

## KLASIFIKASI AL – QUR’AN TERJEMAHAN BAHASA INDONESIA DENGAN MENGGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE* (SVM)

Moch. Fauzan<sup>1,\*</sup>, Hartarto Junaedi<sup>1</sup>, Endang Setyati<sup>1</sup>

<sup>1</sup>Departemen Informatika, Institut Sains dan Teknologi Terpadu Surabaya

E-mail: fauzan@umsida.ac.id\*

### ABSTRACT

*The classification of verses of the Koran in Indonesian translation aims to classify verses of the Koran that have the same meaning on certain topics. In this study, the labeling of translated Qur'anic verses is grouped into 6 categories including education, motivation, social, history, politics and science (mathematics). The method proposed in this study uses Chi Square feature selection and Principal Analysis with the application of a classification model using the Support Vector Machine (SVM) algorithm to group the translated verses of the Koran into 6 categories. The initial stage is preprocessing, which aims to find the weighting value for each document using TF-IDF. After getting the weighting value for each document, a search for the best classification model is carried out to label the verses of the Qur'an by using feature selection and without using feature selection. In this study, the best classification model results without using feature selection in the SVM algorithm, the AUC value is 83.3%, while using Chi Square feature selection, the AUC is 73.3%, while the PCA feature selection is 63.3%. So that this research is the best model in classifying the Indonesian translation of the Qur'anic verses without using feature selection with the highest AUC value of 83.3%.*

**Keywords:** *Feature Selection Techniques; Holy Qur'an; Algorithm Support Vector Machine (SVM); AUC; f1-score.*

### ABSTRAK

Klasifikasi ayat al-qur'an terjemahan Bahasa Indonesia bertujuan untuk mengelompokkan ayat alqur'an yang mempunyai makna yang sama pada topik tertentu. Pada penelitian ini pelabelan dokumen ayat al - qur'an terjemahan dikelompokkan menjadi 6 kategori diantaranya pendidikan, motivasi, sosial, sejarah, politik dan sains (matematika). Metode yang diusulkan dalam penelitian ini menggunakan *feature selection Chi Square* dan *Principal Component Analysis (PCA)* dengan penerapan model klasifikasi menggunakan algoritma *Support Vector Machine (SVM)* untuk mengelompokkan ayat al - qur'an terjemahan ke dalam 6 kategori. Tahap awal yang dilakukan adalah *preprocessing* bertujuan untuk mencari nilai pembobotan pada setiap dokumen dengan menggunakan TF-IDF. Setelah mendapatkan nilai pembobotan pada setiap dokumen dilakukan pencarian model klasifikasi terbaik untuk melabeli ayat al-qur'an dengan menggunakan *feature selection* dan tanpa menggunakan *feature selection*. Pada penelitian ini didapatkan hasil model klasifikasi terbaik tanpa menggunakan *feature selection* pada algoritma SVM didapatkan nilai AUC 83.3% sedangkan dengan menggunakan *feature selection Chi Square* mendapatkan nilai AUC 73.3 % sedangkan dengan pada *feature selection PCA* mendapatkan nilai AUC 63.3 %. Sehingga penelitian ini model yang terbaik dalam mengklasifikasi ayat al-qur'an terjemahan Bahasa Indonesia tanpa menggunakan *feature selection* dengan nilai AUC tertinggi 83.3 %.

**Kata Kunci:** *Teknik feature selection; Al-Qur'an; Algoritma Support Vector Machine (SVM); AUC; f1-score*

## 1. Pendahuluan

Klasifikasi teks otomatis adalah cara sistematis untuk mengatur dokumen ke dalam kategori yang telah ditentukan [1]. Nomor dokumen yang diproses selama bertahun-tahun terus meningkat sebagai hasil dari pertumbuhan dan perkembangan teknologi di bidang kecerdasan buatan (AI) dan pembelajaran mesin (ML). Dengan semakin meningkatnya permintaan untuk pemrosesan data dokumen yang besar dalam waktu singkat, maka mengotomatiskan pengolahan dokumen menjadi sebuah kebutuhan.

