

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pandemi COVID – 19 telah berdampak pada seluruh aspek kehidupan bagi warga dunia, tidak terkecuali Indonesia. Di Indonesia saat ini tercatat lebih dari 418.000 terinfeksi COVID – 19 dengan angka kematian 3.4% angka kematian tersebut lebih tinggi dari global yaitu 2.6% dan wilayah Asia Tenggara 1.6% [1].

Saat ini, Indonesia sedang melakukan program vaksinasi untuk mengendalikan pandemi COVID – 19, vaksin yang digunakan Indonesia beragam namun yang paling banyak digunakan vaksin jenis sinovac, pemerintah Republik Indonesia telah berupaya untuk melakukan sosialisasi program vaksinasi kepada seluruh warga Republik Indonesia baik pada media cetak maupun pada media sosial salah satunya *Twitter*[2] .

Informasi pada sosial media yang beragam, tentunya menjadi tantangan tersendiri untuk mengolah informasi. Salah satu teknik mengolah informasi dengan menggunakan pendekatan *Text Mining*. *Topic Modelling* yang menjadi salah satu teknik dalam *Text Mining*, *Topic Modelling* merupakan model statistik untuk menentukan inti dari topik pada sekumpulan dokumen. Penggunaan identifikasi *Topic Modelling* untuk mengetahui *trending* topik yang menjadi perhatian masyarakat saat ini dengan menggunakan data *Twitter*[3].

Pada penelitian sebelumnya tentang implementasi tatap muka *e-learning* mengadakan tes esai, dimulai dengan memberi pertanyaan, menjawabnya dan menandainya menggunakan metode *Latent Semantic Analysis* (LSA) dengan mengukur relevan kunci jawaban dan jawaban dari siswa[4]. Selanjutnya penelitian membangun mesin pencari khusus untuk mencari informasi tentang sebelas hadis dan lengkap dalam Bahasa Indonesia menggunakan metode *Latent Semantic Analysis* (LSA) dan *cosine similarity* untuk berdasarkan kata kunci yang dimasukkan[5]. Penelitian menganalisa untuk menilai jawaban singkat menggunakan metode *Latent Semantic Analysis* (LSA) dengan mendapatkan nilai korelasi yang tinggi[6].

Pada penelitian ini tahapan *Topic Modelling* terdiri dari *Crawler* data menggunakan *Application Programming Interface* (API) yang disediakan oleh *Twitter*. *Preprocessing* terdiri dari *Cleansing Data*, *Case Folding*, *Tokenizing*, *Stopword Removal*, dan *Stemming* [7] sebagai *Text Preprocessing*, *Term Frequency-Inverse Frequency* (TF – IDF) bertujuan untuk membentuk *vector space model* dari hasil data *Text Preprocessing* pada bagian *cleansing* dan seleksi fitur menggunakan perhitungan *Pearson Correlation* dengan nilai *threshold* lebih dari

0.6 bertujuan untuk menghilangkan *term* yang memiliki nilai korelasi dibawah 0.6. *Topic Modelling* dengan metode *Latent Semantic Analysis* (LSA) dan hasil *Topic Modelling* berupa topik yang ditemukan menggunakan metode tersebut (LSA) [8] dengan topik yang sudah diberi label oleh pakar. Evaluasi dari metode *Latent Semantic Analysis* (LSA) menggunakan *Topic Coherence* untuk mengetahui kualitas topik yang dihasilkan.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, maka masalah yang diteliti pada Skripsi ini adalah membangun sistem yang dapat menganalisis topik pada *platform* sosial media *Twitter* dengan *Topic Modeling*.

1.2.2 Metode Usulan

Metode yang digunakan skripsi ini yaitu, *Latent Semantic Analysis* untuk mencari topik dari tanggapan masyarakat mengenai vaksinasi pada *platform* sosial media *Twitter*.

1.2.3 Pertanyaan Penelitian

- a) Bagaimana pengaruh seleksi fitur menggunakan *Pearson Correlation*?
- b) Berapa nilai *Coherence Score* yang dihasilkan dari *Topic Coherence*?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Penelitian ini memiliki tujuan untuk menganalisa atau mencari informasi berbentuk abstrak berupa topik dari kumpulan dokumen yang diperoleh dari *tweet* pada *platform* sosial media *Twitter* menggunakan metode *Latent Semantic Analysis*.

