

Klasifikasi Anjuran, Larangan dan Informasi pada Hadis Sahih Al-Bukhari berdasarkan Model Unigram menggunakan *Artificial Neural Network* (ANN)

Eliza Riviera Rachmawati Jasin¹, Said Al-Faraby², Adiwijaya³

Prodi S1 Teknik Informatika, School of Computing, Telkom University

¹eliza@jasin.net, ²saidalfaraby@telkomuniversity.ac.id, ³adiwijaya@telkomuniversity.ac.id

Abstrak

Umat Islam memiliki kepercayaan pedoman dalam hidupnya yang merujuk kepada Al-quran dan Hadits. Ada beberapa hadits yang sering kita dengar, namun para ulama sepakat hadits yang diriwayatkan oleh Imam Bukhari berpegang teguh pada tingkat kesahihan paling tinggi. Untuk itu banyak umat islam yang mempelajari dan mendalami ilmu hadits. Dalam upaya pembelajaran ilmu hadits dalam era yang serba teknologi ini tentunya tidak luput dari perkembangan dunia digital ini. Oleh karena itu penulis ingin melakukan sebuah penelitian klasifikasi anjuran, larangan, dan informasi pada Hadits Sahih Al-Bukhari yang diharapkan dapat mengembangkan aplikasi yang ada dan membantu masyarakat untuk lebih mudah mempelajari ilmu hadits. Klasifikasi akan terbagi menjadi tiga buah kelas, yaitu hadits yang merupakan kalimat anjuran, larangan, dan informasi. Metode yang akan digunakan dalam klasifikasi ini ialah menggunakan n-gram model dan *Artificial Neural Network* (ANN). Metode tersebut dipilih berdasarkan penelitian yang telah dilakukan sebelumnya mengenai klasifikasi hadis menggunakan ANN yang di lakukan oleh M. Arshi Shaloot, Norisma Idris, Rohana Mahmud yang menghasilkan hasil akurasi sebesar 94%. Pada penelitian ini dilakukan beberapa scenario pengujian dengan memodifikasi tahapan *preprocessing*, *feature extraction*, dan parameter ANN. Pengujian tersebut menghasilkan hasil evaluasi terbesar yaitu 85% dengan menggunakan precision, recall, dan f1-score.

Kata kunci : *Klasifikasi teks, Hadits Sahih Al-Bukhari, Unigram, ANN*

Abstrack

Muslims have confidence in their guidance that refers to the Qur'an and Hadith. There are some traditions that we often hear, but the Islamic scholar agreed hadith narrated by Imam Bukhari has the highest validity. Therefore, many Muslims who study and explore the science of hadith. In an effort to study the science of hadith in the era of the all the technology is certainly not miss from the the digital world. Therefore, the authors wanted to do a study about suggestion, prohibition, and information of Hadith Sahih Al-Bukhari which is expected to develop existing applications and help people to more easily learn the science of hadith. Classification will be divided into three classes, suggestion, prohibition, and information sentences. The method will be used in this classification is unigram analysis for the feature extraction and Artificial Neural Network (ANN). The method is chosen based on research that has been done before by M. Arshi Shaloot, Norisma Idris, Rohana Mahmud about hadith classification using Artificial Neural Network (ANN) with 94% result for accuracy. In this research, we used several scenario by adding some modification in preprocessing, feature extraction, and ANN parameter steps. The test resulted the greatest evaluation is 85% using precision, recall, and f1-score

Keyword : *Text classification, Hadith Sahih Al-Bukhari, Unigram, ANN*

1. Pendahuluan

Salah satu pencerahan yang dibawa oleh Islam bagi kemanusiaan adalah pemikiran secara ilmiah, masyarakat Arab dan Timur Tengah pra Islam tidak memperdulikan persoalan-persoalan mengenai alam semesta, bagaimana alam tercipta dan bagaimana alam bekerja, maka dari sinilah mereka belajar merenungi pertanyaan-pertanyaan ini dan untuk mencari jawaban tentang itu

semua, mereka merujuk kepada Al-quran dan Hadis [1]. Ada beberapa hadis yang sering kita dengar, namun para ulama sepakat hadis yang diriwayatkan oleh Imam Bukhari berpegang teguh pada tingkat kesahihan paling tinggi dan tidak akan turun dari tingkat tersebut, kecuali terhadap beberapa hadis yang bukan merupakan materi pokok dari sebuah bab [2].

Untuk itu banyak umat islam yang mempelajari dan mendalami ilmu hadis. Dalam upaya pembelajaran ilmu hadis dalam era yang serba teknologi ini tentunya tidak luput dari perkembangan dunia digital ini. Kumpulan hadis-hadis Islam sangat mudah sekali untuk dijumpai di internet, baik berbentuk *file*, ataupun artikel dalam sebuah laman. Begitu juga dengan mobile application dan desktop application yang membantu umat Islam untuk mempelajari ilmu hadis sudah banyak sekali dan tidak jarang orang menggunakannya. Bukan hal yang aneh apabila banyak sekali aplikasi-aplikasi teknologi yang bermunculan dengan berbagai kekurangan dan kelebihannya yang menawarkan kemudahan untuk umat Islam dalam mempelajari ilmu agama karena hal tersebut sangat diperlukan di era modern ini.

