

Klasifikasi Teks Terjemahan Ayat Al-Qur'an dalam Bahasa Indonesia dengan *Machine Learning*

Classification of Translated Texts of Al-Qur'an Verses in Indonesian with Machine Learning

Fajar Deny Elfianita*¹, Oddy Virgantara Putra², Triana Harmini³, Asif Trisnani⁴, Taufiqurrahman⁵

^{1,2,3,4,5} Universitas Darussalam Gontor

^{1,2,3,4,5} Ponorogo, Indonesia

e-mail: *fajarelfianita@mhs.unida.gontor.ac.id, oddy@unida.gontor.ac.id,
triana@unida.gontor.ac.id, asiftrisnani@unida.gontor.ac.id,
taufiqurrahman@unida.gontor.ac.id

Abstrak - Al-Qur'an telah diterjemahkan ke dalam lebih dari ratusan bahasa di dunia. Di Indonesia terjemahan Al-Qur'an muncul ketika zaman Walisongo atau sekitar abad ke 15 untuk kepentingan penyebaran ajaran agama Islam. Saat ini Al-Qur'an terjemahan sudah banyak diterbitkan di Indonesia, namun Al-Qur'an terjemahan belum dikategorikan sesuai dengan kandungan utama Al-Qur'an. penelitian ini mengusulkan model berbasis pembelajaran mesin untuk mengklasifikasi konten terjemahan Al-Qur'an Bahasa Indonesia agar ayat dikelompokkan sesuai kandungan utama Al-Qur'an. Penelitian ini mengklasifikasikan ayat menjadi 5 kelas yaitu Akidah, Akhlaq, Ibadah, Kisah dan Hukum dengan dataset yang berasal dari website Kementerian Agama Republik Indonesia. Data menggunakan teks terjemahan dari juz 5 hingga juz 7 yang terdiri dari Surat An-Nisa', Al-Maidah dan Al-An'am. Sebelum proses pemodelan dilakukan preprocessing berupa Case Folding, Tokenization, Stopword Removal dan Stemming, dan Feature Extraction. Pada tahap klasifikasi menggunakan perbandingan model Naive Bayes dan Support Vector Machine (SVM). Hasil penelitian ini, pada model Naive Bayes mampu melakukan klasifikasi teks dengan hasil akurasi 80,20%. Sedangkan Support Vector Machine (SVM) memiliki tingkat akurasi 70,83%. Untuk penelitian selanjutnya diperlukan pendekatan deep learning dan dapat menggunakan model lain sebagai bahan perbandingan dengan model penelitian saat ini.

Kata kunci – Al-Qur'an, Klasifikasi, Naive Bayes, Terjemahan .

Abstract - The Qur'an has been translated into over a thousand languages around the world. For the purpose of spreading Islam's teachings, a translation of the Qur'an appeared in Indonesia during the Walisongo era or during the 15th century. Many Qur'an translations have been released in Indonesia recently, but the translated Qur'an has not been categorized according to the Qur'an's basic content. This work presents a machine learning-based methodology for classifying the translated text of the Indonesian Qur'an, with verses classified according to the Qur'an's main content. Using statistics from the Ministry of Religion of the Republic of Indonesia's website, this study divides verses into five categories: Faith, Morality, Worship, History, and Law. The data is based on a translation of Surah An-Nisa', Al-Maidah, and Al-An'am, which includes chapters 5 to 7. Preprocessing included Case Folding, Tokenization, Stopword Removal and Stemming, and Feature Extraction prior to the modeling step. At the classification step, the Naive Bayes model and the Support Vector Machine (SVM) were compared. According to the findings of this research, the Naive Bayes model can categorize text with an accuracy of 80.20 %. The Support Vector Machine (SVM) has a 70.83 % accuracy rate. A deep learning technique is required for further study, and other models might be used as material for comparison with the current research model.

Keywords – Al-Qur'an, Classification, Naive Bayes, Translation.

