

**Klasifikasi *Multi-Label* pada Hadis Bukhari dalam  
Terjemahan Bahasa Indonesia Menggunakan *Information  
Gain* dan *Backpropagation Neural Network***

**Tugas Akhir**  
**diajukan untuk memenuhi salah satu syarat**  
**memperoleh gelar sarjana**  
**dari Program Studi Teknik Informatika**  
**Fakultas Informatika**  
**Universitas Telkom**

**1301141071**  
**Muhammad Yuslan Abubakar**



**Program Studi Sarjana Teknik Informatika**  
**Fakultas Informatika**  
**Universitas Telkom**  
**Bandung**

**2018**

# LEMBAR PENGESAHAN

**Klasifikasi *Multi-Label* pada Hadis Bukhari dalam Terjemahan Bahasa Indonesia  
Menggunakan *Information Gain* dan *Backpropagation Neural Network***

***Multi-Label Topic Classification for Hadith of Bukhari in Indonesian Translation using  
Information Gain and Backpropagation Neural Network***

**NIM: 1301141071**

**Muhammad Yuslan Abubakar**

Tugas akhir ini telah diterima dan disahkan untuk memenuhi sebagian syarat memperoleh  
gelar pada Program Studi Sarjana Teknik Informatika

Fakultas Informatika

Universitas Telkom

Bandung, 29 Juni 2018

Menyetujui

Pembimbing I

Pembimbing II

Prof. Dr. Adiwijaya, S.Si., M.Si.

NIP: 00740046

Said Al Faraby, S.T., M.Sc.

NIP: 15890019

Ketua Program Studi  
Sarjana Teknik Informatika,

Said Al Faraby, S.T., M.Sc.

NIP: 15890019

## LEMBAR PERNYATAAN

Dengan ini saya, Muhammad Yuslan Abubakar, menyatakan sesungguhnya bahwa Tugas Akhir saya dengan judul ” **Klasifikasi *Multi-Label* pada Hadis Bukhari dalam Terjemahan Bahasa Indonesia Menggunakan *Information Gain* dan *Backpropagation Neural Network*** ” beserta dengan seluruh isinya adalah merupakan hasil karya sendiri, dan saya tidak melakukan penjiplakan yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan. Saya siap menanggung resiko/sanksi yang diberikan jika dikemudian hari ditemukan pelanggaran terhadap etika keilmuan dalam buku TA atau jika ada klaim dari pihak lain terhadap keaslian karya.

Bandung, 29 Juni 2018

Yang Menyatakan,

Muhammad Yuslan Abubakar

# Klasifikasi *Multi-Label* pada Hadis Bukhari dalam Terjemahan Bahasa Indonesia Menggunakan *Information Gain* dan *Backpropagation Neural Network*

Muhammad Yuslan Abubakar<sup>1</sup>, Adiwijaya<sup>2</sup>, Said Al Faraby<sup>3</sup>

<sup>1,2,3</sup>Fakultas Informatika, Universitas Telkom, Bandung

<sup>1</sup>muhammadyuslan@students.telkomuniversity.ac.id, <sup>2</sup>adiwijaya@telkomuniversity.ac.id,

<sup>3</sup>saidalfaraby@telkomuniversity.ac.id

---

## Abstrak

Hadis merupakan sumber hukum dan pedoman kedua bagi umat Islam setelah Al-Qur'an dan banyak sekali hadis yang telah diriwayatkan oleh para ahli hadis selama ini. Penelitian ini membangun sebuah sistem yang dapat melakukan klasifikasi hadis Bukhari terjemahan berbahasa Indonesia. Topik ini diangkat untuk memenuhi kebutuhan umat Islam dalam mengetahui apa saja informasi mengenai anjuran dan larangan yang terdapat dalam suatu hadis. *Backpropagation neural network* digunakan karena metode ini dapat melakukan klasifikasi dengan jumlah fitur yang sangat banyak dan beragam. Dibantu dengan *information gain* sebagai metode seleksi dalam memilih fitur-fitur yang berpengaruh untuk setiap label kelas pada data *multi-label* dan *single label* yang mana hal ini belum pernah dilakukan sebelumnya. Hasil yang didapatkan menunjukkan bahwa sebesar 88.42% data hadis *multi-label* dapat diklasifikasikan dengan benar. Sementara itu sebesar 65.275% dapat diklasifikasikan dengan benar pada data hadis *single label* menggunakan bantuan *information gain* dalam pemilihan fitur.

**Kata kunci :** hadis, *backpropagation neural network*, *information gain*, *multi-label*, seleksi fitur, klasifikasi teks

---

## Abstract

Hadith is the second source of law and guidance for Moslems after the Qur'an and there are so many hadith that have been narrated by the expert of hadith so far. This research builds a system that can classify shahih hadith of Bukhari in Indonesian language. This topic is raised to meet the needs of importance for Moslems so they know about the suggestions and restrictions contained in a hadith. Backpropagation neural network is chosen because this method can do classification with a huge number of varied features. Assisted with information gain as feature selection in order to select influential features for each class label on multi-label data and single label which has never been done before. The results show that 88.42% of multi-label hadith data can be correctly classified. Meanwhile, 65.275% of the single label hadith data can be classified correctly using information gain as feature selection.

**Keywords:** hadith, backpropagation neural network, information gain, multi-label, feature selection, text classification

---

## 1. Pendahuluan

Hadis merupakan sumber hukum kedua setelah Al-Qur'an sebagai pedoman hidup umat Islam [1]. Banyak sekali hadis yang telah diriwayatkan oleh para ahli hadis dan telah dikelompokkan kedalam beberapa kategori untuk memudahkan orang lain memahami hadis-hadis tersebut. Salah satu ahli hadis yang telah meriwayatkan banyak hadis dari Rasulullah SAW adalah Bukhari.