1.3.2 Manfaat Penelitian

- a) Memudahkan untuk menganalisis teks dari sumber dokumen yang akan disajikan dalam bentuk informasi yang bisa dipergunakan sehingga analisa dan masyarakat mendapatkan informasi penting.

- b) Dapat menilai kinerja teknik *Topic Modelling* terhadap analisis opini dari berbasis data *Twitter* dengan menggunakan *Latent Semantic Analysis*.

1.4 Batasan Masalah

Untuk lebih memusatkan pada ruang lingkup penelitian dan menghindari meluasnya permasalahan serta untuk mengatasi keterbatasan yang dihadapi oleh penulis maka disusun beberapa batasan masalah sebagai berikut:

1. Penelitian difokuskan pada data *twitter* dengan satu *keyword* yakni “vaksinasi” dengan jumlah data 6888 *tweet* dan lebih dari satu *keyword* yakni “kebijakan vaksinasi” dengan jumlah data 50 *tweet* yang diambil dalam rentang waktu antara 12:03 sampai 12:05 pada 06 April 2021.
2. Nilai *input* merupakan nilai *threshold* dengan nilai lebih dari 0.6 pada proses *text-preprocessing* dibagian seleksi fitur menggunakan metode *Pearson Correlation*.
3. Banyaknya topik yang ditentukan oleh pakar yang dihasilkan oleh metode *Latent Semantic Analysis*.

1.5 Metodologi Penelitian

Metodologi penelitian yang digunakan dalam Skripsi ini adalah:

1. Studi Pustaka

Pada tahap ini dilakukan pengumpulan informasi dan referensi yang diperlukan untuk pembuatan perangkat lunak. Informasi yang diperoleh kemudian dipelajari dengan mencari literatur – literatur yang dibutuhkan mengenai permasalahan, untuk menemukan topik yang tersembunyi pada dokumen dengan menggunakan metode LSA untuk mereview opini masyarakat tentang vaksinasi.

2. Perancangan Sistem

Pada perancangan sistem meliputi analisa dan desain sistem yang akan dikembangkan dengan mengacu pada hasil pemahaman sistem dan studi pustaka serta data yang telah diperoleh sebelumnya. Pada tahap ini, desain model basis data dan alur kerja sistem serta beberapa perhitungan metode yang akan digunakan agar dapat memenuhi kebutuhan sistem.

3. Implementasi Sistem

Pada tahap implementasi sistem akan dibangun berbasis web dengan tujuan memberikan akses yang mudah dan lebih memiliki penyimpanan yang lebih luas.

4. Evaluasi Sistem

Untuk evaluasi sistem dapat dilihat pada *output* hasil perhitungan dari *Topic Coherence*.

1.6 Sistematika Penulisan

Secara garis besar materi laporan Skripsi ini terbagi dalam beberapa bab yang tersusun berikut:

BAB I PENDAHULUAN

Bab ini berisikan latar belakang mengenai skripsi, rumusan masalah, tujuan dan manfaat, batasan masalah, metodologi penelitian, dan sistematika penulisan laporan.

BAB II STUDI PUSTAKA

Bab ini menjelaskan tentang teori – teori yang digunakan untuk menjelaskan beberapa tahapan *Topic Modeling*, seperti *Term Frequency-Inverse Document Frequency (TF-IDF)*, *Pearson Correlation*, *Topic Modeling* dan *Topic Coherence*.

BAB III PERANCANGAN SISTEM

Bab ini mendeskripsikan rancangan sistem dipresentasikan melalui diagram alir, termasuk analisa sistem yang didalamnya ada Analisa input, proses, dan output. Menjelaskan tentang dataset yang digunakan, tahapan yang dilalui beserta perhitungan manual metode *Latent Semantic Analysis* dan *Topic Coherence* dan mencantumkan *library* yang disediakan oleh bahasa pemrograman *python* pada proses *Text Preprocessing* bagian seleksi fitur dan *vector space model*.

BAB IV IMPLEMENTASI SISTEM

Bab ini melakukan uji coba menggunakan nilai threshold dengan rentang nilai 0.6 sampai dengan 0.9, pada proses *Text Preprocessing* bagian seleksi fitur dan menghitung evaluasi dari topik yang sudah diberi label oleh pakar.

BAB V PENUTUP

Bab berisi kesimpulan dari hasil uji coba *Topic Modeling* dan saran untuk pengembangan skripsi.