

**PENERAPAN METODE *SUPPORT VECTOR MACHINE* (SVM) UNTUK  
MENGKLASIFIKASI TEKS TERHADAP HADITS PALSU**

**PROPOSAL TUGAS AKHIR**

**NPM** : 0617124001  
**Nama** : Asep Maulana Ismail  
**Konsentrasi** : *Information Technology*



**SK.Ketua Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT)**

**Nomor : 2035/SK//BAN-PT/Akred/S1/IX/2016**

**PROGRAM STUDI INFORMATIKA – FAKULTAS TEKNIK  
UNIVERSITAS WIDYATAMA  
BANDUNG**

### REVIEW PROPOSAL PENELITIAN

NIM : 0617124001  
Nama : Asep Maulana Ismail  
Konsentrasi : *Information Technology*  
Judul : Penerapan Metode *Support Vector Machine* (SVM) Untuk Mengklasifikasi Teks Terhadap Hadits Palsu

Aspek Penilaian	Hasil Review
Originalitas	
Kelengkapan informasi	
Kelayakan ilmu dan waktu pelaksanaan	

Kesimpulan:

Diterima / diperbaiki / ditolak

Alasan :

Bandung, 23 Maret 2019	
Menyetujui, Ka. Lab. <i>Information Technology</i> ,	Mengetahui, Sek. Prodi Informatika
<b>Ardiles Sinaga S.T., M.T.</b>	<b>Ari Purno Wahyu, S.Kom, M.Kom.</b>

## 1. Latar Belakang Masalah

Hadits merupakan sumber dasar hukum kedua setelah Al-Quran. Oleh karena itu mengetahui suatu hadits tersebut adalah hadits sahih dan non-sahih yang termasuk di dalamnya yaitu *dhaif* (lemah) dan *maudhu* (palsu) sangatlah penting. Hadits itu adalah seluruh perkataan, perbuatan, ketetapan, serta persetujuan Nabi Muhammad S.A.W yang menjadi salah satu dasar pengambilan hukum dalam Islam [1].

Berkembangnya teknologi saat ini berdampak kepada derasny arus informasi yang kita dapatkan, terutama melalui sosial media. Salah satu informasi yang mudah menyebar yaitu informasi mengenai sebuah hadits. Dengan berharap mendapatkan nilai ibadah, banyak yang menyebarkan kembali informasi hadits yang ia terima. Namun bila tidak diimbangi dengan ketelitian dalam memilih dan memilah hadits mana saja yang bersifat sahih dan tidak, dapat beresiko kepada menyebarnya hadits palsu di masyarakat.

Mempelajari dan mengamalkan sebuah hadits sangatlah penting bagi seorang Muslim, karena hadits merupakan salah satu sumber hukum dalam kehidupan sehari-hari. Beberapa peneliti telah melakukan penelitian tentang penerapan hadits pada kehidupan manusia. Bahkan ada beberapa peneliti yang membandingkan buku hadits yang diterbitkan oleh ulama. Tujuannya adalah untuk mencari kebenaran hadits-hadits tersebut dan menerapkannya dalam kehidupan sehari-hari [2].

Studi tentang hadits ini telah dilakukan oleh para peneliti secara konvensional dengan membandingkan satu hadits dengan yang lain, atau bahkan perbandingan silang antara hadits yang ada [3]. Penelitian pun dilakukan untuk meneliti perbedaan antara Al-Quran dan hadits [4].

Mengklasifikasikan sebuah hadits sangatlah penting untuk mengetahui tingkat hadits apakah itu hadis yang sahih atau tidak. Penelitian yang dilakukan dengan menggunakan literatur tentang hadits telah dilakukan dengan menerapkan metode sosiologis untuk memeriksa penyebaran dari sebuah hadits dan untuk mendokumentasikan apa yang membuat suatu hadits lebih menonjol di masanya [5].