Oleh karena itu penulis ingin melakukan sebuah penelitian klasifikasi anjuran, larangan dan informasi pada Hadis Sahih Al-Bukhari yang diharapkan dapat mengembangkan aplikasi yang ada dan membantu masyarakat untuk lebih mudah mempelajari ilmu hadis. Klasifikasi akan terbagi menjadi tiga buah kelas, yaitu hadis yang merupakan kalimat anjuran, hadis yang merupakan kalimat larangan, dan hadis yang merupakan kalimat informasi.

Metode yang akan digunakan dalam klasifikasi ini ialah *Artificial Neural Network (ANN)* berdasarkan model unigram. Metode tersebut dipilih berdasarkan penelitian yang telah dilakukan sebelumnya oleh Fawaz Al Zaghoul dan Sami Al Dhaheri mengenai *Arabic text classification* menggunakan *Artificial Neural Network (ANN)* [3] dan juga penelitian mengenai klasifikasi hadis menggunakan ANN yang dilakukan oleh M. Arshi Shaloot, Norisma Idris, Rohana Mahmud [4] yang menghasilkan hasil akurasi sebesar 94%. Seperti *classifier* yang sudah sering digunakan yaitu *Naïve Bayes, classifier* ini tidak berusaha memahami arti sebuah kalimat, namun mencoba untuk mengklasifikasikannya. Pada penelitian ini ada sedikit modifikasi yang akan dilakukan yaitu perbandingan metode ANN menggunakan model unigram dengan model bigram dan trigram.

2. Landasan Teori dan Perancangan

2.1. Hadis Sahih Al-Bukhari

Umat Islam mempercayai bahwa Al-Qur'an dan Hadits adalah pegangan umat Islam untuk menuntun kepada kehidupan yang lebih baik, agar tidak tersesat selama di dunia. Ada beberapa hadits yang diriwayatkan, diantaranya Hadits Bukhari, Hadits Muslim, Hadits Tirmidzi, Hadits Nasa'I, Hadits Ibnu Majah, dan Hadit Ibnu Daud.

Para ulama menyepakati hadits yang diriwayatkan oleh Imam Bukhari berpegang teguh pada tingkat kesahihan paling tinggi dan tidak akan turun dari tingkat tersebut, kecuali terhadap beberapa hadits yang bukan merupakan materi pokok dari sebuah bab [2]. Hal ini dianggap oleh sebagian besar umat Islam sebagai pegangan agaman yang paling otoritatif setelah Al-Qur'an. Bahkan Hadits Bukhari ini sekarang dijadikan sebagai patokan keunggulan ilmiah dalam bidang studi hadits [5].

Menurut Ibnu Sholah, hadits sahih harus memenuhi syarat sebagai berikut [6]:

1. Kandungan isinya tidak bertentangan dengan Al-Qur'an.
2. Harus bersambung sanadnya.
3. Diriwayatkan oleh orang / perawi yang adil.
4. Diriwayatkan oleh orang yang dhobit (kuat ingatannya).
5. Tidak syadz (tidak bertentangan dengan hadits lain yang lebih shahih).
6. Tidak cacat walaupun tersembunyi.

Terkait dengan penelitian ini, terdapat salah satu penelitian sebelumnya yang juga mengangkat tema mengenai ensiklopedia hadis yang dilakukan oleh Nanang Fakhur Rozi. Penelitian tersebut berjudul **"Identifikasi Jenis Hadits Menggunakan Beberapa Kombinasi Metode Learning"** dengan berdasarkan kualitas sanad dan matan yang membagi hadits ke dalam dua kelompok yaitu hadis *Maqbul* (diterima) dan *Mardud* (ditolak) untuk mengecek keabsahan sebuah hadis [7]. Sanad merupakan jalan matan (isi hadits baik berupa sabda maupun perbuatan Nabi Muhammad Shallallahu 'Alaihi Wasallam yang diceritakan oleh sahabat) dari Nabi Muhammad Shallallahu 'Alaihi Wasallam sampai kepada orang yang

mengeluarkan (mukhrij) hadits itu atau mudawwin (orang yang menghimpun atau membukukan) hadits [6].

2.2. Klasifikasi Dokumen

Klasifikasi dokumen adalah masalah pembelajaran mendasar yang merupakan bagian utama dari banyak manajemen informasi dan pengambilan tugas [8]. Dalam konteks pembangunan, klasifikasi dokumen memainkan peran penting karena beberapa aplikasi untuk mengatur, mengelompokkan, mencari dan meringkas membutuhkan informasi ini [8]. Penggunaan klasifikasi ini banyak sekali digunakan oleh banyak peneliti untuk mengetahui berbagai opini, atau statistik dari sekumpulan data *tex*. Klasifikasi dokumen dapat digunakan dalam berbagai jeni dokumen, salah satunya dokumen *tex*.

Beberapa penelitian yang telah mengggukan klasifikasi dokumen diantaranya penelitian yang berjudul **“Tagging documents using neural networks based on local word features”** dengan menggunakan *Artificial Neural Network (ANN)* sebagai *classifier*, penelitian *local word* sebagai *feature extraction* untuk menentukan *tagging/label* dari masing-masing dokumen [9]. Klasifikasi dokumen ini banyak memanfaatkan metode-metode *machine learning* dalam proses pengklasifikasiannya.

