

1. Pendahuluan

Latar Belakang

Penelitian akan keabsahan data hadits merupakan salah satu bagian penting dalam studi hadits. Saat ini, jumlah data hadits yang terekam cukup besar, diperkirakan antara empat ribu hingga tiga puluh ribu[1], dengan keabsahan dari masing-masing data hadits masih perlu didiskusikan. Untuk memeriksa keabsahan hadits, suatu hadits perlu diperiksa dari kelengkapan rantai sanad, yang menceritakan rantai urutan periwayatan, serta konsistensi konten matan (isi dari hadits) dengan hadits yang lain. Setiap kelompok ulama yang mempelajari hadits memiliki kriteria masing-masing dalam menentukan keabsahan tersebut. Terkadang satu kriteria yang digunakan memiliki perselisihan dengan kriteria yang lain, dan tidak semua kriteria telah diaplikasikan pada seluruh data hadits yang tercatat[2].

Besarnya jumlah data hadits, serta pertimbangan bahwa tidak semua kriteria telah diaplikasikan ke suatu data hadits, menyebabkan adanya alat bantu untuk mengklasifikasi keabsahan hadits berdasarkan kriteria tertentu dinilai bisa membantu studi keabsahan hadits. Seiring berkembangnya teknologi informasi, teknologi untuk menyusun alat bantu tersebut sudah bermunculan. Dan mengingat bahwa tidak semua kriteria penilaian telah terdokumentasi dengan baik, alat prediksi semacam *machine learning* dinilai dapat membantu penyusunan alat bantu tersebut. Yang menjadi masalah adalah metode *machine learning* masih belum bisa dijamin hasilnya dan penelitian untuk implementasi *machine learning* untuk memprediksi klasifikasi keabsahan data hadits masih perlu dilanjutkan, untuk mencapai hasil yang lebih optimal dan dapat dijamin[3].

Pada sejumlah penelitian terkait pemrosesan data hadits sebelumnya, telah dilakukan penelitian oleh Altammami, dkk. [4] untuk memisahkan sanad dan matan dengan menggunakan n-gram. Selain itu, telah dilakukan penelitian untuk melakukan klasifikasi keabsahan dengan metode machine learning yang dilakukan oleh Aldhlan, dkk [5][6] dengan metode *decision tree* untuk mengklasifikasi hadits ke dalam empat kelas, yaitu “shahih”, “hasan”, “dhaif”, dan maudhu. Performa pengklasifikasi Hadits hasil penelitian tersebut memiliki Correct Classification Rate akhir 97.597%. Haque, dkk. [7] juga melakukan klasifikasi keabsahan sanad dengan metode sentiment analysis dan machine learning, dengan hasil akurasi tertinggi dengan menggunakan Linear SVC dengan nilai 79,5%.

Pada penelitian ini, akan dilakukan representasi fitur hadits, dengan nama lengkap sebagai token-nya, ke dalam bentuk n-gram. Representasi fitur tersebut kemudian diujikan ke dalam sistem machine learning untuk diperiksa performa dan pengaruhnya dalam mengklasifikasi keabsahan sanad hadits.

Topik dan Batasannya

Ada beberapa macam metode dalam membentuk model *machine learning*. Metode yang cukup sering digunakan antara lain Naïve Bayes, *decision tree*, dan Support Vector Machine. Masing-masing metode memiliki kelebihan dan kekurangannya masing-masing, dan kelebihan yang menyebabkan ketiga metode tersebut cenderung sering digunakan adalah karena kesederhanaannya dalam implementasi, dan performanya yang baik dalam sejumlah penelitian. Ketiga metode tersebut juga termasuk metode *machine learning* yang paling sering digunakan untuk penelitian klasifikasi teks.

Pada penelitian ini, dipilih studi sanad hadits untuk implementasi *machine learning* karena bentuknya yang terstruktur dan relatif konsisten pada berbagai sumber kumpulan hadits. Salah satu masalah dalam studi sanad hadits adalah berpengaruhnya urutan penutur hadits serta kelengkapan rantai penutur hadits dalam menentukan keabsahan hadits tersebut. Dari ketiga metode *machine learning* yang telah disebutkan, hanya metode *decision tree* yang mempertimbangkan urutan. Selain itu, untuk mempertimbangkan kelengkapan rantai penutur hadits, ketiga metode tersebut memerlukan bantuan khusus. Metode bantuan yang akan dilakukan dalam percobaan ini adalah dengan merepresentasikan fitur ke dalam bentuk n-gram.

Metode n-gram merupakan metode dalam *Natural Language Processing* berbasis *Machine learning*, yang memperhatikan urutan yang n-buah kata dalam kalimat, yang dalam kasus ini, metode n-gram akan digunakan untuk mengelompokkan n-buah penutur. Metode ini dirasa dapat menggambarkan keterkaitan antar penutur hadits dalam bentuk yang sederhana.

Dengan merepresentasikan fitur ke dalam bentuk n-gram, diharapkan dapat membantu model berbasis ketiga metode tersebut untuk memperoleh hasil yang lebih optimal. Pada penelitian ini, akan diteliti pengaruh

penggunaan fitur n-gram untuk membantu sistem yang disusun berdasarkan masing-masing metode *machine learning* tersebut.

Tujuan

Tujuan dari penelitian ini adalah untuk menyusun sistem berbasis *machine learning* untuk mengklasifikasi hadits ke dalam kategori “shahih”, “hasan” dan “dhaif”, yang memanfaatkan representasi fitur ke dalam bentuk n-gram. Dari hasil penggunaan sistem akan diperiksa performa dari sistem klasifikasi pada masing-masing metode *machine learning* yang telah ditentukan, yaitu Naïve Bayes, SVM dan *decision tree* dan kemudian diperiksa pengaruh penggunaan fitur n-gram dalam melakukan klasifikasi tersebut.

Organisasi Tulisan

Pada jurnal ini akan disampaikan sejumlah studi terkait yang telah dilakukan, penjelasan mengenai metode percobaan yang digunakan, analisis dari hasil percobaan yang dilakukan, serta kesimpulan yang didapatkan dari hasil percobaan.