**PENERAPAN *TEXT MINING* UNTUK MELAKUKAN KLASIFIKASI PADA KORPUS AYAT AL-QUR’AN**

Laporan Tugas Akhir I

**Disusun sebagai syarat kelulusan mata kuliah**

**IF4091/Tugas Akhir I dan Seminar**

**Oleh**

**Habibina Arif Muzayyan**

**NIM : 13519125**



**PROGRAM STUDI TEKNIK INFORMATIKA**

**SEKOLAH TEKNIK ELEKTRO & INFORMATIKA**

**INSTITUT TEKNOLOGI BANDUNG**

**2022**

**PENERAPAN *TEXT MINING* UNTUK MELAKUKAN KLASIFIKASI PADA KORPUS AYAT AL-QUR’AN**

**Laporan Tugas Akhir I**

**Oleh**

**Habibina Arif Muzayyan**

**NIM : 13519125**

**Program Studi Teknik Informatika**

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung

Bandung, 21 Desember 2022

Mengetahui,

Pembimbing,

Dr. Techn. Muhammad Zuhri Catur Candra, S.T, M.T.

NIP. 197709212010121002

# DAFTAR ISI

[**DAFTAR ISI**](#_heading=h.30j0zll) **2**

[**DAFTAR GAMBAR**](#_heading=h.3znysh7) **3**

[**DAFTAR TABEL**](#_heading=h.tyjcwt) **5**

[**BAB I PENDAHULUAN**](#_heading=h.4d34og8) **6**

[I.1 Latar Belakang](#_heading=h.2s8eyo1) 6

[I.2 Rumusan Masalah](#_heading=h.17dp8vu) 7

[I.3 Tujuan](#_heading=h.3rdcrjn) 7

[I.4 Batasan Masalah](#_heading=h.26in1rg) 8

[I.5 Metodologi](#_heading=h.lnxbz9) 8

[I.6 Jadwal Pelaksanaan Tugas Akhir](#_heading=h.35nkun2) 9

[**BAB II STUDI LITERATUR**](#_heading=h.44sinio) **10**

[II.1 Al-Qur'an](#_heading=h.2jxsxqh) 10

[II.1.1 Struktur Al-Qur'an](#_heading=h.jjc677lpv3u8) 10

[II.2 Hadis](#_heading=h.rdch04g07yql) 10

[II.2.1 Struktur Hadis](#_heading=h.z337ya) 11

[II.2.2 Klasifikasi Hadis](#_heading=h.pgi6wmgx7g1l) 11

[II.3 Text Mining](#_heading=h.u4ftlqcjufl6) 12

[II.3.1 Natural Language Processing (NLP)](#_heading=h.mox2pqvrmtbh) 12

[II.3.2 Machine Learning (ML)](#_heading=h.9uk4n4t3gaa2) 12

[II.4 Text Classification](#_heading=h.37e6st9t4egi) 13

[II.4.1 Naive Bayes](#_heading=h.xadu0dk1kor6) 13

[II.4.2 Support Vector Machine (SVM)](#_heading=h.ni3eorc0f957) 14

[II.4.3 Deep Learning](#_heading=h.rhw1d5hfi53b) 14

[II.5 Penelitian Terkait](#_heading=h.7csmlipvnv2h) 15

[II.5.1 Vector Space Model for Arabic Information Retrieval — Application to “Hadith” Indexing](#_heading=h.f9vgw2ckfbl5) 15

[II.5.2 Penerapan Text Mining untuk melakukan Klasifikasi pada Korpus Al-Hadits](#_heading=h.2y1ampwsu20i) 15

[II.5.3 Comparative Analysis of Text Classification Algorithms for Automated Labelling of Quranic Verses](#_heading=h.tpw3v9906x2a) 16

[**BAB III RENCANA PENYELESAIAN MASALAH**](#_heading=h.4i7ojhp) **17**

[III.1 Analisis Masalah](#_heading=h.ud9dxbcndcj2) 17

[III.2 Analisis Solusi](#_heading=h.p49jkezdp9n2) 17

[III.3 Rancangan Pembangunan Solusi](#_heading=h.geu01pfcgqak) 18

# DAFTAR GAMBAR

[Gambar II.1. Struktur Hadis](#_heading=h.3j2qqm3) 11

[Gambar II.4.2. Ilustrasi Support Vector Machine (SVM)](#_heading=h.xzxria2otq9i) 14

