

# BAB 1

## PENDAHULUAN

### 1.1. Latar Belakang

Manusia dan informasi merupakan dua hal yang tak dapat dipisahkan. Informasi digunakan manusia untuk mengembangkan ide-ide baru, mendalami sebuah ilmu tertentu, membuat keputusan, berinteraksi, berkomunikasi, dan menjawab pertanyaan. Kebutuhan manusia akan informasi ini kemudian memunculkan permintaan akan informasi. Di era ini, permintaan akan informasi dapat dipenuhi dengan sangat mudah dan cepat dengan hadirnya berbagai macam teknologi informasi seperti koran, tv, radio, dan internet.

Internet yang menjadi gerbang perlintasan informasi tanpa memperhitungkan letak geografis dan waktu, menyediakan banyak sekali informasi tak terbatas yang dapat dicari dengan menggunakan mesin pencari seperti *google*. Melimpahnya informasi yang dihasilkan oleh mesin pencarian tersebut mengakibatkan potensi pengguna dalam mendapatkan informasi yang tidak relevan dengan apa yang sedang dicari. Terlalu banyaknya hasil pencarian juga mengharuskan pengguna untuk meluangkan waktu lebih lama untuk memilah informasi yang tepat sesuai dengan apa yang sedang dicari.

Permasalahan yang sama juga dialami oleh *chat bot*. selain, mesin pencari, penggunaan *chat bot* saat ini juga dapat dikategorikan sebagai sebuah teknologi informasi yang membantu pengguna untuk mendapatkan informasi. Indonesian Contact Center Association (2017) menjelaskan bahwa *Chat bot* merupakan sebuah program komputer yang dirancang untuk melakukan komunikasi dan interaksi dengan pengguna. Ada dua jenis *chat bot* saat ini ditinjau dari

teknologi yang digunakan yakni *chat bot* dengan sistem aturan dan *chat bot* dengan implementasi kecerdasan buatan. *Chat bot* dengan kecerdasan buatan memungkinkan sistem untuk dapat berkomunikasi dengan pengguna selayaknya manusia. *Chat bot* bekerja dengan memberikan respon jawaban dari pertanyaan yang dimasukkan oleh pengguna. Respon ini dihasilkan dengan cara memindai kata kunci masukkan dan kemudian mencari padanan balasan yang paling cocok atau pola kata-kata yang paling mirip dengan apa yang sudah dideklarasikan pada database sistem. Untuk membantu komputer dalam memahami, menginterpretasi, dan memanipulasi bahasa manusia sehingga mendapatkan informasi tertentu, maka digunakan pendekatan bernama *Natural Language Processing (NLP)*. Saat ini, *chat bot* sudah banyak digunakan untuk membantu mempermudah pekerjaan manusia seperti *command control*, *customer care*, *messaging application*, dan *virtual assistance*.

Dalam penelitian yang dilakukan oleh Fatroni (2011), *Text Mining* adalah proses menambang data berupa teks dengan sumber data biasanya dari dokumen dan tujuannya adalah mencari kata-kata yang mewakili dalam dokumen sehingga dapat dilakukan analisa keterhubungan dalam dokumen. *Text Mining* banyak digunakan untuk menyelesaikan masalah yang membutuhkan pencarian informasi atau biasa dikenal dengan *Information Retrieval (IR)*. *Vector Space Model (VSM)* adalah suatu model yang digunakan untuk mengukur kemiripan antara suatu dokumen dengan suatu *query*. *VSM* merupakan salah satu model matematika yang dapat digunakan pada *IR* untuk menentukan bahwa sebuah dokumen itu relevan terhadap sebuah informasi.

Dalam penelitian yang dilakukan oleh Heninggar Septiantri yang membandingkan metode Latent Semantic Analysis (LSA) dengan VSM mengenai sistem penilai jawaban esai otomatis Bahasa Indonesia, didapatkan dari hasil uji coba bahwa secara keseluruhan rata-rata korelasi nilai VSM – manusia lebih tinggi dibanding LSA-manusia (Septiantri,2009).

Berdasarkan pemaparan latar belakang permasalahan di atas, penulis merasa perlu diadakan penelitian guna menyelesaikan permasalahan terkait dengan sistem temu kembali dengan mengimplementasikan metode *Text Mining* dan algoritma *vector space model* agar dapat memperoleh jawaban atau informasi yang relevan. Oleh karena itu, penulis berencana melakukan penelitian dengan judul “Identifikasi Kecocokan Dokumen Tanya Jawab Menggunakan Metode *Text Mining* dan *Vector Space Model*.”

## **1.2. Rumusan Masalah**

Dari uraian latar belakang di atas maka rumusan masalah dari penelitian ini antara lain:

1. Bagaimana membangun sistem temu kembali dengan menggunakan metode *text mining* dan *Vector Space Model*?
2. Bagaimana tingkat *precision* dan *recall* sistem temu kembali menggunakan metode *text mining* dan *Vector Space Model*?

## **1.3. Batasan Masalah**

Dalam penelitian ini, batasan masalah yang ditentukan penulis adalah:

1. Dokumen yang digunakan adalah dokumen dengan Bahasa Indonesia.

2. Daftar pertanyaan dan jawaban yang dijadikan objek penelitian telah didefinisikan terlebih dahulu pada database oleh penulis.
3. Dokumen pertanyaan dan jawaban yang digunakan adalah hasil wawancara kuisioner mengenai PKL dan skripsi di Teknik Informatika UPN “Veteran” Jawa Timur.

#### **1.4. Tujuan Penelitian**

Berdasarkan rumusan masalah yang dirumuskan sebelumnya, maka tujuan dari penelitian ini dapat dituliskan antara lain:

1. Untuk mengetahui implementasi metode *text mining* dan *Vector Space Model* dalam membangun sistem temu kembali.
2. Untuk mengetahui tingkat *precision* dan *recall* dari sistem temu kembali yang dibangun menggunakan metode *text mining* dan *Vector Space Model*

#### **1.5. Manfaat Penelitian**

Dari penelitian ini, penulis berharap dapat menghasilkan sebuah ilmu baru khususnya pada bidang data mining dalam menerapkan metode *text mining* dan *vector space model* yang digunakan dan dikembangkan dalam pembuatan sistem temu kembali seperti program tanya jawab, *chat bot*, ini.