

Identifikasi Jenis Hadits Menggunakan Beberapa Kombinasi Metode Learning

Nanang Fakhrrur Rozi¹, Entin Martiana K., M.Kom.², Nur Rosyid M., S.Kom.²

¹ Mahasiswa Jurusan Teknik Informatik PENS-ITS Surabaya
nfrozy@gmail.com

² Dosen Pembimbing Jurusan Teknik Informatika PENS-ITS Surabaya
{entin, rosyid}@eepis-its.edu

Abstrak

Hadis merupakan salah satu sumber hukum Islam yang cukup diperhatikan. Namun, banyak didapatkan hadis palsu maupun dhaif yang beredar luas di masyarakat, sehingga menimbulkan berbagai permasalahan yang terkadang sampai memuncak pada munculnya pemahaman-pemahaman yang tidak sesuai dengan syariat Islam. Penelitian tersebut biasanya dilakukan secara manual oleh para pakar di bidang ilmu hadis yang juga biasanya diajarkan di pondok pesantren atau perguruan tinggi Islam. Oleh karena itu, penelitian ini mencoba mengotomatiskan proses pengecekan keabsahan hadis tersebut dengan memanfaatkan metode machine learning 'Naïve Bayes' berdasarkan kaidah-kaidah yang ada dalam ilmu hadis, sehingga pada akhirnya dapat menghasilkan sebuah sistem yang mampu menjadi salah satu referensi awal dalam meneliti keabsahan hadis sebagai hujah sebelum akhirnya masuk pada proses-proses penelitian yang lebih jauh dan mendalam. Dilengkapi dengan fasilitas pencarian hadis dengan berbagai mode pencarian, salah satunya berdasarkan relevansi yang dibangun dengan metode Vector Space Model, membuat sistem ini dapat pula difungsikan sebagai aplikasi ensiklopedia hadis yang terdiri atas koleksi-koleksi hadis yang telah terklasifikasi tingkat kredibilitasnya.

Kata kunci: hadis, hujah, machine learning, klasifikasi, naive bayes, ensiklopedia, pencarian, vector space model

1. Pendahuluan

Melihat banyaknya hadis yang beredar, baik melalui tulisan maupun acara kajian dan diskusi publik, hal itu sering menjadikan masyarakat agak kebingungan untuk menerima atau tidak hadis yang diutarakan. Hal itu sebagai wacana karena masih sangat minimnya kajian hadis di Indonesia, yang pada akhirnya berakibat mudahnya hadis-hadis lemah dan palsu tersebar di Indonesia (Fathullah, 2010). Lebih spesifik, hal tersebut akhirnya dapat memicu perpecahan dan perselisihan.

Dengan membandingkan hadis dari berbagai sumber periwayatan untuk kemudian diteliti kredibilitasnya menggunakan ilmu-ilmu hadis (*'ulum al-hadits*) maka akan dapat diketahui tingkat kredibilitas hadis-hadis tersebut.

Berdasarkan hal tersebut, penelitian ini mencoba untuk mengembangkan sebuah ensiklopedia hadis berbahasa Indonesia yang memiliki kemampuan untuk mengidentifikasi secara otomatis dapat-tidaknya suatu hadis dijadikan sebagai hujah menurut ilmu-ilmu hadis yang ada dengan memanfaatkan disiplin ilmu *machine learning*.

2. Tentang Hadis

Hadis merupakan sumber berita yang datang dari Nabi SAW dalam segala bentuk, baik berupa perkataan, perbuatan, maupun sikap persetujuan (Khon, 2008:3). Sebagai dasar hukum kedua setelah Al-Quran bagi umat Islam (Khon, 2008:22), maka hadis penting untuk dijaga dan dimurnikan dari segala pemalsuan. Beberapa sebab terjadinya pemalsuan antara lain karena munculnya perpecahan umat yang sudah terasa pada masa-masa akhir kepemimpinan *khulafa ar-rasyidin* ke-4 ('Ali bin Abi Thalib r.a.) (Khon, 2008:200). Sehingga, para ulama di kalangan sahabat berusaha menjaga hadis dengan serius melalui berbagai cara, salah satunya dengan perlawatan ke berbagai daerah Islam (*rihlah Islamiyah*) untuk mengecek kebenaran hadis yang sampai kepada mereka baik dari segi *matan* maupun *sanad* (Khon, 2008:51). Karena ketersebaran itu, muncullah berbagai hadis yang datang dari berbagai sumber periwayatan.