[Gambar III.3.1. Proses rancangan pembangunan solusi](#_heading=h.uq6fnue6pj7w) 18

# DAFTAR TABEL

[Tabel I.6.1. Jadwal Pelaksanaan Tugas Akhir tahun 2022](#_heading=h.by3yq960n7vz) 9

[Tabel I.6.2. Jadwal Pelaksanaan Tugas Akhir tahun 2023](#_heading=h.cjpbscm2qjf8) 9

[Tabel III.2.1. Perbandingan antar metode Text Classification](#_heading=h.a8dkiiz0fxpy) 18

# PENDAHULUAN

## Latar Belakang

Al-Qur’an merupakan kitab suci umat Islam yang berisi firman Allah yang diturunkan kepada Nabi Muhammad saw. dengan perantaraan malaikat Jibril untuk dibaca, dipahami, dan diamalkan sebagai petunjuk atau pedoman hidup bagi umat manusia. Al-Qur’an adalah kitab yang tidak pernah berubah isinya sepanjang masa. Terdapat 114 Surah dan 6236 Ayat di dalam Al-Qur’an. Masing-masing ayat yang tersebar dalam Al-Qur’an memiliki makna yang perlu dipahami lebih lanjut. Makna dalam ayat Al-Qur’an dapat berupa petunjuk, cerita, atau pujian kepada Allah swt.

Kondisi yang terjadi saat ini diantaranya adalah sebagian banyak umat manusia tidak memahami secara lebih dalam makna yang terkandung dalam ayat Al-Qur’an. Selain itu ayat-ayat Al-Qur’an juga banyak ditafsirkan dengan berbagai pemahaman yang berbeda.

Akibat dari kondisi yang terjadi saat ini dapat menimbulkan beberapa masalah seperti kesulitan dalam memahami agama Islam melalui ayat Al-Qur’an. Umat manusia mencari solusi permasalahan kehidupan sehari-hari melalui kata kunci yang dibutuhkan kemudian mencari ayat atau hadis yang terkait. Hal ini juga dapat menimbulkan kesulitan dalam mencari dalil-dalil yang berkaitan dengan suatu tafsir tentang ayat atau hadis tertentu.

Banyak upaya yang telah dilakukan untuk mengatasi masalah tersebut. Beberapa upaya yang telah dilakukan adalah seperti melakukan beberapa penelitian serta mengembangkan aplikasi tafsir Al-Qur’an dan Ensiklopedia Hadis. Harrag et al. (2008) telah melakukan penelitian tentang klasifikasi hadis menggunakan metode model vektor spasial. Adeleke et al. (2007) telah melakukan penelitian terkait klasifikasi label ayat Al-Qur’an menggunakan metode k-NN, SVM, dan Naive Bayes. Namun upaya-upaya tersebut masih belum cukup untuk mengatasi masalah yang terjadi karena klasifikasi ayat Al-Qur’an tersebut masih belum dapat menjelaskan makna ayat Al-Qur’an lebih dalam.

Oleh karena itu, penelitian ini dilakukan untuk klasifikasi korpus ayat Al-Qur’an menggunakan penerapan *Text Mining* yaitu *Text Classification.* Alasan dilakukan penelitian ini adalah untukuntuk mempermudah memahami makna ayat-ayat dalam korpus Al-Qur’an dan mempermudah mencari dalil-dalil yang berkaitan dengan suatu tafsir tentang ayat Al-Qur’an tertentu*.* Pada penelitian ini klasifikasi ayat Al-Qur’an akan dilakukan berdasarkan dari kategori bab pada Kitab Hadis Shahih Bukhari. Hal ini dapat menyebabkan kategori korpus ayat Al-Qur’an yang diklasifikasikan semakin bervariasi sehingga lebih mudah untuk memahami kategori korpus ayat Al-Qur’an.

## Rumusan Masalah

Rumusan masalah yang didapat dari penjelasan latar belakang pada subbab 1.1 pada tugas akhir ini adalah sebagai berikut:

1. Bagaimana penerapan *Text Mining* untuk melakukan klasifikasi pada korpus ayat Al-Qur’an?
2. Bagaimana hasil evaluasi dari penerapan *Text Mining* untuk melakukan *Text Classification* pada korpus ayat Al-Qur’an?