I. PENDAHULUAN

Al-Qur'an menjadi pedoman hidup manusia karena di dalamnya terdapat berbagai penjelasan sebagai petunjuk hidup umat manusia. Kitab suci Al-Qur'an pertama kali diturunkan menggunakan Bahasa Arab[1]. Di Indonesia Al-Qur'an mulai diterjemahkan ke dalam Bahasa Indonesia pada abad ke 15[2]. Saat itu merupakan zaman Walisongo sehingga Al-Qur'an banyak diterjemahkan untuk kepentingan penyebaran ajaran agama Islam di Indonesia. Teks terjemahan Al-Qur'an sangat diperlukan untuk memahami kandungan Al-Qur'an, karena Al-Qur'an sejatinya menggunakan Bahasa Arab saat diwahyukan kepada nabi Muhammad SAW. Sehingga arti terjemahan Al-Qur'an dibutuhkan oleh orang yang belum mengerti Bahasa Arab untuk memahami Al-Qur'an.

Al-Qur'an dan terjemahnya sangat penting bagi umat Islam. Membaca Al-Qur'an tanpa mengetahui maknanya maka tidak dapat mengamalkannya. Terjemahan Al-Qur'an juga sangat berperan penting dalam dakwah Islam. Al-Qur'an merupakan sumber dakwah yang harus disampaikan kepada siapa saja dengan metode penyampaian yang berbeda-beda[3]. Terjemahan Al-Qur'an saat ini sudah banyak diterbitkan di Indonesia. Berbagai jenis Al-Qur'an telah diterbitkan beserta terjemahan Bahasa Indonesia bahkan Bahasa daerah[4]. Namun, terjemahan Al-Qur'an belum dilengkapi dengan pengelompokan ayat sesuai dengan kandungan pokok Al-Qur'an. Sehingga belum tersedianya pengelompokan ayat terjemahan sesuai dengan kandungan pokok Al-Qur'an.

Menurut buku Akidah Akhlaq yang diterbitkan oleh Kementerian Agama Republik Indonesia, pengelompokan kandungan utama Al-Qur'an diantaranya yaitu Akidah, Ibadah, Muamalah, Akhlak karimah, Tarikh dan Syariat[5]. Pengelompokan ayat untuk memudahkan pencarian ayat sesuai dengan yang diinginkan. Langkah awal yang harus dilakukan untuk memudahkan pencarian yaitu pengelompokan ayat sesuai dengan kandungan pokok ayat di Al-Qur'an.

Semakin berkembangnya kemajuan teknologi saat ini menjadikan pengelompokan teks terjemahan Al-Qur'an dapat diimplementasikan dengan metode *text mining*[6]. Metode *text mining* banyak diterapkan ke berbagai bidang seperti jurnalistik dan berita[7], pendidikan[8], jasa industri[9], keamanan[10] dan tentunya bidang Al-Qur'an[11]. Berbagai penelitian mengenai *text mining* yang diimplementasikan di berbagai bidang memungkinkan *text mining* dapat digunakan untuk mengelompokkan ayat terjemahan Al-Qur'an sesuai dengan kandungan pokok ayat Al-Qur'an.

Pengelompokan ayat menggunakan model *text mining* dapat diterapkan dengan berbagai model. *Text mining* dapat diterapkan dengan metode *clustering* maupun klasifikasi. Beberapa penelitian pernah menerapkan metode pengelompokan seperti *clustering* yang dilakukan dengan permodelan K-Means *clustering* pada ayat Al-Qur'an[12]. Sedangkan penerapan klasifikasi pada ayat Al-Qur'an dapat dilakukan dengan berbagai model algoritma klasifikasi seperti metode Naïve Bayes, Support Vector Machine, KNN dan Decision Tree[13][14]. Klasifikasi dilakukan dengan menentukan suatu dokumen ke dalam kelas-kelas tertentu.

