

BABI

PENDAHULUAN

1.1 Latar Belakang

Al-Quran memuat 1.108 ayat yang mengandung kata bagian dari alam, seperti air, awan, besi, bintang, burung, cahaya, darah, emas, jahe, kapal, langit, dan zarah. Jika diperhatikan secara lebih seksama, ayat-ayat tersebut secara tegas menuntun pada konstruksi ilmu alam dan ilmu-ilmu lainnya. Informasi Al-Quran tentang alam dapat diklasifikasi menjadi 2 kelompok besar. Pertama, informasi langsung secara tekstual sehingga tidak memerlukan penafsiran atau pemahaman lebih lanjut. Kedua, informasi secara implisit dan memerlukan penafsiran lebih jauh atas redaksional ayat tersebut[1].

Umumnya buku tafsir terdiri dari beberapa jilid seperti pada tafsir Quraish Shihab dengan *Tafsir al-Miṣbāh* yang terdiri dari 30 juz sebanyak 15 jilid dalam satu set[2]. Namun, untuk mempermudah dalam menemukan dokumen/tafsir sesuai dengan kategorinya seperti kategori sains, maka diperlukan teknik pengolahan teks atau lebih dikenal Text Categorization (TC) yang dapat mengklasifikasikan tafsir/dokumen sesuai dengan kategorinya

Dengan dilakukannya proses klasifikasi tafsir ini bertujuan untuk mengetahui cara suatu teknologi dalam mengklasifikasikan teks tafsir ayat sains pada Al-Qur'an, serta untuk mengetahui tingkat akurasi yang dihasilkan pada saat melakukan klasifikasi.

Salah satu teknologi yang dapat digunakan adalah dengan memanfaatkan pendekatan kecerdasan buatan yaitu *Machine Learning*. Tujuan machine learning sendiri adalah untuk belajar dari data yang ada. *Machine Learning* digunakan untuk mempelajari bagaimana menangani data menjadi lebih efisien dan membantu dalam proses pengelompokan dan prediksi dalam suatu data atau biasa disebut klasifikasi[3]. Klasifikasi adalah pendekatan data mining (*machine learning*) yang digunakan untuk memperkirakan jumlah anggota dalam instance data dan memberi label pada data tersebut[4].

Salah satu teknik dalam Machine Learning yang bisa digunakan untuk melakukan klasifikasi adalah dengan menggunakan algoritma Support Vector Machine (SVM). SVM adalah algoritma supervised yang berupa klasifikasi dengan cara membagi data menjadi dua kelas menggunakan garis vector yang disebut hyperplane[5]. SVM umumnya mampu memberikan kinerja yang lebih tinggi dalam hal akurasi klasifikasi daripada algoritma klasifikasi data lainnya. SVM telah digunakan dalam berbagai masalah seperti pengenalan tulisan tangan, pengenalan nada, klasifikasi gambar dan deteksi objek, klasifikasi data dan klasifikasi teks[6].

Dalam proses klasifikasi teks peningkatan jumlah dokumen akan mempengaruhi kinerja klasifikasi, hal ini dapat membuat kinerja sistem klasifikasi menjadi berat. Hal ini disebabkan sistem akan mengambil isi dari uraian tafsir pada setiap dokumen. Karena itu diperlukan cara untuk meningkatkan kinerja sistem klasifikasi yaitu dengan melakukan text preprocessing untuk menghilangkan kata-kata yang tidak terstruktur pada dokumen dan menerapkan teknik pemilihan fitur. Pemilihan fitur merupakan suatu metode yang digunakan untuk mengurangi jumlah kata yang digunakan untuk meningkatkan akurasi hasil klasifikasi.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, maka rumusan masalah sebagai berikut:

1. Bagaimana implementasi Algoritma Support Vector Machine (SVM) yang tepat untuk melakukan klasifikasi pada teks tafsir ayat sains?
2. Bagaimana tingkat akurasi yang dihasilkan oleh Algoritma Support Vector Machine (SVM) dalam melakukan klasifikasi pada teks tafsir ayat sains?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

1. Mengetahui cara kerja Algoritma Support Vector Machine (SVM) dalam melakukan klasifikasi pada teks tafsir ayat sains.
2. Mengetahui tingkat akurasi Algoritma Support Vector Machine (SVM) dalam melakukan klasifikasi pada teks tafsir ayat sains.