Ada beberapa metode klasifikasi yang biasa digunakan dalam klasifikasi dokumen baik berupa *text* maupun *speech*, antara lain *multinomial naive bayes* pada penelitian berjudul **“Aspect-based sentiment analysis to review products using Naïve Bayes”** [11] dan **“Klasifikasi Topik pada Lirik Lagu dengan Metode Multinomial Naive Bayes”** [12], penelitian menggunakan *bayessian network* yaitu **“Learning Struktur Bayesian Networks menggunakan Novel Modified Binary Differential Evolution pada Klasifikasi Data”** [13], dan penelitian menggunakan *Hidden Markov Model* **“Fuzzy Hidden Markov Models for Indonesian Speech Classification”** [14]

2.3. Klasifikasi Hadis

Terkait dengan penelitian ini, klasifikasi hadis dibagi menjadi tiga kelompok, yaitu anjuran, larangan dan informasi. Beberapa contoh dari masing-masing kelompok tersebut adalah sebagai berikut.

1. Klasifikasi Anjuran

Pada penelitian ini terdapat tiga buah klasifikasi, diantaranya adalah klasifikasi anjuran. Salah satu contoh data yang termasuk dalam kategori anjuran adalah sebagai berikut.

“Berbicaralah dengan manusia sesuai dengan kadar pemahaman mereka, apakah kalian ingin jika Allah dan rasul-Nya didustakan?” –Hadis Bukhari 124

2. Klasifikasi Larangan

Pada penelitian ini terdapat tiga buah klasifikasi, diantaranya adalah klasifikasi larangan. Salah satu contoh data yang termasuk dalam kategori larangan adalah sebagai berikut.

“Janganlah kalian kembali menjadi kafir, sehingga kalian saling membunuh satu sama lain.” – Hadis Bukhari 118

3. Klasifikasi Informasi

Pada penelitian ini terdapat tiga buah klasifikasi, diantaranya adalah klasifikasi informasi. Salah satu contoh data yang termasuk dalam kategori informasi adalah sebagai berikut.

“Seorang muslim adalah orang yang Kaum Muslimin selamat dari lisan dan tangannya, dan seorang Muhajir adalah orang yang meninggalkan apa yang dilarang oleh Allah.” – Hadis Bukhari 9

2.4. N-gram Model

Tidak dapat dipungkiri bahwa dalam membuat sebuah kalimat, mendengar, dan membaca manusia tanpa disadari telah melakukan sebuah prediksi dari satu per satu kata yang telah dibacanya dan membuat kesimpulan arti dari kalimat tersebut. Hal ini juga yang dilakukan oleh mesin translator. Mesin translator akan melihat satu atau lebih dari satu kata dari kalimat yang di *inputkan* dan mengambil kesimpulan arti dari kalimat tersebut.

Model probabilistik dibutuhkan dalam membuat sebuah prediksi kata pada sebuah kalimat. Pada pengurutan kata, model statistic yang digunakan biasa disebut language models or LMs. Salah satu tekniknya adalah dengan menggunakan N-gram. Nilai N pada N-gram tergantung pada berapa kata terurut yang ingin diambil. N=1 biasa disebut unigram contohnya seperti “*please*”, “*turn*”, “*your*”, “*homework*”, untuk N=2 disebut *bigram* contohnya seperti “*please turn*”, “*turn your*”, “*your homework*”, untuk N=3 disebut *trigram* contohnya “*please turn your*”, “*turn your homework*” [11].

Contoh dari pemodelan n-gram model berkaitan dengan penelitian klasifikasi hadis ini adalah sebagai berikut.

Tabel 1. Contoh model n-gram

Contoh hadis	<s> akan datang dari keturunanku suatu kaum mereka melepaskan din mereka </s>
Model unigram	['akan', 'datang', 'dari', 'keturunanku', 'suatu', 'kaum', 'mereka', 'melepaskan', 'din', 'mereka']
Model bigram	['akan <s>', 'datang akan', 'dari datang', 'keturunanku dari', 'suatu keturunanku', 'kaum suatu', 'mereka kaum', 'melepaskan mereka', 'din melepaskan', 'mereka din']
Model trigram	['datang <s>,akan', 'dari akan,datang', 'keturunanku datang,dari', 'suatu dari,keturunanku', 'kaum keturunanku,suatu', 'mereka suatu,kaum', 'melepaskan kaum,mereka', 'din mereka,melepaskan', 'mereka melepaskan,din']

Terkait dengan penelitian klasifikasi dokumen teks menggunakan n-gram model, terdapat salah satu penelitian yang dilakukan oleh M. Ghiassi, J. Skinner, and D. Zimbira yang berjudul “**Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic Artificial Neural Network**” pada penelitian tersebut n-gram digunakan sebagai ekstrasi fitur [12].

2.5. TF-IDF

Term Frequency Inverse Document Frequency (TF-IDF) merupakan salah satu metode *feature extraction* yang banyak digunakan dalam proses pengklasifikasian data teks. TF-IDF memiliki persamaan sebagai berikut.

$$w_{ij} = tf \times idf = tf_{ij} \times \log \frac{D}{df_i}$$

Dimana :

w_{ij} = bobot kata t_j terhadap dokumen d_i .

tf_{ij} = jumlah kemunculan kata t_j dalam d_i .