Berdasarkan latar belakang tersebut maka permasalahan yang ditemukan yaitu belum adanya pengklasifikasian teks terjemahan ayat Al-Qur'an sesuai dengan kandungan utama kitab tersebut. Sehingga perlu dilakukan pengklasifikasian teks terjemahan ayat Al-Qur'an menjadi beberapa kelas sesuai dengan kandungan pokok ayatnya. Penelitian ini akan menggunakan teks terjemahan Al-Qur'an Bahasa Indonesia dan mengklasifikasikan terjemahan ayat menjadi lima kelompok yaitu akidah, akhlak, ibadah, kisah dan hukum. Penelitian ini membandingkan model Naïve Bayes dan Support Vector Machine untuk mengukur kinerja terbaik model. Klasifikasi teks bertujuan untuk menghasilkan kelas-kelas teks dokumen ayat terjemahan Al-Qur'an sesuai dengan kandungan ayat di dalam Al-Qur'an

II. LANDASAN TEORI

2.1 Text Mining

Tersebar jutaan bahkan milyaran data di suni ini baik terstruktur maupun tidak terstruktur baik teks atau angka. *Text mining* merupakan penggalian data yang difokuskan pada dokumen

teks. Data yang digali dan menghasilkan informasi dari teks tersebut nantinya dibutuhkan untuk pengambilan keputusan dan informasi serta referensi penting[15].

Teknik dalam proses pengolahan teks bermacam-macam diantaranya *clustering* dan klasifikasi. Baik *clustering* maupun klasifikasi telah banyak digunakan pada penelitian tentang pengolahan dokumen teks[16]. Teknik juga menggunakan beberapa model algoritma yang berbeda-beda antara *supervised learning* dan *unsupervised learning*.

2.2 Klasifikasi

Klasifikasi merupakan proses pengolahan data dengan memisahkan atau membagi objek tersebut menjadi beberapa kelas. Pembagian disesuaikan dengan jumlah kelas yang diinginkan[17]. Komponen utama dari klasifikasi diantaranya kelas, prediktor, set data latih dan set data uji.

Kelas merupakan variabel tidak bebas, kelas adalah hasil klasifikasi. Contoh kelas diantaranya positif dan negatif atau kelas tumor ganas dan tumor jinak. Sedangkan prediktor yaitu variabel bebas sebuah model tergantung dari atribut data proses klasifikasi. Contohnya jika klasifikasi tumor, maka prediktornya seperti merokok, tekanan darah dan sebagainya. Data latih merupakan sekumpulan prediktor dan kelas yang telah dilatih dipersiapkan untuk pelatihan. Selanjutnya data uji adalah data yang baru dan siap diklasifikasi untuk menguji performa model[18].

2.3 Naïve Bayes

Algoritma Naïve Bayes merupakan algoritma dengan proses pembelajaran yang menggantungkan pada metode probabilitas. Keuntungan menggunakan model Naïve Bayes yaitu memiliki performa klasifikasi yang akurat menurut penelitian yang telah dilakukan. Algoritma Naïve Bayes juga memiliki peluang kecil dalam membuat kesalahan dalam klasifikasi[19].

Metode klasifikasi pada Naïve Bayes menggolongkan probabilitas sederhana. Naïve Bayes juga merancang asumsi bahwa data antar kelas tidak saling bergantung. Proses pembelajaran algoritma ditekankan untuk tujuan mengestimasi probabilitass.

2.4 Support Vector Machine

Support Vector Machine merupakan sistem pembelajaran yang menerapkan ruang hipotesis yang berbentuk fungsi-fungsi linier dalam suatu fitur yang berdimensi tinggi serta dilatih dengan algoritma pembelajaran dengan teori optimasi yang didasarkan. Support Vector Machine dahulu pertama kali diperkenalkan pada tahun 1992 oleh Vapnik. Model ini dikenalkan sebagai rangkaian berupa beberapakonsep-konsep yang sangat unggul dalam bidang *pattern recognition*[20].

Support Vector Machine atau SVM merupakan kategori model yang masuk dalam supervised learning. Model SVM sangat terkenal dapat menyelesaikan masalah khususnya pada regresi atau klasifikasi. Model SVM akan memplot terhadap setiap item data sebagai titik dalam sebuah ruangan dimensi[21].