Guna memfilter keotentikannya, para ulama melakukan riset terhadap hadis yang beredar dan meletakkan dasar kaidah atau peraturan ketat bagi periwayat hadis sehingga muncullah ilmu hadis. Ilmu hadis inilah yang menjadi dasar penelitian hadis oleh berbagai peneliti dan pengumpul hadis.

2.1 Bagian-Bagian Hadis

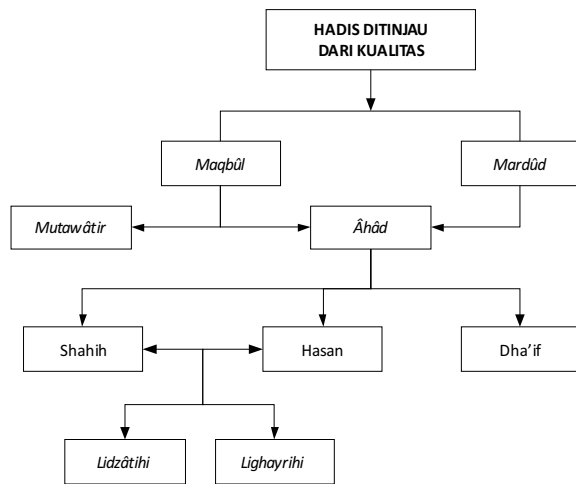
Setiap hadis terbagi atas dua bagian (Azami, 1977:32), yaitu sanad dan matan. Untuk lebih jelasnya, dapat dilihat contoh hadis yang dikeluarkan oleh Imam Bukhari r.a berikut.

Telah menceritakan kepada kami Sulaiman Abu Ar-Rabi', ia berkata bahwa Isma'il bin Ja'far bercerita kepadanya, ia berkata bahwa Nafi' bin Malik bin Abi 'Amir Abu Suhail bercerita kepadanya dari ayahnya dari Abu Hurairah dari Nabi SAW, beliau bersabda: "Tanda-tanda orang munafik ada tiga, bila berbicara dusta, bila berjanji tidak ditepati, dan bila diamanati dia berkhianat".

Hadis tersebut memuat urutan nama rawi hingga sampai kepada Nabi SAW yang disebut sebagai sanad (*isnad*). Sedangkan perkataan/informasi yang disampaikan (dalam tanda petik) merupakan matan atau isi hadis.

2.2 Pengelompokan Hadis

Jika dilihat dari berbagai tinjauan, hadis dapat dibagi ke dalam beberapa kelompok. Namun, dalam penelitian ini, pengelompokan hadis didasarkan pada kualitas sanad dan matan hadis, yang membagi hadis ke dalam dua kelompok (kejuhahan) hadis, yakni hadis *Maqbul* (diterima) dan *Mardud* (ditolak). Gambar 1 berikut menjelaskan skema pembagian hadis ditinjau berdasarkan kualitas sanad dan matan hadis (Khon, 2008:148).



Gambar 1. Pengelompokan Hadis Berdasarkan Kualitas

2.3 Penelitian Kredibilitas Hadis

Secara umum, terdapat dua cara (Athaya, 2007:2) yang bisa dilakukan untuk meneliti atau mengkritik kredibilitas (*maqbul/mardud*)—sering dikatakan oleh kebanyakan masyarakat sebagai *shahih/dha'if*—hadis:

1. menghukumi *zhahir* sanad saja tanpa menilai matannya;
2. menghukumi sanadnya secara *bathin*, dimana matan hadis juga dihukumi (dengan kata lain, menghukumi hadis secara keseluruhan).