D = jumlah dokumen

df_i = jumlah kemunculan term dalam D

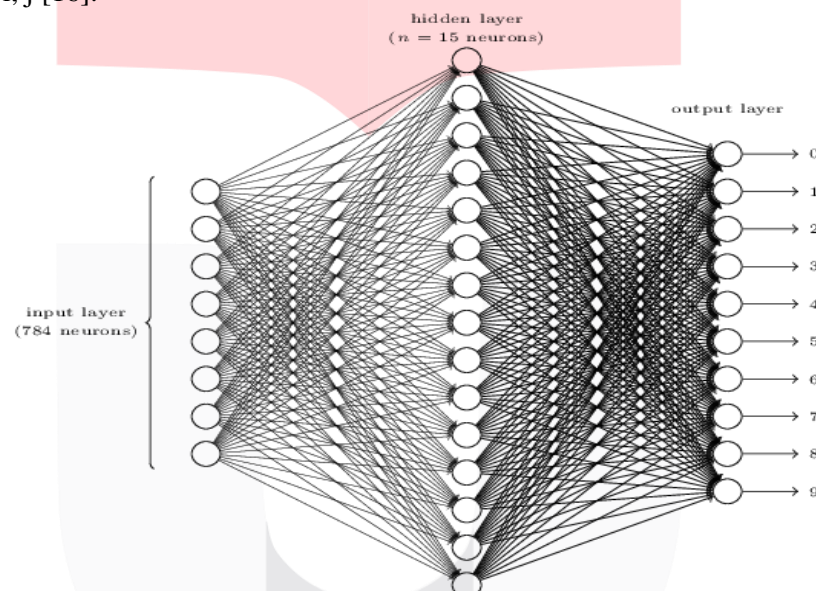
Ada beberapa penelitian terkait yang memanfaatkan TD-IDF sebagai metode ekstraksi fitur, salah satunya adalah yang dilakukan oleh Cahyo Darujati dan Agustinus Bimo yang

berjudul “Pemanfaatan Teknik Supervised untuk Klasifikasi Teks Bahasa Indonesia” [10].

2.6. Artificial Neural Network

Hingga kini belum ada definisi ANN secara universal yang bisa diterima oleh semua pakar secara universal. Beberapa ahli memiliki definisi masing-masing mengenai ANN, menurut DARPA, *neural network* adalah sistem yang terdiri dari banyak proses elemen sederhana yang beroperasi parallel berfungsi untuk menentukan struktur jaringan, kekuatan koneksi, dan proses akan keluar pada elemen komputasi atau nodes[13]. Menurut Zurada, *artificial neural systems* atau *neural networks* merupakan *physical cellular systems* yang bisa memperoleh, menyimpan, dan memanfaatkan *experiential knowledge*[14]. Menurut Nigrin, jaringan syaraf adalah sirkuit yang terdiri dari proses elem sederhana yang sangat besar yang berbasis neurologi[15].

Jaringan syaraf tiruan merupakan urutan *triple* (N, V, w) dengan dua himpunan N, V dan fungsi w , di mana N adalah himpunan neuron dan V aset $\{(i, j) \mid i, j \in N\}$ yang elemennya merupakan koneksi antara neuron i dan neuron j . Fungsi $w: V \rightarrow R$ simpul. Jaringan = neuron + bobot koneksi, dimana $w((i, j))$, bobot terkoneksi antara neuron i dan neuron j , disingkat menjadi $w_{i, j}$ [16].



Gambar 1. Arsitektur Artificial Neural Network [17]

Artificial Neural Network mempunyai struktur yang tersebar parallel yang sangat besar dan mempunyai kemampuan belajar, sehingga bisa melakukan generalisasi. Artinya, ANN bisa menghasilkan output yang benar untuk input yang belum pernah dipelajari (dilatih kepadanya). Dengan kedua kemampuan pemrosesan informasi ini, ANN mampu menyelesaikan berbagai permasalahan kompleks berskala besar[18].

Salah satu penelitian yang menggunakan metode ANN sebagai *classifier* pada klasifikasi dokumen ialah penelitian yang berjudul “**Arabic Tex Classification Based on Features Reduction Using Artificial Neural Networks**” yang dilakukan oleh F. A. Zaghoul dan S. Al-Dhaheri [3] dan “**Hadith Data Mining and Classification: a Comparative Analysis**” yang dilakukan oleh M. Arshi Shaloot, Norisma Idris, Rohana Mahmud [4] yang menghasilkan hasil akurasi sebesar 94%.

3. Pengujian dan Analisis

3.1. Tujuan Pengujian

Sistem yang dibangun pada tugas akhir ini adalah sistem yang mampu mengenali kategori dari masing-masing hadis, apakah hadis tersebut termasuk hadis anjura, larangan, atau informasi. Oleh karena itu, pengujian pada sistem dilakukan agar mendapatkan hasil

ekstraksi fitur atau atribut yang baik bagi *classifier*, sehingga dapat mencapai tujuan penelitian ini.

3.2. Data Set

Data set yang digunakan pada penelitian ini yaitu sebanyak 1650 data, dengan proporsi yang seimbang, yaitu 550 data anjuran, 550 data larangan, dan 550 data informasi.

3.3. Skenario Pengujian

Berdasarkan karakteristik data yang telah didapatkan dari eksplorasi data, maka scenario pengujian dilakukan untuk mengobservasi pengaruh dari perbedaan scenario dengan performa sistem yang dihasilkan.

Adapun scenario pengujian pada penelitian menggunakan *classifier* ANN ini adalah sebagai berikut.

