

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Al-quran adalah kitab suci terakhir yang juga merupakan pedoman untuk umat islam dan juga sumber utama ajaran islam. Bukan hanya mengajarkan petunjuk hubungan manusia dengan tuhan, tetapi juga mengajarkan hubungan manusia dengan sesama manusia (*hablum min Allah wa hablum min an-nas*). Didalam al-quran diajarkan juga hubungan manusia dengan alam. Agar bisa mengamalkannya dalam kehidupan sehari-hari, dibutuhkan pemahaman tentang kandungan al-quran seraca konsisten dan sungguh-sungguh[1].

Al-quran juga merupakan pedoman hidup bagi seluruh umat islam, maka dari itu seluruh umat islam memiliki kewajiban untuk mempelajari Al-quran. Pada saat mempelajari Al-quran, terkadang hanya melakukan pemahaman yang terdapat pada ayat-ayat tertentu tanpa memahami ayat-ayat yang terdapat dalam surat yang lainnya.

Keterkaitan antara ayat satu dengan ayat yang lainnya dapat memudahkan dalam memahami secara tepat ayat-ayat dan surat yang, hal tersebut menyebabkan diperlukannya pengelompokan ayat-ayat Al-quran yang memiliki topik yang serupa. Adapun salah satu cara untuk melakukan pengelompokan ayat tersebut adalah dengan menggunakan algoritma *clustering*. Algoritma *clustering* merupakan sebuah proses untuk melakukan pengelompokan data, kedalam beberapa *cluster* sehingga data didalam *cluster* tersebut memiliki karakteristik yang sama dan memiliki karakteristik yang berbeda pada antar *cluster*. Adapun algoritma untuk melakukan pengelompokan ayat-ayat al-quran yang mempunyai tema yang

sama kedalam beberapa *cluster* adalah menggunakan algoritma K-means++ *clustering*.

*Clustering* adalah sebuah proses untuk mempartisi suatu set objek data kedalam bagian yang disebut dengan *cluster*. Algoritma yang sering digunakan untuk mengklaster objek adalah K-means *Clustering*[2]. *Clustering* dapat mengelompokkan data kedalam beberapa klaster, termasuk mengelompokkan data data terjemahan Al-quran kedalam beberapa kelompok. Dimana didalam suatu kelompok tersebut, memiliki karakteristik yang sama dan memiliki karakteristik yang berbeda pula dengan data yang berada di kelompok lain.

Cara kerja k-means *clustering* yaitu data atau obyek dimulai dengan kelompok centroid pertama dipilih secara random (acak), yang bertujuan untuk titik awal pada setiap cluster. Langkah selanjutnya adalah dengan melakukan perhitungan secara terus menerus untuk mengoptimalkan posisi centroidnya. Proses tersebut akan berhenti ketika centroidnya telah stabil dan jumlah iterasi yang ditentukan sebelumnya sudah tercapai[3]. Algoritma ini termasuk kedalam algoritma pengelompokan yang paling banyak digunakan karena pengerjaannya yang sederhana dan cepat, tetapi terkadang menghasilkan hasil yang tidak optimal atau jauh dari optimal karena didalam K-means tidak ada jaminan akurasi.

Rujukan penelitian sebelumnya[4], menggunakan algoritma k-means untuk mengelompokkan terjemahan ayat al-quran dalam Bahasa Indonesia. Adapun percobaan yang dilakukannya menghasilkan nilai *Silhouette Koefisien* sebesar 0.3777 pada surat Al-baqarah. Akurasi klaster yang tidak maksimal dikarenakan salah satunya pada pendefinisian awal pusat klasternya yang secara acak.

Dalam penelitian sekarang, akan mencoba pengelompokan data menggunakan algoritma k-means++ *clustering*. K-means++ adalah pengembangan dari k-means. Salah satu kelemahan algoritma k-means adalah tidak dapat memberikan akurasi yang tepat meskipun perhitungan tersebut termasuk sangat cepat dan juga sederhana. Menurut jurnal K-Means++ *The Advantages of Careful seeding* jika k-means ditambahkan *randomized seeding technique* hal ini dapat

Meningkatkan akurasi pada algoritma k-means[5]. Akurasi tersebut akan bergantung pada nilai centroid yang ditentukan diawal perhitungan hasilnya jika menggunakan nilai centroid yang berbeda maka akan memerikan hasil yang berbeda pula dan bahkan membutuhkan banyak perulangan untuk menentukan anggota suatu cluster jika nilai centroidnya tidak tepat. Dengan itu penambahan rumus *randomized seeding technique* maka akan memudahkan nilai centroid ketika awal perhitungan.