2.5 Al-Qur'an

Al-Qur'an merupakan kitab pedoman hidup umat Islam. Al-Qur'an mengatur segala hubungan hidup antara manusia dan Allah juga manusia dengan sesama manusia dan makhluk hidup lain ciptaan Allah. Klasifikasi perlu dilakukan untuk memudahkan manusia memahami semua ayat-ayat Al-Qur'an. Klasifikasi juga mempersingkat waktu untuk menemukan ayat Al-Qur'an yang dimaksud. Al-Quran dapat diklasifikasikan menjadi beberapa pembahasan seperti sesuai tempat diturunkannya ayat, sesuai dengan topik ayat, atau sesuai dengan kandungan pokok ayatnya[14].

Al-Qur'an sangat lekat keterkaitannya dengan hubungan sesame makhluk hidup maupun hubungan manusia dengan Rabbnya. Al-Qur'an berisi pedoman petunjuk, rahmat serta sebagai pembenar kitab-kitab sebelumnya[22]. Pembelajaran Al-Qur'an digunakan untuk memahami kalam Ilahi dan sebagai pembelajaran hikmah kehidupan.

2.6 Penelitian Terdahulu

Beberapa penelitian mengenai pengolahan teks terjemahan telah dilakukan. Penelitian yang berkaitan dengan penelitian saat ini dipaparkan untuk memberikan gambaran dan pemahaman lebih mendalam terhadap penelitian. Beberapa tahun sebelumnya penelitian mengenai dokumen teks telah dilakukan oleh banyak peneliti termasuk dengan berbagai model algoritma yang berbeda-beda. Studi Pustaka berikut akan mengidentifikasi metode yang pernah digunakan oleh peneliti sebelumnya. Berikut penyajian pustaka terdahulu dapat dirangkum dalam Tabel 1.

Tabel 1. Penelitian Terdahulu

No	Judul	Tahun, Penulis	Metode	Hasil/Kesimpulan
1	Document Classification using <i>Naïve Bayes</i> for Indonesian Translation of the Quran	2019, Syopiansyah Jaya Putra	<i>Naïve Bayes</i>	klasifikasi teks terjemahan Bahasa Indonesia dengan metode Naive Bayes memiliki rata-rata akurasi 95%. hal ini membuktikan bahwa Naive Bayes masih relevan untuk digunakan sebagai permodelan klasifikasi meski memiliki kekurangan seperti saat menambah topik atau kelas klasifikasi maka akurasi akan menurun.
2	The Qur'anic Uses <i>Algorithm C4.5</i>	2020, Mohamad Irfan	<i>Algoritma C4.5</i>	Penelitian menjalankan proses klasifikasi ayat terjemahan Al-Qur'an menjadi kelas surat Makkiyah dan Madaniyah. Pengujian model menghasilkan akurasi 95.6%. Algoritma C4.5 terbukti dapat membantu pengguna untuk mengelompokkan surat-surat Makkiyah dan Madaniyah dalam Al-Qur'an.
3	The Multi-Class Classification for the First Six Surats of the Holy Quran	2020, Noh Sabri Elmitwally	<i>Naïve Bayes, KNN, Decision Tree J48 dan Support Vector Machine</i>	Penelitian membahas mengenai klasifikasi surat dalam Al-Qur'an. Korpus dilatih dengan beberapa permodelan. Semakin meningkat ukuran korpus dan fitur dari klasifikasi maka akan meningkat juga kualitas dan akurasi. Hasil dengan akurasi terbaik yaitu pada model SVM dengan rata-rata score 88% f1-score.

III. METODE

Penelitian menggunakan data dari website Kementerian Agama Republik Indonesia. Pengumpulan data dilakukan dengan cara mengunduh langsung dari website. File berupa format ZIP yang segera diekstrak menjadi file-file dokumen surat. Data yang digunakan dalam penelitian ini yaitu data teks terjemahan Al-Qur'an Bahasa Indonesia dari juz 5 hingga juz 7. Surat yang terdapat pada juz 5 hingga juz 7 diantaranya Surat An-Nisa', Surat Al-Maidah dan Surat Al-An'am.