Hadis digunakan untuk membantu umat Islam khususnya dalam memahami ilmu-ilmu agama Islam. Banyak hadis-hadis yang menganjurkan kita untuk berbuat dalam kebaikan. Ada juga yang melarang dalam hal berbuat dosa kepada Allah SWT. Untuk dapat mengetahui hal tersebut, dibutuhkan suatu sistem untuk mengidentifikasi apakah hadis tersebut bersifat anjuran, larangan atau sekedar informasi saja. Dari hal itu, timbulah pertanyaan bagaimana sebuah hadis terjemahan berbahasa Indonesia dapat diklasifikasikan kedalam bentuk yang lebih khusus yaitu berupa hadis anjuran, hadis larangan dan hadis informasi. Suatu hadis bisa saja merupakan salah satu jenis dari ketiga kelas tersebut namun tidak menutup kemungkinan bahwa termasuk gabungan dari ketiganya.

Namun saat ini belum ditemukan adanya penelitian yang melakukan klasifikasi hadis kedalam beberapa kelompok klasifikasi atau yang biasa disebut dengan *multi-label classification*. Hal ini diperlukan mengingat bahwa

suatu hadis tidak hanya mengandung informasi saja melainkan adanya pesan khusus yang ingin disampaikan seperti bentuk anjuran dan larangan mengenai suatu topik keilmuan tertentu.

Atas dasar itulah penulis akan memfokuskan penelitian dalam membuat model klasifikasi berdasarkan hadis Bukhari sebanyak 1064 data hadis kedalam bentuk klasifikasi *multi-label*. Hadis-hadis tersebut telah diberikan label sebelumnya dengan tujuan untuk melakukan proses pelatihan pada sistem. *Information Gain* dipilih sebagai seleksi fitur karena metode ini bekerja dengan cara memberikan nilai kepada setiap fitur dan hanya akan dipilih fitur-fitur yang berkorelasi terhadap kelas yang sesuai. Adapun metode yang akan digunakan adalah *Backpropagation Neural Network* untuk menyelesaikan masalah yang telah dipaparkan sebelumnya. Metode *Backpropagation Neural Network* dipilih berdasarkan penelitian yang telah dilakukan sebelumnya oleh Min-Ling Zhang dan Zhi-Hua Zhou mengenai *multi-label Neural Networks with Applications to Functional Genomics and Text Categorization* [13]. Hasil yang didapatkan dari penelitian itu menunjukkan bahwa metode *Backpropagation Neural Network* dapat melakukan klasifikasi teks *multi-label* lebih baik dibandingkan dengan beberapa metode lain. Fitur yang telah didapatkan pada proses seleksi dan ekstraksi fitur/ciri akan digunakan sebagai masukan (*input*) pada sistem klasifikasi *Backpropagation Neural Network*. Sementara itu, hasil yang didapatkan dari proses klasifikasi merupakan label kelas dari masing-masing data.

Beberapa masalah yang dibatasi pada penelitian ini adalah penggunaan *dataset* sebesar 1064 data hadis terjemahan berbahasa Indonesia dengan tiga label yaitu anjuran, larangan dan informasi. Proporsi data untuk masing-masing kelas juga tidak merata, terbukti dengan adanya satu kelas yang memiliki jumlah data sekitar 75% dari keseluruhan data. Sulitnya mencari data dengan 3 jenis label tersebut juga menjadi salah satu hambatan dalam penelitian ini.

Metode evaluasi diperlukan untuk mengukur seberapa baik sistem yang telah dibangun dan diterapkan pada penelitian ini. *Hamming Loss* dipilih sebagai metrik evaluasi untuk mengukur performansi dari hasil klasifikasi *multi-label* yang telah didapatkan. *Hamming Loss* merupakan salah satu metrik evaluasi yang umum digunakan untuk pengujian data *multi-label* dengan cara menghitung *error rate* dari hasil klasifikasi.

Beberapa bagian selanjutnya pada jurnal ini adalah sebagai berikut. Pada bagian kedua, terdapat studi terkait yang menjelaskan mengenai penelitian sebelumnya yang berkaitan dengan penelitian yang sedang dikerjakan saat ini serta beberapa tinjauan pustaka terkait penelitian. Pada bagian ketiga, akan dijelaskan mengenai sistem yang akan dibangun pada penelitian ini yaitu klasifikasi pada hadis Bukhari terjemahan berbahasa Indonesia dengan menggunakan metode *Information Gain* dan *Backpropagation Neural Network*. Kemudian pada bagian keempat, akan dipaparkan hasil pengujian dan analisis hasil yang telah didapatkan. Selanjutnya pada bagian kelima, akan dijelaskan mengenai kesimpulan yang didapatkan dari penelitian ini.

## 2. Studi Terkait

Hadis dapat diklasifikasikan ke banyak kelompok khusus. Seperti yang telah dijelaskan pada [1], hadis dapat dikategorikan sebagai hadis shahih, maudho', dhaif dan hasan. Hadis juga dapat dikategorikan sebagai hadis anjuran, hadis larangan dan hadis informasi seperti yang dijelaskan oleh Jasin [3] dan Kusumaningrum [4]. Klasifikasi hadis kedalam beberapa kelompok diperlukan untuk bisa mendapatkan informasi yang lebih khusus dari hadis tersebut. Dengan mengetahui kelompok suatu hadis, diharapkan pencarian informasi terkait hadis tersebut dapat dengan mudah dilakukan.

Penelitian terkait klasifikasi hadis dalam terjemahan berbahasa Indonesia sebelumnya adalah untuk melakukan klasifikasi apakah hadis tersebut merupakan hadis kuat (shahih) atau lemah. Sejauh ini klasifikasi hadis dengan kelas hadis informasi, hadis larangan dan hadis anjuran telah dilakukan pada penelitian [3] dan [4]. Penelitian tersebut melakukan klasifikasi hanya kedalam satu kelas dari tiga kelas yang ada atau dengan kata lain model klasifikasi yang dibangun adalah klasifikasi *multiclass*. Adapun kebanyakan penelitian serupa yang sebelumnya telah dilakukan masih dalam bentuk Bahasa Arab.