Penelitian sebelumnya [6][7] mengelompokan surat-surat yang ada didalam al-quran, hasilnya dari kedua penelitan tersebut menghasilkan kelompok sebanyak 3 cluster dan menghasilkan kelompok sebesar 4 cluster. Efisiensi algoritmanya dibatasi karena pemilihan k centroid pertama dilakukan secara acak. Adapun penelitian [8] menggunakan K-NN untuk pengeopokan surat-surat dalam al-quran, K-NN mampu mengelompokan surat-surat didalam Al-quran secara baik. Tetapi penentuan awal nilai k sangat mempengaruhi keakuratan klasifikasi.

Adapun pada penelitian [9] menggunakan Algoritma C4.5 untuk mengelompokan Surat dalam Al-quran, pengelompokan tersebut berdasarkan tempat diturunkannya (*Makkiyah dan Madaniyah*) dan golongan suratnya. Didapatkan akurasi hasil pengujian pada aplikasi yang dibuat, dari 114 data surat

dengan menggunakan algoritma C4.5 yaitu 95,6% dari hasil algoritma ini. Terbukti bahwa algoritma C4.5 dapat membantu user dalam mengelompokkan surat-surat di dalam Al-quran. Adapun penelitian sebelumnya [10] menggunakan algoritma K-means pada data Ekspor Buah-Buahan algoritma ini terbukti dapat mengelompokkan data tersebut dan menghasilkan tiga kelompok cluster. Tetapi, dibutuhkannya peningkatan akurasi clustering pada data agar mendapatkan hasil yang lebih akurat pada proses clustering. Rujukan penelitian sebelumnya [11] menggunakan perbandingan algoritma Fuzzy C-means dan K-Means++ menggunakan Dataset Iris, *Soybean-small* dan wine menghasilkan tingkat akurasi yang sama pada kedua algoritma. Tetapi algoritma K-means++ lebih unggul tingkat akurasinya dibandingkan dengan algoritma Fuzzy C-means.

Dari hasil rujukan penelitian diatas terbukti bahwa adanya perbedaan hasil akurasi penelitian ketika pengelompokan data dengan tipe yang berbeda. Penentuan awal *cluster* sangat berpengaruh agar dapat menghasilkan akurasi yang tepat pada proses pengelompokan, karena pada dasarnya penelitian yang dilakukan sebelumnya penentuan titik awal *cluster* dilakukan secara acak sehingga menghasilkan akurasi *cluster* yang kurang maksimal. Pada penelitian sekarang akan mencoba mengelompokkan terjemahan ayat-ayat al-quran dengan menggunakan algoritma K-means++ untuk penentuan awal *cluster* yang lebih akurat agar menghasilkan akurasi *cluster* yang lebih maksimal dengan judul “**Klasterisasi Terjemahan Ayat-ayat Al-quran Menggunakan Algoritma K-means++**”.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang diatas, dapat dirumuskan permasalahan sebagai berikut :

1. Bagaimana Pengelompokan Terjemahan ayat – ayat Al-quran Menggunakan Metode Algoritma K-means++ ?
2. Apa saja interpretasi atau makna hasil klusterisasi terjemahan ayat-ayat Al-quran ?

### 1.3 Tujuan Penelitian

Adapun tujuan dari pembangunan system clusterisasi terjemahan ayat-ayat al-quran ini adalah sebagai berikut :

1. Untuk mengelompokan terjemahan ayat-ayat al-quran yang memiliki kemiripan tema yang sama kedalam beberapa klaster
2. Untuk menengetahui atau menginterpretasikan hasil klusterisasi terjemahan ayat-ayat al-quran dalam Bahasa indonesia