Al-Quran adalah teks yang unik dan sumber informasi yang baik dengan 78.000 kata yang disusun secara sistematis menjadi beberapa bagian seperti ayat, bab, perempat, bagian dll, oleh para ulama [1]. Pada ayat Al-Qur'an memiliki makna yang saling berkaitan diantara surah yang satu dengan surah yang lain sehingga diperlukan suatu analisis khusus untuk pelabelan terjemahan ayat Al-Qur'an. Pada penelitian ini tidak cukup hanya melakukan pelabelan saja tetapi juga akan melakukan analisis pada pengaruh *feature selection* terhadap klasifikasi ayat Al-Qur'an. Pada terjemahan Al-Qur'an tersebut akan dikelompokkan kedalam 6 kategori diantaranya Pendidikan, Motivasi, Sosial, Sejarah, Politik dan Sains (Matematika).

Pada penelitian ini proses pengklasifikasian teks terjemahan Al-Qur'an menggunakan algoritma *Support Vector Machine (SVM)*, algoritma ini dipilih karena dari berbagai sumber referensi yang ada mampu menyelesaikan proses klasifikasi yang *multi class* dan menghasilkan nilai *accuracy* yang tinggi. Proses klasifikasi ini pertama dilakukan teknik pembobotan teks dokumen dengan menggunakan metode TF-IDF, setelah melakukan pembobotan pada dokumen dilakukan pencarian model terbaik dalam pengklasifikasian teks menggunakan

SVM. Untuk menentukan model klasifikasi dilakukan menggunakan *feature selection Chi Square* dan *Principal Component Analysis (PCA)* serta tanpa menggunakan *feature selection*. Hal ini dilakukan untuk mencari sejauh mana pengaruh penggunaan *feature selection* pada klasifikasi teks terjemahan Al-Qur'an Bahasa Indonesia.

## 2. Tinjauan Pustaka

Bidang studi Al-Qur'an telah menyaksikan cukup banyak karya penelitian di antaranya adalah aplikasi klasifikasi teks pada Al-Qur'an [2]–[7]; aplikasi berbasis ontologi [8]–[11]; aplikasi Alquran digital [12]–[15]. Dari sekian teknik yang telah banyak diterapkan pada masalah klasifikasi teks, diantaranya adalah *Naïve Bayes* [4], *Decision Tree* [5], *Neural Network* [6], *Support Vector Machine* [7], dan *K-NN* [3]. Pada penelitian sebelumnya tentang klasifikasi ayat Al-Qur'an terjemahan Bahasa Inggris menggunakan algoritma SVM dan *feature selection* menghasilkan nilai *accuracy* 90% [3] dan pada penelitian sebelumnya klasifikasi ayat Al-Qur'an terjemahan Bahasa Inggris dengan menggunakan algoritma *Neural Network* menghasilkan nilai *accuracy* 85% [4].

Pada penelitian ini akan dilakukan percobaan klasifikasi ayat Al-Qur'an terjemahan Bahasa Indonesia menggunakan *feature selection* dan tanpa menggunakan *feature selection* pada algoritma SVM. *Text Mining* merupakan salah satu bidang khusus dari data mining. *Text mining* juga dapat diartikan sebagai penggalian suatu informasi pada teks tertentu yang mempunyai karakteristik dalam kategori yang sama sehingga bisa tersajikan suatu informasi yang memiliki tujuan yang sama. *Text mining* dapat didefinisikan sebagai suatu proses menggali informasi dimana seorang *user* berinteraksi dengan sekumpulan dokumen menggunakan *tools* analisis yang merupakan

komponen – komponen dalam *data mining* yang salah satunya adalah kategorisasi [3].

### 2.1. Tahapan *Text Mining*

Dalam *text mining*, proses klasifikasi merupakan inti dari proses untuk menemukan pola (*pattern discovery*), tetapi secara lengkap proses *text mining* dibagi menjadi 3 (tiga) tahapan utama, yaitu proses awal terhadap teks untuk mengubah dokumen teks yang awalnya berupa kalimat – kalimat, menjadi vektor yang sesuai dengan algoritma yang digunakan (*text preprocessing*), transformasi teks ke dalam bentuk antara (*text transformation / feature generation*), dan penemuan pola (*pattern discovery*) [3].

