

**Klasifikasi Multilabel pada Hadis Bukhari Terjemahan  
Bahasa Indonesia Menggunakan *Mutual Information* dan  
*Support Vector Machine***

**Tugas Akhir**

**diajukan untuk memenuhi salah satu syarat**

**memperoleh gelar sarjana**

**dari Program Studi S1 Informatika**

**Fakultas Informatika**

**Universitas Telkom**

**1301154424**

**Mohamad Irwan Afandi**



**Program Studi Sarjana S1 Informatika**

**Fakultas Informatika**

**Universitas Telkom**

**Bandung**

**2019**

## LEMBAR PENGESAHAN

**Klasifikasi Multilabel pada Hadis Bukhari Terjemahan Bahasa Indonesia  
Menggunakan *Mutual Information* dan *Support Vector Machine***

***Multi-Label Classification in Bukhari's Hadith Translated to Indonesia Language  
Using *Mutual Information* and *Support Vector Machine****

**NIM : 1301154424**

**Mohamad Irwan Afandi**

Tugas akhir ini telah diterima dan disahkan untuk memenuhi sebagian syarat memperoleh gelar pada Program Studi Sarjana S1 Informatika

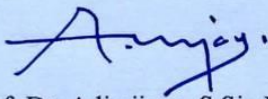
Fakultas Informatika

Universitas Telkom

Bandung, 27 Mei 2019

Menyetujui

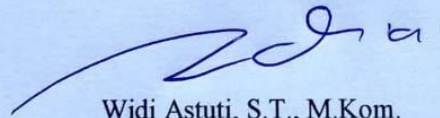
Pembimbing I,



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

00740046-1

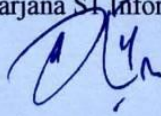
Pembimbing II,



Widi Astuti, S.T., M.Kom.

18910106-3

Ketua Program  
Studi Sarjana S1 Informatika,



Niken Dwi Cahyani, S.T., M.Kom.  
NIP: 00750199-1



## LEMBAR PERNYATAAN

Dengan ini saya, Mohamad Irwan Afandi, menyatakan sesungguhnya bahwa Tugas Akhir saya dengan judul Klasifikasi Multilabel pada Hadis Bukhari Terjemahan Bahasa Indonesia Menggunakan *Mutual Information* dan *Support Vector Machine* 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 di kemudian hari ditemukan pelanggaran terhadap etika keilmuan dalam buku TA atau jika ada klaim dari pihak lain terhadap keaslian karya,

Bandung, 01/Mei/2019

Yang Menyatakan

Mohamad Irwan Afandi



# Klasifikasi Multilabel pada Hadis Bukhari Terjemahan Bahasa Indonesia Menggunakan *Mutual Information* dan *Support Vector Machine*

Mohamad Irwan Afandi<sup>1</sup>, Adiwijaya<sup>2</sup>, Widi Astuti<sup>3</sup>

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

<sup>1</sup>irwanafandi@student.telkomuniversity.ac.id, <sup>2</sup>adiwijaya@telkomuniversity.ac.id,

<sup>3</sup>astutiwidi@telkomuniversity.ac.id,

---

## Abstrak

Hadis merupakan sumber hukum kedua bagi umat Islam setelah Al-Quran. Banyak sekali hadis yang telah diriwayatkan, namun Hadis Bukhari memiliki tingkat kesahihan paling tinggi menurut para ulama. Seiring dengan perkembangan teknologi, hadis sangatlah mudah didapatkan melalui dunia digital. Akan tetapi untuk mempelajari hadis tidak semudah yang kita bayangkan. Banyaknya hadis yang ada dan juga belum dikategorikan membuat belajar hadis dengan kategori tertentu sangat sulit dilakukan. Oleh sebab itu penulis melakukan penelitian klasifikasi anjuran, larangan dan informasi pada Hadis Sahih Al-Bukhari terjemahan Bahasa Indonesia yang diharapkan dapat mempermudah masyarakat dalam mempelajari hadis. Proses klasifikasi menggunakan model unigram/bigram dengan *Mutual Information* (MI) sebagai seleksi fitur dan *Support Vector Machine* (SVM) sebagai metode klasifikasi. Pada penelitian ini dilakukan beberapa skenario pengujian dengan memodifikasi *term model*, *preprocessing*, *feature selection* dan menggunakan beberapa metode klasifikasi untuk membuktikan bahwa SVM merupakan salah satu metode klasifikasi teks yang cocok digunakan. Pengujian dengan menggunakan model unigram, tidak menggunakan stopword/stemming, menggunakan MI dan menggunakan SVM memberikan nilai hamming loss terbaik yaitu 0.0686. Hasil penelitian yang diperoleh juga menunjukkan bahwa metode SVM dengan menggunakan MI lebih baik daripada metode klasifikasi teks yang lain.

**Kata kunci:** Hadis Bukhari, *Hamming loss*, Klasifikasi, *Mutual Information*, *Preprocessing*, *Support Vector Machine*.

---

## Abstract

Hadith is the second source of law and guidance for Muslims after the Qur'an. There are many hadiths that have been narrated, but Bukhari's hadith has the highest level of validity according to the Islamic scholar. Along with the development of technology, hadith is very easy to obtain through the digital world. However, learning hadith is not as easy as we imagine. The number of hadiths that exist and also not have been categorized make learning hadith with specific categories very difficult to do. Therefore, the author conducted a research about classification of suggestion, prohibition and information on Bukhari's hadith that was translated to the Indonesian language which is expected to facilitate the public in learning hadith easily. The classification process uses unigram/bigram model with *Mutual Information* (MI) as feature selection and *Support Vector Machine* (SVM) as classification method. This research used several scenarios of testing by modifying the term model, preprocessing, selection feature and some kind of classification method to prove that SVM is one of the text classification method that suitable to use. Testing using a unigram model, without stopword removal or stemming, using MI and also SVM gives the best hamming loss value, it is 0.0686. The research also indicates that SVM with MI produces higher accuracy than other text classification methods.

**Keyword:** Bukhari's Hadith, Classification, Hamming loss, Mutual Information, Preprocessing, Support Vector Machine.

---

## 1. Pendahuluan

Pada bagian ini berisi empat sub-bagian yaitu: Latar Belakang, Topik dan Batasannya, Tujuan dan Organisasi Tulisan. Di bawah ini akan dijelaskan dari masing-masing sub-bagian tersebut.

### Latar Belakang

Hadis merupakan segala bentuk ucapan, tindakan, dan ketetapan Rasulullah SAW yang dijadikan sebagai petunjuk dan sumber hukum kedua bagi umat Islam setelah Al-Quran [1,2]. Hadis disampaikan turun temurun oleh



ahli hadis dengan tujuan agar generasi selanjutnya bisa meneladani tindakan dan kebiasaan Rasulullah SAW semasa hidupnya. Banyak sekali ahli hadis yang telah meriwayatkan berbagai hadis, namun sebagian besar para ulama sepakat bahwa hadis yang diriwayatkan oleh Imam Bukhari memiliki tingkat kesahihan paling tinggi [24,29]. Oleh sebab itu banyak umat Islam yang mempelajari Hadis Bukhari dalam upaya mendekatkan diri kepada sang pencipta.

Kenyataannya mempelajari ilmu hadis tidak semudah yang kita bayangkan. Meskipun dengan bantuan teknologi membuat hadis semakin mudah didapatkan, misalkan melalui artikel di internet ataupun melalui aplikasi *mobile*. Akan tetapi banyaknya hadis yang ada dan juga belum dikategorikan membuat belajar hadis dengan kategori tertentu sangat sulit dilakukan. Pada dasarnya hadis berisi tentang anjuran, larangan dan informasi. Hal yang paling menarik adalah bahwasannya sebuah hadis bisa diklasifikasikan ke salah satu kelas atau kombinasi dari ketiga kelas tersebut. Tujuan dari penelitian ini adalah melakukan pengelompokan Hadis Bukhari terjemahan Bahasa Indonesia sesuai dengan kelasnya. Diharapkan mampu mempermudah pembaca dalam mempelajari dan menemukan hadis sesuai dengan kelas yang mereka inginkan. Sebenarnya klasifikasi Hadis Bukhari ke dalam kelas tersebut sudah pernah dilakukan pada penelitian [24,26,27], namun penelitian tersebut masih bersifat *single label*. [3,4,5,6] Penelitian ini telah melakukan klasifikasi multilabel, tetapi mereka menggunakan data yang berbeda.