1.4 Batasan Masalah

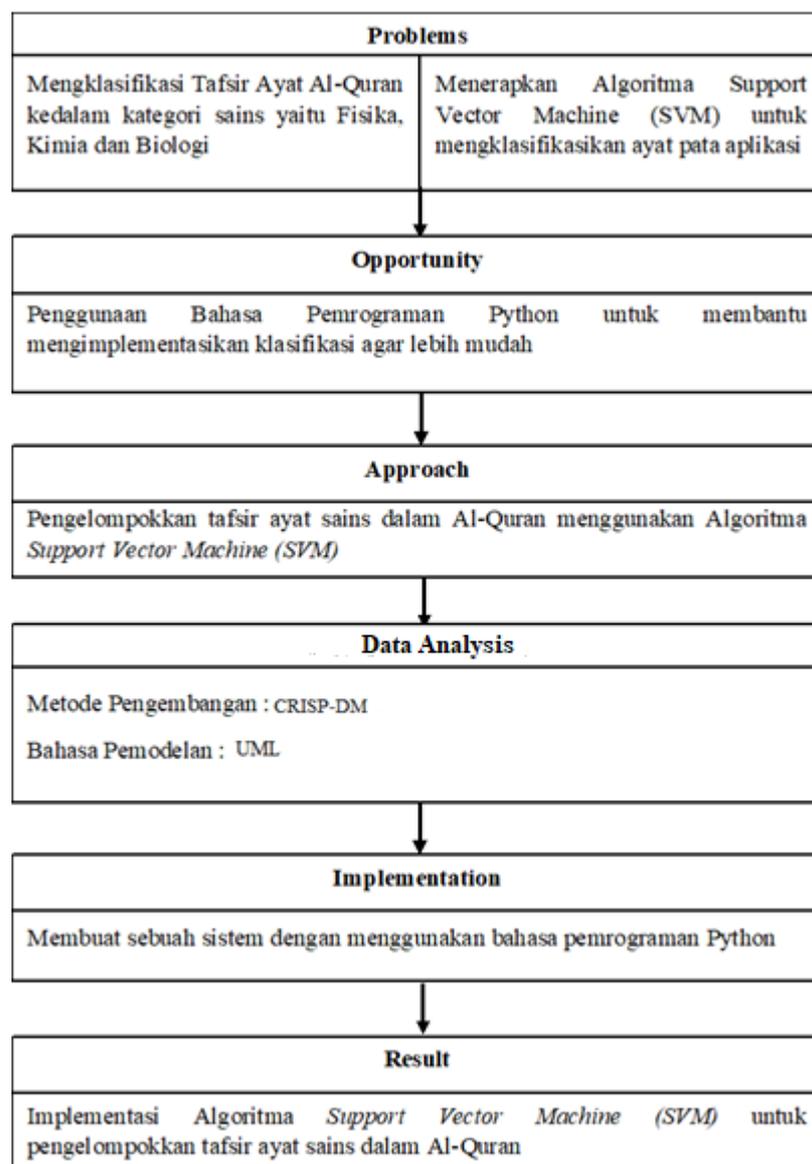
Terdapat beberapa batasan masalah dalam penelitian ini, agar sesuai rencana dan tidak keluar dari pokok permasalahan yang telah dirumuskan.

Adapun batasan masalahnya yaitu:

1. Proses pengelompokan terjemahan tafsir ayat sains dalam Al-Quran menggunakan algoritma *Support Vector Machine (SVM)*.
2. Data yang digunakan berupa teks dalam Bahasa Indonesia
3. Tafsir diklasifikasikan menjadi beberapa kategori yaitu Fisika, Biologi dan Kimia.
4. Aplikasi ini dibangun menggunakan Bahasa Pemrograman Python
5. Metode pengembangan data mining menggunakan CRISP-DM

1.5 Kerangka Pemikiran

Diperlukannya kerangka pemikiran dalam membuat aplikasi Klasifikasi Terjemahan Tafsir Ayat Al-Quran tentang Ilmu Sains untuk menggambarkan secara global alur logika pada suatu penelitian. Adapun kerangka pemikiran dari penelitian ini digambarkan pada Gambar 1.1.



