

BAB I PENDAHULUAN

1.1 Latar Belakang Masalah

Situasi sosial budaya dan alam maupun kebiasaan manusia sudah berkembang, ditambah lagi saat ini ada hal yang bernama teknologi, dan teknologi saat ini berkembang sangat pesat, dari yang bertujuan membantu manusia dalam kehidupan sehari-hari maupun dalam pekerjaannya. Pengaruh teknologi juga masuk kedalam kepercayaan atau agama yang dianut oleh setiap manusia, sebagai contoh dari agama Islam sudah banyak mengambil bagian pada teknologi yang berkembang, dari Al-Qur'an yang dijadikan aplikasi, banyaknya website panduan mengenai Hadits, mengaji, sholat dan game edukasi secara Islamiyah.

Hadits merupakan salah satu dasar dari agama Islam selain Al-Qur'an, Hadits juga merupakan makna dari seluruh sikap, perkataan, dan perbuatan Rasulullah SAW dalam mengajarkan agama Islam[1]. Hadits memiliki berbagai macam makna, ada hadits dengan makna yang sangat jelas dan mudah dipahami sampai yang tidak dapat dipahami maknanya, Hadits juga berisi anjuran serta larangan yang ditunjukkan untuk seluruh umat Islam. Dari penjelasan singkat mengenai Hadits, untuk dapat memudahkan umat Islam dalam mengenali makna dari Hadits, maka perlu dilakukan pengklasifikasian berbagai macam topik yang ada pada Hadits, untuk menyelesaikan masalah klasifikasi topik dalam Hadits, dapat mengangkat penyelesaian menggunakan metode *multi-label classification* untuk menjelaskan bahwa banyak makna yang terkandung dalam teks hadits.

Penelitian ini dibuat untuk melanjutkan penelitian sebelumnya dengan masalah yang sama yaitu, pengklasifikasian topik pada teks Hadits dengan penyelesaian masalah dengan *multi-label classification*, yang telah dilakukan oleh Abu Bakar, dkk pada penelitian sebelumnya menggunakan algoritma *backpropagation neural network* untuk melakukan klasifikasi dan menggunakan algoritma *information gain* untuk mendapatkan hasil akurasi dari algoritma yang digunakan pada model[2].

Pada penelitian ini *dataset* yang digunakan adalah *dataset* dari bukhari yang telah dibuat pada penelitian sebelumnya dengan topik atau *class* pada hadits dalam *dataset* dibagi menjadi 3 yaitu nasihat, larangan, dan informasi dengan 1 teks dari hadits dapat masuk diantara ke 3 topik atau *class* yang telah dibuat.

Tabel 1.1 *Dataset* dari penelitian sebelumnya

Data	Nasihat	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

Pada tabel 1.1 menunjukkan bahwa untuk 1 teks Hadits dapat memiliki lebih dari satu *class* atau topik dengan direpresentasikan dengan nilai 0 untuk menunjukkan bahwa teks tersebut bukan bagian dari topik bahasan yang dinyatakan, dan 1 untuk menyatakan bahwa teks tersebut merupakan bagian dari topik bahasan yang dinyatakan.

Untuk penerapan algoritma pada *multi-label classification* ada berbagai macam algoritma yang dapat digunakan, salah satunya menggunakan algoritma *Tree Augmented Naive Bayes*[3], dan algoritma lain yang juga digunakan pada penerapan *multi-label classification* yaitu, *KNN*[4], *SVM*[5], *Bayesian Network*[6], *Naive Bayes*[6], juga *Multinomial Naive Bayes*[7].

Pada dasarnya algoritma *SVM* dan *KNN* merupakan algoritma yang berfokus kepada kasus pengenalan pola, tetapi dapat juga digunakan untuk *classification single-label*. untuk *SVM* diperlukan algoritma tambahan untuk merubah *multiclass-*

problem menjadi *multi-label classification*[5], kelemahan terbesar dari *SVM* adalah dimana algoritma *SVM* sulit dipakai untuk jumlah sample data yang begitu besar[5].

