

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dunia internet semakin berkembang, begitupula penggunaan *email* dalam kehidupan sehari-hari menjadi meningkat. Kita cenderung menganggap *email* sebagai sarana komunikasi baik dalam hal pribadi maupun hal bisnis. Telah ada industri yang berkembang dengan pesat dalam menggunakan email sebagai alat bantu pemasaran beberapa tahun belakangan ini. Faktanya, saat ini *email-email* yang tidak diharapkan seperti iklan, pornografi, jasa dan lain sebagainya menjadi hal yang sangat umum saat ini. Email ini dianggap sebagai *spam* dan dapat meluas dari puluhan hingga ratusan dalam sehari. Dengan begitu *spam* menjadi suatu hal yang mengganggu dalam *inbox email* kita. Sebagai user, biasanya menginginkan email yang berguna, relevan dan dalam waktu yang sesingkat mungkin, tanpa harus melalui kerumitan dalam *browsing inbox* yang diserang *spam*. Pemecahan masalah ini adalah dengan menggunakan *tool email filtering* yang dapat menyaring *email-email* yang tidak diinginkan ini.

Saat ini sudah banyak teknik spam filtering yang dibuat untuk mengatasi spam email ini, seperti rule based filtering, naïve bayesian filtering dan support vector machine. Akan tetapi, *spam* dapat sampai ke *inbox email user* meskipun telah menggunakan *tool* tersebut. Dan dapat mengakibatkan *user* perlu menghabiskan waktu yang lama dalam menghapus *spam email* dari *inbox email*. Masing-masing memiliki kelebihan dan kekurangan, bahkan dengan teknik seperti itu bisa saja email penting tidak bisa masuk ke dalam *inbox user*.

Tugas akhir ini menganalisis dan mengimplementasikan sebuah perangkat *spam email filtering* yang dapat mengkategorisasikan *email* dengan mengidentifikasi *content* dari *email* untuk menentukan *email* mana yang merupakan *spam email*. Tugas akhir ini berfokus pada implementasi dari algoritma *Vektor Space Model* dalam mengkategorisasikan *content* dari *email*. *Vector Space Model* adalah salah satu model *Information Retrieval* {IR}, yang merepresentasikan sebuah dokumen sebagai sekumpulan kata-kata beserta

frekuensinya. Dengan *email filter*, *user* dapat memfokuskan diri pada *email* yang penting, dan mengabaikan *spam email*.

Disamping itu, Vector Space Model memiliki keuntungan berupa skema pembobotannya yang meningkatkan performansi pencarian, strategi pencocokannya mengijinkan pencarian dokumen yang cocok dengan kondisi query, dan rumus ranking kosinus mengurutkan dokumen menurut tingkat kemiripannya dengan *query*.

1.2 Perumusan Masalah

Adapun perumusan masalah pada Tugas Akhir ini adalah :

1. Bagaimana mengimplementasikan spam email filtering menggunakan salah satu teknik information retrieval yaitu Vector Space Model
2. Bagaimana mengimplementasikan spam email filtering yang dapat mengenali jenis dari spam yang terus berkembang.

1.3 Batasan Masalah

Pembahasan masalah pada tugas akhir ini akan dibatasi pada ruang lingkup :

1. Menggunakan salah satu teknik *Information Retrieval* yaitu *Vector Space Model*.
2. Sekumpulan dokumen email spam dan non-spam dari inbox user digunakan sebagai data *indexing* dan data *testing* pada aplikasi *spam email filtering* ini dan sudah dalam bentuk file yang disimpan dalam folder tersendiri.
3. Bagian yang dianalisis adalah bagian *header* dan *body email* yang berupa teks bukan gambar.
4. Menggunakan inputan email dalam bahasa Inggris.
5. Tidak membahas pengintegrasian ke dalam program email client.

1.4 Tujuan Penelitian

Tujuan penelitian dari tugas akhir ini adalah:

1. Mengimplementasikan salah satu teknik *Information Retrieval* yaitu *Vector Space Model* untuk membangun sebuah perangkat *spam email filtering*.
2. Menganalisis keakuratan dari hasil implementasi *spam email filtering* tersebut dalam memfilter email-email spam.

1.5 Metodologi Pemecahan Masalah

Metode yang digunakan dalam penyelesaian Tugas Akhir ini yaitu:

1. Studi Literatur
Mempelajari literatur-literatur tentang email, teknik-teknik *email filtering* dan konsep pada *Information Retrieval* khususnya *Vector Space Model*.
2. Pengumpulan dan Analisis Data.
Mengumpulkan dokumen *email spam* dan *non-spam* yang akan digunakan sebagai data dalam proses indexing dan proses querying untuk aplikasi. Data tersebut diperoleh dari email yang terdapat pada *inbox outlook express* user tertentu.
3. Analisis dan Perancangan Sistem
Menganalisis kebutuhan sistem. Dari hasil analisis tersebut dibuat perancangan aplikasinya.
4. Implementasi
Mengimplementasikan hasil dari perancangan sistem dengan membangun aplikasi *spam email filtering* yang dapat memfilter dokumen email menjadi *spam* dan *non-spam* dengan menggunakan teknik *Vector Space Model*.
5. Pengujian Sistem dan Analisis Fungsi Hasil Implementasi.
Melakukan pengujian terhadap parameter performansi *email filtering*, hasil pengujian kemudian dianalisis untuk mendapatkan kesimpulan akhir.
6. Pengambilan Kesimpulan
7. Pendokumentasian Tugas Akhir.

1.6 Sistematika Penulisan

Tugas Akhir ini akan disusun dengan sistematika sebagai berikut :

Bab I Pendahuluan.

Bab ini memberikan gambaran secara garis besar tentang Tugas Akhir yang dilakukan penulis. Mencakup latar belakang pembuatan Tugas Akhir, perumusan masalah, pembatasan masalah, tujuan, metodologi pemecahan masalah dan sistematika penulisan.

Bab II Landasan Teori.

Bab ini menjelaskan seluruh teori yang menjadi landasan konseptual dan pendukung penyelesaian Tugas Akhir ini yaitu teori tentang mekanisme *Vektor Space Model*

Bab III Analisis Kebutuhan dan Perancangan Sistem.

Menjelaskan tentang proses analisis masalah dan kebutuhan perangkat lunak. Berisi rancangan yang meliputi proses perancangan perangkat lunak dengan menggunakan *Vektor Space Model*.

Bab IV Implementasi dan Analisis Pengujian.

Mengimplementasikan dan analisis pengujian dari proses *filtering email* dengan menggunakan *Vektor Space Model*.

Bab V Kesimpulan dan Saran.

Bab ini berisi kesimpulan dari keseluruhan sistem yang dibuat serta saran yang berkaitan dengan aplikasi Tugas Akhir ini dan kemungkinan pengembangan selanjutnya.