membuat klusterisasi dapat dipakai untuk memberikan label pada kelompok yang belum diketahui sebelumnya, hal ini berbeda dengan klasifikasi, dimana label kelas telah diketahui[3].

Algoritma K-Means merupakan algoritma *clustering* partisi yang paling dikenal dalam melakukan klusterisasi. Kelebihan metode K-Means ini adalah sederhana, efisien, mudah dipahami, dan cepat dalam proses klusterisasi nya[4]. Dengan memberikan kumpulan data point dan jumlah k yang diinginkan, algoritma ini akan mengulangi data ke k kluster berdasarkan fungsi jarak. Rujukan penelitian sebelumnya [4][5], pada penelitian [4] dilakukan pengujian menggunakan metode K-Means untuk mengelompokkan ayat-ayat Al-Qur'an terjemahan bahasa Indonesia, dengan menggunakan beberapa percobaan menghasilkan nilai Silhouette Koefisien sebesar 0.3777 untuk surat al-baqarah, tidak maksimalnya akurasi kluster diakibatkan salah satunya karena pendefinisian awal pusat kluster secara acak, dengan adanya penelitian ini diharapkan untuk penelitian selanjutnya, dapat menggunakan metode lain. Dan pada penelitian [5] melakukan perbandingan

algoritma klusterisasi yaitu K-Means, SOM, dan *Hierarchical* dalam mengelompokkan data set waktu yang menghasilkan Algoritma SOM lebih baik dari K-Means dan *Hierarchical* namun K-Means sangat baik ketika diberi dataset yang besar. Pada penelitian sekarang, peneliti akan mencoba uji perbandingan menggunakan K-Means dan Fuzzy C-Means.

Fuzzy C-Means Clustering (FCM) adalah suatu teknik klusterisasi fuzzy dimana keberadaan tiap titik data dalam suatu kluster ditentukan oleh derajat keanggotaan. Dengan metode ini, data akan dikelompokkan ke dalam kluster-kluster yang berbeda sesuai persentase keanggotaannya di masing-masing kluster. Kluster yang dipilih adalah kluster yang mempunyai persentase keanggotaan terbesar. Rujukan penelitian sebelumnya [6][7][8], pada penelitian [6] melakukan pengelompokan data nilai siswa SMA untuk penentuan jurusan menggunakan metode Fuzzy C-Means, FCM menghasilkan akurasi sebesar 72,26% untuk 3 kluster. Pada penelitian [7] yaitu melakukan pengelompokan data curah hujan kota Bengkulu untuk membantu prediksi hujan berdasarkan wilayah menghasilkan bahwa metode FCM dapat menentukan lokasi terbaik dalam cluster berdasarkan proses iterasi nya. Rujukan lainnya [8] melakukan pengelompokan pada berita twitter terhadap layanan berita pada akun portal, didapatkan hasil kategori berupa Kriminal, politik, umum, religi, olahraga, ekonomi, kesehatan, edukasi, teknologi, makanan dan minuman. Dengan akurasi mencapai 80.95%, untuk ketidak akuratan sisanya diakibatkan oleh penentuan derajat keanggotaan awal yang kurang tepat sehingga menimbulkan pusat kluster awal kurang tepat.

Ada beberapa rujukan penelitian sebelumnya mengenai perbandingan metode K-Means dan Fuzzy C-Means yaitu pada penelitian [9][10][11][12], pada

penelitian [9][10] menggunakan data yang berbeda yaitu data kemiskinan dan data nilai akademik siswa. Dimana hasil yang didapat dari perbandingan kedua algoritme adalah K-Means yang lebih unggul dan cocok dibandingkan dengan algoritma Fuzzy C-Means. Namun, pada penelitian [11][12] yang masing-masing menggunakan *data user knowledge modeling* dan data kepuasan siaran televisi menghasilkan bahwa algoritma Fuzzy C-Means lebih unggul dan lebih cocok dibandingkan dengan algoritme K-Means.