Kawther A. Aldhlan [1], telah melakukan suatu mekanisme untuk meningkatkan performansi pada klasifikasi hadis. Metode yang digunakan adalah *Decision Tree* karena menurutnya merupakan pendekatan yang tepat untuk klasifikasi hadis dikarenakan kemudahan induksi aturan dan hasil interpretasi. Metode ini juga digunakan karena memiliki kemampuan yang dapat mengatasi nilai (*value*) yang hilang namun usaha yang dibutuhkan untuk mencapai hal itulah yang dianggap sebagai kelemahannya. Mengabaikan nilai yang hilang ini dapat menyebabkan keputusan yang kritis seperti kesalahan klasifikasi hadis. Untuk mengatasi hal ini, mereka menggunakan suatu mekanisme yang disebut dengan *Missing Data Detector* (MDD).

Sedangkan Al-Kabi [6], menggunakan metode TF-IDF dengan bantuan Microsoft Visual Basic Programming Language. Alasan penggunaan bahasa pemrograman ini adalah karena mendukung tulisan/teks Arab dan data dalam bentuk dokumen/file. Adapun klasifikasi yang dilakukan adalah kedalam 8 bentuk kelas yaitu "*Knowledge*", "*Praying*", "*Eclipse*", "*Call to Prayer*", "*Faith*", "*Good Manners*", "*Fasting*" dan "*Almsgiving*". Hadis yang

digunakan dalam klasifikasi ini adalah hadis Bukhari dengan rata-rata akurasi yang didapat adalah sebesar 83.2%. Hasil akurasi yang didapatkan ini bergantung penuh pada proses *stopwords* dan *stemming*.

E. Jasin [3] melakukan klasifikasi hadis anjuran, hadis larangan dan hadis informasi pada hadis shahih Bukhari berdasarkan model Unigram menggunakan *Artificial Neural Network* (ANN). Ia juga menggunakan bantuan metode ekstraksi fitur TF-IDF untuk mendapatkan nilai-nilai pada setiap kata. Hasil performansi yang didapatkan dengan *F1-Score* sebesar 85%. Penelitian yang serupa juga telah dilakukan oleh A. Kusumaningrum [4] pada tahun 2017 dengan menggunakan bantuan metode *Support Vector Machine*. *K-Fold* digunakan untuk melakukan pembagian data menjadi data latih dan data uji. Rata-rata performansi yang didapatkan dengan menggunakan metode ini adalah sebesar 88%.

Untuk kasus *multi-label* pada teks, sudah ada beberapa penelitian terkait hal tersebut. Min-Ling Zhang dan Zhi-Hua Zhou [13] pada tahun 2006 telah melakukan penelitian terkait *multi-label* pada kategorisasi teks. Metode yang digunakan berupa *neural network* yang bernama BP-MLL (*Backpropagation for Multi-Label Learning*). Metode ini berasal dari algoritma *backpropagation* yang populer dengan menggunakan fungsi *error* yang menangkap karakteristik pembelajaran *multi-label*, yaitu label dari suatu *instance* (fitur) harus diberi peringkat lebih tinggi daripada yang tidak termasuk dalam *instance* tersebut. Mereka menggunakan metode tersebut dikarenakan hasil akurasi yang didapat lebih tinggi daripada metode *multi-label-learning* lainnya seperti BoosTexter, ADTBoost.MH, Rank-SVM dan BasicBP.

Pada [2] menggunakan metode JSFC dengan memanfaatkan *Fisher Discriminant-Based* pada seleksi fitur untuk meminimalkan jarak antar kelas yang sama dan memaksimalkan jarak antar kelas yang berbeda sehingga data dapat lebih dipisahkan. Dengan adanya metode tersebut, fitur yang saling berpengaruh untuk masing-masing label kelas dapat dengan mudah dipilih dan pembelajaran untuk model klasifikasi dapat dibangun lebih efektif. Namun kekurangan dari metode ini adalah sulitnya konfigurasi parameter dalam pemilihan fitur sehingga akan memerlukan banyak waktu untuk data dengan jumlah fitur yang besar. Adapun pada [5], metode yang dibangun dapat mengurangi masalah *overfitting* dan *noisy* pada data dengan mengadopsi regresi PLS dan  $l_1$ -norm untuk memodelkan hubungan antar data. Kesulitan yang dihadapi adalah sulitnya mendapatkan cara untuk menentukan nilai variabel yang optimal yang akan memberikan pengaruh yang besar pada performansi.

Soleimani, H [8] memodelkan beberapa kelas secara bersama-sama sehingga didapat kemungkinan ketergantungan antar kelas yang berbeda serta menggunakan metode Markov Chain Monte Carlo yang dapat mengatasi hubungan antar kumpulan label dan fitur yang berubah-ubah. Dengan adanya hal ini maka model yang dibangun dapat mempelajari kelas topik hanya dengan sejumlah kecil dokumen latih yang sudah dilabeli.

Tsoumakas dan Katakis [10] mengelompokkan metode klasifikasi *multi-label* kedalam dua kategori utama, yaitu metode transformasi masalah yang merupakan metode-metode yang mengubah masalah klasifikasi *multi-label* menjadi satu atau lebih klasifikasi *single* label atau masalah regresi, karena keduanya terdapat bibliografi besar dalam algoritma pembelajaran dan metode adaptasi algoritma yang merupakan metode-metode yang diperluas khusus pada pembelajaran algoritma untuk menangani data *multi-label* secara langsung. Mereka mengimplementasi metode PT3, PT4 dan PT6 pada data "*genbase*", "*yeast*" dan "*scene*" dengan *classifier* kNN, C4.5, Naïve Bayes dan SMO.