## Tujuan

Adapun tujuan yang akan dicapai dalam pelaksanaan tugas akhir ini adalah untuk mengembangkan *Text Classification* pada korpus ayat Al-Qur’an dengan menerapkan *Text Mining*. Pada penelitian ini, digunakan teknik terbaik yang mungkin untuk pembangunan *Text Classification* pada korpus ayat Al-Qur’an.

## Batasan Masalah

Batasan-batasan masalah pada pelaksanaan tugas akhir ini adalah sebagai berikut:

1. Korpus Al-Qur’an yang digunakan merupakan terjemahan Bahasa Indonesia.
2. Kategori *Text Classification* korpus ayat Al-Qur’an berasal dari kategori bab pada Kitab Hadis Shahih Bukhari.

## Metodologi

Berikut ini adalah metode-metode yang digunakan selama pelaksanaan tugas akhir ini:

1. Tahapan Analisis Masalah

Tahap analisis masalah adalah tahap yang pertama pada tugas akhir ini. Pada tahap ini dilakukan analisis mengenai rumusan masalah. Analisis masalah dilakukan dengan studi literatur.

1. Tahapan Analisis Pilihan Solusi

Setelah analisis masalah, dilakukan tahap analisis pilihan solusi. Pada tahap ini dilakukan analisis mengenai pilihan solusi terbaik untuk menyelesaikan rumusan masalah.

1. Tahapan Perancangan Solusi

Perancangan solusi dilakukan setelah analisis pilihan solusi. Pada tahap ini dilakukan dengan merancang solusi yang telah dipilih.

1. Tahapan Pembangunan Solusi

Tahap pembangunan solusi dilakukan setelah perancangan solusi. Pada tahap ini dilakukan pembangunan solusi berdasarkan rancangan solusi yang telah dibuat.

1. Tahapan Pengujian dan Evaluasi

Tahap pengujian dan evaluasi merupakan tahap terakhir pada tugas akhir ini. Pengujian dan evaluasi dilakukan dengan menguji sistem yang dibangun terhadap permasalahan yang dihadapi serta mengevaluasi performansi dari solusi yang telah dibangun.

## Jadwal Pelaksanaan Tugas Akhir

Berikut ini adalah jadwal pelaksanaan tahun 2022 dan 2023 pada tugas akhir ini:

##### Tabel I.6.1. Jadwal Pelaksanaan Tugas Akhir tahun 2022

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Kegiatan** | **September** | | | **Oktober** | | | | | **November** | | | | **Desember** | | | |
| **11** | **18** | **25** | **2** | **9** | **16** | **23** | **30** | **6** | **13** | **20** | **27** | **4** | **11** | **18** | **25** |
| Studi Literatur |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Penulisan hasil studi literatur |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Penulisan rencana tugas akhir |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Penulisan rencana penyelesaian masalah |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Revisi Laporan TA1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

##### 

##### Tabel I.6.2. Jadwal Pelaksanaan Tugas Akhir tahun 2023

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Kegiatan** | **Januari** | | | **Februari** | | | | **Maret** | | | | **April** | | | | |
| **15** | **22** | **29** | **5** | **12** | **19** | **26** | **5** | **12** | **19** | **26** | **2** | **9** | **16** | **23** | **30** |
| Studi Literatur |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Penulisan penyelesaian masalah |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Penulisan hasil pengujian |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Penulisan simpulan tugas akhir |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Revisi Laporan TA2 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

# STUDI LITERATUR

## Al-Qur'an

Al-Qur’an berasal dari Bahasa Arab yaitu kata “*qara’a*” (قرء) yang artinya bacaan atau yang dibaca. Menurut Kamus Besar Bahasa Indonesia (KBBI), Al-Qur'an adalah kitab suci umat Islam yang berisi firman Allah yang diturunkan kepada Nabi Muhammad saw. dengan perantaraan malaikat Jibril untuk dibaca, dipahami, dan diamalkan sebagai petunjuk atau pedoman hidup bagi umat manusia. Al-Qur’an diturunkan kepada Nabi Muhammad saw. selama 22 tahun 2 bulan 22 hari.