Ayat terjemahan Al-Qur'an Bahasa Indonesia diklasifikasikan menjadi 5 kelas yaitu Akidah, Akhlaq, Ibadah, Kisah dan Hukum. Model algoritma yang digunakan dalam penelitian ini adalah algoritma Naïve Bayes dan Support Vector Machine. Distribusi data pada setiap kelas berdasarkan lima kelas disajikan dalam Tabel 2.

Tabel 2. Distribusi Jumlah Label Kelas

No	Label	Jumlah
1	Akidah	253
2	Akhlaq	64
3	Ibadah	7
4	Kisah	26
5	Hukum	33
Total		383

3.1 Preprocessing Data

Dokumen teks sebelum dilakukan proses klasifikasi perlu dilakukan *preprocessing*. Fungsinya untuk menghilangkan beberapa karakter dan kata yang tidak diperlukan. Preprocessing data meliputi *case folding*, *tokenization*, *stopword removal* dan *stemming* serta *feature extraction*.

a) Case Folding

Kata pada setiap kalimat dalam dokumen teks perlu untuk disamakan formatnya. Penyamaan format huruf dapat dilakukan dengan teknik *case folding*. Teknik ini dibutuhkan untuk mengubah format huruf menjadi sama rata dan membuang komponen yang tidak perlu ditampilkan saat proses karena tidak dibutuhkan. Contoh penggunaan yaitu pada fungsi *lower* untuk menghilangkan huruf kapital dan menggantinya dengan huruf kecil. Pada Tabel 3 akan dipaparkan contoh penggunaan *case folding*.

Tabel 3. Proses *Case Folding*

Data Input	Data Output
Maka, berperanglah engkau (Nabi Muhammad) di jalan Allah. Tidaklah engkau dibebani (tanggung jawab), kecuali (yang terkait) dengan dirimu sendiri. Kobarkanlah (semangat) orang-orang mukmin (untuk berperang). Semoga Allah menolak serangan orang-orang yang kufur itu. Allah sangat dahsyat kekuatan-Nya dan sangat keras siksaan-Nya.	maka, berperanglah engkau (nabi muhammad) di jalan allah. tidaklah engkau dibebani (tanggung jawab), kecuali (yang terkait) dengan dirimu sendiri. kobarkanlah (semangat) orang-orang mukmin (untuk berperang). semoga allah menolak serangan orang-orang yang kufur itu. allah sangat dahsyat kekuatan-nya dan sangat keras siksaan-nya.

b) Tokenization

Tokenization berfungsi untuk mengubah kalimat dalam teks dokumen menjadi bentuk token atau kata. Sebelum proses klasifikasi kalimat harus bersih tanpa tanda baca. Kata yang belum bersih tidak dapat diproses karena harus diubah menjadi bentuk token yang siap dieksekusi. Contoh proses tokenization ditunjukkan pada Tabel 4. Kalimat akan diubah menjadi bentuk kata atau bisa disebut dengan token.

Tabel 4. Proses Tokenization

Data Input	Data Output
Allah hendak menerangkan (syariat-Nya) kepadamu, menunjukkan kepadamu berbagai jalan (kehidupan) orang yang sebelum kamu (para nabi dan orang-orang saleh), dan menerima tobatmu. Allah Maha Mengetahui lagi Mahabijaksana.	['allah', 'hendak', 'terang', '', 'syariat', '', 'kepada', 'tunjuk', 'kepada', 'bagai', 'jalan', '', 'hidup', '', 'orang', 'belum', 'kamu', '', 'para', 'nabi', 'orang', 'saleh', '', 'terima', 'tobat', '', 'allah', 'maha', 'tahu', 'mahabijaksana', '']

c) Stopword Removal dan Stemming

Tidak semua kata akan digunakan dalam proses klasifikasi, sehingga *stopword removal* digunakan untuk menghilangkan kata yang tidak digunakan atau bisa disebut dengan kata yang tidak terlalu penting dalam kalimat. Kami menggunakan library Python Sastrawi. Sastrawi dapat berfungsi untuk melakukan *stemming* dan *stopword removal*. Tabel 5 menunjukkan contoh proses *stopword removal* dan *stemming* menggunakan Python Sastrawi.