### 3. Sistem yang Dibangun

Penelitian ini membangun 2 (dua) buah sistem yang dapat melakukan klasifikasi pada data *multi-label* dan *single* label. Kedua sistem menggunakan *dataset* berupa hadis shahih Bukhari terjemahan berbahasa Indonesia yang telah diberi label sebelumnya. Sistem pertama terdiri dari 1064 data *multi-label* dengan 289 diantaranya memiliki lebih dari satu label. Adapun sistem kedua terdiri dari 455 data *single* label. Label kelas pada data *multi-label* terdiri dari 3 (tiga) kelas yaitu kelas anjuran, kelas larangan dan kelas informasi sedangkan pada data *single* label terdiri dari 5 (lima) kelas yaitu kelas 1 adalah tentang "iman", kelas 2 adalah tentang "ilmu", kelas 3 adalah tentang "wudhu", kelas 4 adalah tentang "shalat", dan kelas 5 adalah tentang "waktu-waktu shalat". Adapun masing-masing data hadis *multi-label* dapat berupa salah satu dari ketiga kelas tersebut maupun gabungan dari beberapa kelas. Representasi data akan ditampilkan seperti pada Tabel 1 dan Tabel 2.

Pada pengerjaan penelitian ini, sistem akan mampu melakukan klasifikasi hadis terjemahan berbahasa Indonesia terhadap 3 (tiga) kelas data *multi-label* dan 5 (lima) kelas data *single* label seperti yang telah didefinisikan sebelumnya. Sistem akan terbagi kedalam beberapa proses utama yaitu proses *preprocessing* yang akan menghasilkan data bersih, dilanjutkan dengan proses *feature selection* yang akan memilih fitur, proses ekstraksi fitur yang akan memberikan nilai untuk tiap-tiap fitur dan yang terakhir adalah proses klasifikasi yang akan menentukan kelas dari tiap hadis. Secara umum, sistem akan ditunjukkan pada Gambar 1.

Sebelumnya, dilakukan tahap *preprocessing* untuk menghilangkan *noise* yang terdapat pada *dataset*. Proses lengkap dari *preprocessing* digambarkan pada Gambar 2.

**Tabel 1.** Representasi Data *Multi-Label*

Data	Anjuran	Larangan	Informasi
Janganlah kalian berdusta terhadapku (atas namaku), karena barangsiapa berdusta atas namaku dia akan masuk neraka.	0	1	1
Kami pernah shalat Maghrib bersama Nabi ketika matahari sudah tenggelam tidak terlihat.	0	0	1
Apabila seorang dari kalian memperbaiki keIslamannya maka dari setiap kebaikan akan ditulis baginya sepuluh (kebaikan) yang serupa hingga tujuh ratus tingkatan, dan setiap satu kejelekan yang dikerjakan akan ditulis satu kejelekan saja yang serupa dengannya.	1	0	1

**Tabel 2.** Representasi Data *Single Label*

Data	Kelas Kitab				
	Iman	Ilmu	Wudhu	Shalat	Waktu Shalat
Janganlah kalian berdusta terhadapku (atas namaku), karena barangsiapa berdusta atas namaku dia akan masuk neraka.	0	1	0	0	0
Kami pernah shalat Maghrib bersama Nabi ketika matahari sudah tenggelam tidak terlihat.	0	0	0	0	1
Apabila seorang dari kalian memperbaiki keIslamannya maka dari setiap kebaikan akan ditulis baginya sepuluh (kebaikan) yang serupa hingga tujuh ratus tingkatan, dan setiap satu kejelekan yang dikerjakan akan ditulis satu kejelekan saja yang serupa dengannya.	1	0	0	0	0

1. *Case folding*, merupakan proses untuk merubah setiap kata menjadi huruf kecil dan untuk menyeragamkan setiap kata yang ada. *Case folding* pada penelitian ini menggunakan bantuan fungsi *str.lower()* yang ada pada bahasa pemrograman Python.
2. *Cleaning*, merupakan proses untuk menghilangkan tanda baca dan atau simbol yang ada pada data. Proses ini memerlukan bantuan *library* yang ada pada bahasa pemrograman Python yaitu *Regular Expression*.
3. Tokenisasi, merupakan proses yang akan memenggal setiap kata yang terdapat pada setiap data masukan. Proses tokenisasi pada penelitian ini menggunakan fungsi *str.split()* yang ada pada bahasa pemrograman Python.
4. *Stopword removal*, merupakan proses untuk menghilangkan kata-kata yang dianggap kurang penting pengaruhnya. Biasanya ciri-ciri dari kata-kata ini adalah frekuensi kemunculannya yang cukup sering dibandingkan dengan kata-kata lain.