Berdasarkan permasalahan yang telah diuraikan di atas, penulis melakukan klasifikasi multilabel pada Hadis Bukhari terjemahan Bahasa Indonesia menggunakan *Mutual Information* (MI) dan *Support Vector Machine* (SVM). MI merupakan seleksi fitur yang menghitung seberapa banyak informasi yang terkandung dalam *term* kata, dan *term* tersebut berkontribusi untuk membuat keputusan klasifikasi yang tepat pada suatu kelas [30]. MI dipilih sebagai seleksi fitur karena memiliki titik fokus terhadap hubungan *term* kata dengan suatu kelas. Sehingga fitur yang dihasilkan dari proses ini merupakan fitur yang sangat berpengaruh dalam penentuan kelas dari hadis yang dapat meningkatkan akurasi klasifikasi dari sistem yang dikembangkan. Sementara SVM dipilih sebagai metode klasifikasi karena menghasilkan nilai akurasi lebih tinggi dibandingkan metode lain berdasarkan penelitian [7,8,9,27]. Penelitian ini akan menggunakan model *unigram* dan *bigram* untuk melihat pengaruh fitur yang dihasilkan terhadap akurasi klasifikasi. Selain itu juga akan dilakukan perbandingan penggunaan metode SVM dengan metode lain untuk membuktikan bahwa SVM memiliki akurasi yang lebih baik dalam klasifikasi teks.

### **Topik dan Batasannya**

Banyaknya hadis yang ada dan belum dikategorikan membuat belajar hadis dengan kategori tertentu sulit dilakukan. Pada penelitian ini penulis mengangkat permasalahan terkait bagaimana cara membangun klasifikasi multilabel pada Hadis Bukhari terjemahan Bahasa Indonesia secara tepat. Ketepatan pada proses klasifikasi bisa dipengaruhi oleh proses *preprocessing*, *term model* yang digunakan, pemilihan fitur pada kata ataupun metode klasifikasi yang diterapkan. Keempat hal tersebut lah yang akan menjadi titik fokus pada penelitian ini. Penelitian akan menggunakan *stopword removal* dan *stemming* pada tahap *preprocessing*, penggunaan model *unigram* atau *bigram*, penggunaan MI sebagai seleksi fitur dan penggunaan beberapa metode klasifikasi seperti SVM, Naïve Bayes (NB), K-Nearest Network (KNN) serta *Neural Network* (NN) untuk memperoleh akurasi klasifikasi terbaik. Selain itu terdapat beberapa hal yang dijadikan sebagai batasan masalah pada penelitian ini. Pertama, banyak kelas yang digunakan untuk klasifikasi multilabel terdiri dari tiga kelas yaitu larangan, anjuran dan informasi. Jumlah Hadis Bukhari yang digunakan sebanyak 1064 hadis dengan proses pelabelan yang dilakukan dengan metode *hand labelling*. Jumlah hadis yang digunakan memang belum terlalu banyak karena sejauh ini baru 1064 hadis yang telah selesai diberi label.

### **Tujuan**

Penelitian ini dilakukan dengan tujuan mempermudah masyarakat dalam mempelajari hadis dengan cara mengelompokkan hadis yang ada kedalam tiga kelas yaitu anjuran, larangan dan informasi. Penelitian ini juga bertujuan untuk menganalisis pengaruh penggunaan *stopword removal* dan *stemming* pada tahap *preprocessing* serta penggunaan model *unigram* dan *bigram* terhadap akurasi klasifikasi multilabel pada Hadis Bukhari terjemahan Bahasa Indonesia. Selain itu penggunaan MI sebagai seleksi fitur diharapkan mampu meningkatkan nilai akurasi pada klasifikasi. Serta untuk membuktikan bahwa SVM merupakan salah satu metode klasifikasi teks yang menghasilkan nilai prediksi lebih baik dari pada klasifikasi yang lain.

### **Organisasi Tulisan**

Selanjutnya, jurnal tugas akhir ini akan menjelaskan penelitian yang sudah pernah dilakukan sebelumnya serta hal-hal yang terkait dengan penelitian ini pada bagian II. Metode penelitian yang digunakan dan sistem yang dibangun pada proses klasifikasi akan dijelaskan pada bagian III. Pada bagian IV akan dijelaskan hasil yang diperoleh serta evaluasi dari penelitian yang telah dilakukan. Terakhir yaitu penarikan kesimpulan dan pemberian saran untuk penelitian kedepannya yang akan dijelaskan pada bagian V.

### **2. Studi Terkait**

Klasifikasi teks merupakan topik yang sering dibicarakan beberapa tahun terakhir. Klasifikasi teks digunakan untuk mengelompokkan teks terkait ke dalam beberapa kelas yang telah ditentukan, misalnya pengelompokan pada data *twitter*, berita, buku, *email* dan sebagainya [6,8,10,11]. Dalam melakukan klasifikasi teks, permasalahan yang paling sering dijumpai biasanya berkaitan dengan jumlah dimensi yang terlalu tinggi. Hal ini terjadi karena banyaknya *vocabulary* yang dihasilkan oleh dataset, sehingga dapat memperlambat kinerja sistem dan menurunkan akurasi klasifikasi itu sendiri. Seleksi fitur merupakan salah satu cara yang sering digunakan untuk mengatasi permasalahan jumlah dimensi yang tinggi pada klasifikasi teks [12]. Pada proses seleksi fitur akan dipilih fitur (kata) yang unik serta memiliki pengaruh besar terhadap penentuan kelas dari suatu data. Sementara itu untuk fitur yang tidak terpilih akan diabaikan dan tidak lagi digunakan pada tahap selanjutnya. Hal ini tentu dapat mengurangi jumlah fitur yang ada pada dataset, sehingga dimensi dari dataset yang digunakan bisa direduksi.

Terdapat beberapa metode seleksi fitur yang sering digunakan pada kasus klasifikasi teks seperti penelitian [13,14,15,16], salah metode tersebut adalah MI. [17] Penelitian ini menggunakan MI untuk menyeleksi *term* yang paling berpengaruh dalam mengidentifikasi email penting dan *spam*. Selain itu MI juga digunakan untuk mengurangi jumlah dimensi fitur pada email, sehingga fitur yang dihasilkan dapat berkontribusi dengan baik dalam mengidentifikasi kelas dari setiap email. Hasil akurasi klasifikasi terbaik untuk menentukan email penting dan spam pada penelitian ini sebesar 97,3% dengan menggunakan *regression* sebagai metode klasifikasinya. Penelitian yang dilakukan oleh Fahmi [10], dimana MI digunakan untuk menyeleksi fitur pada dataset berita. Fitur dengan nilai MI tinggi berpengaruh besar terhadap penentuan kelas dari data berita. Adapun hasil dari penelitian ini menunjukkan bahwa dengan metode yang sama, klasifikasi menggunakan MI menghasilkan akurasi F1-score sebesar 75,34% sementara tanpa menggunakan MI hanya sebesar 45,95%.

Pembobotan fitur merupakan proses yang harus dilalui setelah fitur-fitur yang ada pada dataset didapatkan. Pada proses ini setiap data akan dihitung bobotnya dimana hasil dari pembobotan akan menjadi inputan untuk sistem klasifikasi yang dibangun. Penelitian Lin Wu [15], TF-IDF digunakan untuk melakukan pembobotan kata pada korpus dokumen Bahasa China. Hasil dari pembobotan berupa vector representasi korpus yang kemudian digunakan untuk proses klasifikasi teks. Penelitian [18] menggunakan TF-IDF untuk pembobotan pada dokumen dengan Bahasa Bengali, kemudian dokumen tersebut diklasifikasikan menggunakan metode SVM. Hasil klasifikasi menunjukkan bahwa pembobotan menggunakan TF-IDF dengan *term* model *unigram* menghasilkan akurasi sebesar 88,13%, sementara dengan *Chi Square Distribution* sebesar 82,82%.