Tabel 5. Proses *Stopword Removal* dan *Stemming*.

Data Input	Data Output
kebaikan (nikmat) apa pun yang kamu peroleh (berasal) dari allah, sedangkan keburukan (bencana) apa pun yang menimpamu itu disebabkan oleh (kesalahan) dirimu sendiri. kami mengutus engkau (nabi muhammad) menjadi rasul kepada (seluruh) manusia. cukuplah allah sebagai saksi.	baik nikmat apa yang kamu oleh asal allah keburu bencana apa yang timpa sebab salah diri sendiri utus engkau nabi muhammad jadi rasul seluruh manusia cukup allah saksi

d) Feature Extraction

Kami menggunakan ekstraksi fitur dalam penelitian ini. Ekstraksi fitur menerapkan TF-IDF untuk meningkatkan akurasi setiap model. Metode TF-IDF atau *Term Frequency-Inverse Document* dapat digunakan untuk *information retrieval* maupun *feature extraction*.

Proses eksekusi menggunakan Google Collaboratory, dengan perangkat komputer kapasitas RAM 4 GB dan processor intel inside Core i3. Selanjutnya data yang sudah bersih dilakukan klasifikasi menggunakan model Naïve Bayes dan Support Vector Machine.

3.2 Naïve Bayes

Penelitian menggunakan algoritma klasifikasi *supervised learning* Naïve Bayes. Naïve Bayes merupakan algoritma yang menggunakan metode probabilitas. Permodelan Naïve Bayes menggunakan 383 dataset ayat. Ayat dibagi menjadi 5 kelas dengan data latih sebanyak 80% dan data testing sebanyak 20%. Dari 383 ayat sejumlah 287 ayat menjadi data latih dan data testing menggunakan 96 ayat.

3.3 Support Vector Machine

Algoritma Support Vector Machine menggunakan regresi. Penelitian ini menggunakan algoritma SVM dengan membagi data menjadi 80% data latih dan 20% data uji. Data uji dilakukan dengan permodelan SVM dan mendapatkan hasil dengan presisi, recall, f-1 score dan akurasi sebagai bahan perbandingan.

IV. HASIL DAN PEMBAHASAN

Model yang telah diimplementasikan tentu mendapatkan hasil klasifikasi yang berbeda-beda. Masing-masing memiliki performa yang berbeda sesuai dengan proses klasifikasi yang telah dijalankan. Dalam penelitian ini menggunakan algoritma Naïve Bayes dan Support Vector Machine untuk klasifikasi. Pengukuran performa model menggunakan precision, recall, F1-score, dan akurasi.

a) Naïve Bayes

Algoritma Naïve Bayes merupakan salah satu algoritma untuk pembelajaran berbasis teori. Algoritma ini merupakan model *supervised learning* pada pembelajaran mesin. Hasil dari klasifikasi menggunakan model Naïve Bayes ditunjukkan pada Tabel 6.

Tabel 6. Evaluasi Model Naïve Bayes

Label	Precision	Recall	F1-score	Support
Hukum	0.75	0.33	0.4	6
Aqidah	0.8	0.99	0.91	72
Akhlaq	0.43	0.31	0.4	13
Ibadah	0	0	0	2
Kisah	0	0	0	3

b) Support Vector Machine

Model Support Vector Machine merupakan kategori model yang menggunakan regresi atau data klasifikasi dengan metode supervised learning. Sehingga data dilatih dahulu dan baru akan diuji setelah dilatih. SVM merupakan metode untuk membandingkan kandidat set parameter standar nilai diskrit. Hasil evaluasi menggunakan model SVM ditunjukkan pada Tabel 7.