### Struktur Al-Qur'an

Al-Qur’an terdiri dari 30 bab atau bagian (juz). Al-Qur'an terdiri dari 114 Surah dan 6236 Ayat. Surah Al-Qur’an terdiri dari dua macam yaitu surah *makkiyah* dan surah *madaniyah*. Surah *makkiyah* adalah surah yang diturunkan kepada Nabi Muhammad saw. sebelum peristiwa hijrah ke Madinah, sedangkan surah *madaniyah* diturunkan setelah peristiwa Nabi Muhammad saw. hijrah ke Madinah.

## Hadis

Hadis berasal dari Bahasa Arab yaitu kata “*hadatsa*” (حدث) yang artinya kabar atau peristiwa. Menurut Kamus Besar Bahasa Indonesia (KBBI), Hadis adalah sabda, perbuatan, takrir (ketetapan) Nabi Muhammad saw. yang diriwayatkan atau diceritakan oleh sahabat untuk menjelaskan dan menentukan hukum Islam. Hadis merupakan sumber ajaran Islam yang kedua setelah Al-Qur'an.

### Struktur Hadis

Hadis terdiri dari 3 bagian yaitu penutur hadis (sanad), isi hadis (matan), dan periwayat hadis terakhir (rawi). Sanad adalah rentetan periwayat hadis dari Nabi Muhammad saw sampai periwayat hadis terakhir. Matan adalah materi atau isi dari hadis yang dapat berisi ucapan, perbuatan, atau takrir (ketetapan) dari Nabi Muhammad saw. Rawi adalah orang meriwayatkan hadis yang terakhir.

|  |
| --- |
| Hadis  Sanad  Matan  Rawi |

##### Gambar II.1. Struktur Hadis

### Klasifikasi Hadis

Hadis dapat diklasifikasikan berdasarkan kebenaran hadis tersebut. Hal ini dapat terjadi karena Nabi Muhammad saw. telah wafat 14 abad yang lalu dan ingatan beberapa sanad yang tidak terlalu bagus. Hadis diklasifikasikan menjadi 4 yaitu hadis *shahih*, hadis *hasan*, hadis *dhaif*, dan hadis *maudu*. Hadis *shahih* adalah hadis yang dapat dikonfirmasi kebenaran isinya. Hadis *hasan* adalah hadis yang kebenaran isinya masih dapat diterima. Hadis *dhaif* adalah hadis yang lemah atas kebenaran isinya. Hadis *maudu* dapat dikatakan hadis palsu karena isi hadisnya tidak dibenarkan.

Selain itu, hadis juga dapat diklasifikan berdasarkan perawi hadis tersebut. Hadis yang diriwayatkan oleh Imam Bukhari dapat disebut Hadis Bukhari, hadis yang diriwayatkan oleh Imam Tirmidzi dapat disebut Hadis Tirmidzi, dan sebagainya.

## *Text Mining*

*Text Mining* merupakan salah satu cabang disiplin ilmu komputer sains dan sistem informasi yang mempelajari tentang penambangan teks. *Text* *mining* (juga disebut sebagai *text analytics*) adalah teknologi kecerdasan buatan yang menggunakan *Natural Language Processing* (NLP) untuk mengubah teks bebas (tidak terstruktur) dalam dokumen dan basis data menjadi data terstruktur yang dinormalisasi yang cocok untuk analisis atau untuk menggerakkan algoritma *Machine Learning* (ML).

### *Natural Language Processing* (NLP)

*Natural language processing* (NLP) adalah studi tentang bahasa alami manusia sehingga komputer dapat “membaca” teks bahasa alami mirip seperti yang dilakukan manusia. NLP terdiri dari berbagai macam *task* bagian untuk memproses teks, diantaranyaadalah *Sentence Splitter, Tokenization, Stemming, Word Stop Elimination.* NLP mencakup juga *Natural Language Understanding* (NLU) dan *Natural Language Generation* (NLG)yang mensimulasikan kemampuan manusia untuk membuat teks bahasa alami, misal untuk meringkas informasi atau mengambil bagian dalam suatu dialog bahasa alami.

### *Machine Learning* (ML)

*Machine Learning* (ML) adalah teknologi kecerdasan buatan yang menyediakan sistem dengan kemampuan untuk belajar secara otomatis dari pengalaman tanpa memerlukan pemrograman eksplisit, dan dapat membantu memecahkan masalah kompleks dengan akurasi yang dapat menyaingi atau bahkan terkadang melampaui manusia. ML yang arsitekturnya menirukan jaringan saraf manusia disebut *Deep Learning* (Pembelajaran Mendalam).