Pada penelitian sebelumnya yang dilakukan oleh Agung B. Prasetyo yang berjudul “*Hoax Detection System on Indonesian News Sites Based on Text Classification using SVM dan SGD*” menyebutkan bahwa *Support Vector Machine* atau SVM merupakan salah satu algoritma *supervised learning* yang baik digunakan untuk klasifikasi sebuah teks yang memiliki dimensi input yang banyak [6]. Sedangkan pada penelitian lain juga pernah dilakukan oleh Ina Najiyah yang berjudul “*Hadith Degree Classification for Shahih Hadith Identification Web Based*” memperkuat permasalahan yang sudah diutarakan diatas merupakan permasalahan yang harus dan mampu dipecahkan oleh teknologi.

Berdasarkan permasalahan dan penelitian terkait diatas maka penulis menyusun laporan tugas akhir yang berjudul “**PENERAPAN METODE *SUPPORT VECTOR MACHINE* (SVM) UNTUK MENGLASIFIKASI TEKS TERHADAP HADITS PALSU**”.

## **2. Rumusan Masalah**

Berdasarkan pada permasalahan yang telah disebutkan pada bagian latar belakang masalah, maka dapat dirumuskan permasalahan sebagai berikut:

1. Bagaimana membuat sistem klasifikasi teks yang dapat menentukan apakah suatu hadits termasuk hadits palsu atau bukan berdasarkan karakteristik dari teks hadits dengan menggunakan algoritma SVM?
2. Bagaimana tingkat akurasi algoritma SVM dalam mengklasifikasi hadits palsu ?

## **3. Tujuan**

Penelitian pada permasalahan yang telah dijelaskan pada bagian latar belakang masalah, bertujuan sebagai berikut :

1. Untuk membuat sistem klasifikasi teks yang dapat menentukan apakah suatu hadits termasuk hadits palsu atau bukan berdasarkan karakteristik dari teks hadits dengan menggunakan algoritma SVM.
2. Untuk mengetahui tingkat akurasi algoritma SVM dalam mengklasifikasi hadits palsu.

## 4. Batasan Masalah

Adanya batasan masalah dalam pembuatan sistem klasifikasi hadits palsu, yaitu :

1. Klasifikasi hadits terbagi menjadi dua, yaitu hadits palsu dan bukan hadits palsu.
2. Data hadits asli menggunakan hadits-hadits yang diambil dari buku hadits dan internet, sedangkan hadits palsu menggunakan hadits-hadits yang diambil dari internet dan sosial media.
3. Validasi hadits dilakukan berdasarkan keterangan di buku, internet dan juga dilakukan validasi secara manual oleh orang yang memiliki pengetahuan tentang ilmu hadits.
4. Dalam menentukan rasio yang tepat antara *data training* dan *data test*, percobaan akan dilakukan menggunakan rasio 70:30, 75:25, dan 80:20.

## 5. Metodologi

Metode penelitian yang dilakukan dalam penyusunan laporan tugas akhir ini adalah eksperimen. Metodologi penelitian yang diterapkan dalam penulisan tugas akhir ini meliputi :

### 5.1. Studi Literatur

Pada tahap studi literatur, mempelajari literatur dan mengumpulkan data yang berkaitan dengan Tugas Akhir. Pengumpulan data tersebut dilakukan dengan beberapa cara diantaranya mencari referensi di perpustakaan maupun di internet yang berupa *paper*, jurnal, dan buku-buku.

### 5.2. Analisis dan Perancangan

Dalam usaha untuk mendapatkan data-data yang benar sehingga tercapai maksud dari tujuan penyusunan Tugas Akhir ini, penulis menggunakan metode pengumpulan data secara langsung dari buku Mukhtarul Hadits untuk hadits asli, dan referensi dari internet untuk hadits palsu. Dan juga penulis akan mengklarifikasi validitas hadits tersebut kepada orang yang kompeten di bidang hadits.