Tabel 7. Evaluasi Model SVM

Label	Precision	Recall	F1-score	Support
Hukum	0.5	0.17	0.25	6
Aqidah	0.79	0.92	0.85	72
Akhlaq	0.12	0.08	0.1	13
Ibadah	0	0	0	2
Kisah	0	0	0	3

Permodelan Naïve Bayes dan Support Vector Machine menghasilkan hasil yang bervariasi. Dengan 383 ayat terjemahan Al-Qur'an masing-masing model mendapatkan hasil tergantung pada pembangunan model *machine learning*. Perbandingan hasil klasifikasi pada model Naïve Bayes dan Support Vector Machine ditunjukkan pada Tabel 8.

Tabel 8. Performa Model

Measures	Naïve Bayes	SVM
Accuracy	80.2	70.83
Precision	74.25	70.83
Recall	80.2	63.74
F-Score	76.18	66.31

Naïve Bayes memiliki hasil akurasi 80.2%. Model Naïve Bayes meraih nilai recall 80.2% sama seperti akurasi yang didapatkan. Namun SVM mendapat akurasi lebih rendah dibandingkan dengan Naïve Bayes yaitu 70.83%. Secara umum Naïve Bayes memiliki akurasi, presisi, recall dan f-score tertinggi dibandingkan dengan model Support Vector Machine.

V. KESIMPULAN

Implementasi *text mining* untuk klasifikasi dokumen terjemahan Al-Qur'an Bahasa Indonesia dengan model algoritma Naïve Bayes dan Support Vector Machine dapat digunakan untuk klasifikasi dokumen terjemahan Al-Qur'an Bahasa Indonesia menjadi 5 kelas yaitu Akidah, Akhlaq, Ibadah, Kisah dan Hukum. Algoritma pengujian data yang memiliki performa terbaik dengan hasil akurasi 80% yaitu Naïve Bayes. Dibandingkan dengan model Support Vector Machine yang mendapatkan akurasi 70% menjadikan Naïve Bayes sebagai model dengan performa terbaik dalam kasus klasifikasi dokumen terjemahan. Dengan begitu algoritma Naïve Bayes dan Support Vector Machine dapat digunakan sebagai model klasifikasi teks. Selanjutnya untuk mengembangkan penelitian dapat dilakukan dengan permodelan lain dan menggunakan konsep *deep learning*.