Inti dari proses klasifikasi teks adalah metode yang digunakan dalam melakukan klasifikasi teks tersebut ke dalam beberapa kelas. SVM merupakan salah satu metode yang sering digunakan dalam berbagai kasus klasifikasi. Pada penelitian [19], metode SVM digunakan untuk melakukan klasifikasi rekaman suara seseorang yang berbicara dalam Bahasa Mandarin. Sistem klasifikasi yang dibangun harus mampu mengenali kategori emosi dari suara yang diucapkan. Kemudian sistem akan menentukan apakah suara diucapkan ketika seseorang dalam keadaan marah, senang, sedih, bosan ataupun netral. Nilai akurasi terbaik yang dihasilkan dari penelitian ini sebesar 84,2%. Nilai tersebut lebih bagus dibandingkan dengan penggunaan metode NN yang hanya 80,8%. Penelitian [7] menggunakan metode SVM, KNN dan NB untuk mengelompokkan dokumen ke dalam empat kategori yaitu lingkungan, olahraga, politik dan kesenian. Hasil klasifikasi menunjukkan bahwa rata-rata nilai F1-Score dengan menggunakan metode SVM mencapai 86,26%. Nilai akurasi tersebut lebih bagus dibandingkan dengan nilai akurasi yang dihasilkan dengan menggunakan metode KNN ataupun NB. Penelitian yang dilakukan oleh Priyadarshini Patil [20], mendeteksi apakah tanaman kentang terserang penyakit atau tidak berdasarkan foto daun dari tumbuhan kentang itu sendiri. Adapun metode klasifikasi yang digunakan adalah Artificial Neural Network (ANN), SVM dan *Random Forest* (RF). Akurasi klasifikasi terbaik dihasilkan dengan menggunakan metode ANN yaitu sebesar 92%, diikuti oleh SVM dengan 84% dan RF dengan 79%. Adapun klasifikasi multilabel dengan menggunakan metode SVM juga pernah dilakukan pada penelitian [21]. Penelitian ini menggunakan pendekatan *Binary Relevance* (BR), tujuannya adalah untuk mentransformasikan masalah klasifikasi multilabel menjadi beberapa masalah klasifikasi biner. Kemudian untuk setiap klasifikasi biner yang terbentuk akan diklasifikasikan secara *single-label* menggunakan algoritma klasifikasi SVM. Hasil prediksi yang dihasilkan selanjutnya akan diperbaiki lagi dengan mempertimbangkan probabilitas relasi antar label untuk meningkatkan akurasi klasifikasi. Hasil klasifikasi ini menunjukkan bahwa penggunaan metode SVM dalam kasus multilabel lebih bagus dibandingkan penggunaan metode NB dan C4.5. Berkaitan dengan dimensi data yang tinggi, metode SVM tetap bisa bekerja dengan baik pada data dengan jumlah dimensi yang tinggi. Hal ini karena persebaran data yang bersifat multidimensi tetap beracut pada satu sisi koordinat dan tidak akan bergeser jauh dari posisi awal terutama untuk data teks. Selain itu sebagian besar data berupa teks akan terpisah secara linear ketika berubah menjadi data vector. Sehingga bisa dengan mudah ditentukan mana data yang memiliki kelas positif dan negatif dengan menggunakan *hyperplane*, khususnya kernel *linear* [22]. Metode SVM juga mampu bekerja dengan baik pada data dengan jumlah yang tidak terlalu banyak. Hal ini karena setelah *hyperplane* data ditemukan, maka

secara otomatis data baru akan diklasifikasikan hanya dengan melihat posisi dimana data itu berada. Tentu dengan metode klasifikasi yang seperti ini kesalahan prediksi data bisa diminimalisir.

Perhitungan akurasi pada klasifikasi multilabel bisa dilakukan dengan beberapa cara, salah satunya adalah dengan menggunakan *hamming loss*. Pada proses evaluasi, *hamming loss* [21] digunakan untuk mengukur seberapa tinggi nilai dari misclassification yang terbentuk berdasarkan kecocokan antara pasangan label prediksi dengan label data yang asli. Semakin rendah nilai *hamming loss* yang diperoleh, maka performansi dari model yang dibangun akan semakin bagus begitu juga sebaliknya.

### 3. Sistem yang Dibangun

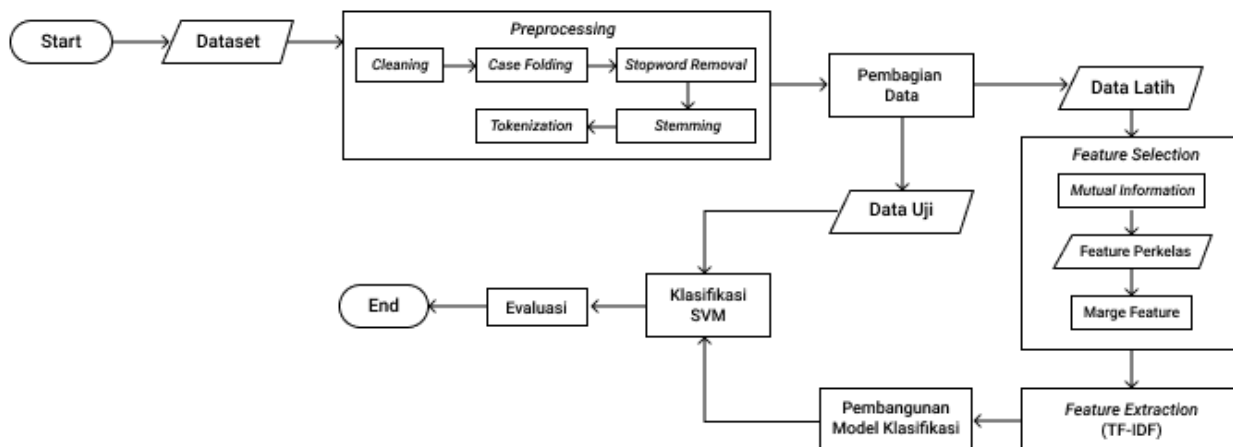
Sebelum membahas mengenai sistem yang dikembangkan, berikut penjelasan tentang data yang digunakan dalam penelitian ini. Penelitian ini dirancang untuk membangun sistem yang mampu mengklasifikasi multi-label data. Data yang digunakan merupakan Hadis Bukhari sejumlah 1064 hadis, dimana setiap hadis telah diberi label kelas sebelumnya. Kelas yang digunakan pada dataset terdiri dari tiga jenis yaitu anjuran, larangan dan informasi. Pada data multi-label, sebuah hadis bisa tergolong ke dalam satu jenis kelas namun tidak menutup kemungkinan kombinasi dari ketiga kelas tersebut. Representasi dari data hadis yang digunakan dalam penelitian ini ditunjukkan pada Tabel 1.

**Tabel 1.** Representasi Data Hadis

Hadis Bukhari	Kelas Anjuran	Kelas Larangan	Kelas Informasi
Barang siapa berwudlu hendaklah mengeluarkan (air dari hidung), dan barang siapa beristinja' dengan batu hendaklah dengan bilangan ganjil.	1	0	1

Pada Tabel 1 bisa dilihat contoh dari *dataset* yang digunakan pada penelitian ini. Hadis tersebut memiliki nilai 1 pada kelas anjuran dan informasi serta nilai 0 pada kelas larangan. Itu berarti bahwa hadis tersebut termasuk ke dalam hadis anjuran dan informasi.

Pada penelitian ini, dibangun sebuah sistem yang mampu melakukan klasifikasi teks Hadis Bukhari secara otomatis ke dalam ketiga kelas tersebut. Secara garis besar sistem akan terbagi menjadi beberapa tahapan yaitu *preprocessing*, *feature selection*, *feature extraction*, klasifikasi data dan diakhiri dengan proses evaluasi terhadap model klasifikasi. Adapun gambaran dari sistem yang dibangun secara detail dapat dilihat pada Gambar 1.



**Gambar 1.** Gambaran Umum Sistem

Pada tahap awal, dataset yang masuk ke dalam sistem akan melalui proses *preprocessing*. Proses ini dilakukan dengan tujuan untuk menghilangkan *noise* yang terdapat pada dataset, sehingga data menjadi lebih ideal serta optimal ketika diolah pada proses selanjutnya. Disini dataset yang berupa teks akan mengalami proses *cleaning*, *case folding*, *stopword removal*, *stemming* dan *tokenization*. *Cleaning* merupakan proses menghilangkan tanda baca, angka dan simbol pada kalimat. *Case folding* merupakan proses mengubah semua huruf pada dataset menjadi huruf kecil semua, sehingga setiap kata dalam kalimat tersebut akan menjadi seragam. *Stopword Removal* merupakan proses menghilangkan kata-kata yang dianggap tidak penting dalam kalimat. *Stemming* merupakan proses untuk

mengembalikan sebuah kata dalam kalimat ke dalam bentuk kata dasarnya. *Tokenization* merupakan proses pemotongan kalimat menjadi kata-kata dari kalimat tersebut. Adapun contoh proses *preprocessing* mulai dari data yang diinputkan hingga menghasilkan data yang ideal bisa dilihat pada Tabel 2.