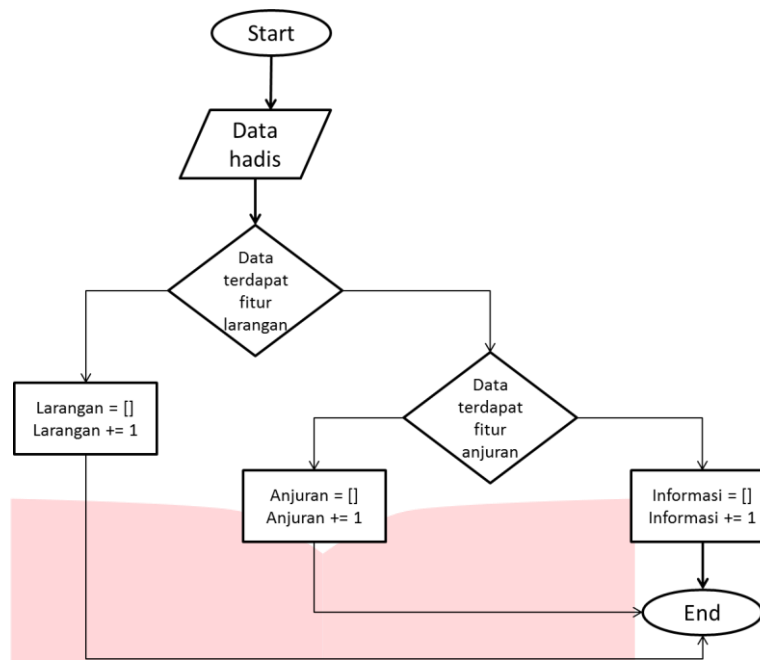
1. Model n-gram dengan menggunakan *punctuation removal*, *case folding*, stopword removal, lemmatization, dengan bag of words sebagai *feature extraction* dan 10 *hidden layer* (masing-masing 10 unit neuron) pada arsitektur ANN
2. Model unigram dengan menggunakan *punctuation removal*, *case folding*, stopword removal, lemmatization, dengan TF-IDF sebagai *feature extraction* dan 10 *hidden layer* (masing-masing 10 unit neuron) pada arsitektur ANN
3. Model unigram dengan menggunakan *punctuation removal*, *case folding*, stopword removal, lemmatization, non-KBBI removal dengan bag of words sebagai *feature extraction* dan 10 *hidden layer* (masing-masing 10 unit neuron) pada arsitektur ANN

Model unigram tanpa menggunakan *punctuation removal*, *case folding*, stopword removal, lemmatization, dengan bag of words sebagai *feature extraction* dan 20 *hidden layer* (masing-masing 10 unit neuron) pada arsitektur ANN

3.4. Hasil Pengujian dan Evaluasi

3.4.1. Pengujian dan Evaluasi menggunakan *Pattern Matching*

Selain melakukan penelitian pengujian menggunakan metode klasifikasi ANN, pengujian juga menggunakan *pattern matching* untuk menguji sistem yang dibangun. Pengujian ini bertujuan untuk memastikan bahwa untuk melakukan klasifikasi terhadap data teks hadis ini tidak cukup apabila hanya menggunakan *pattern matching*. Hal ini terjadi karena ada beberapa hadis yang memiliki makna ambigu. Secara keseluruhan logika dari sistem *pattern matching* yang dibangun adalah seperti pada flow chart dibawah ini.



Gambar 2. Skenario *pattern matching*

Pada gambar di atas, dapat dilihat terdapat prioritas pada proses pengklasifikasian menggunakan *pattern matching* ini. Data pertama di cek apakah merupakan data larangan atau tidak, apabila iya data termasuk kelas larangan, apabila tidak maka data di cek kembali apakah data tersebut mengandung string kata anjuran atau tidak, apabila iya maka data termasuk kelas anjuran, apabila tidak maka data akan masuk ke kelas informasi.

Beberapa contoh string yang merepresentasikan kelas anjuran dan larangan adalah sebagai berikut.

Tabel 2. Contoh string unik kelas anjuran dan larangan

String unik kelas anjuran		String unik kelas larangan
Bershadaqahlah	Shalatlal	Jangan
Hendaklah	Berwudhulah	Larang
Ajaklah	Bersihkanlah	Tidak boleh
Beritahukanlah	Sujudlah	
Mulailah	Rukuklah	
Lakukanlah	Takbirlah	
Berpuasalah	Jagalah	
Permudahlah	Jawablah	

Berdasarkan sistem klasifikasi yang dibangun menggunakan *pattern matching*, hasil evaluasi yang didapatkan adalah sebesar **69,35%**. Apabila dilakukan analisis dari data hadis, ditemukan beberapa hadis yang memiliki fitur-fitur yang berorientasi pada kelas lain, hal tersebut menyebabkan ketidak tepatan pada proses prediksi kelas. Contoh dari hadis tersebut adalah sebagai berikut.

- a. Dibawah ini merupakan data hadis yang memiliki string anjuran dan larangan.