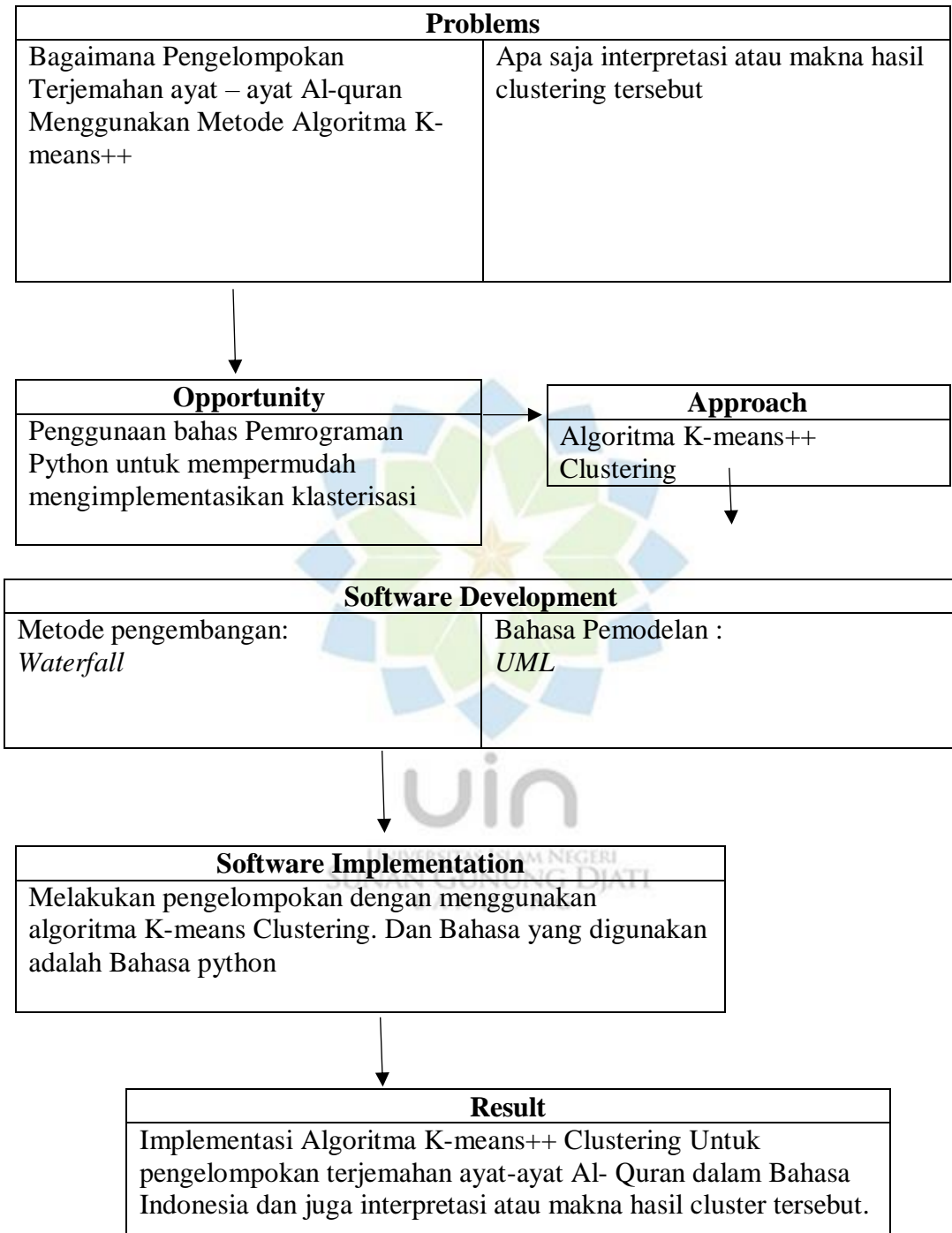
### 1.4 Batasan Masalah

Agar penelitian ini lebih terarah dan sesuai dengan tujuan yang diinginkan maka berikut adalah beberapa Batasan masalah dari pembangunan Sistem klusterisasi terjemahan ayat-ayat al-quran menggunakan algoritma K-means++

1. Input dari sistem ini berupa terjemahan ayat al-quran dalam bahasa indonesia
2. Sistem dirancang menggunakan bahasa pemrograma *Python*
3. Pengembangan sistem ini menggunakan bahasa pemrograman *python*
4. Metode *Text Processing* menggunakan *Tokenize* (Pemisahan kata), *Stopword* (penghapusan kata), *Stemming* (Perubahan kata dasar menggunakan algoritma nazief), dan Perhitungan TF-IDF (pembobotan kata).
5. Sistem ini menggunakan Algoritma K-Means++ *Clustering*