### 5.3. Implementasi / Coding

Implementasi metodologi penelitian ini terbagi menjadi beberapa langkah. Langkah pertama yaitu preprocessing, dimana setiap hadits akan diproses terlebih dahulu agar melalui tahap tokenizing, stemming, dan vectorizing. Langkah kedua yaitu menghitung bobot tiap fitur dengan menggunakan metode TF-IDF. Langkah terakhir adalah klasifikasi hadits menggunakan

metode SVM yang diimplementasikan oleh program berbasis Script dengan bahasa pemrograman Python.

#### **5.4. Pengujian**

Melakukan percobaan dengan variasi rasio data training antara 70:30, 75:25, dan 80:20 untuk melihat perbedaan tingkat akurasi dari sistem klasifikasi ini.

#### **5.5. Kesimpulan**

Membuat kesimpulan berdasarkan hasil dari penelitian, yaitu berupa tingkat akurasi klasifikasi hadits dengan menggunakan metode SVM.

### **6. Landasan Teori**

#### **6.1. Artificial Intelligence (AI)**

*Artificial Intelligence* (AI) merupakan sebuah teknik mengadaptasi pola pikir manusia atau tingkah laku manusia kepada sistem komputer. AI memiliki empat teknik dasar dalam pemecahan suatu masalah, yaitu: *Searching*, *Reasoning*, *Planning*, dan *Learning*. Dari masing-masing teknik tersebut terdapat berbagai macam metode yang dapat digunakan yang telah diusulkan oleh para pakar AI. Masing-masing metode memiliki kelebihan dan kekurangan. Sebuah metode yang bekerja sangat baik untuk permasalahan A belum tentu dapat digunakan untuk permasalahan B. Dengan kata lain, penggunaan metode ini sangat bergantung kepada permasalahan apa yang sedang terjadi [7].

##### **6.1.1. Machine Learning (ML)**

*Machine Learning* (ML) merupakan sebuah bidang pada *Computer Science* dalam memberikan kemampuan belajar kepada sebuah komputer tanpa diprogram secara eksplisit. Dalam mempelajari sebuah permasalahan, ML memerlukan sebuah model yang sudah didefinisikan berdasarkan parameter-parameter tertentu. Proses learning pada ML adalah proses dimana *program* mengoptimasi parameter-parameter pada model tersebut dengan memanfaatkan *data training* atau *past experiences* [8].

ML melibatkan berbagai macam bidang ilmu seperti statistika, ilmu komputer, matematika, dan bahkan neurologi. ML menggunakan teori-teori statistika untuk membuat model matematika. Model dapat bersifat *predictive* yaitu untuk memprediksi masa depan, *descriptive* yaitu untuk memperoleh pengetahuan baru dari data, atau bahkan gabungan dari keduanya. Inti dari ML adalah bagaimana membuat sebuah komputer dapat menyelesaikan berbagai macam permasalahan dengan mempelajari data (*past experiences*) sebagaimana manusia mempelajari sesuatu [8].

### 6.1.2. Supervised Learning

*Supervised Learning* adalah sebuah cara belajar dengan mencari hubungan (*mapping function*) dari *input data* ( $x$ ) yang sudah diberi *label* ( $y$ ). Permasalahan yang berkaitan dengan *supervised learning* dapat dikategorikan menjadi dua jenis, yaitu:

#### 1. Classification

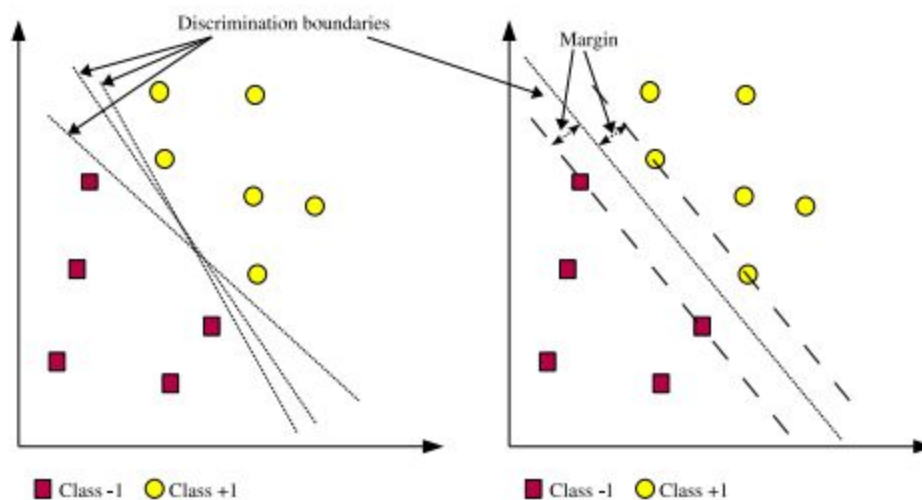
*Classification* bertujuan untuk mencari *outcome* dari *input* yang sudah diberikan, dimana output berupa kategori-kategori dari input data. Contoh: pria/wanita, tinggi/rendah, baik/buruk, dan sebagainya.

#### 2. Regression

*Regression* bertujuan untuk mencari *outcome* dari *input* yang sudah diberikan, dimana output berupa prediksi yang berbentuk nilai aktual. Contoh: tinggi badan seseorang, curah hujan, dan sebagainya [8].

### 6.1.3. Support Vector Machine (SVM)

*Support Vector Machine* (SVM) adalah salah satu algoritma *supervised learning* yang bekerja dengan baik untuk kasus klasifikasi teks yang memiliki banyak input berdasarkan pada teks sebagai fiturnya. Ide utama dari algoritma SVM adalah untuk membuat *hyper plane* yang memisahkan area menjadi beberapa bagian [6]. *Hyper plane* adalah sebuah garis yang memisahkan dua kategori klasifikasi.



Gambar 1. SVM berusaha menemukan *hyper plane* terbaik dalam memisahkan kedua *class -1* dan *+1* [9].

### 6.1.3.1. Linear Support Vector Machine

Seperti yang kita lihat pada Gambar 1, terlihat bahwa terdapat dua kelas yang dipisahkan oleh sebuah garis *hyper plane*. Masalah yang dihadapi dalam klasifikasi ini yaitu bagaimana menemukan garis (*hyper plane*) yang memisahkan antara kedua kelompok tersebut secara sempurna. Akan banyak alternatif garis yang memisahkan kedua kelas tersebut, namun *hyper plane* terbaik yang dapat memisahkan kedua kelas tersebut adalah *hyper plane* yang memiliki *margin* paling besar [9].

Margin adalah jarak antara hyper plane dengan vektor terdekat dari masing-masing kelas. Vektor yang paling dekat dengan hyper plane disebut dengan *support vector*. Pada Gambar 1 sebelah kanan dapat dilihat bahwa terdapat garis solid yang merupakan hyper plane terbaik, karena tepat berada di tengah antara kedua kelas. Usaha dalam mencari lokasi hyper plane ini merupakan inti dari proses pembelajaran pada SVM [9]

Setiap data dinotasikan sebagai  $x_i \in \langle D, i = 1, 2, \dots, N \rangle$ .  $N$  adalah banyaknya data. Kelas positif dinotasikan sebagai  $+1$ , dan kelas negatif dinotasikan sebagai  $-1$ . Dengan demikian, tiap data dan label kelas nya dinotasikan sebagai  $y_i \in \{-1, +1\}$  [9].