Tabel 3. Contoh hadis yang memiliki string anjuran dan larangan

Nabi Shallallahu ‘alaihi wasallam bersabda: "Seandainya tiba kepada kita harta dari negeri Bahrain aku pasti memberikan kepadamu sekian, sekian dan sekian". namun harta dari Bahrain tidak kunjung datang hingga Nabi Shallallahu ‘alaihi wasallam wafat. Ketika harta dari Bahrain datang Abu Bakar memerintahkan dan berseru:
--

"Siapa yang telah dijanjikan sesuatu atau dihutangi oleh Nabi Shallallahu 'alaihi wasallam **hendaklah** menemui kami". Maka aku mendatangnya dan aku katakan bahwa Nabi Shallallahu 'alaihi wasallam telah berkata kepadaku begini dan begitu, lalu ia (Abu Bakar) memberiku setangkep, lalu aku menghitungnya ternyata ia berjumlah lima ratus, lalu ia berkata; "Ambillah dua kali lagi seperti itu".

"Pada hari Perang Uhud, bapakku didatangkan dalam kondisi sudah terbunuh dengan bagian anggota badannya ada yang terpotong hingga diletakkan di hadapan Rasulullah Shallallahu'alaihiwasallam, sedangkan jasadnya sudah ditutup dengan kain. Maka aku menghampiri untuk membukanya namun kaumku mencegahku. Aku coba sekali lagi untuk membukanya namun kaumku tetap mencegahku hingga akhirnya Rasulullah Shallallahu'alaihiwasallam memerintahkan untuk diangkat (dibawa). Saat itu Beliau mendengar ada suara teriakan. Maka Beliau bertanya: "(Suara) siapakah itu?". Orang-orang menjawab: "Putri dari 'Amru" atau "saudara perempuan 'Amru. Kemudian Beliau berkata,: "Mengapa kamu menangis?" atau "**Janganlah** kamu menangis, karena malaikat senantiasa akan menaunginya dengan sayap-sayapnya hingga (jenazah) ini diangkat".

- b. Dibawah ini merupakan data hadis yang memiliki string anjuran dan larangan sekaligus.

Tabel 4. Contoh hadis yang memiliki string anjuran dan larangan sekaligus

Rasulullah Shallallahu 'alaihi wasallam ditanya tentang pakaian yang dibolehkan dipakai oleh seorang yang sedang berihram. Maka Beliau bersabda: "**Janganlah** dia mengenakan gamis (baju), sorban, celana, mantel (pakaian yang menutupi kepala) dan jangan pula dia memakai pakaian yang diberi minyak wangi atau wewangian yang terbuat dari daun tumbuhan dan apabila dia tidak memiliki sepasang sandal, **hendaklah** dia mengenakan sepatu tapi dipotongnya hingga berada dibawah mata kaki".

Jika salah seorang dari kalian masuk ke dalam WC untuk buang hajat, maka **janganlah** menghadap ke arah kiblat membelakanginya. **Hendaklah** ia menghadap ke arah timurnya atau baratnya.

Apa ini? Orang-orang menjawab: Tali ini milik Zainab, bila dia shalat dengan berdiri lalu merasa letih, dia berpegangan tali tersebut. Maka Nabi bersabda: **Jangan** ia lakukan sedemikian itu. **Hendaklah** seseorang dari kalian tekun dalam ibadah shalatnya dan apabila dia merasa letih, **shalatlah** sambil duduk.

Jika salah seorang dari kalian meludah **janganlah** ia meludah ke arah kiblat atau ke sebelah kanannya, tapi **hendaklah** ia lakukan ke arah kiri atau di bawah kaki kirinya.

Janganlah kalian melaksanakan puasa wishal, maka siapa dari kalian yang mau melakukan puasa wishal **hendaklah** dia melakukannya hingga (makan) sahur. Orang-orang berkata: "Bukankah anda sendiri melakukan puasa wishal, wahai Rasulullah?" Beliau bersabda: "Aku tidak sama dengan keadaan seorang kalian karena saat aku tidur akan ada pemberi makan yang datang kepadaku lalu memberi aku makan dan datang pemberi minum lalu memberi aku minum".

- c. Dibawah ini merupakan data hadis yang memiliki string anjuran namun bila dilihat secara intuisi, hadis ini merupakan hadis larangan.

Tabel 5. Contoh hadis larangan yang memiliki string anjuran

Barangsiapa berdusta terhadapku maka **hendaklah** ia persiapkan tempat duduknya di neraka.

3.4.2. Pengujian dan Evaluasi menggunakan *Artificial Neural Network (ANN) classifier*

a. Skenario I

Pada scenario ini dilakukan tiga kali percobaan, yaitu dengan menggunakan model unigram, bigram dan trigram. Pada model unigram, didapatkan hasil sebagai berikut.

Tabel 6. Hasil pengujian unigram dengan 10 *hidden layer* dan 10 unit neuron

	Precision	Recall	F1-Score	Jumlah
Informasi	0.91	0.72	0.80	110
Anjuran	0.81	0.79	0.80	110
Larangan	0.70	0.85	0.77	110
Avg/total	0.80	0.79	0.79	330

Dapat dilihat dari tabel di atas, hasil pengujian dari model unigram dengan menggunakan *punctuation removal*, *case folding*, *stopword removal*, *lemmatization*, dengan bag of words sebagai *feature extraction* dan 10 *hidden layer* (masing-masing 10 unit neuron) pada arsitektur ANN menghasilkan hasil evaluasi terbesar pada presision, yaitu **0.80 atau 80%**. Artinya skenario ini berjalan dengan baik untuk klasifikasi hadis. Hal ini dapat disebabkan karena model unigram tidak membuat data kehilangan arti aslinya karena modelnya bag of words.