Gambar 1. 1 Kerangka Pemikiran

1.6 Metode Penelitian

Metode penelitian dibutuhkan untuk memperoleh data yang diperlukan dalam melakukan penelitian, untuk itu diperlukan metode sebagai berikut.

1.6.1. Pengumpulan Data

Data yang dipakai dalam penelitian ini merupakan data berupa teks tafsir Al-Qur'an ayat sains berbahasa Indonesia, untuk itu diperlukan metode sebagai berikut:

1. Metode Kepustakaan

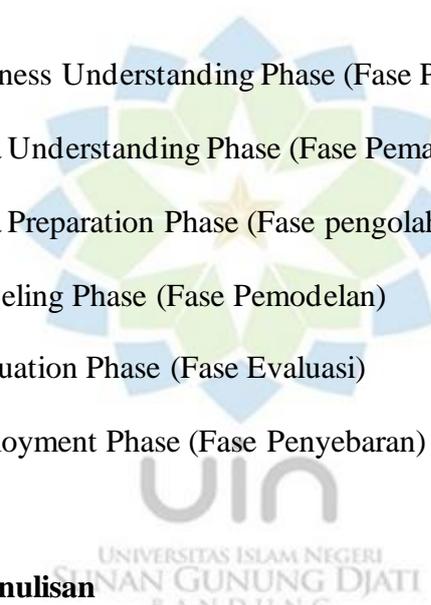
Metode kepustakaan merupakan metode yang digunakan untuk memperoleh konsep teoritis dengan cara melakukan pencarian data dalam jurnal-jurnal maupun buku-buku yang dapat membantu proses penyelesaian masalah yang serupa. Dalam penelitian ini digunakan buku karya Agus Purwanto, D.Sc. yang berjudul Nalar Ayat-ayat Semesta

2. Web Scraping

Web Scraping merupakan teknik guna memperoleh informasi dari website secara otomatis tanpa harus menyalinnya secara manual. Dalam penelitian ini website yang digunakan dalam teknik web scraping adalah tafsiirq.com. Dalam website tafsiirq.com diambil data tafsir karya M. Quraish Shihab.

1.6.2. Metode Pengembangan Data Mining

Tahapan yang diperlukan dalam proses pengembangan data mining pada penelitian ini yaitu menggunakan CRISP-DM. CRISP-DM merupakan standarisasi pada proses data mining sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian. Dalam CRISP-DM sebuah proyek data mining memiliki siklus yang terbagi ke dalam enam fase. Diantaranya adalah sebagai berikut[7]:

- 
- a. Business Understanding Phase (Fase Pemahaman Bisnis)
 - b. Data Understanding Phase (Fase Pemahaman Data)
 - c. Data Preparation Phase (Fase pengolahan Data)
 - d. Modeling Phase (Fase Pemodelan)
 - e. Evaluation Phase (Fase Evaluasi)
 - f. Deployment Phase (Fase Penyebaran)

1.7 Sistematika Penulisan

Sistematika penulisan perangkat lunak ini dipecah menjadi lima bab, yang disusun sebagai berikut :

BAB I: PENDAHULUAN

Pada bab ini diuraikan tentang Latar Belakang Masalah, Rumusan Masalah, Tujuan Penelitian, Batasan Masalah, Manfaat Penelitian, Kerangka Pemikiran, Metode Penelitian dan Sistematika Penulisan.

BAB II: STUDI PUSTAKA

Dalam bab ini menjelaskan mengenai tahapan dalam proses pengerjaan perangkat lunak dan membahas teori-teori yang relevan untuk digunakan dalam penelitian ini.

BAB III: ANALISIS DAN PERANCANGAN

Dalam bab ini menjelaskan tentang analisis dan kebutuhan sistem serta perancangan sistem mulai dari perancangan umum, perancangan prosedur, perancangan kode dan perancangan basis data.

BAB IV: IMPLEMENTASI APLIKASI

Pada bab ini menjelaskan implementasi perangkat lunak yang telah dianalisa dan dirancang sebelumnya serta hasil dari pengujian sistem yang memperlihatkan berjalan atau tidaknya program yang telah dibuat.

BAB V: KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan yang didapat dari pembahasan dan saran yang berguna untuk pengembangan skripsi ini.