.Bila diasumsikan kedua kelas tersebut dipisahkan secara sempurna oleh *hyper plane* di  $D$ -dimensional feature space. *Hyper plane* tersebut dapat didefinisikan sebagai berikut:

$$w \cdot x_i + b = 0 \quad (1)$$

Data  $x_i$  yang tergolong kepada data kelas negatif adalah data yang memenuhi kriteria sebagai berikut:

$$w \cdot x_i + b \leq -1 \quad (2)$$

Data  $x_i$  yang tergolong kepada data kelas positif adalah data yang memenuhi kriteria sebagai berikut:

$$w \cdot x_i + b \geq 1 \quad (3)$$

Selanjutnya, masalah ini diformulasikan ke dalam *Quadratic Programming* (QP) problem, dengan meminimalkan persamaan 4 dibawah constraint persamaan 5.

Meminimalkan:

$$\|w\|^2 \quad (4)$$

\Bergantung pada:



$$y_i (w \cdot x_i + b) - 1 \geq 0, \forall i \quad (5)$$

Optimasi ini dapat diselesaikan dengan Lagrange Multipliers:

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i y_i (w \cdot x_i + b - 1) \quad (6)$$

$\alpha_i$  adalah *Lagrange multiplier* yang berkorespondensi dengan  $x_i$ . Nilai  $\alpha_i$  adalah nol atau positif. Untuk menyelesaikan masalah tersebut, pertama-tama meminimalkan  $L$  terhadap  $w$ , dan memaksimalkan  $L$  terhadap  $\alpha_i$ . Dengan memodifikasi persamaan 6, *maximization problem* di atas dapat direpresentasikan dalam  $\alpha_i$ :

Memaksimalkan:

$$\sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j x_i \cdot x_j \quad (7)$$

Bergantung pada:

$$\alpha_i \geq 0, \sum_{i=1}^N \alpha_i y_i = 0 \quad (8)$$

Solusi dari *problem* ini akan menghasilkan banyak  $\alpha_i$  dengan nilai nol. Data berkorespondensi dengan  $\alpha_i$  yang tidak nol, merupakan *support vectors*, yaitu data yang memiliki jarak terdekat dengan *hyper plane* [9].

### 6.1.3.2. Soft Margin

Penjelasan yang sudah diutarakan diatas merupakan asumsi bahwa hyper plane dapat memisahkan kedua kelas dengan sempurna. Akan tetapi, seringkali kita temukan bahwa kedua buah kelas tersebut tidak terpisah secara sempurna. Hal ini menyebabkan optimasi tidak dapat diselesaikan, karena tidak ada  $w$  dan  $b$  yang memenuhi pertidaksamaan 5 [9].

Maka dari itu, pertidaksamaan 5 dimodifikasi dengan memasukkan slack variable  $\xi_i$  ( $\xi_i \geq 0$ ), menjadi:

$$y_i (w \cdot x_i + b) \geq 1 - \xi_i, \forall i \quad (9)$$

Demikian juga dengan 4, sehingga diperoleh:

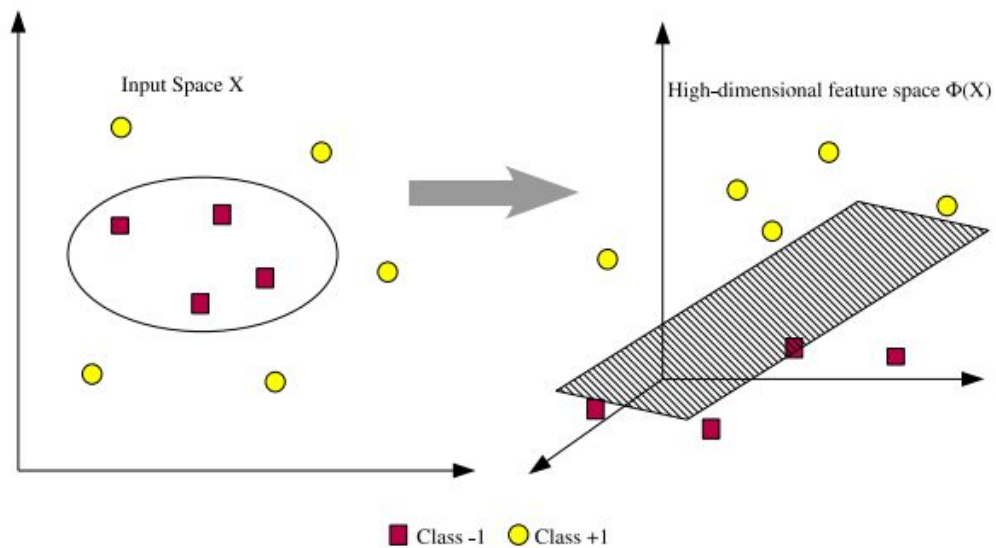
Minimize

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (10)$$

Parameter  $C$  bertugas mengontrol tradeoff antara margin dan *classification error*. Semakin besar nilai  $C$ , semakin besar penalty yang dikenakan untuk tiap *classification error* [9].

### 6.1.3.3. Non Linear Support Vector Machine

Seperti yang sudah dijelaskan diatas, SVM adalah salah satu bentuk dari linear machine, sehingga hanya dapat digunakan untuk menyelesaikan permasalahan-permasalahan linear. Namun tak jarang permasalahan yang kita hadapi bukanlah permasalahan linear (non-linear). Untuk mengatasi permasalahan non-linear, seluruh data yang ada pada ruang vektor awal harus dipindahkan ke ruang vektor baru yang memiliki dimensi lebih tinggi [9], seperti yang kita lihat pada Gambar 2:



Gambar 2. Memetakan data pada ruang vektor awal ke ruang vektor baru berdimensi lebih tinggi [9]

Selanjutnya adalah melakukan proses training sama seperti yang dilakukan pada Linear SVM. Proses optimasi pada fase ini memerlukan perhitungan dot product dua buah vektor  $x_i$  dan  $x_j$  pada ruang vektor baru yang dinotasikan sebagai  $\phi(x_i) \cdot \phi(x_j)$ . Teknik komputasi ini dikenal dengan sebutan Kernel Trick, yaitu menghitung dot product dua buah vektor di ruang vektor baru menggunakan komponen kedua buah vektor tersebut di ruang vektor asal [9].

Klasifikasi pada non linear SVM dapat dirumuskan sebagai berikut [9]:

$$f(\phi(x)) = w \cdot \phi(x) + b \quad (11)$$

Terdapat beberapa jenis fungsi yang dapat digunakan sebagai fungsi Kernel K, diantaranya adalah sebagai berikut:

Tabel 1. Fungsi kernel yang sering digunakan pada SVM [9]

Nama Kernel	Definisi
Polynomial	$K(x_i, x_j) = (x_i \cdot x_j + 1)^p$
Gaussian	$K(x_i, x_j) = \exp(-(\ x_i - x_j\ ^2 / 2\sigma^2))$
Sigmoid	$K(x_i, x_j) = \tanh(\alpha x_i \cdot x_j + \beta)$

$$= \sum_{i=1, x_i \in SV}^N \alpha_i y_i \phi(x) \cdot \phi(x_i) + b \quad (12)$$

$$= \sum_{i=1, x_i \in SV}^N \alpha_i y_i K(x, x_i) + b \quad (13)$$

*Support vector* SV adalah subset dari *data training*, dengan nilai  $\alpha$  tidak negatif  $\{x_i \mid \alpha_i \neq 0\}$  ( $i = 1, 2, \dots, l$ ) [9].

## 6.2. Hadits

Al-Qur'an dan Hadits merupakan inti dari ajaran Islam. Sunnah memainkan peran penting dalam pengembangan seluruh kehidupan manusia dan peradaban [10]. Hadits merupakan segala sesuatu yang berasal dari Nabi SAW baik itu berupa perbuatan, ucapan, persetujuan yang menjadi dasar pengambilan hukum dalam Islam [1].