Pada model n-gram lainnya, yaitu bigram pengujian juga melakukan percobaan, dengan parameter yang sama dengan model unigram. Hasil dari pengujian tersebut dipaparkan pada tabel di bawah ini.

Tabel 7. Hasil pengujian bigram dengan 10 *hidden layer* dan 10 unit neuron

	Precision	Recall	F1-Score	Jumlah
Informasi	0.57	0.94	0.71	110
Anjuran	0.82	0.43	0.56	110
Larangan	0.90	0.75	0.82	110
Avg/total	0.77	0.71	0.70	330

Pada percobaan dengan menggunakan model bigram, didapatkan hasil evaluasi yang maksimal dengan precision yaitu sebesar **0.77 atau 77%**. Hasil ini dibawah nilai evaluasi dari percobaan menggunakan model unigram, akan tetapi masih lebih baik daripada hasil pengujian menggunakan pattern matching. Artinya skenario ini masih berjalan dengan baik untuk klasifikasi hadis. Hal ini dapat disebabkan karena model bigram membuat data kehilangan arti aslinya karena modelnya diambil dari dua buah kata yang posisinya sudah tidak berurutan karena adanya stopwords.

Percobaan selanjutnya dengan n-gram model yaitu dengan menggunakan model trigram, hasil percobaan tersebut dijabarkan pada tabel dibawah ini.

Tabel 8. Hasil pengujian trigram dengan 10 *hidden layer* dan 10 unit neuron

	Precision	Recall	F1-Score	Jumlah
Informasi	0.45	0.94	0.60	110
Anjuran	0.66	0.40	0.50	110
Larangan	0.72	0.21	0.32	110
Avg/total	0.61	0.52	0.48	330

Pada model trigram, hasil evaluasi semakin menurun, yaitu yang paling maksimal pada precision, sebesar **0.61 atau 61%**. Artinya skenario ini kurang berjalan dengan baik untuk klasifikasi hadis. Sama seperti bigram, hal ini dapat disebabkan karena modelnya diambil dari lebih dari satu kata buah kata yang posisinya sudah tidak berurutan karena adanya stopwords.

b. Skenario II

Pada scenario ini, yaitu model unigram tanpa menggunakan *punctuation removal*, *case folding*, stopword removal, lemmatization, dengan TF-IDF sebagai *feature extraction* dan 10 *hidden layer* (masing-masing 10 unit neuron) pada arsitektur ANN didapatkan hasil pengujian seperti yang dipaparkan pada tabel dibawah ini.

Tabel 9. Hasil pengujian unigram menggunakan TF-IDF dengan 10 *hidden layer* dan 10 unit neuron

	Precision	Recall	F1-Score	Jumlah
Informasi	0.90	0.63	0.74	110
Anjuran	0.72	0.80	0.76	110
Larangan	0.73	0.87	0.80	110
Avg/total	0.78	0.77	0.76	330

Dapat dilihat pada hasil pengujian di atas, scenario ini menghasilkan evaluasi terbesar di angka **0.78 atau 78%** hasil ini lebih baik dari *pattern matching* akan tetapi masih dibawah evaluasi scenario pertama yang menggunakan TF sebagai ekstrasi fiturnya. Artinya skenario ini masih berjalan dengan baik untuk klasifikasi hadis. Hal ini dapat disebabkan karena dengan TF-IDF data tidak memperdulikan urutan dari kata-kata tersebut.

c. Skenario III

Pada skenario ini, yaitu dengan menggunakan model unigram tanpa menggunakan *punctuation removal*, *case folding*, stopword removal, lemmatization, non-KBBI removal dengan bag of words sebagai *feature extraction* dan 10 *hidden layer* (masing-masing 10 unit neuron) pada arsitektur ANN didapatkan hasil percobaan seperti yang dipaparkan pada tabel di bawah ini.

Tabel 10. Hasil pengujian unigram menggunakan non-KBBI removal dengan 10 *hidden layer* dan 10 unit neuron

	Precision	Recall	F1-Score	Jumlah
Informasi	0.83	0.77	0.80	110
Anjuran	0.76	0.76	0.76	110
Larangan	0.79	0.84	0.81	110
Avg/total	0.79	0.79	0.79	330

Dapat dilihat pada skenario III ini hasil evaluasi pada presicion, recall, dan f1-score memiliki hasil evaluasi yang sama, yaitu **0.79 atau 79%**. Artinya skenario ini masih berjalan dengan baik untuk klasifikasi hadis. Hal ini dapat disebabkan karena dengan menggunakan stopword KBBI ada beberapa data yang hilang tanpa memperdulikan arti sebenarnya apakah atribut itu penting atau tidak.

d. Skenario IV

Pada scenario ini, yaitu model unigram tanpa menggunakan *punctuation removal*, *case folding*, stopword removal, lemmatization, dengan bag of words sebagai *feature extraction* dan 20 *hidden layer* (masing-masing 10 unit neuron) pada arsitektur ANN didapatkan hasil pengujian seperti yang di jabarkan pada tabel di bawah ini.

Tabel 11. Hasil pengujian unigram menggunakan dengan 20 *hidden layer* dan 10 unit neuron

	Precision	Recall	F1-Score	Jumlah
Informasi	0.86	0.89	0.88	110
Anjuran	0.79	0.84	0.81	110
Larangan	0.90	0.82	0.86	110
Avg/total	0.85	0.85	0.85	330

Dapat dilihat pada tabel di atas, hasil dari pengujian dengan mengubah jumlah *hidden layer* mendapatkan hasil evaluasi sebesar **0.85 atau 85%** pada precision, recall dan juga f1-score. Artinya skenario ini berjalan dengan baik untuk klasifikasi hadis. Hal ini dapat disebabkan karena pengaruh model arsitektur dari *classifier*.

