

BAB I

PENDAHULUAN

Pada bagian ini berisi bahasan tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, ruang lingkup penelitian, dan sistematika penulisan.

1.1 Latar Belakang Masalah

Pada masa sekarang ini berkembangnya teknologi yang semakin meningkat mengakibatkan peningkatan informasi dalam jumlah yang besar. Peningkatan aliran informasi ini menyebabkan banyaknya penumpukan data. Ketersediaan data yang semakin meningkat setiap harinya salah satunya dihasilkan dari penggunaan teknologi informasi di berbagai bidang kehidupan. Selain data yang bersumber dari teknologi dan informasi juga terdapat data yang merupakan sumber hukum syariat Islam, yaitu hadis. Berbagai macam hadis dengan topik yang berbeda dan perawi yang berbeda menyebabkan banyaknya data hadis. Hal ini menimbulkan kebutuhan untuk memanfaatkan informasi dan pengetahuan yang terkandung di dalam data tersebut. Karena data di dunia nyata semakin berkembang hari demi hari sehingga kumpulan data yang sangat besar dengan sedikit atau tanpa pengetahuan latar belakang dapat diidentifikasi menjadi pola yang menarik dengan pengelompokan atau *clustering*[1].

Clustering teks adalah bagian yang penting dalam metode *text mining*, dan juga merupakan bagian dari data mining. *clustering* teks adalah klasifikasi dokumen tanpa pengawasan, yang membagi koleksi teks menjadi beberapa bagian yang disebut *cluster*, masing-masing *cluster* memiliki kesamaan yang lebih besar daripada yang berada dalam *cluster* yang berbeda[2]. Terdapat berbagai macam metode *clustering*, salah satunya yang digunakan pada penelitian ini adalah metode *clustering* berbasis partisi yaitu metode CLARANS (*Clustering Algorithm Based on Randomized Search*).

Salah satu yang menyebabkan performa buruk dari algoritma *clustering* adalah karena terdapat banyak fitur yang tidak relevan. Karenanya dimensi ruang

fitur yang tinggi merupakan salah satu masalah utama yang harus dipertimbangkan dalam proses pengelompokan teks. Setelah dilakukan preprocessing setiap dokumen direpresentasikan menjadi vektor menggunakan Vector Space Model (VSM). Pada VSM setiap term yang terdapat di dalam dokumen merupakan representasi dari fitur yang berbeda. Semakin besar dokumen maka akan menghasilkan fitur yang semakin banyak, ratusan bahkan ribuan fitur[3]. Banyaknya fitur yang tidak relevan ini dapat mengakibatkan kinerja yang buruk untuk algoritma yang digunakan.

Terdapat banyak teknik reduksi dimensi yang telah dikembangkan oleh para peneliti. Pada penelitian ini digunakan teknik reduksi berupa *feature selection* dan *feature extraction*. Agar mendapatkan fitur yang lebih relevan dilakukan penggabungan dengan metode reduksi yang berbeda yaitu menggunakan metode *modified union* yang merupakan gabungan dari metode *union* dan *intersection*. Sehingga pada penelitian ini dilakukan penggabungan metode reduksi pada *feature selection* dengan *modified union* dan mereduksi kembali fitur dengan *Principal Component Analysis* (PCA) pada *feature extraction* untuk mengurangi tingginya ruang dimensi fitur.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan didapatkan rumusan masalah sebagai berikut:

1. Bagaimana cara mereduksi fitur data terjemahan hadis dengan memadukan metode *modified union feature selection* dan metode *feature extraction*?
2. Bagaimana cara *clustering* data terjemahan hadis menggunakan algoritma CLARANS (*Clustering Algorithm Based On Randomized Search*)?
3. Bagaimana perbandingan hasil *clustering* data terjemahan hadis menggunakan algoritma CLARANS sebelum dan sesudah direduksi?