Hadits dibagi menjadi dua jenis yaitu hadits *Mutawatir* dan hadits *Ahad*. Hadits *Mutawatir* adalah hadits yang keasliannya sudah tidak diragukan lagi, sedangkan hadits *Ahad* adalah hadits yang derajatnya belum mencapai level seperti hadits *Mutawatir* [2].

Derajat dari sebuah hadits dibagi menjadi tiga, yaitu Hadits *Shahih*, *Dhaif*, dan *Maudhu*. Hadits *Shahih* adalah hadits yang diriwayatkan oleh seorang rawi (periwayat) dengan *sanad* yang tidak terputus. Sedangkan hadits *Dhaif* yaitu hadits yang tidak memiliki syarat-syarat hadits *Hasan* (baik) dikarenakan hilangnya beberapa kondisi seperti *sanad*-nya terputus (ujung *sanad* tidak merujuk kepada Nabi SAW) atau terdapat faktor lain seperti *matan* (redaksi) hadits

itu sendiri. Selain itu ada juga yang disebut dengan hadits *Maudhu* (palsu), yaitu hadits yang dibuat oleh seseorang selain Nabi SAW dengan mengatasnamakan Nabi SAW [2].

## 7. Jadwal

Tugas akhir ini direncanakan dilaksanakan selama 4 bulan. Berikut adalah rincian waktu pelaksanaan.

Tabel 3.1. Jadwal Penelitian Tugas Akhir

No	Kegiatan	Minggu Ke-															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Studi literatur dan pengumpulan data																
2	Analisis dan perancangan																
3	Implementasi																
4	Pengujian																
5	Sidang																
6	Pembuatan Laporan																

## 8. Usulan Pembimbing

Berikut adalah pembimbing yang diusulkan untuk penelitian yang akan dilakukan:

1. Iwa Ovyawan Herlistiono, S.T., M.T.
2. Ardiles Sinaga, S.T., M.T.

## 9. Daftar Pustaka

- [1]. Terjemah Kitab Bulughul Maram: Kumpulan Hadits Hukum Panduan Hidup Muslim Sehari-hari. Al-Hafizh Ibnu Hajar Al-Asqalani, Jakarta 2016
- [2]. Najiyah, Ina, Sari Susanti, Dwiza Riana, Mochammad Wahyudi, "Hadith Degree Classification for Shahih Hadith Identification Web Based", STMIK Nusa Mandiri Jakarta, 2017.
- [3]. E., Wakil, "The Prophet's Treaty with the Christians of Najran: An Analytical Study to Determine the Authenticity of the Covenants", 2016 Journal of Islamic Studies, 2016
- [4]. H., Sayoud, "Author discrimination between the Holy Quran and Prophet's statements", Lit Linguist Computing, 2012
- [5]. C. Melchert, "Narrative Social Structure: Anatomy of the Ḥadīth Transmission Network", Journal of Islamic Studies, 2008
- [6]. B. Prasetyo, Agung, R. Rizal Isnanto, Dania Eridani, "Hoax Detection System on Indonesian News Sites Based on Text Classification using SVM and SGD", Proc. of 2017 4th Int. Conf. on Information Tech., Computer, and Electrical Engineering (ICITACEE), 2017
- [7]. Suyanto, "Artificial Intelligence: Searching, Reasoning, Planning, Learning - Revisi Kedua", Informatika, 2014
- [8]. Primartha, Rifkie, "Belajar Machine Learning Teori dan Praktik", Informatika, 2018
- [9]. Satriyo Nugroho, Anto, "Pengantar Support Vector Machine", Anto Satriyo Nugroho, 2007
- [10]. Khilwani Ibrahim, Nuzulha, Mohamad Fauzan Noordin, Suhaila Samsuri, "Isnad Al-hadith Computational Authentication: An Analysis Hierarchically", 6th International Conference on Information and Communication Technology for The Muslim World (ICT4M), 2016