Daftar Pustaka

- [1] S. Qutub, "Sumber-sumber Ilmu Pengetahuan dalam Al Qur'an dan Hadits," *Binus Univ.*, vol. 2, no. 2, pp. 1339–1350, 2011.
- [2] "Sejarah Singkat Imam Bukhari," *Kumpulan Hadist Bukhari Muslim*, 31-Jan-2008. .
- [3] F. A. Zaghoul and S. Al-Dhaheri, "Arabic Text Classification Based on Features Reduction Using Artificial Neural Networks," in *2013 UKSim 15th International Conference on Computer Modelling and Simulation*, 2013, pp. 485–490.
- [4] M. A. Saloot, N. Idris, R. Mahmud, S. Ja'afar, D. Thorleuchter, and A. Gani, "Hadith data mining and classification: a comparative analysis," *Artif. Intell. Rev.*, vol. 46, no. 1, pp. 113–128, Jun. 2016.
- [5] S. O. Subedar, *Commentary On Sahih Al-Bukhari*, vol. 1. Bukhari Publications.
- [6] S. Efendi, *Ringkasan Shahih Bukhari*. .
- [7] N. F. Rozi, E. M. Kusumaningtyas, and N. R. Muhtada'i, "Identifikasi Jenis Hadits Menggunakan Beberapa Kombinasi Metode Learning," *EEPIS Final Proj.*, 2011.
- [8] R. Power, J. Chen, T. K. Kuppusamy, and L. Subramanian, "Document Classification for Focused Topics.," in *AAAI Spring Symposium: Artificial Intelligence for Development*, 2010.
- [9] A. P. Azcarraga, P. Tensuan, and R. Setiono, "Tagging documents using neural networks based on local word features," in *2014 International Joint Conference on Neural Networks (IJCNN)*, 2014, pp. 724–731.
- [10] "PEMANFAATAN TEKNIK SUPERVISED UNTUK KLASIFIKASI TEKS BAHASA INDONESIA (PDF Download Available)," *ResearchGate*. [Online]. Available: https://www.researchgate.net/publication/303304818_PEMANFAATAN_TEKNIK_SUPERVISED_UNTUK_KLASIFIKASI_TEKS_BAHASA_INDONESIA. [Accessed: 26-Jul-2017].
- [11] M. S. Mubarak, Adiwijaya, and M. D. Aldhi, "Aspect-based sentiment analysis to review products using Naïve Bayes," *AIP Conf. Proc.*, vol. 1867, no. 1, p. 20060, Aug. 2017.
- [12] R. A. Aziz, M. S. Mubarak, and A. Adiwijaya, "Klasifikasi Topik pada Lirik Lagu dengan Metode Multinomial Naive Bayes," in *Indonesia Symposium on Computing (IndoSC) 2016*, 2016.
- [13] "Learning Struktur Bayesian Networks menggunakan Novel Modified Binary Differential Evolution pada Klasifikasi Data," *ResearchGate*. [Online]. Available: https://www.researchgate.net/publication/308831480_Learning_Struktur_Bayesian_Networks_menggunakan_Novel_Modified_Binary_Differential_Evolution_pada_Klasifikasi_Data. [Accessed: 23-Aug-2017].
- [14] U. N. Wisesty, M. S. Mubarak, and Adiwijaya, "A classification of marked hijaiyah letters' pronunciation using hidden Markov model," *AIP Conf. Proc.*, vol. 1867, no. 1, p. 20036, Aug. 2017.
- [15] D. Jurafsky and J. H. Martin, *Speech and Language Processing (2Nd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2009.

- [16] M. Ghiassi, J. Skinner, and D. Zimbra, "Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network," *Expert Syst. Appl.*, vol. 40, no. 16, pp. 6266–6282, Nov. 2013.
- [17] D. N. N. Study (U.S.), *DARPA Neural Network Study*. AFCEA Intl, 1988.
- [18] J. M. Zurada, *Introduction to artificial neural systems*. West, 1992.
- [19] "Book Review: Neural Networks for Pattern Recognition by Albert Nigrin (MIT Press, 1993)," *SIGART Bull*, vol. 5, no. 2, pp. 11–12, Apr. 1994.
- [20] A. Waibel, A. A. Alshehri, and S. Ezekiel, "Multi-perspective anomaly prediction using neural networks," in *2013 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, 2013, pp. 1–6.
- [21] gk_, "Text Classification using Neural Networks," *Machine Learnings*, 26-Jan-2017. [Online]. Available: <https://machinelearnings.co/text-classification-using-neural-networks-f5cd7b8765c6>. [Accessed: 05-Jul-2017].
- [22] Suyanto, *Soft Computing Membangun Mesin ber-IQ tinggi Edisi Cetakan Pertama*. Informatika Bandung, 2011.
- [23] W. E. Deming and S. L. Morgan, "Preface to the Second Edition," *Data Handl. Sci. Technol.*, vol. 11, p. XV, 1993.
- [24] P. Mishra and P. Lotia, "Comparative Performance Analysis of SVM Speaker Verification System using Confusion Matrix."