Berdasarkan rujukan penelitian diatas membuktikan bahwa adanya perbedaan hasil akurasi ketika penelitian pengelompokan menggunakan tipe data yang berbeda. Serta cukup berpengaruh nya penentuan pusat klaster awal dan derajat keanggotaan pada hasil klaster yang pada dasarnya dilakukan secara acak pada masing-masing algoritma. Penelitian sekarang akan mencoba mengelompokkan hadits secara otomatis menggunakan metode/algoritma berdasarkan kemiripan matan hadits. Selain mengelompokkan hadits, penelitian ini juga berfokus untuk membandingkan algoritma K-Means yang pada dasarnya *clustering* tegas (*Crisp Clustering*) dan Fuzzy C-Means yang pada dasarnya *clustering fuzzy* (*Soft Clustering*) menggunakan data teks yaitu Hadits terjemahan bahasa Indonesia. Atribut yang akan dipakai yaitu matan atau isi dari hadits tersebut. Dimana isi hadits tersebut akan di kelompokkan berdasarkan kemiripan isi topiknya kedalam beberapa bab dan akan dihitung berapa akurasi ketepatan klasterisasi nya. Ada beberapa parameter yang akan ditekankan atau difokuskan pada penelitian ini, yaitu parameter yang dapat mempengaruhi hasil klaster dan akurasi pada masing-masing algoritma diantaranya jumlah klaster, jumlah iterasi, waktu eksekusi, jumlah dimensi data, awal pusat klaster, awal derajat keanggotaan,

serta kombinasi pengujian awal pusat klaster dan awal derajat keanggotaan. Diharapkan dengan adanya pertimbangan parameter diatas dapat diketahui faktor apa saja yang mempengaruhi hasil klasterisasi data menjadi lebih baik.

Maka dari itu, penulis tertarik untuk melakukan penelitian analisis perbandingan dengan judul “**Perbandingan Analisis Algoritma K-Means Dan Fuzzy C-Means Untuk Pengelompokan Hadits Terjemahan Bahasa Indonesia**”.

## 1.2 Rumusan Masalah

Berdasarkan dari latar belakang diatas maka dapat dirumuskan menjadi beberapa masalah diantaranya:

- a. Bagaimana kinerja metode K-Means dan Fuzzy C-Means pada pengelompokan Terjemahan Hadits Bahasa Indonesia?
- b. Bagaimana pengaruh pendefinisian awal pusat klaster awal, derajat keanggotaan awal, jumlah klaster, dan jumlah dimensi data pada hasil pengelompokan?
- c. Metode manakah yang lebih cocok untuk pengelompokan Terjemahan Hadits Bahasa Indonesia?

## 1.3 Tujuan Penelitian

Adapun tujuan penelitian dari proposal penelitian ini antara lain:

- a. Untuk mengetahui kinerja metode K-Means dan Fuzzy C-Means pada pengelompokan data menggunakan beberapa pengujian untuk menentukan kinerja masing-masing metode

- b. Untuk mengetahui pengaruh pendefinisian pusat kluster awal, derajat keanggotaan awal, jumlah kluster, dan jumlah dimensi data pada hasil pengelompokan
- c. Untuk menentukan metode yang cocok untuk data Terjemahan Hadits Bahasa Indonesia

#### 1.4 Batasan Masalah

Pada tugas akhir ini terdapat beberapa batasan masalah agar hasil dari tugas akhir ini sesuai dengan apa yang diharapkan. Adapun batasan masalah dalam tugas akhir ini yaitu:

- a. Data yang digunakan berupa teks (data tidak terstruktur) Bahasa Indonesia
- b. Data yang akan digunakan untuk pengelompokan yaitu isi (matan) dari hadits Bukhari berbahasa Indonesia
- c. Bab hadits yang digunakan untuk percobaan yaitu bab wudlu, shalat, zakat, haji, dan puasa
- d. Metode *Text Preprocessing* menggunakan *Stop word* (penghapusan kata), *Tokenize* (Pemisahan kata), *Stemming* (Perubahan kata dasar), dan penghitungan TF-IDF (Pembobotan kata)
- e. Penelitian menggunakan bahasa pemrograman Python untuk mempermudah perhitungan klasterisasi dan implementasi metode
- f. Metode pengembangan perangkat lunak menggunakan metode *Rapid Application Development (RAD)*.
- g. Pengujian hasil kluster menggunakan *Silhouette Coefficient* dan pengukuran lain seperti *Precision*, *Recall*, *Accuracy*, dan *F-Measure*