**Tabel 2.** *Preprocessing Dataset Hadis*

Nama Proses	Input Kalimat	Output Kalimat
Cleaning	Tanda iman adalah mencintai (kaum) Anshar dan tanda nifaq adalah membenci (kaum) Anshar.	Tanda iman adalah mencintai kaum Anshar dan tanda nifaq adalah membenci kaum Anshar
Case Folding	Tanda iman adalah mencintai kaum Anshar dan tanda nifaq adalah membenci kaum Anshar	tanda iman adalah mencintai kaum anshar dan tanda nifaq adalah membenci kaum anshar
Stopword Removal	tanda iman adalah mencintai kaum anshar dan tanda nifaq adalah membenci kaum anshar	tanda iman mencintai kaum anshar tanda nifaq membenci kaum anshar
Stemming	tanda iman mencintai kaum anshar tanda nifaq membenci kaum anshar	tanda iman cinta kaum anshar tanda nifaq benci kaum anshar
Tokenization	tanda iman cinta kaum anshar tanda nifaq benci kaum anshar	tanda, iman, cinta, kaum, anshar, tanda, nifaq, benci, kaum, anshar

Hasil dari proses *preprocessing* berupa potongan kata masing-masing hadis yang disimpan pada sebuah *array*. Data hadis tersebut kemudian akan dipecah menjadi dua bagian, yaitu data latih dan data uji dengan menggunakan *k-fold*. Setelah data dipecah, proses selanjutnya adalah penyeleksian kata pada *array* data latih untuk diambil fitur kata yang paling relevan terhadap masing-masing kelas, sementara fitur yang tidak relevan akan dibuang [25]. Metode yang digunakan untuk proses seleksi fitur pada penelitian ini adalah *Mutual Information*. Pada dasarnya metode MI memiliki konsep menghitung seberapa banyak informasi yang terkandung dalam *term*, dan *term* tersebut berkontribusi untuk membuat keputusan klasifikasi yang tepat pada suatu kelas [30]. Misalkan term “Janganlah” memiliki nilai MI tertinggi untuk kelas larangan. Itu berarti bahwa jika dalam suatu hadis terdapat kata “Janganlah”, maka besar kemungkinan hadis tersebut tergolong ke dalam hadis larangan. Berdasarkan penelitian sebelumnya [28], untuk memperoleh nilai MI dibutuhkan beberapa nilai pendukung. Nilai tersebut diantaranya frekuensi *term* x pada kelas A, frekuensi *term* x namun berada di kelas yang berbeda, frekuensi *trem* lain pada kelas A, frekuensi *term* lain di kelas yang berbeda dan terakhir yaitu jumlah semua *term* yang ada. Untuk memperoleh nilai tersebut penulis menggunakan proses *Language Modelling* khususnya dalam penghitungan *term*, yaitu menggunakan model *unigram* dan *bigram*. Model *unigram* merupakan proses perhitungan sebuah *term* untuk mengetahui seberapa sering *term* tersebut muncul dalam keseluruhan korpus, sedangkan *bigram* adalah dua *term*. Setelah frekuensi diperoleh, tahap selanjutnya adalah menghitung nilai MI dari masing-masing *term* yang ada. Perhitungan nilai MI bisa dilihat pada persamaan 1.

$$I(U, C) = \sum_{et \in \{1,0\}} \sum_{ec \in \{1,0\}} P(U = et, C = ec) \log_2 \frac{P(U = et, C = ec)}{P(U = et)P(C = ec)} \quad (1)$$

Berdasarkan persamaan 1, variabel *U* merupakan variabel acak dengan nilai *et*=1 (mengandung *term* t) dan *et*=0 (tidak mengandung *term* t). Sementara untuk *C* merupakan variabel acak dengan nilai *ec*=1 (kata berada di kelas c) dan *ec*=0 (kata tidak berada di kelas c). Jika dikaitkan dengan seleksi fitur untuk hadis yang akan kita pakai *U* merupakan nilai dari *term* sedangkan *C* merupakan kelasnya. Persamaan 1 bisa diuraikan menjadi lebih detail seperti pada persamaan 2.

$$I(U, C) = \frac{N_{11}}{N} \log_2 \frac{N \cdot N_{11}}{N_1 \cdot N_1} + \frac{N_{01}}{N} \log_2 \frac{N \cdot N_{01}}{N_0 \cdot N_1} + \frac{N_{10}}{N} \log_2 \frac{N \cdot N_{10}}{N_1 \cdot N_0} + \frac{N_{00}}{N} \log_2 \frac{N \cdot N_{00}}{N_0 \cdot N_0} \quad (2)$$

Berdasarkan persamaan 2, variabel *Ns* merupakan jumlah kalimat yang bernilai *ec* dan *et*. Misalkan *N*<sub>10</sub> merupakan kalimat yang mengandung *term* t (*et*=1) namun tidak di kelas c (*ec*=0). *N*<sub>1.</sub> = *N*<sub>10</sub>+*N*<sub>11</sub> adalah jumlah kalimat yang mengandung *term* t (*et*=1) dan jumlah kalimat independen anggota kelas tersebut (*ec* ∈ {1,0}). *N* = *N*<sub>00</sub>+*N*<sub>01</sub>+*N*<sub>10</sub>+*N*<sub>11</sub>. Merupakan total semua *term*. Berdasarkan persamaan tersebut akan diperoleh nilai MI untung masing-masing *feature*. Semakin tinggi nilai MI dari *feature* tersebut, maka semakin besar pula *feature* tersebut mempengaruhi suatu kelas.

Proses selanjutnya adalah mendapatkan nilai bobot dari setiap fitur yang telah diperoleh pada tahap seleksi fitur.



*Term Frequency-Inverse Document Frequency (TF-IDF)* merupakan salah satu metode ekstraksi fitur yang populer dan sering digunakan dalam kasus klasifikasi teks. Hasil dari proses ini berupa sebuah matriks dimana setiap baris matriks adalah data sedangkan setiap kolom yang ada adalah fitur. Matriks ini akan diisi dengan bobot nilai dari setiap fitur terhadap dokumen atau mengalikan antara nilai TF dengan nilai IDF. TF (Term Frequency) adalah jumlah kemunculan kata pada sebuah dokumen dan IDF (Inverse Document Frequency) adalah jumlah dokumen terkait yang mengandung kata tertentu. Bobot yang telah diperoleh akan menjadi inputan untuk sistem klasifikasi yang digunakan. Perhitungan bobot dalam TF-IDF bisa dilakukan dengan persamaan 3.

$$W_{i,j} = tf_{i,j} \times \log \left( \frac{N}{df_j} \right) \quad (3)$$

Berdasarkan persamaan 3,  $W_{i,j}$  merupakan bobot kata  $t_j$  terhadap dokumen  $d_i$ ,  $tf_{i,j}$  merupakan jumlah kemunculan kata  $t_j$  terhadap dokumen  $d_i$ ,  $N$  merupakan jumlah dokumen keseluruhan dan  $df_j$  adalah jumlah dokumen keseluruhan yang memuat kata  $j$ .

Pada penelitian ini, data hadis yang ada akan diklasifikasikan menggunakan metode *Support Vector Machine*. Secara sederhana SVM bekerja dengan cara mencari nilai *hyperplane* pemisah yang optimal antara sampel positif dan negatif [23]. *Hyperplane* dikatakan optimal jika margin yang terbentuk antara *hyperplane* dengan sampel positif dan negatif memiliki jarak yang paling dekat. Setelah letak *hyperplane* ditemukan, maka kelas dari sampel baru yang masuk otomatis bisa ditentukan. Jika sampel berada di sisi positif, berarti kelas dari data tersebut positif begitu juga sebaliknya. Diberikan data uji,  $T = \{(x_1, y_1), \dots (x_n, y_n)\}$ , dimana  $x_i \in R^n$  adalah pola inputan, dan  $y_i \in \{-1, 1\}$  adalah label kelas untuk dua kelas. Masalah utama adalah, SVM mencoba menemukan *classifier*  $f(X)$  yang meminimalkan tingkat kesalahan dari klasifikasi. Adapun persamaan untuk *Linear classifier*  $f(X)$  pada SVM bisa dilihat pada persamaan 4.

$$\min \left( \frac{1}{2} ||W||^2 + C \sum_{i=1}^t \xi_i \right) \quad (4)$$

Berdasarkan persamaan 4,  $y_i((W, X_i) + b) \geq 1 - \xi_i$  dan  $\xi_i \geq 0$  dimana  $C$  merupakan parameter regularisasi yang digunakan untuk menyeimbangkan kompleksitas dan akurasi klasifikasi pada data uji  $T$ . Tidak hanya dengan kernel *linear*, penentuan *hyperplane* juga bisa dilakukan dengan fungsi *non-linear kernel* untuk mengubah SVM *linear* menjadi SVM *non-linear* yang lebih fleksibel.