Digunakannya algoritma *K-NN* dalam penelitian ini, untuk dilakukan percobaan penggunaan algoritma *K-NN* yang diintegrasikan dengan *label powerset* guna untuk mengetahui akurasi. Sebelumnya sudah ada penelitian yang menggunakan *K-NN* dengan *problem transformation* yang berbeda yaitu *Binary Relevance* dengan nilai *F1-Score* sebesar 0.8539 [4] dan *label powerset* namun dengan model algoritma *CNN-KIM* dengan nilai *hamming loss* dengan melakukan *preprocessing stemming* dan dilakukan pembobotan kata dengan *word embedding* sebesar 0.178403 [8] yang pernah diterapkan pada *dataset* kali ini. Karena sifat dari model algoritma *K-NN* yang hanya dapat melakukan klasifikasi data berdasarkan *single-label*[9] membuatnya cocok untuk diintegrasikan dengan *label powerset* yang melakukan *problem transformation* untuk diubah menjadi *multi-class*[8] dan diolah kembali agar menjadi hasil klasifikasi *multi-label*.

Berdasarkan penjabaran latar belakang yang sudah dijabarkan, maka penelitian ini akan menggunakan *dataset* dari hadis shahih bukhari yang kemudian akan dilakukan klasifikasi *multi-label* terhadap 3 topik yang ada yaitu anjuran, larangan dan informasi, dengan menggunakan algoritma *K-NN* sebagai model klasifikasi dan *label-powerset* sebagai *problem transformation* untuk melakukan perubahan pada label agar dapat dilakukan klasifikasi dengan algoritma *K-NN*, dan metode yang digunakan untuk mengetahui hasil evaluasi performansi akhir pada model yang dibuat yaitu dengan metode *hamming loss* untuk pengukuran performansinya. Tujuan dari penelitian ini untuk mengetahui sejauh mana akurasi yang dapat diberikan oleh model algoritma *K-NN* jika diintegrasikan dengan *label powerset* dalam mengklasifikasikan teks hadis yang diusulkan pada penelitian ini.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana menerapkan algoritma *K-NN* yang diintegrasikan dengan *label powerset* pada metode *multi-label classification* yang mampu mengenali label pada sebuah teks Hadits ?
2. Sejauh mana algoritma *K-NN* dapat mengenali label pada sebuah teks Hadits pada metode *multi-label classification* dengan integrasi *label powerset* ?

1.3 Tujuan Penelitian

Berdasarkan batasan masalah diatas maka tujuan dari pelaksanaan penelitian ini adalah sebagai berikut:

1. Membangun sebuah model untuk mengenali label dari sebuah teks Hadits dengan menggunakan algoritma *K-NN* yang diintegrasikan dengan *label powerset*.
2. Menunjukkan tingkat keakuratan algoritma *K-NN* pada *multi-label classification* untuk mengenali label pada sebuah teks Hadits dengan integrasi *label powerset*.

1.4 Batasan Penelitian

Berdasarkan rumusan masalah dan tujuan penelitian maka untuk mewujudkan penelitian yang sesuai dengan masalah yang ada diperoleh batasan-batasan penelitian sebagai berikut:

1. Penelitian ini hanya membangun sebuah model untuk melakukan *multi-label classification* dari teks Hadits.
2. *Dataset* yang digunakan pada penelitian ini merupakan hadist bukhari yang didapatkan dari penelitian sebelumnya.
3. Penelitian ini hanya akan menggunakan penentu akurasi *hamming loss* sebagai penilaian performansi dari model yang dibuat.
4. Penelitian ini menggunakan *label powerset* sebagai *problem transformation* dari *dataset*, agar dapat diolah menjadi *multi-label classification* pada algoritma *K-NN*.

5. Penelitian ini menggunakan algoritma *K-NN* sebagai model untuk *multi-label classification*.

1.5 Manfaat Penelitian

Berdasarkan tujuan penelitian maka dapat disimpulkan manfaat dari penelitian ini sebagai berikut:

1. Bagi keilmuan

Memberikan kontribusi ilmiah di bidang *natural language processing* tentang analisis *multi-label classification* menggunakan *K-NN* yang telah diintegrasikan *label powerset*

2. Bagi akademik

Berkontribusi untuk penelitian di bidang *natural language processing*, membantu memudahkan dalam menganalisis *multi-label classification*.

3. Bagi peneliti.

- 1) Mendapatkan nilai akurasi kemiripan antar teks dan label pada teks hadits dengan algoritma *K-NN* dan telah diintegrasikan dengan *label powerset* yang telah diusulkan.
- 2) Menambah wawasan bagi pembaca yang ingin mempelajari model algoritma *K-NN* dengan diintegrasikan *label powerset* dalam pembuatan *multi-label classification*