- h. Penelitian akan membandingkan kinerja masing-masing algoritma secara internal menggunakan pengujian *silhouette coefficient* dengan memperhatikan variabel sebagai berikut: jumlah kluster, waktu eksekusi, dimensi data, iterasi, pusat kluster awal, dan derajat keanggotaan awal. Adapun secara eksternal menggunakan membandingkan hasil *Precision*, *Recall*, *Accuracy*, dan *F-Measure* sesuai dengan data aslinya.
- i. Pada pengujian tingkat *Precision*, *Recall*, *Accuracy*, dan *F-Measure*, kluster yang digunakan pada pengujian ini yaitu kluster yang berjumlah 5 kluster dan kombinasi yang digunakan diambil berdasarkan nilai *Silhouette Coefficient* terbaik serta yang memiliki jumlah iterasi paling minimum.

### 1.5 Manfaat Penelitian

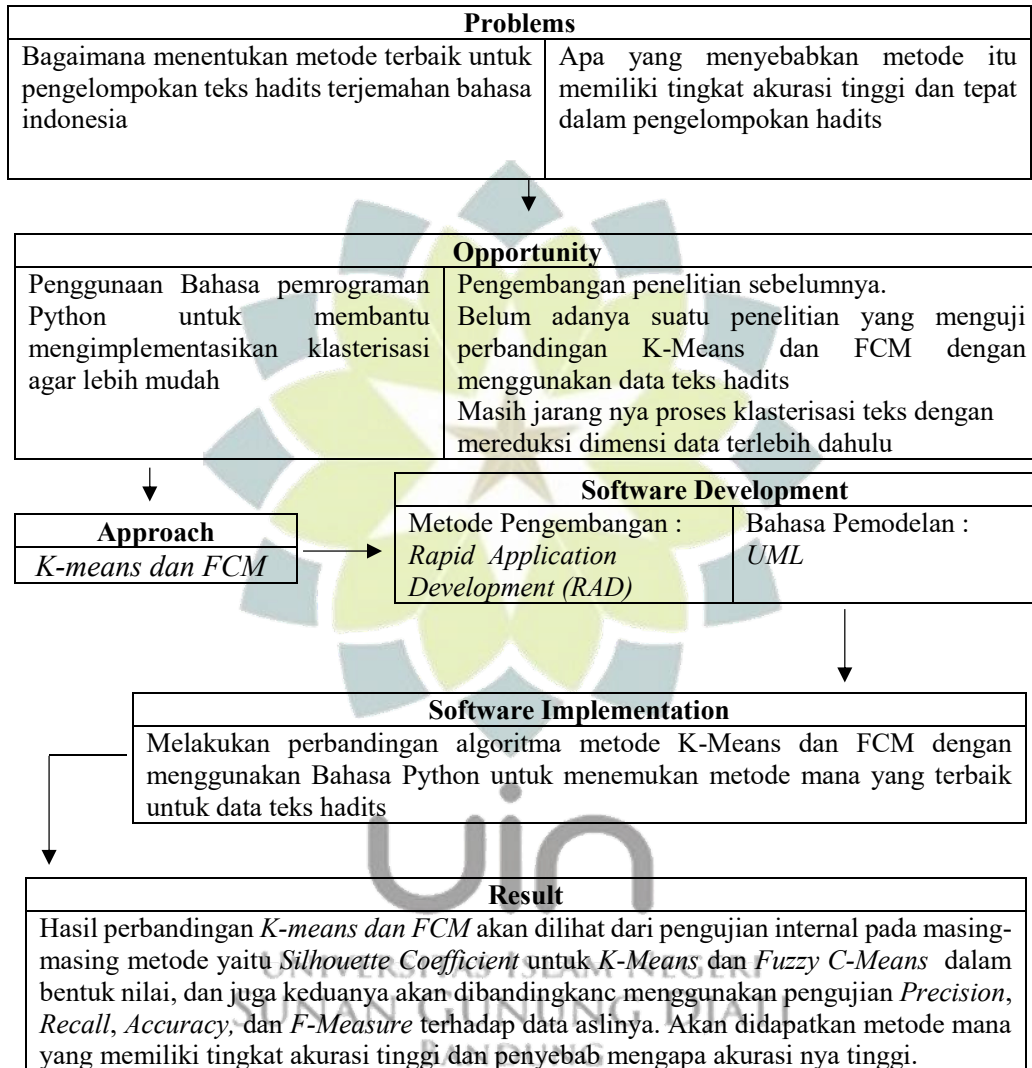
Manfaat yang diharapkan pada penelitian ini adalah:

- a. Mengelompokkan hadits secara otomatis berdasarkan kemiripan isi hadits.
- b. Menganalisis kinerja K-Means dan Fuzzy C-Means terhadap data teks hadits.
- c. Memberikan gambaran mengenai metode klusterisasi dengan metode K-Means dan Fuzzy C-Means.
- d. Menambah rekomendasi skenario percobaan untuk dimanfaatkan dalam pengujian lain.
- e. Sebagai referensi bagi peneliti lain yang ingin mengembangkan metode klusterisasi.

## 1.6 Kerangka Pemikiran

Adapun kerangka pemikiran dari Penelitian ini yang di gambarkan pada

Gambar 1.1.



Gambar 1. 1 Kerangka Pemikiran



## 1.7 Metodologi Penelitian

Berikut adalah metodologi yang digunakan pada penelitian ini.

### 1.7.1 Metodologi Penelitian

Metode yang digunakan dalam penelitian kali ini adalah sebagai berikut:

#### a. Data

Data yang digunakan pada penelitian ini adalah teks hadits imam Bukhari terjemahan bahasa Indonesia. Data yang digunakan berasal dari sebuah aplikasi *desktop* yang membahas lengkap tentang ilmu hadits terutama dari 9 imam, aplikasi tersebut bernama HaditsSoft. Isi atau database di ambil dari aplikasi Ensiklopedi Hadits Kitab 9 Imam oleh Lidwa Pusaka, ditambah data-data lain seperti derajat hadits yang di ambil dari *ebook* terjemahan inggris oleh Darussalam juga *ebook* lain untuk derajat Musnad Ahmad oleh Syu'aib Al Arnauth. Adapun tambahan lima kitab lainnya kami ambil dari ebook Kampungsunah dan website [shirathal-mustaqim.org](http://shirathal-mustaqim.org).

Peneliti telah meminta izin kepada pengembang aplikasi HaditsSoft selaku pembuat aplikasi dan membuat struktur *database* hadits yang diambil dari berbagai sumber-sumber yang ada, agar dapat menggunakan *database* hadits nya untuk keperluan penelitian ini.

#### b. Analisis

Analisis data dalam penerapan data mining ini menggunakan tahapan Knowledge Discovery in Database (KDD) yang terdiri dari beberapa tahapan, yaitu data *selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation*.

c. Perancangan

Tahapan ini dilakukan setelah analisis. Dalam tahapan ini dilakukan perubahan kebutuhan-kebutuhan menjadi bentuk karakteristik yang mudah dipahami serta desain mekanik.

d. Penulisan Program

Setelah melakukan perancangan maka tahapan selanjutnya diterapkan pada penulisan program. Penulisan program ini digunakan bahasa pemrograman *Python*.

e. Pengujian

Setelah selesai dibuat programnya, selanjutnya tahap pengujian menggunakan teknik *black box testing*.

## 1.8 Sistematika Penulisan

Penulisan laporan kerja praktik ini di bagi ke dalam enam bab, yang disusun berdasarkan sistematika sebagai berikut:

### **BAB I            PENDAHULUAN**

Bab ini berisi mengenai latar belakang masalah, perumusan masalah, batasan masalah, tujuan kerja praktik, manfaat kerja praktik, metodologi kerja praktik, dan sistematika penyusunan.

### **BAB II            LANDASAN TEORI**

Bab ini menjelaskan tentang teori-teori yang akan digunakan untuk memecahkan permasalahan dan mendukung penelitian yang menjadi tinjauan utama untuk pemrogram yang akan dibuat.

### **BAB III ANALISIS DAN PERANCANGAN SISTEM**

Bab ini membahas mengenai kajian website yang akan dibuat dari mulai melakukan analisis sistem sampai perancangan sistem.

### **BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM**

Pada bab keempat menerangkan tentang hasil implementasi dan pengujian dari sistem yang telah dibangun.

### **BAB V PENUTUP**

Pada bab kelima berisi pernyataan singkat berupa kesimpulan dari pembahasan sistem yang dibuat secara keseluruhan dan saran untuk mengembangkan sistem yang lebih baik lagi untuk ke depannya