1.3 Batasan Masalah

Pada skripsi ini terdapat beberapa batasan masalah, batasan masalah yang digunakan diantaranya yaitu:

1. *Dataset* yang digunakan yaitu berupa data terjemahan Bahasa Indonesia dari hadis sahih Bukhori dan Muslim, sebanyak 2175 hadis yang terdiri dari lima kategori hadis yaitu: iman, shalat, zakat, puasa, dan haji.
2. Metode yang digunakan untuk mereduksi fitur adalah metode *Term Variance* (TV) dan metode *Document Frequency* (DF) pada *feature selection* dan metode *Principal Component Analysis* (PCA) pada *feature extraction*.
3. Untuk menggabungkan fitur yang telah direduksi pada *feature selection*, digunakan metode *modified union*.
4. Metode *clustering* yang digunakan adalah metode CLARANS.
5. Metode yang digunakan untuk mengevaluasi hasil *cluster* adalah metode *Davies-Bouldin Index* (DBI) dan *Silhouette Coefficient* (SH).

1.4 Tujuan Penelitian

Adapun tujuan dan manfaat dari skripsi ini antara lain:

1. Dapat mereduksi fitur data terjemahan hadis dengan memadukan metode *modified union feature selection* dan metode *feature extraction*.
2. Dapat melakukan *clustering* data terjemahan hadis menggunakan algoritma CLARANS.
3. Mendapatkan perbandingan hasil *clustering* menggunakan algoritma CLARANS sebelum dan sesudah direduksi.

1.5 Metode Penelitian

Metode yang ditempuh oleh penulis dalam menyelesaikan skripsi adalah sebagai berikut:

1. Studi Literatur

Tahap Studi Literatur merupakan tahap penulis mengumpulkan dan memahami materi yang terkait dengan penelitian ini. Diantaranya *clustering* menggunakan algoritma *Partitioning Around Medoid* (PAM) yang merupakan dasar dari algoritma CLARANS.

2. Penelitian

Pada tahap penelitian penulis melakukan reduksi fitur data terjemahan hadis berupa *feature selection* yang digabungkan dengan metode *modified union* dan direduksi kembali dengan *feature extraction* untuk *clustering* menggunakan algoritma CLARANS, kemudian dilakukan perbandingan hasil *clustering* data terjemahan hadis sebelum dan sesudah dilakukan reduksi.

1.6 Sistematika Penulisan

Berdasarkan sistematika penulisan, skripsi ini terdiri atas lima bab serta daftar pustaka dimana dalam setiap bab terdapat beberapa subbab.

BAB I PENDAHULUAN

Bab ini berisi beberapa hal tentang pendahuluan diantaranya berupa latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, dan sistematika penulisan dari masalah yang dikaji.

BAB II LANDASAN TEORI

Bab ini berisi tentang landasan teori berupa penjelasan tentang teori-teori yang berkaitan dan menunjang dalam penulisan skripsi, seperti *text mining*, *text clustering*, *text preprocessing*, *feature selection*, *document frequency*, *term variance*, *feature extraction*, *principal component analysis*, CLARANS serta metode evaluasi meliputi *davies-bouldin index*, dan *silhouette coefficient*.

BAB III MODIFIED UNION EXTRAKSI FITUR UNTUK ALGORITMA CLARANS PADA DATA TERJEMAH HADITS

Bab ini berisi pembahasan tentang penelitian yang di lakukan dari pengambilan *dataset*, lalu tahap *text preprocessing*, kemudian melakukan reduksi fitur menggunakan metode *Document Frequency* (DF) dan metode *Term Variance* (TV), kemudian dilakukan penggabungan menggunakan metode *modified union*, setelah itu melakukan reduksi lagi menggunakan metode *Principal Component Analysis* (PCA) lalu dilakukan *clustering* menggunakan metode *clustering* CLARANS, dan terakhir melakukan evaluasi dengan metode *Davies-Bouldin Index* (DBI) dan *Silhouette Coefficient* (SC).

BAB IV ANALISIS HASIL *CLUSTERING* DAN REDUKSI DIMENSI FITUR UNTUK *CLUSTERING* CLARANS

Bab ini berisi pemaparan mengenai analisis hasil *clustering* yang sudah dilakukan pada bab sebelumnya. Analisis hasil *clustering* yang dilakukan berupa membandingkan hasil *clustering* sebelum dan sesudah direduksi. Kemudian juga dilakukan analisis pada *dataset* yang digunakan.

BAB V PENUTUP

Bab ini berisi penjelasan mengenai beberapa hal yang menjadi kesimpulan atas penelitian yang telah dilakukan serta beberapa saran pengembangan tulisan ini.

DAFTAR PUSTAKA