Evaluasi merupakan tahap akhir dari proses klasifikasi yang digunakan untuk melihat seberapa akurat sistem yang telah dibangun. Salah satu metode evaluasi untuk kasus klasifikasi multilabel adalah *hamming loss*, dimana metode ini menghitung seberapa besar misclassification yang terjadi berdasarkan sistem yang dibangun [21]. Semakin kecil nilai *hamming loss* maka sistem akan semakin akurat dalam mengklasifikasikan data begitu juga sebaliknya. Adapun cara perhitungan *hamming loss*, bisa dilihat pada persamaan 5.

$$hloss(h) = \frac{1}{p} \sum_{i=1}^p \frac{1}{Q} |h(x_i) \Delta Y_i| \quad (5)$$

Dimana  $p$  merupakan jumlah dokumen keseluruhan,  $Q$  adalah jumlah label pada kelas yang digunakan dan  $|h(x_i) \Delta Y_i|$  merupakan banyaknya *missclassification* yang terjadi setelah prediksi kelas dokumen didapatkan.

## 4. Evaluasi

### 4.1. Skenario Pengujian

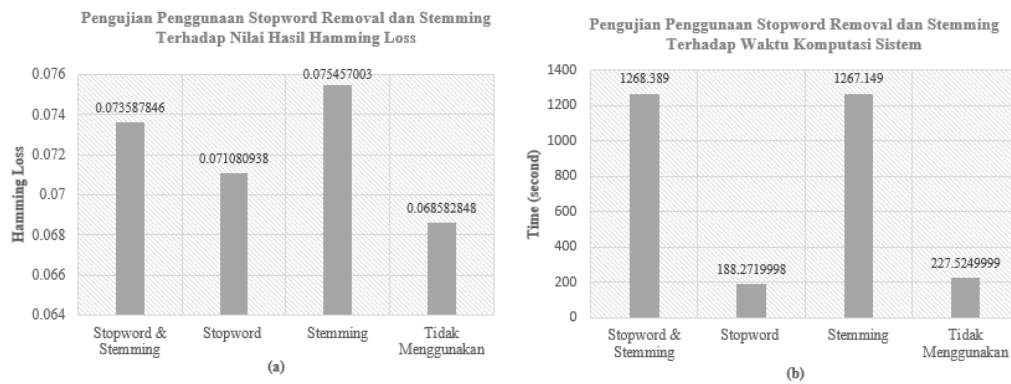
Skenario pengujian penelitian Tugas Akhir (TA) ini memiliki tiga titik fokus yaitu pada tahap *preprocessing*, tahap seleksi fitur dan tahap klasifikasi. Tujuan utamanya adalah untuk mendapatkan parameter yang optimal serta membuktikan bahwa metode SVM adalah salah satu metode yang cocok diterapkan dalam kasus klasifikasi teks. Skenario pengujian yang dilakukan merupakan kombinasi dari ketiga tahap tersebut, dimana menghasilkan 18 skenario pengujian. Penelitian ini menerapkan metode *k-fold cross validation* dalam melakukan pengujian, dataset  $X$  akan dibagi menjadi  $k$ -bagian dengan jumlah yang sama,  $X = x_1, x_2, x_3, \dots, x_k$ . Untuk setiap  $k$  bagian pada satu kali perulangan, data akan digunakan sebagai data uji dan sisanya sebagai data latih dalam membangun model klasifikasi. Maksudnya jika  $x_1$  digunakan sebagai data uji maka  $x_2, x_3, \dots, x_k$  akan digunakan sebagai data latih, jika  $x_2$  data uji maka  $x_1, x_3, \dots, x_k$  akan digunakan sebagai data latih begitu seterusnya hingga semua bagian menjadi data uji. Adapun akurasi dari klasifikasi didapatkan dengan menghitung rata-rata nilai *hamming loss* pada setiap *fold* yang dihasilkan.

Penelitian ini menggunakan 10-fold yang diterapkan pada 1064 data Hadis Bukhari terjemahan Bahasa Indonesia.

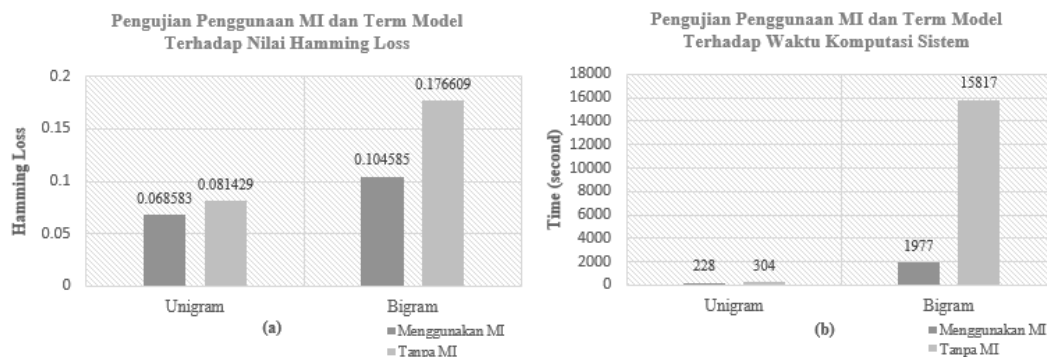
#### 4.2. Hasil Pengujian

Skenarion pengujian pertama yaitu penggunaan *stopword removal* dan *stemming*. *Stopword removal* merupakan proses menghilangkan kata yang dianggap tidak penting sementara *stemming* merupakan proses mengubah kata dalam kalimat ke bentuk kata dasarnya. Untuk menghasilkan proses *stopword removal* dan *stemming* pada data hadis, penulis menggunakan *library* Sastrawi. Pada pengujian ini sistem menggunakan MI sebagai fitur seleksi dan SMV dengan kernel *linear* sebagai metode klasifikasi. Hasil pengujian dan waktu komputasi sistem bisa dilihat pada Gambar 2.

Skenario pengujian kedua adalah menguji kinerja sistem klasifikasi yang dibangun dengan membandingkan hasil klasifikasi menggunakan MI dan tanpa MI. Sementara itu, untuk memperoleh nilai MI sendiri diperlukan nilai pendukung yaitu jumlah masing-masing *term* baik di kelas yang sama atau pun kelas yang berbeda. Sehingga setiap kalimat yang ada pada hadis harus dilakukan tokenisasi untuk menghasilkan *term* kata yang diperlukan. Pada penelitian ini ada dua model dalam pembentukan term yaitu *unigram* dan *bigram*. Pada skenario ini pengujian dilakukan dengan mengkombinasikan penggunaan MI dan model. Adapun untuk proses *preprocessing* yang digunakan dilihat dari hasil terbaik pada percobaan pertama yaitu tanpa menggunakan *stopword removal* dan *stemming*. Untuk metode klasifikasinya sendiri tetap menggunakan SVM dengan kernel *linear* seperti pada percobaan pertama. Hasil pengujian kedua dan waktu komputasi sistem bisa dilihat pada Gambar 3.



**Gambar 2.** (a) Hasil pengujian pengaruh penggunaan *stopword removal* dan *stemming* terhadap hasil *hamming loss*, (b) Hasil pengujian pengaruh penggunaan *stopword removal* dan *stemming* terhadap waktu komputasi sistem

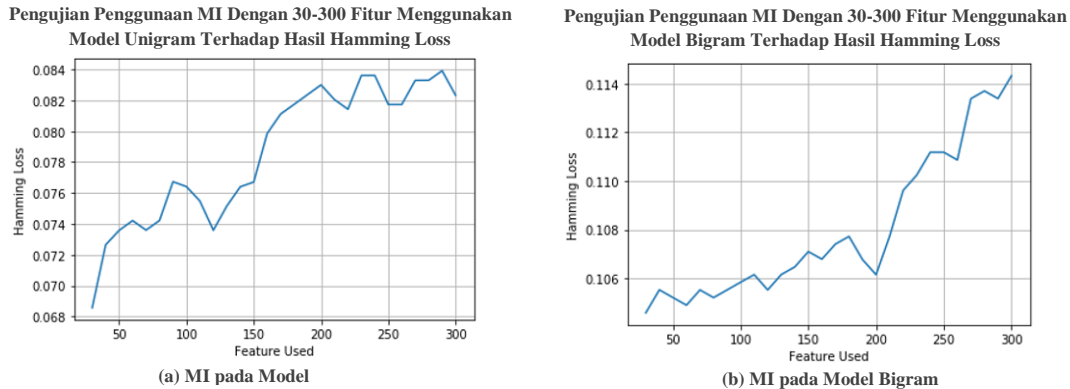


**Gambar 3.** (a) Hasil pengujian pengaruh penggunaan MI dan term model terhadap hasil *hamming loss*, (b) Hasil pengujian pengaruh penggunaan MI dan term model terhadap waktu komputasi sistem