Tahapan *preprocessing* ini dilakukan agar dalam klasifikasi dapat diproses dengan baik. Tahapan dalam *preprocessing text* adalah sebagai berikut:

- a. *Case Folding*, merupakan proses untuk mengubah semua karakter pada teks menjadi huruf kecil. Karakter yang diproses hanya huruf ‘a’ hingga ‘z’ dan selain karakter tersebut akan dihilangkan seperti tanda baca titik (.), koma (,), dan angka.[7]
- b. *Tokenizing*, merupakan proses memecah yang semula berupa kalimat menjadi kata-kata atau memutus urutan string menjadi potongan-potongan seperti kata-kata berdasarkan tiap kata yang menyusunnya.
- c. *Stopwords*, merupakan kosakata yang bukan merupakan kata unik atau ciri pada suatu dokumen atau tidak menyampaikan pesan apapun secara signifikan pada teks atau kalimat [8]. Kosakata yang dimaksudkan adalah kata penghubung dan kata keterangan yang bukan merupakan kata unik misalnya “sebuah”, “oleh”, “pada”, dan sebagainya.
- d. *Stemming*, yakni proses untuk mendapatkan kata dasar dengan cara menghilangkan awalan, akhiran, sisipan,

dan *confixes* (kombinasi dari awalan dan akhiran).

### 2.2. *Text Transformation*

Pemilihan fitur kata merupakan tahap lanjut dari pengurangan dimensi pada proses transformasi teks. Walaupun tahap sebelumnya sudah melakukan penghapusan kata-kata yang tidak deskriptif *stopwords*, namun tidak semua kata-kata didalam dokumen memiliki arti penting.

- *Term Frequency (TF-IDF)*

TF-IDF merupakan metode yang digunakan untuk menentukan nilai frekuensi sebuah kata di dalam sebuah dokumen atau artikel dan juga frekuensi di dalam banyak dokumen, dengan Persamaan (1).

$$W_{d,t} = tf_{d,t} * IDF_t \quad (1)$$

dimana,  $d$  adalah dokumen ke –  $d$ ;  $t$  adalah kata ke –  $t$  dari kata kunci;  $W$  merupakan bobot dokumen ke –  $d$  terhadap kata ke –  $t$ ; dan  $tf$  adalah *term* frekuensi/frekuensi kata.

- *Chi Square*

*Chi-Square* merupakan salah satu metode *feature selection* yang termasuk metode filter. Metode *feature selection* terbukti dapat meningkatkan akurasi pada beberapa penelitian sebelumnya.

**Tabel 1.** Kontigensi Perhitungan Nilai *Chi-Square*

Kata	Kelas	
	C	c*
t	A	B
t*	C	D

Nilai yang terdapat pada Tabel 1 merupakan nilai frekuensi observasi dari suatu kata terhadap kelas.

Pada tahap ini, tiap kata yang diperoleh dihitung menggunakan Persamaan (2).

$$X^2(t, c) = \frac{N(AD-CB)^2}{(A+C)(B+D)(A+B)(C+D)} \quad (2)$$

• *Principal Component Analisis (PCA)*

PCA dapat digunakan untuk mereduksi dimensi suatu data tanpa mengurangi karakteristik data tersebut secara signifikan. Hal ini dapat dipandang sebagai suatu teknik independen domain untuk ekstraksi fitur, yang berlaku untuk berbagai macam data. Dalam rangka untuk melakukan analisis komponen utama pada set dokumen pelatihan, akan diwakili set fitur vektor acak oleh vektor berdimensi  $N(x)$ . *Dataset* matrik  $x$  berukuran  $(M \times N)$  yang dimisalkan  $x_1, x_2, \dots, x_M$  adalah  $N \times 1$  *vectors*. Algoritma dari analisis komponen utama adalah sebagai berikut:

- a. Hitung vektor rata-rata menggunakan Persamaan (3)

$$\bar{x} = \frac{\sum_{i=1}^M x_i}{M} \quad (3)$$

- b. Hitung mean dengan Persamaan (4).