DAFTAR PUSTAKA

- [1] Syaamil Al-Qur'an Al-Kariim, 1st ed. Bandung: PT. Sygma Examedia Arkanleema, 2009.
- [2] N. Baidan, "Problematika Penerjemahan Al-Qur'an dalam Bahasa Indonesia," *Indones. J. Islam. Lit. Muslim Soc.*, vol. 2, pp. 1–20, 2017.
- [3] Maskur, "Program Terjemahan Al-Qur'an Sistem 40 Jam sebagai Metode Dakwah Alternatif di Makassar," *J. Kaji. Keislam.*, vol. 1, no. 2, pp. 105–118, 2020.
- [4] H. Nst, "Metodologi Terjemahan Al-Qur'an dalam Al-Qur'an dan Terjemahannya Bahasa Batak Angkola," *J. Ilmu-Ilmu Ushuluddin*, vol. 07, no. 01, pp. 2–18, 2019.
- [5] Y. Hasyim, *Akidah Akhlak*, 1st ed. Jakarta: Direktorat KSKK Madrasah Direktorat Jenderal pendidikan Islam Kementerian Agama RI, 2020.
- [6] M. D. R. Wahyudi, "Penerapan Algoritma Cosine Similarity pada Text Mining Terjemah Al-Qur'an Berdasarkan Keterkaitan Topik," *Semesta Tek.*, vol. 22, no. 1, pp. 41–50, 2019, doi: 10.18196/st.221235.
- [7] G. L. Yovellia Londo, D. H. Kartawijaya, H. T. Ivaryani, P. W. P. Yohanes Sigit, A. P. Muhammad Rafi, and D. Ariyandi, "A Study of Text Classification for Indonesian News Article," *Proceeding - 2019 Int. Conf. Artif. Intell. Inf. Technol. ICAIIT 2019*, pp. 205–208, 2019, doi: 10.1109/ICAIIIT.2019.8834611.
- [8] A. Hermanto, "Implementasi Text Mining Menggunakan Naive Bayes untuk Penentuan Kategori Tugas Akhir Mahasiswa Berdasarkan Abstraksinya," *KONVERGENSI*, vol. 12, no. 2, pp. 1–10, 2016.
- [9] A. F. Kamal, "Text Mining untuk Analisa Sentiment Ekspedisi Jasa Pengiriman Barang Menggunakan Metode Naive Bayes pada Aplikasi J&T Express," Dian Nuswantoro, 2017.
- [10] P. Padli and Ependi, "Analisa Pola Penyelaguna Facebook yang menjadi Alat Kejahatan Trafficking dengan Data Mining," in *Bina Darma Conference an Computer Science*, 2020, pp. 339–344.
- [11] T. H. Putrisanni, Adiwijaya, and S. Al Faraby, "Klasifikasi Ayat Al-Qur'an Terjemahan Bahasa Inggris menggunakan K-Nearest Neighbor (KNN) dan Information Gain," *KOMIK Konf. Nas. Teknol. Inf. dan Komput.*, vol. 3, pp. 362–369, 2019, doi: DOI: 10.30865/komik.v3i1.1614.
- [12] M. Robani and A. Widodo, "Algoritma K-Means Clustering untuk Pengelompokkan Ayat pada Terjemahan Bahasa Indonesia," *J. Sist. Inf. Bisnis*, vol. 2, no. 164–176, 2016.
- [13] R. Hidayat and S. Minanti, "Comparative Analysis of Text Mining Classification Algorithms for English and Indonesian Qur'an Translation," *IJID Int. J. Informatics Dev.*, vol. 8, pp. 48–51, 2019.
- [14] M. Irfan, W. Uriawan, N. Lukman, O. T. Kurahman, and W. Darmalaksana, "The Qur'anic Classification Uses Algorithm C4.5," *EAI*, 2018, doi: <http://dx.doi.org/10.4108/eai.2-10-2018.2295558>.
- [15] W. C. Xia Xie, YuFu, Hai Jin, Yaliang Zhao, "A novel text mining approach for scholar information extraction from web content in Chinese," *Futur. Gener. Comput. Syst.*, pp. 859–872, 2020.
- [16] C. Slamet, A. Rahman, M. A. Ramdhani, and W. Darmalaksana, "Clustering the Verses of the

- Holy Qur'an Using K-Means Algotirhm," *Asian J. Inf. Technol.*, vol. 15, no. 4, pp. 5160–5162, 2016.
- [17] I. Romli, "Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5," *J. Sains Komput. Inform.*, vol. Volume 4 n, p. 696, 2020.
- [18] P. Widodo, "Konsep Data Mining," in *Penerapan Data Mining dengan MATLAB*, 1st ed., Bandung: Rekayasa Sains, 2013, p. 2.
- [19] R. E. Putri, "Perbandingan Metode Klasifikasi Naive Bayes dan K-Nearest Neighbor pada Analisa Data Status Kerja di Kabupaten Demak Tahun 2012," *J. GAUSSIAN*, vol. 3, pp. 831–838, 2014.
- [20] D. Irawan, E. B. Perkasa, D. Wahyuningsih, E. Helmud, and Yurindra, "Perbandingan Klassifikasi SMS Berbasis Support Vector Machine , Naive Bayes Classifier , Random Forest dan Bagging Classifier," *J. SISFOKOM*, vol. 10, pp. 432–437, 2021.
- [21] M. I. Rahman, N. A. Samsudin, A. Mustapha, and A. Abdullahi, "Comparative Analysis for Topic Classification in Juz Al-Baqarah," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 12, no. 1, pp. 406–411, 2018, doi: 10.11591/ijeecs.v12.i1.pp406-411.
- [22] A. Sari, "Kisah-kisah dalam Al-Qur'an," *J. Pengetah. tentang Ilmu dan Hikmah*, vol. Volume 1, pp. 89–112, 2016.