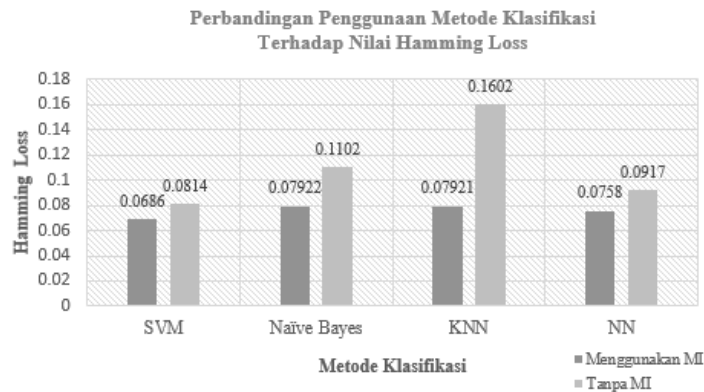
Selain menguji penggunaan MI dan model, penulis juga melakukan pengujian terhadap penggunaan jumlah fitur yang berbeda pada pada skenario pengujian dua. Fitur disini merupakan hasil perhitungan MI, dimana semakin tinggi nilai MI maka pengaruh fitur tersebut dalam penentuan kelas suatu hadis akan semakin tinggi. Jumlah fitur yang digunakan sebanyak 30-300 fitur diamana untuk setiap perulangan yang dilakukan, jumlah fitur yang digunakan akan bertambah sebanyak 10 fitur. Hasil percobaan ini berupa grafik nilai *hamming loss* untuk setiap fitur yang digunakan. Hasil grafik pengujian fitur bisa dilihat pada gambar 4.

Skenario pengujian ketiga adalah membandingkan hasil klasifikasi hadis menggunakan metode SVM dengan metode klasifikasi NB, KNN dan NN. Untuk klasifikasi yang menggunakan metode SVM, kernela yang digunakan

adalah kernel *linear*. Hal ini karena dengan menggunakan kernel *linear*, hasil akurasi klasifikasi yang dihasilkan lebih bagus dibandingkan dengan menggunakan kernel *rbf*, *polynomial* atau *sigmoid*. Sementara itu untuk metode klasifikasi NB, KNN dan NN *parameter* yang digunakan merupakan *parameter default* bawaan dari *library*. Adapun tahap *preprocessing* pada pengujian ketiga tidak menggunakan *stopword removal* dan *stemming* serta untuk *term model* yang digunakan merupakan model *unigram*. Hasil dari pengujian ketiga bisa dilihat pada gambar 5.



**Gambar 4.** Pengujian pengaruh penggunaan MI dengan jumlah fitur diantara 30-300 dengan model unigram dan bigram terhadap nilai hamming loss



**Gambar 5.** Perbandingan penggunaan metode klasifikasi terhadap nilai hamming loss yang dihasilkan

#### 4.3. Analisis Pengujian

Skenario pengujian pertama dilakukan untuk mengamati pengaruh dari penggunaan *stopword removal* dan *stemming* terhadap nilai *hamming loss* dan waktu komputasi sistem. Hasil dari pengujian bisa dilihat pada gambar 2. Berdasarkan pengujian yang dilakukan, nilai *hamming loss* terbaik dihasilkan dari pengujian tanpa menggunakan *stopword removal* dan *stemming*. Nilai *hamming loss* yang dihasilkan sebesar 0,06858 atau sekitar 93,14% data multilabel terklasifikasi dengan benar. Sementara untuk waktu komputasi terbaik dihasilkan dari proses pengujian dengan menggunakan *stopword removal* namun tanpa *stemming*, yaitu selama 188 detik. Ditinjau berdasarkan nilai *hamming loss* yang dihasilkan, proses *stemming* tidak bisa dilakukan pada setiap kata dalam hadis. Hal ini karena dengan dilakukan proses *stemming*, setiap kata dalam hadis akan diubah ke bentuk kata dasarnya. Sehingga makna kalimat pada hadis pun ikut berubah. Contoh pada kalimat “bangun dan shalatlah”, akan berubah menjadi “bangun dan shalat” jika dilakukan proses *stemming*. Menghilangkan kata “-lah” pada kata shalatlah, bisa merubah makna kalimat yang semula bermakna anjuran akan berubah menjadi informasi. Sementara itu proses *stopword removal* merupakan proses penghapusan kata karena dianggap tidak memiliki makna atau kata tersebut sering muncul. Akan tetapi bukan berarti bahwa kata akan dihapus secara keseluruhan, namun hanya mengurangi jumlah kata yang ada. Ketika menggunakan MI sebagai seleksi fitur, kata yang jarang muncul bisa jadi memiliki nilai MI yang lebih tinggi dibandingkan dengan kata yang sering muncul karena kata dianggap unik. Sehingga kata yang mengalami proses *stopword removal* akan banyak digunakan sebagai fitur pada MI meski kata tersebut tidak memiliki makna sama sekali, contohnya kata dan, di, jika, dll. Hal ini tentu akan menyebabkan menurunnya hasil klasifikasi karena fitur yang

digunakan tidak menggambarkan suatu kelas, bahkan cenderung tidak memiliki makna. Akan tetapi penggunaan *stopword* mampu mempercepat waktu komputasi sistem, karena semakin sedikit kata yang diproses maka waktu sistem dalam mengolah kata yang ada pun semakin cepat.

Skenario pengujian kedua adalah mengamati pengaruh penggunaan MI sebagai seleksi fitur pada proses klasifikasi berdasarkan model yang digunakan. Hasil pengujian bisa dilihat pada gambar 3. Nilai *hamming loss* terbaik serta waktu komputasi tercepat dihasilkan oleh klasifikasi yang menggunakan MI dan dengan model *unigram*. Dimana 93,14% data multi-label bisa diklasifikasikan secara benar serta waktu komputasi sistem yaitu 228 detik, sementara tanpa seleksi fitur sekitar 91,86% dengan waktu 304 detik. Ini terjadi karena dengan menggunakan seleksi fitur, fitur yang digunakan merupakan fitur yang benar-benar menggambarkan kelas dari kalimat, sehingga membuat klasifikasi menjadi efektif. Sementara tanpa seleksi fitur berarti semua fitur digunakan dalam proses klasifikasi, akibatnya data akan semakin menyebar dan sulit dibedakan oleh sistem klasifikasi. Hal ini juga bisa dibuktikan dengan gambar 4, dimana semakin banyak fitur yang digunakan maka nilai *hamming loss* juga akan semakin besar. Adapun untuk waktu komputasi sendiri semakin sedikit fitur yang digunakan maka pengecekan yang dilakukan oleh sistem akan semakin sedikit yang membuat waktu komputasi akan semakin cepat. Disisi lain untuk model *unigram* dan *bigram*, penggunaan model *unigram* lebih bagus karena perhitungan MI berlaku pada tiap kata (satu kata). Dimana jika kata tersebut memiliki nilai MI tinggi, berarti memang kata tersebut unik serta menggambarkan kelas tertentu. Namun untuk model *bigram* tidak, hal ini karena dengan model *bigram* perhitungan MI berlaku pada kombinasi dua buah kata yang berurutan. Akibatnya kata yang awalnya tidak unik bisa berubah menjadi unik dan memiliki nilai MI yang tinggi apabila dikombinasikan dengan kata yang kedua. Meskipun kedua kata tersebut sebenarnya sama-sama tidak menggambarkan suatu kelas. Contoh kombinasi antara kata “jika ia, ke arah, atau di, di bawah, dst” yang sebenarnya tidak memiliki makna sama sekali, namun malah digunakan sebagai seleksi fitur karena memiliki nilai MI yang tinggi. Meski begitu, metode seleksi fitur menggunakan MI juga memiliki banyak kekurangan dalam pemilihan fitur. Fitur yang dihasilkan tidak semuanya menggambarkan kelas dari hadis. Bahkan masih banyak fitur yang seharusnya menggambarkan kelas hadis namun tidak digunakan sebagai fitur utama dikarenakan nilai MI yang cenderung tidak terlalu tinggi misalkan kata berwujudlah, tidurlah, mengharamkan, sebaiknya, dll. Secara tidak langsung karakteristik dari kata yang berpengaruh dalam menentukan kelas anjuran, larangan dan informasi bisa diamati secara manual. Fitur kata dengan kelas anjuran biasanya memiliki karakteristik berupa kata kerja, keterangan atau sifat dengan akhiran “-lah”, mengandung kata “wajib” atau “anjur”, kata “sebaiknya”, dll. Fitur kata dengan kelas larangan biasanya mengandung kata “jangan, larang dan haram”. Sementara untuk kelas informasi biasanya terdapat kata “nisacaya, barangsiapa, sesungguhnya, bahwasannya” serta kata-kata yang identik dengan menyimpulkan sebuah gagasan. Berdasarkan karakteristik fitur kata di setiap kelas yang telah diketahui sebelumnya, maka sebenarnya bisa dengan mudah fitur kata perkelas didapatkan. Hal ini akan membuat akurasi klasifikasi multilabel data bisa ditingkatkan.