## 1.5 Kerangka Pemikiran

Tabel 1. 1 Tabel kerangka pemikiran



Penelitian ini bertujuan untuk mengelompokkan terjemahan ayat-ayat Al-quran menggunakan Bahasa Indonesia sesuai dengan kesamaan kata yang sering muncul

menggunakan algoritma K-means++ dan juga mencari tahu interpretasi atau makna hasil dari klasterisasi tersebut. Bahasa yang akan digunakan dalam penelitian ini adalah Bahasa pemrograman *python* dan juga algoritma K-means++ clustering. Metode pengembangannya menggunakan metode waterfall dan Bahasa pemodelan UML. Diharapkan hasil dari penelitian ini dapat mengelompokan ayat-ayat terjemahan al-quran sesuai dengan kemiripan tema atau kata pada setiap cluster yang telah di bentuk.

## 1.6 Metode Penelitian

### a. Teknik Pengumpulan Data

Untuk bisa mendapatkan data yang lebih akurat untuk bahan penelitian ini, adapun beberapa cara untuk teknik pengumpulan datanya yang bertujuan untuk mendapatkan keterangan yang rinci dan juga jelas tentang masalah yang ada. Berikut merupakan teknik pengumpulan data yang dipakai pada penelitian ini :

#### 1. Studi pustaka

Studi disini dilakukan dengan cara meneliti, menelaah dan juga mempelajari berbagai literature yang sumbernya dari buku, jurnal ilmiah, situs-situs pada internet juga bacaan yang berkaitan dengan topik penelitian.

#### 2. Dokumen

Penelitian disini menggunakan data terjemahan ayat-ayat al-quran dalam Bahasa Indonesia, yang sudah didigitalisasi dalam bentuk basis data versi kementerian agama. Basis data didapatkan dengan cara mengunduhnya dari [qurandatabase.org](http://qurandatabase.org), yang merupakan sebuah web penyedia basis data al-quran yang berisi 114 surat dengan jumlah total ayat 6236 ayat.

### b. Metode Pengembangan Perangkat Lunak

Pengembangan sistem ini menggunakan metode waterfall. Berikut merupakan tahapan-tahapan dari metode waterfall :

#### 1. Analisa Kebutuhan

Komponen apa saja yang diperlukan system dapat dijabarkan sebagai berikut:

- a) Input sistem
- b) Output sistem
- c) Proses yang berjalan dalam sistem
- d) Database yang digunakan

#### 2. Desain sistem

Desain sistem merupakan tahapan dimana dilakukannya perancangan sistem terhadap solusi dari permasalahan yang ada. Desain sistem disini menggunakan perangkat pemodelan sistem seperti *Data Flow Diagram*, *Use Case Diagram*, *class diagram* dan lain lain.

#### 3. Penulisan Kode Program

*Coding* atau penulisan kode program merupakan penerjemahan dari design kedalam Bahasa yang dapat dikenali oleh komputer.

#### 4. Pengujian Program

Merupakan tahapan akhir dimana sebuah sistem yang baru diuji kemampuannya sehingga bisa didapatkan kelemahan dan kekurangan sistem tersebut, kemudian dilakukan perbaikan terhadap aplikasi tersebut sehingga menjadi aplikasi yang lebih baik.



## **1.7 Sistematika Penulisan**

### **BAB I PENDAHULUAN**

Bab ini berisikan tentang latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, metodologi penelitian, dan sistematika penulisan.

### **BAB II LANDASAN TEORI**

Menjelaskan tentang teori-teori yang digunakan dalam tugas akhir ini serta untuk menjelaskan dan menyelesaikan permasalahan yang akan dikaji.

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

Bab ini juga membahas tentang rancangan aplikasi yang nantinya akan dibangun.

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

Membahas proses implementasi dan pengujian perangkat lunak dengan detail. Proses Implementasi terdiri dari Persiapan software dan hardware, instalasi aplikasi, dan juga tampilan akhir aplikasi yang dibangun. sedangkan pengujian meliputi identifikasi software, rencana pengujian, kasus uji dan hasil uji, evaluasi pengujian.

### **BAB V PENUTUP**

Membahas tentang kesimpulan dan saran yang diperoleh dari penulisan tugas akhir ini.