## *Text Classification*

*Text Classification* (Klasifikasi Teks) adalah teknik pembelajaran mesin yang menetapkan serangkaian kategori yang telah ditentukan sebelumnya ke teks terbuka. *Text Classification* dapat digunakan untuk mengatur, menyusun, dan mengkategorikan hampir semua jenis teks – dari dokumen, studi medis, dan file di seluruh web. *Text Classification* adalah salah satu tugas mendasar dalam *Natural Language Processing* (Pemrosesan Bahasa Alami) dengan aplikasi yang sangat beragam seperti analisis sentimen, pelabelan topik, deteksi spam, dan deteksi aspek.

*Text Classification* dapat diimplementasikan dengan metode *Machine Learning* (Pembelajaran Mesin). Beberapa algoritma untuk *Text Classification* adalah algoritma Naive Bayes, *Support Vector Machine* (SVM), dan *Deep Learning* (Pembelajaran Mendalam).

### Naive Bayes

Algoritma Naive Bayes adalah salah satu algoritma yang paling populer dan paling banyak digunakan dalam *Text Classification* dan analisis teks, secara keseluruhan. Naive Bayes didasarkan pada Teorema Bayes, yang membantu untuk menghitung probabilitas bersyarat terjadinya dua peristiwa, berdasarkan probabilitas terjadinya masing-masing peristiwa. Algoritma Naive Bayes menghitung kemungkinan setiap tag untuk teks tertentu, dan kemudian mengeluarkan tag dengan probabilitas tertinggi.

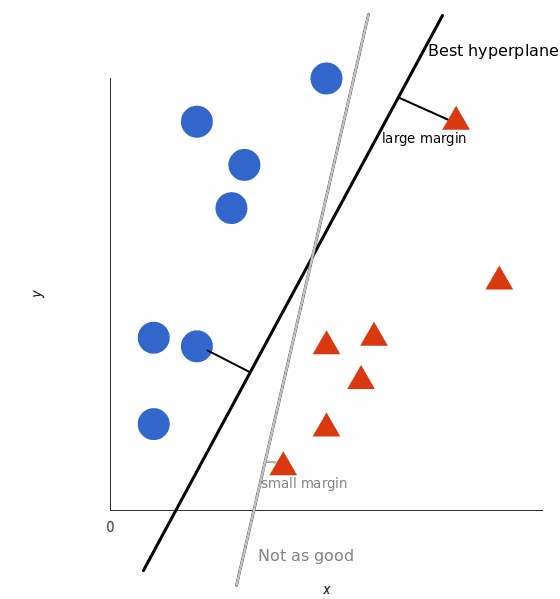
Probabilitas A jika B benar, sama dengan probabilitas B jika A benar, dikalikan probabilitas A benar, dibagi dengan probabilitas B benar.

Ini berarti bahwa setiap vektor yang merepresentasikan teks harus berisi informasi tentang probabilitas kemunculan kata-kata tertentu dalam teks dari kategori tertentu, sehingga algoritma dapat menghitung kemungkinan teks tersebut termasuk dalam kategori tertentu.

### *Support Vector Machine (SVM)*

*Support Vector Machine* (SVM) adalah algoritma pembelajaran mesin klasifikasi teks yang hampir mirip seperti Naive Bayes, SVM tidak memerlukan banyak data pelatihan untuk mulai memberikan hasil yang akurat. SVM memang membutuhkan lebih banyak sumber daya komputasi daripada Naive Bayes, tetapi hasilnya bahkan lebih cepat dan lebih akurat dari Naive Bayes.

Secara singkat, Cara kerja SVM adalah menggambar garis atau "hyperplane" yang membagi ruang menjadi dua subruang. Satu subruang berisi vektor (*tag*) yang termasuk dalam suatu kelas, dan subruang lainnya berisi vektor yang termasuk dalam kelas yang lain.