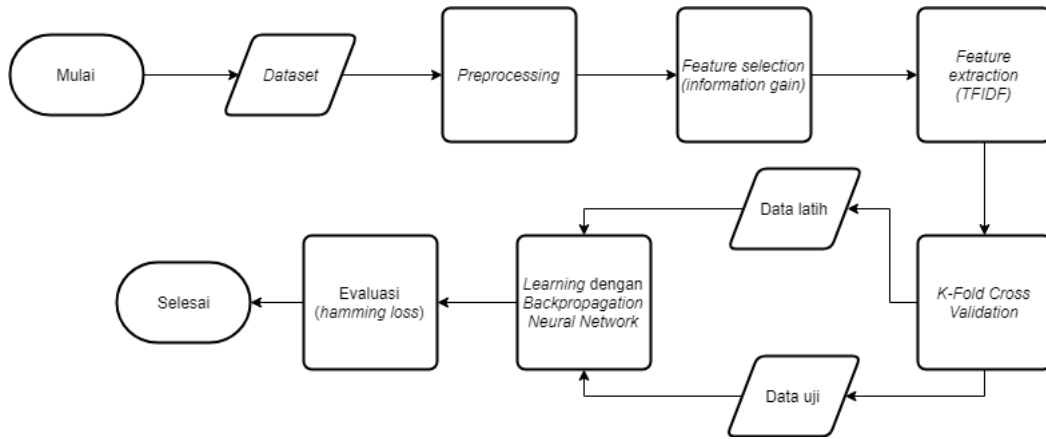
Setelah didapatkan data bersih melalui proses *preprocessing*, langkah selanjutnya adalah memilih dan menentukan fitur-fitur yang dibutuhkan untuk proses klasifikasi. Tahapan ini disebut juga dengan *feature selection*. Tahapan ini akan menentukan fitur/kata apa saja yang relevan dengan masing-masing kelas. Seleksi fitur yang digunakan adalah *information gain* dimana seleksi fitur ini akan menghitung nilai dari setiap kata dalam sekumpulan kata. Kata-kata dengan nilai *information gain* yang tinggi akan menjadi fitur utama untuk setiap data dalam proses klasifikasi. Persamaan (1) berikut ini akan menjelaskan perhitungan untuk menghitung nilai *information gain*.

$$I(C,A) = H(C) - H(C|A), \quad (1)$$

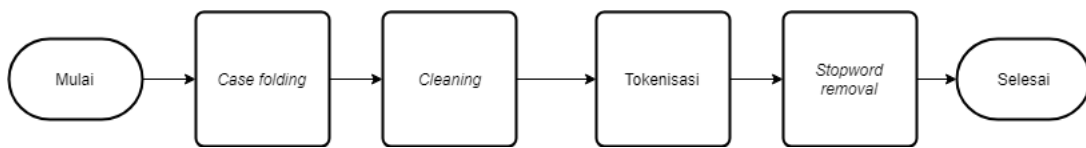
dimana  $H(C) = -\sum_{c \in C} p(C) \log p(C)$  merupakan entropi dari kelas dan  $H(C|A) = -\sum_{c \in C} p(C|A) \log p(C|A)$  merupakan kondisional entropi dari tiap kelas terhadap atribut/kata [7].

Semakin besar nilainya maka semakin menandakan bahwa kata tersebut berpengaruh pada kelas-kelas yang ada. Setelah itu, kata-kata dengan nilai diatas standar (*threshold*) akan dipilih sebagai fitur utama dalam melakukan klasifikasi. Adapun *threshold* ini dapat ditentukan dengan melihat plot dari setiap kata yang didapatkan.

Selanjutnya akan dilakukan tahap ekstraksi fitur untuk mendapatkan nilai dari setiap fitur. *Term Frequency-Inverse Document Frequency* atau biasa disebut dengan TF-IDF adalah salah satu metode ekstraksi fitur yang



**Gambar 1.** Flowchart sistem



**Gambar 2.** Tahapan Preprocessing

populer digunakan khususnya pada klasifikasi teks. TF-IDF merepresentasikan suatu matriks dimana setiap baris pada matriks tersebut sebagai data dan kolom yang ada sebagai kata atau fitur (*term*) [12]. Matriks ini akan diisi oleh suatu nilai atau bobot setiap kata terhadap suatu dokumen. Bobot yang telah didapatkan akan menjadi masukan untuk sistem klasifikasi *neural network*. Adapun TF (*Term Frequency*) merupakan jumlah kemunculan tiap kata pada setiap dokumen dan IDF (*Inverse Document Frequency*) merupakan jumlah dokumen terkait yang mengandung suatu kata tertentu [4]. Perhitungan bobot pada TF-IDF didapatkan seperti pada persamaan (2) berikut.

$$w_{ij} = tf \times idf = tf_{ij} \times \log \frac{D}{df_i}, \quad (2)$$

dimana :

- $w_{ij}$  = bobot kata  $t_j$  terhadap dokumen  $d_i$
- $tf_{ij}$  = jumlah kemunculan kata  $t_j$  dalam dokumen  $d_i$
- $D$  = jumlah dokumen
- $df_i$  = jumlah kemunculan kata dalam dokumen  $D$

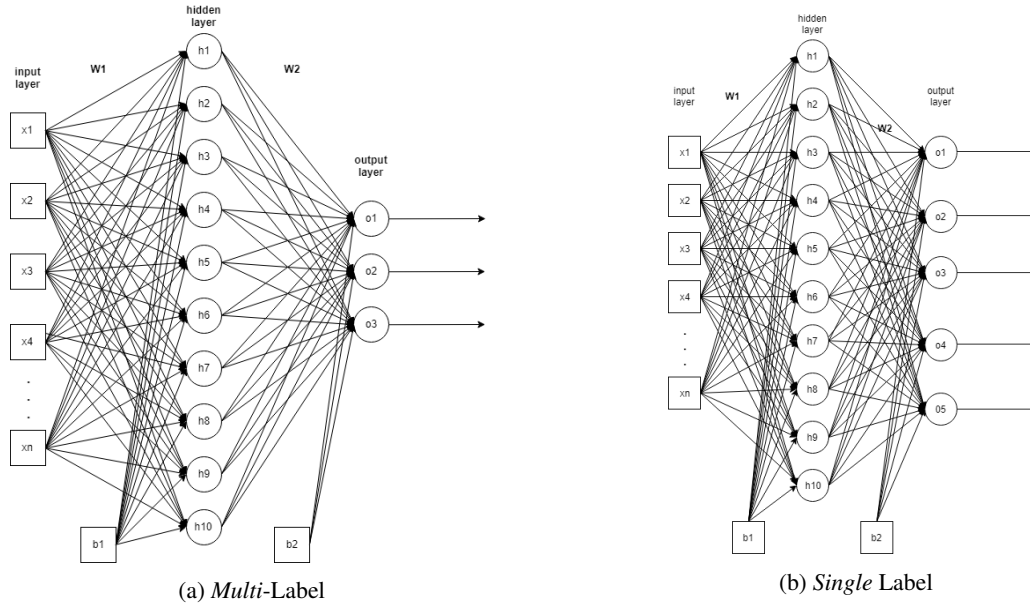
Penelitian ini menggunakan *k-fold cross validation* untuk pengujian. Pada *k-fold cross validation*, kumpulan data  $D$  dipartisi secara acak menjadi  $k$ -bagian dengan ukuran yang kira-kira sama,  $D = \{d_1, d_2, d_3, \dots, d_k\}$ . Setiap  $k$ -bagian pada satu perulangan akan digunakan sebagai data uji dan sisanya akan digunakan sebagai data latih untuk *classifier* yang dibangun. Perkiraan tingkat kesalahan dihitung dari proporsi kesalahan dari seluruh bagian sampel [11]. Pada penelitian kali ini digunakan nilai  $k = 5$ .