$$\Phi_i = x_i - \bar{x} \quad (4)$$

- c. Dari matriks pada Persamaan (5)

$$A = [\Phi_1, \Phi_2, \dots, \Phi_M] \quad (5)$$

Hitung matrix kovariansi C dengan Persamaan (6)

$$C = \frac{1}{M} \sum_{x=1}^M \Phi_n, \Phi_n^T = AA^T \quad (6)$$

Persamaan (7) untuk menghitung *eigenvalue* dari

$$C: \lambda_1 > \lambda_2 > \dots > \lambda_N \quad (7)$$

Dengan Persamaan (8) untuk menghitung *eigenvectors* dari

$$C: u_1 > u_2 > \dots > u_N \quad (8)$$

Reduksi dimensi hanya berhubungan dengan  $K$  *eigenvalue* terbesar seperti pada Persamaan (9).

$$\hat{x} - \bar{x} = \sum_{i=1}^k b_i u_i \text{ where } K \ll N \quad (9)$$

Transformasi  $R^N \rightarrow R^K$  dengan melakukan reduksi dimensi dinyatakan dengan Persamaan (10).

$$(x_i - \bar{x}) = u^T(x_i - \bar{x}) \quad (10)$$

Sehingga banyak principal component pada Persamaan (11) dan Persamaan (12).

$$\sum_{i=1}^K \lambda > \text{threshold (e.g 0.9 or 0.95)} \quad (11)$$

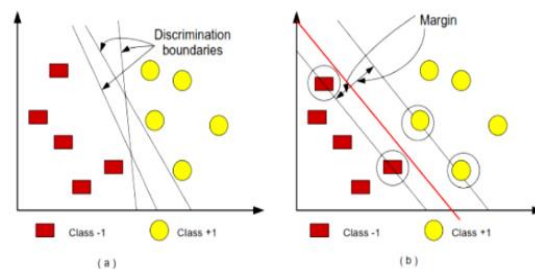
$$\sum_{i=1}^N \lambda > \text{threshold (e.g 0.9 or 0.95)} \quad (12)$$

### 2.3. Klasifikasi

Kategorisasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui.

• *Support Vector Machine (SVM)*

Sebagai salah satu algoritma klasifikasi yang sering digunakan, SVM bekerja dengan cara mencari sebuah *hyperplane* atau garis pembatas pemisah antar kelas yang mempunyai margin atau jarak antar *hyperplane*, seperti pada Gambar 1 dengan data paling terdekat pada setiap kelas yang paling besar yang dihitung dengan Persamaan (13) dan Persamaan (14).



Gambar 1. *Hyperplane* kelas -1 & +1

$$\min \vec{w} \tau(w) = \frac{1}{2} \|\vec{w}\|^2 \quad (13)$$

$$\frac{1}{2} \|w\|^2 = \frac{1}{2} (w_1^2 + w_2^2) \quad (14)$$

Dengan syarat pada Persamaan (15) dan (16):

$$y_i(x_i \cdot w + b) - 1 \geq 0, i = 1, 2, 3, \dots, n \quad (15)$$

$$y_i(x_1 \cdot w_1 + x_2 \cdot w_2 + b) \geq 1 \quad (16)$$

sehingga ditemukan Persamaan (17) sebagai berikut:

$$\begin{aligned} (w_1 + w_2 + b) &\geq 1 \text{ untuk } y_1 = 1, x_1 = 1, x_2 = 1 \\ (-w_1 + w_2 - b) &\geq 1 \text{ untuk } y_2 = -1, x_1 = 1, x_2 = -1 \\ (w_1 - w_2 - b) &\geq 1 \text{ untuk } y_3 = -1, x_1 = -1, x_2 = 1 \\ (w_1 + w_2 - b) &\geq 1 \text{ untuk } y_4 = -1, x_1 = -1, x_2 = -1 \end{aligned} \quad (17)$$

### 3. Metode

Penelitian ini adalah tentang pengaruh penerapan *feature selection* pada teks terjemahan Al-Qur'an Bahasa Indonesia yang akan di uji cobakan untuk mengklasifikasikan ke dalam 6 kelas dengan menggunakan algoritma SVM. Dalam penelitian ini, dua *feature selection* yang diuji coba : *Chi Square* dan *PCA*. Percobaan ini menggunakan algoritma SVM ntuk mengklasifikasikan teks input ke label yang diinginkan. Desain arsitektur kerja pada penelitian ini terdiri dari empat langkah untuk mengklasifikasikan teks Al-Qur'an seperti yang ditunjukkan pada Gambar 2. Sumber data adalah ayat-ayat Al-Qur'an yang dikumpulkan dari sumber gabungan terjemahan dan tafsir Al-Qur'an. Data teks gabungan yang dihasilkan disebut data yang dikelompokkan.