Dalam penelitian ini, akan digunakan cara pertama dalam menilai kualitas hadis, yaitu berdasarkan *zhahir sanad* (perawi-perawi hadis). Kemudian, dalam meneliti kualitas atau sifat seorang perawi, baik dari segi keadilan maupun ke-*dhabit*-an perawi hadis, dalam penelitian ini akan digunakan kaidah yang terdapat dalam ilmu hadis yang disebut dengan ilmu *al-Jarh wa at-Ta'dil*.

3. Rancangan Sistem

Sistem terdiri atas dua fungsi utama, yaitu:

1. *mengidentifikasi hadis*, dalam hal ini menggunakan Naïve Bayes sebagai model klasifikasinya;
2. *mencari hadis*, yaitu dengan memanfaatkan model korelasi *Vector Space Model*.

3.1 Subsistem Identifikasi

Atribut Naïve Bayes yang digunakan adalah sebanyak 2 buah, yaitu berupa tingkat persambungan sanad dan kualitas sanad dihitung berdasarkan kaidah ilmu *Jarh-Ta'dil*.



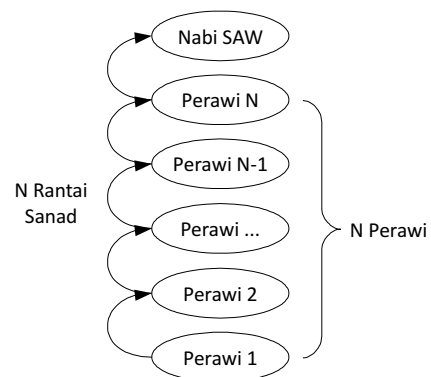
Gambar 2. Diagram Blok Sistem (Klasifikasi)

Nilai Atribut Persambungan Sanad

Dimulai dengan mendapatkan nilai *gp* (*faktor pengali sanad*) dengan formula:

$$gp = 1 / \text{JumlahPerawi}$$

Peningkatan nilai akan didapatkan berdasarkan jumlah mata rantai *sanad* yang tersambung.



Gambar 3. Persambungan Sanad

Dari gambar 3 tersebut, maka setiap hadis yang memiliki sejumlah *N* perawi, maka rantai

persambungan sanad yang harus terbentuk harus sejumlah N pula. Jika ada salah satu rantai sanad yang terputus, maka hal tersebut akan mengurangi nilai persambungan sanadnya. Sehingga didapatkan formula untuk menghitung nilai/grade persambungan sanadnya, yaitu:

$$\text{Grade}_{\text{Sanad}} = \text{JmlTerhubung} \times \text{gp}$$

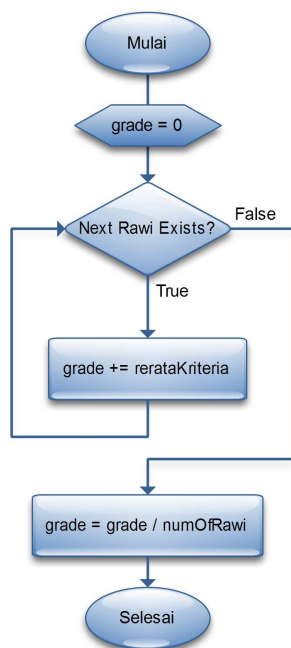
Sehingga, apabila diketahui terdapat 5 perawi ($\text{gp} = 0,2$) dan salah seorang perawi diketahui tidak bertemu secara langsung dengan perawi di atasnya (terputus satu mata rantai), maka grade persambungan sanadnya menjadi:

$$\text{Grade}_{\text{Sanad}} = \text{JmlTerhubung} \times \text{gp} = 4 \times 0,2 = 0,8$$

dari nilai tertinggi 1 yaitu ketika seluruh perawi bersambung secara berurutan.

Nilai Atribut Jarh-Ta`dil

Yaitu dengan menghitung penilaian sifat-sifat perawi berdasarkan data penilaian *jarh-ta`dil* (terdapat pada buku, hal 24-25). Prosedurnya sebagaimana gambar 4 berikut.