Pada penelitian ini, *classifier* yang dibangun adalah *Backpropagation Neural Network*. *Classifier* ini berkerja dengan melakukan dua tahap perhitungan yaitu perhitungan maju yang akan menghitung nilai kesalahan (*error*) antara nilai keluaran sistem dengan nilai yang seharusnya dan perhitungan mundur untuk memperbaiki bobot berdasarkan nilai *error* tersebut [9]. Data latih digunakan pada proses *learning* untuk mendapatkan nilai-nilai parameter yang mengeluarkan hasil terbaik pada sistem *backpropagation neural network*. Hasil terbaik didapatkan ketika nilai *error* yang ada sudah memenuhi standar atau dengan kata lain nilai *error*-nya sangat kecil. Gambar 3 merupakan arsitektur *backpropagation neural network* yang dibangun pada penelitian ini.

Gambar 3(a) merupakan arsitektur *backpropagation neural network* untuk data *multi-label* sedangkan Gambar 3(b) merupakan arsitektur *backpropagation neural network* untuk data *single label*. Setiap data akan dilatih oleh kedua sistem lalu akan ditentukan data tersebut termasuk kedalam kelas apa saja.

Salah satu cara mengetahui apakah sistem yang dibangun sudah baik atau belum adalah dengan mengukur performansi klasifikasi yang dihasilkan. Pada kasus *multi-label*, *hamming loss* merupakan salah satu metrik pengukuran performansi yang sering digunakan. *Hamming loss* akan menghitung nilai kesalahan (*error*) dari hasil





**Gambar 3.** Arsitektur *Backpropagation Neural Network*

klasifikasi. Performansi yang dihasilkan akan sempurna apabila nilai  $hLosss(h) = 0$ . Semakin kecil nilai  $hLosss$  yang dihasilkan, maka semakin baik pula performansi sistem. Persamaan (3) akan menunjukkan perhitungan dari  $hLosss(h)$ .

$$hLoss(h) = \frac{1}{p} \sum_{i=1}^p \frac{1}{Q} |h(x_i) \Delta Y_i| \quad (3)$$

dimana  $p$  merupakan banyak data yang dilakukan klasifikasi dan  $Q$  adalah jumlah label kelas. Sedangkan  $|h(x_i) \Delta Y_i|$  merupakan banyaknya kesalahan klasifikasi yang terjadi.

## 4. Evaluasi

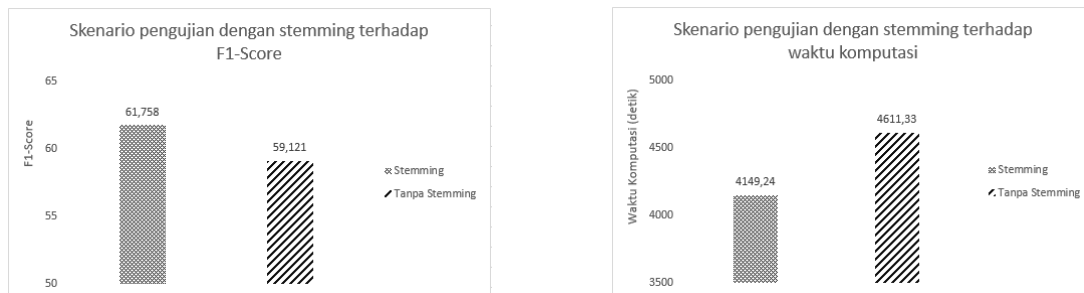
Untuk melihat hasil evaluasi pada *Backpropagation Neural Network* yang dibangun, dilakukan beberapa skenario pengujian. Pengujian pertama dilakukan untuk mengetahui pengaruh dari proses *stemming* pada data hadis Bukhari terjemahan berbahasa Indonesia terhadap nilai *hamming loss* dan *F1-Score* serta waktu komputasi yang dihasilkan. Untuk melakukan proses *stemming*, digunakan bantuan *library* Sastrawi.Stemmer.StemmerFactory(). Pengujian selanjutnya dilakukan untuk melihat pengaruh nilai *threshold* pada *information gain* terhadap nilai *hamming loss* dan *F1-Score* yang didapatkan. Terakhir adalah pengujian dilakukan pada perubahan nilai *learning rate* yang terdapat pada proses *learning* sistem terhadap hasil *hamming loss* dan *F1-Score* yang didapatkan pada data *multi-label* dan data *single label*. Berikut adalah beberapa skenario pengujian untuk data *multi-label* dan *single label*.