#### 3.1. Sumber Data

*Dataset* ini diperoleh dari sebuah website resmi kementerian Agama Republik Indonesia <https://quran.kemenag.go.id/api/v1>. Total data terjemahan Al-Qur'an sebanyak *dataset* terdiri 6236 ayat.

**Tabel 2.** Distribusi Data

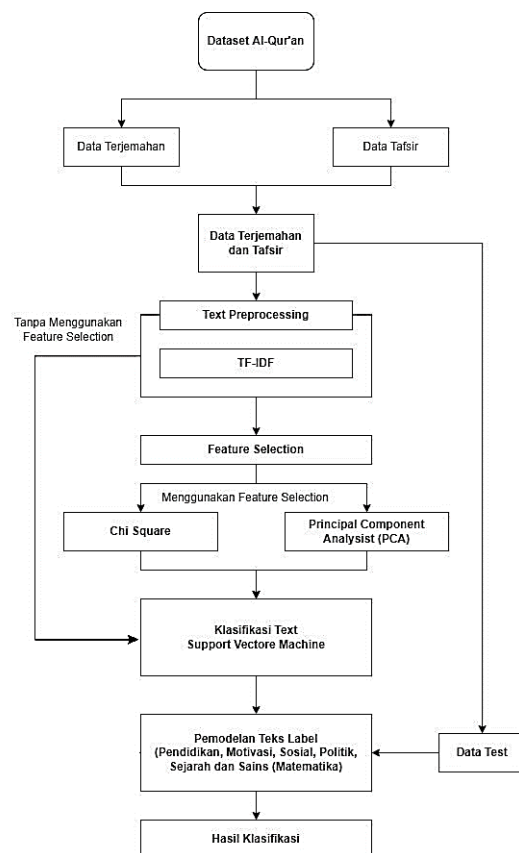
Kelas	Jum. Data
Pendidikan	260
Motivasi	240
Sosial	240
Politik	240
Sejarah	260
Sains (Matematika)	240

Pada penelitian ini distribusi *dataset* terbagi menjadi 1600 sebagai data *training* dan 200 sebagai data *testing* sehingga total keseluruhan data yang terpakai pada penelitian ini sebanyak 1800 data terjemahan.

Data tersebut diambil dari beberapa surah diantaranya Al-Baqarah, Al-Imran, An-nissa, Al-Maidah. Tabel 2 menampilkan distribusi data *training* pada masing-masing kelas yang sudah ditentukan.

#### 3.2. Arsitektur Kerja

Berikut ini merupakan arsitektur sistem kerja klasifikasi teks Al-Qur'an dengan menggunakan *feature selection* dan tanpa menggunakan *feature selection*. Tahap pertama akan melakukan *preprocessing* untuk menghasilkan kata dasar, setelah mendapatkan kata dasar akan dilakukan proses pembobotan dengan menggunakan TF-IDF untuk mencari nilai pada setiap dokumen yang nantinya akan digunakan untuk proses klasifikasi menggunakan algoritma SVM dan akan dilakukan pencarian model terbaik antara menggunakan *feature selection* dan tanpa *feature selection*.



**Gambar 2.** Arsitektur Kerja Sistem

Secara ringkas, algoritma dari kerja sistem dapat dinyatakan dalam urutan sebagai berikut:

- Step 1: Input data (terjemahan dan tafsir Al-Qur'an)
- Step 2: lakukan proses *feature generation* dari data input menggunakan TF-IDF
- Step 3: *preprocessing feature selection* menggunakan *Chi Square*
- Step 4: implementasi *feature selection* dari step 3 dengan klasifikasi SVM
- Step 5: melakukan tahap dari step 2 menggunakan *selection feature* PCA
- Step 6: implementasi *selection feature* dari step 5 dengan klasifikasi SVM
- Step 7: melakukan implementasi proses dari tahap 2 dengan klasifikasi SVM
- Step 8: pembentukan model klasifikasi terbaik antara *feature selection* dan tanpa *feature selection*
- Step 9: melakukan proses data test ke dalam model klasifikasi terbaik
- Step 10: melakukan evaluasi hasil klasifikasi