##### Gambar II.4.2. Ilustrasi *Support Vector Machine* (SVM)

### *Deep Learning*

*Deep Learning* (Pembelajaran mendalam) adalah sekumpulan algoritma dan teknik yang terinspirasi oleh cara kerja otak manusia, yang disebut jaringan saraf. Arsitektur *deep learning* dapat memberikan manfaat yang besar untuk *Text Classification* karena *deep learning* bekerja dengan akurasi sangat tinggi dengan rekayasa dan komputasi yang cukup rendah. Dua arsitektur *deep learning* utama untuk *Text Classification* adalah *Convolutional Neural Network* (CNN) dan *Recurrent Neural Network* (RNN).

## Penelitian Terkait

Banyak penelitian yang sudah dilakukan dan memiliki topik terkait dengan penelitian ini. Berikut ini adalah beberapa penelitian yang memiliki topik berkaitan dengan topik penelitian ini:

### *Vector Space Model for Arabic Information Retrieval — Application to “Hadith” Indexing*

Pada penelitian ini diimplementasikan alat pencarian otomatis dalam database *ldquoHadithrdquo* (Tradisi Nabi, perkataan dan perbuatan Nabi) yang besar untuk mengakses salah satu sumber asli Islam. Untuk tujuan ini, penelitian ini menjelaskan pengembangan *AuthenTique*, alat pencarian penambangan teks otomatis yang diperbarui, berdasarkan *Vector Space Model* (VSM) / model ruang vektor. Tujuannya adalah untuk memungkinkan penyediaan daftar *ldquoHadithsrdquo* yang diklasifikasikan menurut tingkat kemiripannya berdasarkan kueri *user* yang diberikan. (Harrag et al., 2008).

### 

### Penerapan *Text Mining* untuk melakukan Klasifikasi pada Korpus Al-Hadits

Pada penelitian ini, dilakukan *training* menggunakan 3 metode, yaitu SVM, Naive Bayes, dan BERT terhadap dataset kitab hadis Bukhari. Training ini bertujuan agar model yang dihasilkan dapat mengklasifikasikan input hadis dari kitab hadis lain sebagai salah satu dari bab kitab hadis Bukhari. Dengan demikian, setiap hadis dari kitab hadis lain akan memiliki bab klasifikasi baru sesuai bab kitab hadis Bukhari dan memberikan pengetahuan baru bahwa suatu hadis akan masuk kedalam bab spesifik yang tidak ada pada kitab hadis asalnya (Ziad, 2022).

### *Comparative Analysis of Text Classification Algorithms for Automated Labelling of Quranic Verses*

Pada penelitian ini, dilakukan otomatisasi tugas pelabelan ayat Alquran menggunakan algoritma klasifikasi teks menggunakan tiga algoritma klasifikasi teks yaitu, k-Nearest Neighbour, Support Vector Machine, dan Naive Bayes. Terjemahan bahasa Inggris dari ayat-ayat disajikan sebagai fitur pada penelitian ini. Terjemahan bahasa Inggris dari ayat-ayat tersebut kemudian diklasifikasikan sebagai "*Shahada*" (rukun Islam pertama) atau "*Pray*" (rukun kedua dari Islam). Ditemukan bahwa semua algoritma klasifikasi teks mampu mencapai akurasi lebih dari 70% dalam pelabelan ayat Al-Qur’an. (Adeleke et al., 2007).

# RENCANA PENYELESAIAN MASALAH

## Analisis Masalah

Al-Qur’an berisi firman Allah yang memiliki tujuan untuk dibaca, dipahami, dan diamalkan sebagai petunjuk atau pedoman hidup bagi umat manusia. Sedangkan Hadis adalah sabda, perbuatan, takrir (ketetapan) Nabi Muhammad saw. Permasalahan yang terjadi saat ini adalah umat manusia memiliki keterbatasan untuk memahami konsep Al-Qur’an seutuhnya. Untuk itu dapat dibuat suatu penerapan klasifikasi ayat Al-Qur’an agar mempermudah manusia untuk memahami kategori ayat Al-Qur’an berdasarkan pada kategori pada Hadis. Pada penelitian ini, penerapan klasifikasi ayat Al-Qur’an menggunakan penerapan pada *Text Mining.*

Korpus Al-Qur’an yang digunakan untuk *Text Classification* adalah korpus terjemahan bahasa Indonesia. Penerapan ini akan dilakukan dengan harapan manusia khususnya masyarakat Indonesia dapat memahami proses yang dilakukan dalam *Text Classification* pada korpus ayat Al-Qur’an.