### 1. Skenario Pengujian dengan Proses *Stemming*

Pada Gambar 4 dilakukan skenario pengujian untuk melihat apakah ada pengaruh proses *stemming* dan *non-stemming* terhadap hasil *hamming loss* dan waktu komputasi yang didapatkan pada data *multi-label*. Hasil *hamming loss* menunjukkan bahwa pengujian tanpa proses *stemming* mendapatkan hasil yang lebih baik dibandingkan dengan menggunakan proses *stemming* yaitu sebesar 0.119298246 atau dapat dikatakan bahwa sekitar 88.07% data berhasil diklasifikasikan dengan benar. Sedangkan untuk waktu komputasi, pengujian tanpa proses *stemming* juga mendapatkan hasil yang lebih baik dibandingkan dengan menggunakan proses *stemming* yaitu selama 1493.53 detik. Hal ini dikarenakan pada data *multi-label*, setiap kata tidak bisa digeneralisasi dengan menggunakan proses *stemming* karena dapat menghilangkan ciri khusus pada data *multi-label* hadis Bukhari seperti pada label anjuran, kata "-lah" menunjukkan perintah anjuran pada kata "tinggalkan-lah". Sehingga kata ini akan berbeda maknanya jika digeneralisasi dengan menggunakan proses *stemming*.



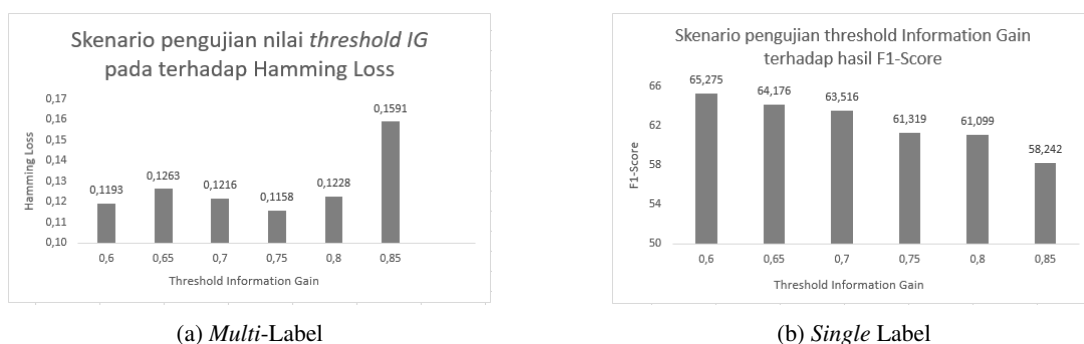
**Gambar 4.** Skenario pengujian dengan proses *stemming* pada *multi-label*



**Gambar 5.** Skenario pengujian dengan proses *stemming* pada *single label*

Adapun untuk pengujian pada data *single label*, didapatkan hasil F1-Score sebesar 61.75% data berhasil diklasifikasikan dengan benar dan waktu komputasi yang lebih baik yaitu selama 4149.24 detik dengan menggunakan proses *stemming* dibandingkan tanpa menggunakan proses *stemming* seperti yang ditunjukkan pada Gambar 5. Ini disebabkan bahwa imbuhan pada suatu kata dapat digeneralisasi menjadi satu kata dengan menggunakan proses *stemming* seperti pada data dengan label shalat, kata "shalat-lah" dan "menyalatkan" dapat digeneralisasi menjadi kata "shalat".

## 2. Skenario Pengujian dengan Threshold Information Gain



(a) *Multi-Label*

(b) *Single Label*

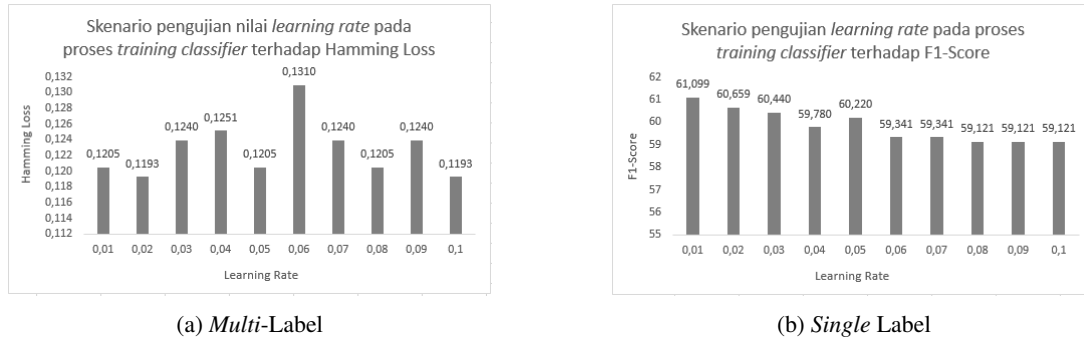
**Gambar 6.** Skenario Pengujian dengan Threshold Information Gain

Gambar 6 memperlihatkan pengaruh nilai *threshold information gain* terhadap data *multi-label* dan *single label*. Pada Gambar 6(a) menunjukkan bahwa skenario pengujian pada data *multi-label* dengan menggunakan *threshold* dimulai dari nilai 0.6 hingga nilai 0.90 dengan perubahan nilai sebesar 0.05. Hasil menunjukkan bahwa nilai *threshold* sebesar 0.75 mendapatkan nilai *hamming loss* yang lebih baik dibandingkan dengan yang lainnya yaitu sebesar 0.1158 atau sebesar 88.42% data dapat diklasifikasikan dengan benar dibandingkan dengan nilai *threshold* yang lain. Hal ini dikarenakan kata-kata atau fitur yang berpengaruh untuk tiap label terdapat pada nilai *information gain* diatas 0.75. Sementara itu, untuk kata-kata yang memiliki nilai *information gain* dibawah 0.75, tidak terlalu mempengaruhi pada penentuan kelas dari tiap data.

Sedangkan untuk data *single label* menunjukkan bahwa hasil F1-Score yang didapatkan lebih baik ketika menggunakan *threshold* sebesar 0.6 dibandingkan dengan menggunakan *threshold* yang lain dengan nilai

F1-Score sebesar 65.275% seperti yang ditunjukkan pada Gambar 6(b). Hal ini dikarenakan kata-kata atau fitur yang terdapat pada masing-masing label memiliki nilai *information gain* yang sama rata dan sama-sama berpengaruh untuk masing-masing label.