### 3.3. Evaluasi

Akurasi adalah salah satu ukuran evaluasi standar yang digunakan untuk memvalidasi hasil pengklasifikasi [5]. Dengan menggunakan rumus matrik, akurasi dapat diperoleh menggunakan Persamaan (18).

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (18)$$

Selain itu, nilai AUC sebagai metode validasi juga mencerminkan kinerja peringkat keseluruhan dari pengklasifikasi.

### 4. Hasil dan Pembahasan

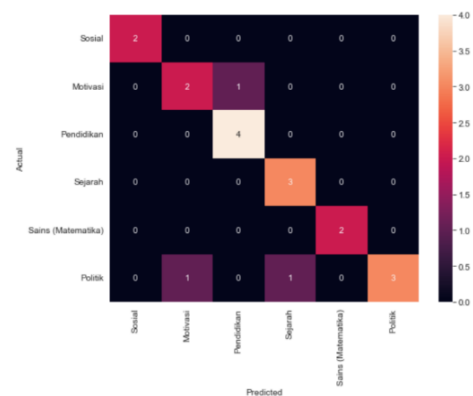
Dari hasil percobaan yang dilakukan pada pengklasifikasian teks Al-Qur'an terjemahan dan tafsir Bahasa Indonesia dengan menggunakan *feature selection* dan

tanpa menggunakan *feature selection* diterapkan pada algoritma SVM didapatkan hasil seperti Tabel 3.

**Tabel 3.** Hasil *f1-score*

FS Algoritma	Waktu	ACC (%)
Tanpa FS	1.65s	85.5
Chi	1.43s	70.5
PCA	1.35s	65.5

Pada hasil diatas, penelitian ini menunjukkan bahwa proses klasifikasi ayat Al-Qur'an terjemahan Bahasa Indonesia dengan menggunakan algoritma SVM mendapat nilai *accuracy* 85.5% tanpa menggunakan *feature selection* tetapi tetapi proses ini memakan waktu yang relatif lebih lama dibandingkan dengan menggunakan *feature selection*. Sementara itu pada penggunaan *feature selection* *Chi Square* didapatkan nilai *accuracy* 70.5% dengan nilai waktu 1.43s relatif lebih cepat dibandingkan tanpa menggunakan *feature selection*. Sedangkan dengan menggunakan *feature selection* PCA mendapatkan nilai *accuracy* hanya 65.5% dengan nilai waktu 1.35s. Analisis hasil ini menunjukkan bahwa pengaruh *feature selection* pada proses klasifikasi menggunakan SVM mencapai hasil *accuracy* yang lebih rendah. Akibatnya tanpa menggunakan *feature selection* terdapat *runtime* komputasi yang lebih lama sehingga proses klasifikasi menggunakan algoritma SVM dinilai kurang efisien.



**Gambar 3.** Hasil *Confusion Matrix*

Pada tahap ini akan dilakukan proses evaluasi *accuracy* dengan menggunakan *Confusion Matrix*. Hasil yang ditunjukkan untuk mengevaluasi nilai *accuracy* seperti yang ditunjukkan pada Gambar 3.

## 5. Simpulan

Penggunaan *feature selection* pada model klasifikasi teks terjemahan ayat Al-Qur'an Bahasa Indonesia tidak memberikan dampak terbaik pada proses klasifikasi dengan menggunakan algoritma SVM. Tujuan akhir dari penelitian ini adalah untuk mengotomatisasi pelabelan ayat-ayat Al-Qur'an ke dalam kelas yang telah ditentukan: Pendidikan, Motivasi, Sosial, Politik, Sejarah dan Sains (Matematika). Secara khusus, penelitian ini bertujuan untuk mencari tahu dampak pengaruh *feature selection* pada proses klasifikasi teks Bahasa Indonesia dengan menggunakan algoritma SVM.