Gambar 4 Penghitungan Atribut Jarh-Ta`dil

3.2 Subsistem Identifikasi

Pencarian pada sistem ini menggunakan mekanisme pengindeksan 'indexing' disamping pencarian standar, yakni penelusuran berdasarkan filter standar basis data menggunakan *LIKE statement*. Sehingga, dalam hal ini, model pencarian memiliki dua tahapan utama, yakni pengindeksan dan pencarian itu sendiri. Proses pengindeksan kata akan dilakukan ketika terjadi proses identifikasi hadis yang juga melibatkan penyimpanan data hadis ke basis data. Sedangkan proses pencarian akan berjalan ketika pengguna memasukkan kata kunci pada laman Pencarian.

Adapun proses korelasi dilakukan dengan memanfaatkan metode *Vector Space Model*. Untuk contoh pengaplikasian metode, terlampir (buku, hal: 30-31).

4. Pengujian Sistem

Pengujian dilakukan pada dua subsistem:

4.1 Subsistem Identifikasi

Pada bagian ini dilakukan sebanyak 5 kali pengujian, dengan rincian:

1. pengujian terhadap data pembelajaran awal;
2. pengujian terhadap data uji (20 data, 10 hadis *maqbul* dan 10 hadis *mardud*), dengan tanpa melakukan adaptasi data;
3. pengujian terhadap data uji sebelumnya dengan melakukan adaptasi terhadap kesalahan identifikasi data;
4. pengujian terhadap data pembelajaran terbaru (dari data pembelajaran awal dan data uji yang telah masuk ke dalam basis data);
5. pengujian terhadap data yang belum lengkap data perawinya (10 data, 5 hadis *maqbul* dan 5 hadis *mardud*).

Pengujian proses identifikasi didasarkan pada 100 data hadis sebagai data awal pembelajaran, yang terdiri atas 50 hadis *maqbul* dan 50 hadis *mardud* yang bersumber dari berbagai kitab hadis.

Selain didasarkan pada data pembelajaran awal tersebut, pengujian juga dilakukan berdasarkan data-data pendukung lain (untuk proses kalkulasi atribut *Naïve Bayes*), yaitu berupa data periwayat hadis (445 data), yang terdiri atas nama perawi dan nomor generasi perawi hadis, termasuk data relasi antarperawi (552 data), data nama panggilan (nama lain/*laqab/kuniah*) rawi (84 data), data kriteria *jarh-ta`dil* (61 data), serta relasi data kriteria dengan perawi (1.123 data).

Pada pengujian nomor 1, galat yang diperoleh dari hasil identifikasi adalah sebesar 3% (3 data salah identifikasi) dari total 100 data (akurasi = 97%). Sedangkan pada pengujian nomor 2, galat yang muncul sebesar 55% dari total 20 data (akurasi = 45%). Kemudian, jika pengujian nomor 2 dibandingkan dengan pengujian nomor 3, nilai galat yang dihasilkan lebih kecil, yakni sebesar 45% (akurasi = 55%). Setelah dilakukan pengujian ulang terhadap data pembelajaran terbaru (pengujian nomor 4), didapatkan bahwa nilai akurasi menurun jika didasarkan pada pengujian nomor 1, yaitu dari 97% menjadi 95,83%. Hal ini diakibatkan oleh tambahan kesalahan identifikasi yang terjadi pada data pembelajaran baru (yang masuk ke basis data setelah diidentifikasi), meskipun pada data pembelajaran sebelumnya terdapat 1 (satu) data yang telah berhasil diidentifikasi dengan baik dimana pada pengujian nomor 1 merupakan data yang diidentifikasi dengan tidak tepat. Adapun ketika dilakukan pengujian nomor 5, yang dilakukan

dengan tanpa melengkapi data perawi pada setiap hadis yang diuji, galat yang muncul yakni sebesar 50% dari total 10 data. Hasil pengujian tersebut dapat dilihat pada tabel 1 berikut.