Skenario pengujian ketiga dilakukan untuk membandingkan hasil klasifikasi teks menggunakan metode SVM dengan metode NB, KNN dan NN. Hasil dari pengujian bisa dilihat pada gambar 5 dimana klasifikasi terbaik dihasilkan oleh sistem klasifikasi yang menggunakan metode SVM dengan 93,14% multi-label data bisa diklasifikasikan secara benar. Kemudian disusul oleh NN dengan nilai 92,42%, KNN 92,079% dan NB 92,078%. Pada penelitian ini setiap data hadis bersifat multi dimensi dimana jumlah dimensi data tergantung dari jumlah fitur yang digunakan. Meski bersifat multi dimensi metode SVM cenderung bekerja dengan baik, hal ini karena metode ini mencari pemisah optimal antar data positif dan negatif terdekat saja, bukan terhadap semua data. Lalu untuk data yang jauh dari pemisah akan ditentukan kelasnya berdasarkan letak dimana dia berada, jika berada di sisi positif maka data tersebut memiliki kelas positif begitu juga sebaliknya. Metode SVM juga menggunakan fitur, dimana fitur akan memberi bobot pada masing-masing data. Semakin banyak fitur berada pada suatu data maka bobot pun akan semakin besar sehingga data akan cenderung berada pada salah satu sisi positif atau negatif. Adapun metode NN merupakan metode yang bersifat *self-learning* dimana semakin banyak data latih yang digunakan, maka sistem akan semakin mudah mengenali kelas dari data uji berdasarkan proses *learning* yang dilakukan oleh sistem. Pada penelitian ini, jumlah data latih untuk setiap *fold* hanya sekitar 957 atau 958 data saja, hal ini kemungkinan membuat proses *learning* algoritma kurang maksimal. Akibatnya, akurasi prediksi yang dihasilkan lebih rendah dibandingkan dengan menggunakan metode svm. Pada metode KNN, jumlah dimensi data yang banyak akan menjadi masalah. Hal ini karena KNN tidak mempelajari bobot data serta tidak tahu fitur mana yang paling berpengaruh dalam penentuan kelas. Algoritma ini hanya mencari sebanyak  $k$ -data latih yang jaraknya paling dekat dengan data uji. Misalnya  $k=5$ , maka dari 5 data latih yang terdekat akan dihitung jumlah kelas positif dan negatif. Kelas dari data uji akan ditentukan dari hasil perhitungan kelas terbanyak dari 5 data tersebut. Jika yang terbanyak kelas positif maka nilai dari data uji adalah positif begitu juga sebaliknya. Sementara pada metode Naïve Bayes, penentuan kelas didasarkan pada perhitungan probabilitas kemunculan katanya saja. Metode ini tidak memperhatikan interaksi antar fitur dengan kelas, sehingga fitur representasi kelas tidak berpengaruh. Jadi apabila probabilitas kelas positif lebih kecil dari pada probabilitas kelas negatif pada data yang bersangkutan, maka data otomatis dikategorikan ke dalam kelas negatif.

Terakhir adalah analisis terhadap data hadis yang digunakan terhadap hasil klasifikasi dari sistem yang dibangun. Dalam konteks klasifikasi, sistem dikatakan baik apabila jumlah data yang digunakan baik positif dan negatif untuk setiap kelasnya harus seimbang. Adapun data hadis yang digunakan pada penelitian ini bersifat *imbalance data*, dimana jumlah label data pada kelas tidak seimbang, terutama untuk data pada kelas larangan dan informasi. Pada kelas anjuran, jumlah rata-rata hadis yang memiliki kelas positif sekitar 207 hadis sementara untuk kelas negatif sebanyak 750 hadis pada setiap foldnya. Adapun untuk kelas larangan sekitar 88 hadis bernilai positif dan 870 hadis bernilai negatif, sementara kelas informasi sekitar 938 data hadis bernilai positif dan 19 hadis bernilai negatif pada setiap foldnya. Berdasarkan data tersebut, sedikitnya jumlah data positif pada hadis anjuran dan larangan akan membuat sistem kesulitan dalam mengidentifikasi kelas anjuran dan larangan pada hadis. Hal ini karena sistem sulit menemukan fitur yang relevan untuk menggambarkan kelas dari hadis tersebut. Akibatnya hanya sedikit sekali fitur yang diperoleh untuk menentukan mana hadis yang tergolong kedalam kelas anjuran dan larangan. Adapun untuk kelas informasi hampir semua data tergolong kedalam data positif yang berarti hampir semua hadis adalah hadis informasi. Akibatnya sistem klasifikasi tidak bisa mengenali mana data positif dan mana data negatif, dalam pembangunan model klasifikasi. Hasilnya semua data akan dikelompokkan ke dalam kelas yang nilainya dominan (positif). Sehingga menyebabkan banya data multilabel salah diklasifikasikan.

## 5. Kesimpulan

Berdasarkan beberapa skenario pengujian dalam penelitian ini, klasifikasi multilabel pada data Hadis Bukhari menggunakan metode SVM terbukti memberikan hasil prediksi yang lebih akurat dari pada metode lain seperti NB, KNN dan NN terutama pada jumlah data yang tidak terlalu banyak. Penggunaan *Mutual Information* sebagai fitur seleksi pun terbukti mampu meningkatkan kebenaran klasifikasi multilabel yang semula 91,86% meningkat menjadi 93,14%. *Mutual Information* akan menyeleksi fitur yang benar-benar menggambarkan kelas dari suatu hadis, sementara fitur yang tidak menggambarkan kelas dari hadis akan diabaikan. Hal ini tentu akan meringankan kinerja sistem sehingga dapat menurunkan waktu komputasi sistem tersebut.

Penggunaan *stemming* pada tahap *preprocessing* harus dihindari, khususnya pada data yang digunakan pada penelitian ini, sebab *stemming* mampu merubah makna dari suatu kalimat. Proses *stopword removal* dan penggunaan model *bigram* pun harus dihindari jika menggunakan MI sebagai seleksi fitur, Hal ini karena *stopword removal* akan mengurangi jumlah kata yang dianggap tidak penting pada kalimat. Padahal pada perhitungan MI, bisa jadi kata yang jumlahnya sedikit malah memiliki nilai MI yang tinggi karena dianggap unik dan menggambarkan suatu kelas. Sehingga kata yang dianggap tidak berguna pun akan menjadi fitur, yang berakibat menurunkan akurasi dari klasifikasi. Sama halnya dengan *stopword removal*, penggunaan model *bigram* dapat membuat kata yang awalnya tidak bermakna menjadi unik dan memiliki nilai MI yang tinggi. Hal karena penggabungan dua buah kata yang tidak memiliki makna sama sekali, membuat kata tersebut bisa dianggap unik dan menggambarkan suatu kelas, padahal tidak.

Beberapa saran yang bisa dilakukan untuk penelitan lebih lanjut diantaranya adalah penambahan jumlah data hadis, dimana jumlah Hadis Bukhari sendiri mencapai 7000 hadis. Data hadis pun harus divalidasi kepada ahli hadis dalam penentuan label anjuran, larangan dan informasi. Perlu diingat bahwasannya dalam penambahan data hadis sendiri, perhatikan proporsi data setiap kelasnya. Usahakan jumlah data hadis tiap kelasnya seimbang agar masalah *imbalance data* bisa dihindari, sehingga sistem berjalan dengan maksimal. Selain itu mencoba mengembangkan ekstraksi fitur berbasis aturan pada sistem yang dibangun. Tujuannya adalah untuk memperoleh fitur-fitur yang sebenarnya relevan namun tidak terpilih sebagai fitur utama dari proses seleksi fitur. Aturan dari penentuan fitur dibuat secara manual dengan inputan berupa *tagger* dari setiap kata memanfaatkan *post tagger* pada *natural language processing*. Berdasarkan proses ini akan diperoleh fitur dari masing-masing kelas hadis yang bisa dijadikan fitur tambahan pada sistem yang bangun. Besar kemungkinan dengan menambah seleksi fitur berbasis aturan ini, akurasi dari sistem akan semakin meningkat.

## Daftar Pustaka

- [1] M. A. Saloot, N. Idris, R. Mahmud, S. Ja'afar, D. Thorleuchter dan A. Gani, "Hadith Data Mining and Classification: a Comparative Analysis," *Journal of International Science and Engineering*, vol. 46, no. 1, pp. 113-128, 2016.