## 6. Daftar Pustaka

- [1] A. O. Adeleke, N. A. Samsudin, A. Mustapha, and N. M. Nawawi, "Comparative analysis of text classification algorithms for automated labelling of Quranic verses," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 7, no. 4, pp. 1419–1427, 2017.
- [2] A. Adeleke and N. Samsudin, "A Hybrid Feature Selection Technique for Classification of Group-based Holy Quran Verses," *International Journal of Engineering & Technology*, no. December, pp. 228–233, 2018.
- [3] A. O. Adeleke, N. A. Samsudin, A. Mustapha, and N. M. Nawawi, "A group-based feature selection approach to improve classification of Holy Quran verses," *Advances in Intelligent Systems and Computing*, vol. 700, no. January, pp. 282–297, 2018.
- [4] A. Adeleke, N. Samsudin, A. Mustapha, and S. Ahmad Khalid, "Automating quranic verses labeling using machine learning approach," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 16, no. 2, pp. 925–931, 2019.
- [5] S. K. Hamed and M. J. Ab Aziz, "Classification of Holy Quran translation using Neural Network technique," *Journal of Engineering and Applied Sciences*, vol. 13, no. 12, pp. 4468–4475, 2018.
- [6] A. Adeleke, N. A. Samsudin, Z. A. Othman, and S. K. Ahmad Khalid, "A two-step feature selection method for quranic text classification," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 16, no. 2, pp. 730–736, 2019.
- [7] A. Ta'a, S. Zainal Abidin, M. S. Abdullah, A. B. Mat Ali, and M. Ahmad, "Al-Quran themes classification using ontology," in *4th International Conference on Computing and Informatics (ICOCI 2013)*, 2013.
- [8] M. A. Siddiqui, S. M. Faraz, and S. A. Sattar, "Discovering the Thematic Structure of the Quran using Probabilistic Topic Model," *Proceedings - 2013 Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences, NOORIC 2013*, no. May 2015, pp. 234–239, 2015.
- [9] M. F. H. Sianturi, S. Al Faraby, S. Ilmu, K. Fakultas, and I. Universitas, "Klasifikasi Dokumen Menggunakan Kombinasi Algoritma Principal Component Analysis Dan Svm Document Classification Using Combination of Principal Component Analysis Algorithm and Svm," *e-Proceeding of Engineering*, vol. 4, no. 3, pp. 5141–5143, 2017.
- [10] S. N. Asiyah, "Klasifikasi berita online menggunakan metode support vector machine dan k-nearest neighbor [skripsi]," *Surabaya: Institut Teknologi*

- Sepuluh Nopember*, vol. 5, no. 2, pp. 1–73, 2016.
- [11] M. I. Rahman, N. A. Samsudin, A. Mustapha, and A. Abdullahi, “Comparative analysis for topic classification in Juz Al-Baqarah,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 1, pp. 406–411, 2018.
- [12] F. Taufiqurrahman, S. Al Faraby, and M. D. Purbolaksono, “Klasifikasi Teks Multi Label pada Hadis Terjemahan Bahasa Indonesia Menggunakan Chi Square dan SVM,” *e-Proceeding of Engineering*, vol. 8, no. 5, pp. 10650–10659, 2021.
- [13] S. Chua and P. N. E. Nohuddin, “Relationship analysis of keyword and chapter in Malay-translated tafseer of Al-Quran,” *Journal of Telecommunication, Electronic and Computer Engineering*, vol. 9, no. 2–10, pp. 185–189, 2017.
- [14] A. Salama, Adiwijaya, and S. Al Faraby, “Klasifikasi Topik Ayat Al-Qur’an Terjemahan Berbahasa Inggris Menggunakan Metode Support Vector Machine Berbasis Vector Space Model dan Word2Vec,” *E-proceeding of Engineering*, vol. 6, no. 2, pp. 9133–9142, 2019.
- [15] T. W. Utami and I. Arianti, “Principal Component Analysis Support Vector Machine (Pca-Svm) Untuk Klasifikasi Kesejahteraan Rumah Tangga Di Kabupaten ...,” *Proceeding SENDIU 2020*, pp. 978–979, 2020.