Tabel 1 Hasil Uji Coba Identifikasi

Nomor/Kode Pengujian	Jumlah Data Uji	Galat (%)	Akurasi (%)
1	100	3	97
2	20	55	45
3	20	45	55
4	120	4,17	95,83
5	10	50	50

Sehingga, dari beberapa hasil pengujian tersebut, dapat ditarik beberapa simpulan mengenai penyebab akurasi identifikasi, antara lain:

- variasi data pembelajaran yang tersimpan; dalam hal ini berhubungan pula dengan banyaknya data yang tersimpan sebagai data pembelajaran yang menghasilkan variasi data yang dapat mencerminkan fitur suatu kelompok data;
- keakuratan data (kelompok hadis) data pembelajaran, hal ini didasarkan pada peningkatan akurasi pada pengujian nomor 3 dari 45% (pengujian nomor 2) menjadi 55% setelah dilakukan adaptasi terhadap data pembelajaran yang salah identifikasi;
- kelengkapan data perawi yang digunakan dalam proses penghitungan nilai atribut Naïve Bayes, hal ini dilihat melalui pengujian nomor 5 yang memberikan kontribusi galat sebesar 50% dari 10 data yang diujikan.

4.2 Subsistem Pencarian

Pengujian proses pencarian berdasarkan relevansi diuji dengan cara memasukkan kata kunci dan membandingkan hasil keluaran dengan hasil yang ditargetkan. Dalam hal ini dilakukan pengujian berdasarkan 32 data hadis. Adapun rincian tema hadis yang tersimpan dalam basis data yaitu tentang hal-hal seputar shalat, puasa, zakat (*shadaqah*), dan lainnya.

Tabel 2 Hasil Uji Coba Pencarian

Percobaan ke-	Kata Kunci	Keterangan Keluaran
1	shalat subuh	6 hasil: 1 hadis shalat subuh, terletak pada urutan ke-1. 5 hadis lain tentang shalat.
2	waktu shalat subuh	6 hasil: 1 hadis shalat subuh, terletak pada urutan ke-1. 5 hadis lain tentang shalat.
3	waktu pelaksanaan shalat subuh	7 hasil: 1 hadis shalat subuh, terletak pada urutan ke-2. Urutan pertama ditempati oleh hadis tentang pelaksanaan haji dan tidak berhubungan dengan shalat.
4	waktu pelaksanaan shalat subuh subuh	7 hasil: 1 hadis shalat subuh, terletak pada urutan ke-1. Hadis tentang haji berada pada urutan ke-2.

Hasil uji coba memberikan hasil yang relevan ketika pemilihan kata kuncinya benar-benar tepat ataupun dengan mencantumkan pengulangan kata sehingga akan diperoleh keluaran yang memiliki kedekatan dengan kata ganda tersebut.

5. Simpulan

1. keakuratan proses identifikasi kejujuran hadis yang didasarkan pada *zhahir sanad* sangat dipengaruhi oleh kelengkapan data perawi hadis pembentuk rantai sanad hadis;
2. keakuratan identifikasi juga sangat dipengaruhi oleh variasi data pembelajaran, sebagaimana sifat dari model klasifikasi Naïve Bayes;
3. pencarian data dengan menggunakan metode *vector space model* dapat memberikan hasil yang cukup akurat, asalkan kata kunci tetap dipilih secara baik.

Daftar Pustaka

- Athaya, Abu 'Umar U. *Kaidah Emas di Dalam Mengetahui Riwayat Hadits Shahih dan Dha'if*. (Tanpa Tahun). Terjemahan oleh Abu Salma al-Atsari. (2007). Maktabah Abu Salma al-Atsari.
- Azami, M. M. 1977. *Studies in Hadith Methodology and Literature*. Indianapolis: Islamic Teaching Center.
- Fathullah, L. 18 Juli, 2010. *Kajian Ilmu Hadis Masih Minim*. Republika, Rubrik Wawancara "Idealnya, Seorang Ulama itu Harus Menguasai Berbagai Disiplin Ilmu".
- Khon, Abdul Majid. 2008. *Ulumul Hadis*. Jakarta: Amzah.