

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Data is New Oil adalah istilah yang sering didengar pada era teknologi saat ini. Hal ini terjadi karena digitalisasi yang dilakukan pada hampir semua aspek kehidupan. Berbagai kegiatan sehari-hari, mulai dari perdagangan, pemerintahan, pembelajaran hingga aktualisasi diri, semuanya dilakukan melalui aplikasi dan akses internet.

Kegiatan digital tersebut meninggalkan jejak informasi atau data yang biasanya disimpan dalam platform penyimpanan berskala besar. Data-data tersebut dikenal sebagai raw data. Sama seperti minyak mentah yang perlu ditambang, diekstraksi dan akhirnya dimurnikan sebelum dapat digunakan, data juga perlu ditambang, dikelompokkan dan dianalisa. Salah satu cara untuk memproses data adalah dengan data mining.

Data mining merupakan suatu proses penambangan informasi dari data. Salah satu fungsi data mining adalah fungsi deskriptif yakni suatu fungsi untuk memahami lebih jauh tentang data yang diamati. Sehingga diharapkan dapat mengetahui karakteristik dari data tersebut berdasarkan pola tertentu yang tersembunyi dalam data.

Salah satu jenis data yang menarik untuk dilakukan data mining adalah teks. Dikenal sebagai text mining, proses data mining pada teks dapat menghasilkan informasi penting yang terkandung dalam teks.

Teks yang populer untuk diteliti adalah teks agama, hal ini jelas penting karena sebagai umat islam kita diwajibkan untuk membaca dan memahami lebih dalam isi Al-Qur'an. Dalam memudahkan memahami isi Al-Qur'an kita dapat mempelajari terjemahannya. Dan dengan menggunakan text mining kita dapat mengambil informasi penting yang terkandung dalam Al-Qur'an.

Adapun metode text mining yang biasa digunakan adalah pengelompokan data atau clustering dimana salah satu clustering yang populer adalah K-means clustering.

K-means clustering merupakan metode unsupervised machine learning yang paling sederhana. Tujuan dari algoritma K-means adalah untuk menemukan

dan mengelompokkan grup dalam data. Sehingga K-means clustering dapat digunakan untuk mengelompokkan kata dalam terjemahan Al-Qur'an.

Oleh karena itu pada tugas akhir ini penulis akan mengkaji mengenai K-means clustering untuk terjemahan ayat Al-Qur'an dengan judul "ANALISA KLASSTERING MENGGUNAKAN PARALEL K-MEANS PADA DATA TERJEMAH AL-QUR'AN BAHASA INDONESIA."

## 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas maka dirumuskan masalah:

1. Bagaimana implementasi *Simple K-mean Clustering* pada pengelompokan terjemahan Ayat Al Qur'an?
2. Bagaimana implementasi *Parallel K-mean Clustering* pada pengelompokan terjemahan Ayat Al Qur'an?
3. Bagaimana analisa perbandingan hasil *Simple K-mean clustering* dan *Parallel K-mean Clustering*?

## 1.3 Batasan Masalah

Batasan masalah pada tugas akhir ini adalah sebagai berikut:

1. Data yang akan digunakan adalah data teks berbahasa Indonesia yang diambil dari basis data terjemahan Al Qur'an berbahasa Indonesia versi Departemen Agama yang sudah dalam bentuk digital.
2. Aplikasi yang digunakan untuk pengolahan data adalah Python.
3. Kelompok dibuat berdasarkan kesamaannya dan bukan berdasarkan tema tertentu.

## 1.4 Tujuan Penelitian

Tujuan dari penulisan tugas akhir ini adalah:

1. Mengetahui hasil Klasterisasi *Simple K-mean Clustering* pada pengelompokan terjemahan Ayat Al Qur'an
2. Mengetahui hasil *Parallel K-mean Clustering* pada pengelompokan terjemahan Ayat Al Qur'an

3. Menganalisa perbandingan hasil *Simple K-mean clustering* dan *Parallel K-mean Clustering*

## 1.5 Sistematika Penulisan

Sistematika penulisan dalam penyusunan Skripsi ini adalah sebagai berikut :

### BAB I PENDAHULUAN

Bab ini berisi tentang Latar Belakang Masalah, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Sistematika Penulisan.

### BAB II LANDASAN TEORI

Bab ini berisi tentang teori-teori yang melandasi pembahasan dalam Skripsi ini. Secara garis besar, bab ini mencakup hal-hal yang berkaitan dengan konsep dasar yaitu *Data Mining*, *Temu Balik Informasi*, *Natural Language Processing*, *Text Mining*, *Clustering*.

### BAB III *K-means*

Bab ini berisi pembahasan utama dari Skripsi ini, yaitu pembahasan mengenai *K-means*, Pre-prosesing data, Pengolahan Data, dan Hasil dan kesimpulan.

### BAB IV KESIMPULAN

Bab ini, berisi kesimpulan dari pembahasan yang telah dikaji. Selain itu juga diberikan saran untuk pengembangan lebih lanjut untuk pembahasan tersebut.