- [2] H. Sayoud, "Automatic Authorship Classification of Two Ancient Books: Quran and Hadith," *2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA)*, pp. 666 - 671, 2014.
- [3] X. Li, H. Xie, Y. Rao, Y. Chen, X. Liu, H. Huang and F. L. Wang, "Weighted Multi-Label Classification Model for Sentiment Analysis of Online News," *2016 International Conference on Big Data and Smart Computing (BigComp)*, pp. 215 - 222, 2016.
- [4] D. C. d. Lucena and R. B. Prudencio, "Semi-Supervised Multi-Label K-Nearest Neighbors Classification Algorithms," *2015 Brazilian Conference on Intelligent Systems (BRACIS)*, pp. 49 - 54, 2015.
- [5] Y.-p. Qin and X.-k. Wang, "Study on Multi-label Text Classification Based on SVM," *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, vol. 1, pp. 300 - 304, 2009.
- [6] A. M. K. Izzaty, M. S. Mubarak, N. S. Huda and Adiwijaya, "A Multi-Label Classification on Topics of Quranic Verses in English Translation Using Tree Augmented Naïve Bayes," *2018 6th International Conference on Information and Communication Technology (ICoICT)*, pp. 103 - 106, 2018.
- [7] Z. Liu, X. Lv, K. Liu and S. Shi, "Study on SVM Compared with the other Text Classification Methods," *2010 Second International Workshop on Education Technology and Computer Science*, vol. 1, pp. 219 - 222, 2010.
- [8] I. Dilrukshi and K. D. Zoysa, "Twitter News Classification: Theoretical and Practical Comparison of SVM Against Naive Bayes Algorithms," *2013 International Conference on Advances in ICT for Emerging Regions (ICTer)*, pp. 278 - 278, 2013.
- [9] Y. Lin and J. Wang, "Research on Text Classification Based on SVM-KNN," *2014 IEEE 5th International Conference on Software Engineering and Service Science*, pp. 842 - 844, 2014.
- [10] F. S. Nurfikri, M. S. Mubarak and Adiwijaya, "News Topic Classification Using Mutual Information and Bayesian Network," *2018 6th International Conference on Information and Communication Technology (ICoICT)*, pp. 162 - 166, 2018.
- [11] A. I. Pratiwi & Adiwijaya, "On the Feature Selection and Classification Based on Information Gain for Document Sentiment Analysis". *Applied Computational Intelligence and Soft Computing*, 2018.
- [12] M. Zareapoor and Seeja, "Feature Extraction or Feature Selection for Text Classification: A Case Study on Phishing Email Detection," *International Journal of Information Engineering and Electronic Business*, vol. 7, pp. 60-65, 2015.
- [13] S. Lei, "A Feature Selection Method Based on Information Gain and Genetic Algorithm," *2012 International Conference on Computer Science and Electronics Engineering*, vol. 2, pp. 355 - 358, 2012.
- [14] N. Rachburee and W. Punlumjeak, "A Comparison of Feature Selection Approach Between Greedy, IG-Ratio, Chi-Square, and MRMR in Educational Mining," *2015 7th International Conference on Information Technology and Electrical Engineering (ICITEE)*, pp. 420 - 424, 2015.
- [15] L. Wu, Y. Wang, S. Zhang and Y. Zhang, "Fusing Gini Index and Term Frequency for Text Feature Selection," *2017 IEEE Third International Conference on Multimedia Big Data (BigMM)*, pp. 280 - 283, 2017.
- [16] A. H. R. Z. Arifin, M. S. Mubarak, & Adiwijaya, "Learning Struktur Bayesian Networks Menggunakan Novel Modified Binary Differential Evolution pada Klasifikasi Data". *In Indonesia Symposium on Computing (IndoSC)*, 2016.
- [17] W. Gad and S. Rady, "Email Filtering Based on Supervised Learning and Mutual Information Feature Selection," *2015 Tenth International Conference on Computer Engineering & Systems (ICCES)*, pp. 147 - 152, 2015.
- [18] M. S. Islam, F. E. M. Jubayer and S. I. Ahmed, "A Support Vector Machine Mixed With TF-IDF Algorithm to Categorize Bengali Document," *2017 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pp. 191 - 196, 2017.
- [19] T. L. Pao, Y. T. Chen, J. H. Yeh and P. J. Li, "Mandarin Emotional Speech Recognition Based on SVM and NN," *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 1, pp. 1096 - 1100, 2006.
- [20] P. Patil, N. Yaligar and S. Meena, "Comparision of Performance of Classifiers - SVM, RF and ANN in Potato Blight Disease Detection Using Leaf Images," *2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pp. 1-5, 2017.
- [21] D. Fu, B. Zhou and J. Hu, "Improving SVM based multi-label classification by using label relationship," *2015 International Joint Conference on Neural Networks (IJCNN)*, pp. 1 - 6, 2015.
- [22] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features,"



*10th European Conference on Machine Learning Chemnitz, Germany, April 21–23, 1998*

- [23] B. Chen, W. Gu and J.Hu, "An Improved Multi-Label Classification Based on Label Ranking and Delicate Boundary SVM," *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-6, 2010.
- [24] S. Al Faraby, E. R. R. Jasin, A. Kusumaningrum and Adiwijaya, "Classification of hadith into positive suggestion, negative suggestion, and information," in *Journal of Physics: Conference Series*, 2018, vol. 971, no. 1, p. 12046.
- [25] M. S. Mubarak, M. D. Purbolaksono, Adiwijaya. *Implementasi mutual information dan bayesian network untuk klasifikasi data microarray. e-Proceeding of Engineering: Vol.4, No.2 Agustus 2017.*
- [26] S. Sacra, S. Al-Faraby and D. Triantoro, "Klasifikasi Informasi, Anjuran dan Larangan pada Hadits Shahih Bukhari Menggunakan Naive Bayes Classifier". *e-Proceeding of Engineering : Vol.4, No.3 Desember 2017 (page 4794)*
- [27] H. Fauzan, Adiwijaya, & S. Al-Faraby, 2018. Pengklasifikasian Topik Hadits Terjemahan Bahasa Indonesia Menggunakan Latent Semantic Indexing dan Support Vector Machine. *Jurnal Media Informatika Budidarma*. 2(4). pp.131-139.
- [28] Septian, Gleen. *Analisis Sentimen terhadap Provider Telekomunikasi pada Twitter dengan Menggunakan Klasifikasi Naïve Bayes*. Undergraduate thesis, Universitas Kristen Maranatha (2015).
- [29] Baiq, Muhammad Fu'ad Abdul. 2015. "Kumpulan Hadist Sahih Bukhari Muslim". Bandung: Insan Kamil.
- [30] Cambridge University Press. (2009). <http://nlp.stanford.edu>. Retrieved October 29, 2018, from <https://nlp.stanford.edu/IR-book/pdf/irbookprint.pdf>.

## Lampiran

**Tabel 3.** Hasil Skenario Pengujian Keseluruhan

No	Feature Word (Terbaik)	Preprocessing		Fitur Seleksi (MI)	Classification	Hamming loss	Time
		Stop Word	Stemming				
1	50	Yes	Yes	Yes	SVM (linear) - unigram	0.0736	1268.389
2	30	Yes	No	Yes	SVM (linear) - unigram	0.0711	188.272
3	60	No	Yes	Yes	SVM (linear) - unigram	0.0755	1267.149
4	30	No	No	Yes	SVM (linear) - unigram	0.0686	227.525
5	2629	Yes	Yes	No	SVM (linear) - unigram	0.0867	1177.607
6	4307	Yes	No	No	SVM (linear) - unigram	0.0805	252.773
7	4320	No	No	No	SVM (linear) - unigram	0.0814	303.762
8	60	No	No	Yes	SVM (rbf) - unigram	0.0717	251.625
9	30	No	No	Yes	SVM (poly) - unigram	0.0999	262.334
10	50	No	No	Yes	SVM (sigmoid) - unigram	0.0723	219.711
11	30	No	No	Yes	SVM (linear) - bigram	0.1046	1977.118
12	21979	No	No	No	SVM (linear) - bigram	0.1766	15816.575
13	40	No	No	Yes	KNN – unigram	0.0792	200.160
14	4320	No	No	No	KNN – unigram	0.1602	248.706
15	30	No	No	Yes	NB – unigram	0.0792	217.621
16	4320	No	No	No	NB – unigram	0.1102	222.705
17	110	No	No	Yes	NN – unigram	0.0758	1130.345
18	4320	No	No	No	NN – unigram	0.0917	522.567