Pada penelitian ini, kategori *Text Classification* korpus ayat Al-Qur’an diklasifikasikan berdasarkan dari kategori bab pada Kitab Hadis Shahih Bukhari. Kitab Hadis Shahih Bukhari memiliki kategori bab terbanyak dibandingkan dengan Kitab Hadis yang lain. Hal ini dapat menyebabkan kategori korpus ayat Al-Qur’an semakin bervariasi sehingga lebih mudah untuk memahami kategori korpus ayat Al-Qur’an.

## Analisis Solusi

Berdasarkan hasil studi literatur, *Text Classification* dapat dilakukan dengan metode pada pembelajaran mesin yaitu, algoritma Naive Bayes, SVM, dan *deep learning*. Metode-metode ini akan diimplementasikan pada penelitian ini untuk mengetahui metode solusi yang terbaik untuk melakukan *Text Classification* korpus ayat Al-Qur’an.

Berikut ini perbandingan antar metode yang akan digunakan untuk *Text Classification*:

##### Tabel III.2.1. Perbandingan antar metode *Text Classification*

|  |  |  |  |
| --- | --- | --- | --- |
| Perbandingan | Naive Bayes | SVM | *Deep Learning* |
| *Supervised / Unsupervised* | *Supervised* | *Unsupervised* | *Hybrid* |
| *Training* | Tidak perlu | Tidak perlu | Perlu |
| *Multiclass classification* | ✓ | ✓ | ✓ |

## Rancangan Pembangunan Solusi

Berikut ini proses rancangan pembangunan solusi yang akan dilakukan pada penelitian ini:

*Text Mining* Al-Qur’an dan Hadis

*Text Preprocessing*

*Feature Engineering*

*Text Classification*

Evaluasi

##### Gambar III.3.1. Proses rancangan pembangunan solusi

Pemilihan metode solusi terbaik dapat dilakukan dengan menggunakan beberapa cara diantaranya adalah menghitung kecepatan komputasi, kebutuhan data pelatihan, atau kompleksitas algoritma.

DAFTAR PUSTAKA

Adeleke, A. O., Samsudin, N. A., Mustapha, A., & Nawi, N. M. (2017). *Comparative analysis of text classification algorithms for automated labelling of Quranic verses*. Int. J. Adv. Sci. Eng. Inf. Technol, 7(4), 1419.

Al-Qur'an. (2016). *Kamus Besar Bahasa Indonesia (KBBI)*. <https://kbbi.kemdikbud.go.id/entri/al-qur'an>. 9 Oktober 2022.

Departemen Agama Republik Indonesia. *Al-Qur'an dan Terjemahannya*.

F. Harrag, A. Hamdi-Cherif and E. El-Qawasmeh. (2008) *Vector space model for Arabic information retrieval — application to “Hadith” indexing.* 2008 First International Conference on the Applications of Digital Information and Web Technologies (ICADIWT), pp. 107-112, doi: 10.1109/ICADIWT.2008.4664328.

Hadis. (2016). *Kamus Besar Bahasa Indonesia (KBBI)*. <https://kbbi.kemdikbud.go.id/entri/hadis>. 9 Oktober 2022.

M. Majdalawieh, F. Marir and I. Tiemsani. (2017). *Developing Adaptive Islamic Law Business Processes Models for Islamic Finance and Banking by Text Mining the Holy Qur'an and Hadith*, 1278-1283.

S. J. Putra, T. Mantoro and M. N. Gunawan. (2017). *Text mining for Indonesian translation of the Quran: A systematic review*, International Conference on Computing, Engineering, and Design (ICCED), pp. 1-5, doi: 10.1109/CED.2017.8308122.

Saloot, M.A., Idris, N., Mahmud, R. et al. (2016). *Hadith data mining and classification: a comparative analysis*. Artif Intell Rev 46, 113–128.

Ziad, R.M. (2022). *Penerapan Text Mining untuk melakukan Klasifikasi pada Korpus Al-Hadits*. Teknik Informatika. Perpustakaan Digital ITB.

Hearst, M. A. (1999, June). Untangling text data mining. In *Proceedings of the 37th Annual meeting of the Association for Computational Linguistics* (pp. 3-10).