### 3. Skenario Pengujian dengan Nilai *Learning Rate*



**Gambar 7.** Skenario Pengujian dengan Nilai *Learning Rate*

Gambar 7 menunjukkan pengaruh nilai *learning rate* pada data *multi-label* dan data *single label* dengan nilai *learning rate* dimulai dari 0.01 hingga 0.1 dengan perubahan sebesar 0.01. Pada Gambar 7(a) terlihat bahwa pengujian dengan nilai *learning rate* 0.02 dan 0.1 mendapatkan hasil *hamming loss* yang terbaik yaitu sebesar 0.1193 atau sebesar 88.07% data dapat diklasifikasikan dengan benar.

Sedangkan untuk data *single label* pada Gambar 7(b) menunjukkan bahwa nilai *learning rate* sebesar 0.01 pada proses *training classifier* mendapatkan hasil F1-Score lebih baik dibandingkan nilai *learning rate* yang lainnya dengan hasil F1-Score sebesar 61.099% data berhasil diklasifikasikan dengan benar. Ini dikarenakan nilai *learning rate* yang besar akan mengakibatkan proses *learning* yang terlalu cepat sehingga tidak mendapatkan nilai perubahan bobot yang optimum.

## 5. Kesimpulan

Dari hasil beberapa skenario pengujian yang telah dilakukan pada penelitian ini, dapat disimpulkan bahwa pengaruh nilai *threshold* pada *information gain* berbeda untuk data *multi-label* dan data *single label*. Pada data *multi-label*, hasil performansi yang dihasilkan lebih baik dengan mengambil beberapa fitur saja yang benar-benar memiliki nilai *information gain* tinggi atau diatas 0.75. Sedangkan pada data *single label*, fitur yang memengaruhi setiap label tidak bisa menghasilkan hasil performansi yang baik pada *classifier* jika hanya mengambil fitur dengan nilai *information gain* diatas 0.75. Pengaruh penggunaan *information gain* bekerja efektif pada data *multi-label*.

Adapun untuk penggunaan proses *stemming*, hasil performansi terbaik didapatkan pada data *single label* dan tidak berlaku untuk data *multi-label*. Hal ini dikarenakan pada data *multi-label*, proses *stemming* akan menghilangkan ciri khas dari fitur untuk setiap label data. Sedangkan untuk data *single label*, generalisasi kata dengan proses *stemming* dapat berpengaruh terhadap hasil performansi sistem yang lebih baik.

Beberapa saran yang dapat diterapkan untuk pengembangan penelitian selanjutnya adalah seperti lebih memperhatikan label data khususnya pada data *single label* yang saat ini masih terdapat banyak label yang kurang sesuai dengan datanya. Selanjutnya adalah mencoba untuk menganalisa makna kata pada data *multi-label* agar hasil yang didapatkan bisa lebih baik.

## Daftar Pustaka

- [1] K. A. Aldhlan, A. M. Zeki, A. M. Zeki, and H. A. Alreshidi. Novel mechanism to improve hadith classifier performance. In *Advanced Computer Science Applications and Technologies (ACSAT), 2012 International Conference on*, pages 512–517. IEEE, 2012.
- [2] J. Huang, G. Li, Q. Huang, and X. Wu. Joint feature selection and classification for multilabel learning. *IEEE transactions on cybernetics*, 48(3):876–889, 2018.

- [3] E. Jasin, S. Faraby, and A. Adiwijaya. Klasifikasi anjuran, larangan dan informasi pada hadis sahih al-bukhari berdasarkan model unigram menggunakan artificial neural network (ann). *eProceedings of Engineering*, 4(3), 2017.
- [4] A. Kusumaningrum, S. Faraby, and A. Adiwijaya. Klasifikasi anjuran, larangan, dan informasi pada hadits shahih bukhari menggunakan metode support vector machine. *eProceedings of Engineering*, 4(3), 2017.
- [5] H. Liu, X. Li, and S. Zhang. Learning instance correlation functions for multilabel classification. *IEEE transactions on cybernetics*, 47(2):499–510, 2017.
- [6] M. Naji Al-Kabi, G. Kanaan, R. Al-Shalabi, S. I. Al-Sinjilawi, and R. S. Al-Mustafa. Al-hadith text classifier. *Journal of Applied Sciences*, 5:584–587, 2005.
- [7] A. I. Pratiwi et al. On the feature selection and classification based on information gain for document sentiment analysis. *Applied Computational Intelligence and Soft Computing*, 2018, 2018.
- [8] H. Soleimani and D. J. Miller. Semisupervised, multilabel, multi-instance learning for structured data. *Neural computation*, 29(4):1053–1102, 2017.
- [9] A. I. Suyanto and A. I. Msc. Searching, reasoning, planning, learning. *Informatika Bandung*, pages 184–186, 2014.
- [10] G. Tsoumakas and I. Katakis. Multi-label classification: An overview. *International Journal of Data Warehousing and Mining (IJDWM)*, 3(3):1–13, 2007.
- [11] K. Yang, H. Wang, G. Dai, S. Hu, Y. Zhang, and J. Xu. Determining the repeat number of cross-validation. In *Biomedical Engineering and Informatics (BMEI), 2011 4th International Conference on*, volume 3, pages 1706–1710. IEEE, 2011.
- [12] M. Zareapoor and K. Seeja. Feature extraction or feature selection for text classification: A case study on phishing email detection. *International Journal of Information Engineering and Electronic Business*, 7(2):60, 2015.
- [13] M.-L. Zhang and Z.-H. Zhou. Multilabel neural networks with applications to functional genomics and text categorization. *IEEE transactions on Knowledge and Data Engineering*, 18(10):1338–1351, 2006.