

Dengan semakin banyak orang yang terhubung ke internet, *electronic mail (email)* menjadi salah satu bentuk komunikasi yang paling cepat dan ekonomis. Selain dimanfaatkan untuk berhubungan antar teman atau kolega, juga digunakan sebagai salah satu media penyebaran berita dalam bidang *electronic commerce (e-commerce)*. Sebuah iklan komersial yang dikirimkan kepada sejumlah besar orang tanpa persetujuan dari orang tersebut disebut sebagai *unsolicited commercial email (UCE)*, *spam email*, *junk mail*, *bulk mail* atau *email sampah*. Mengirimkan *email spam* adalah sebuah pelanggaran terhadap *Acceptable Use Policy (AUP)* atau peraturan penggunaan yang bisa diterima pada hampir semua *Internet Service Providers (ISP)*, dan bisa menyebabkan penghapusan pada *account* pengirim. Pada sebagian besar yurisdiksi, *spamming* adalah sebuah kejahatan atau tindakan ilegal, seperti pada Amerika Serikat, dimana tindakan tersebut diatur dalam *CAN-SPAM Act of 2003* [3]. Di Indonesia sendiri hal ini belum diatur menjadi sebuah perundang-undangan.

Dari *email* kita dapat mengambil beberapa ciri utama yang dapat membedakan antara *email* yang dibutuhkan (*legitimate*) dan *email spam* yang tidak dibutuhkan. *Email spam* sendiri biasanya terdapat kata-kata spesifik yang menandakan bahwa *email* tersebut adalah *email spam*. Tetapi dimungkinkan juga kata-kata tersebut terdapat pada *email-email legitimate*. Dengan bermodalkan kata-kata seperti ini kita dapat membangun sebuah *email filter* dengan menggunakan metode statistik yang dapat secara otomatis mengklasifikasikan *email* sejenis. Banyak terdapat metode-metode *email filtering* yang telah dikembangkan, contohnya adalah *Host-based filtering*, *Rule-based filters*, *Statistical filter*, *White list*, dll. Salah satu yang termasuk dalam *statistical filter* untuk pendeteksian *email spam* adalah metode pembeda markov [11, 8]. Dimana metode ini memiliki salah satu keuntungan yaitu memiliki kemampuan untuk terus belajar dari *email* yang dikategorikan sebagai *email spam* atau *email legitimate*. Metode pembeda markov menjalankan konsep bahwa susunan kata-kata jauh lebih penting dibandingkan dengan kata secara individu.

Dengan menggunakan filter yang otomatis ini membuat *email spam* dapat dipisahkan dengan *email legitimate*. Yang pada akhirnya akan mengurangi beban *mail server* sendiri ataupun pengguna *mail server* tersebut. Aplikasi ini akan menggunakan inputan *email* berupa file teks. Sedangkan *email* yang mengandung gambar, file-file attachment dan file-file lainnya akan dibaca sebagai file teks juga. Pada proses pengujian, *email-email* yang akan digunakan tidak didapat secara langsung dari internet, tetapi dikelompokkan menjadi sekumpulan data set yang disimpan pada folder-folder tertentu.

1.2 Rumusan Masalah

Permasalahan dalam *email filtering* adalah :

- Bagaimana untuk mendapatkan akurasi yang tinggi dalam klasifikasi *email spam*.
- Bagaimana cara menentukan *email* menjadi *spam* atau *legitimate*.
- Bagaimana menentukan *threshold* yang digunakan batas penentu sebuah *email* dikatakan sebagai *email spam*.

Pada tugas akhir ini dibuat aplikasi *spam filtering* dengan mengimplementasikan metode pembeda markov sebagai metode klasifikasinya. Aplikasi ini diharapkan dapat menjawab permasalahan di atas.

1.3 Batasan Masalah

- Menggunakan metode pembeda markov sebagai metode pengklasifikasian.
- Aplikasi ini menggunakan inputan *email* yang berupa teks.
 - Aplikasi tidak menerima *email* secara langsung dari internet, tetapi dari sekumpulan data set yang berupa file teks yang telah disimpan pada folder tertentu.
- Aplikasi ini berdiri sendiri dan tidak menempel pada server *email*.

1.4 Tujuan Tugas Akhir

- Membuat aplikasi *spam filtering* berdasarkan metode pembeda markov yang digunakan sebagai filter *email spam*.
- Menganalisis hasil klasifikasi yang dihasilkan oleh filter *email spam* dengan metode pembeda markov.
- Menganalisis pengaruh perubahan dari nilai *threshold* pada metode pembeda markov.

1.5 Metodologi Penelitian

1.5.1 Studi literatur.

Mengumpulkan bahan-bahan referensi Tugas Akhir yang meliputi topik-topik metode pembeda markov, rekayasa perangkat lunak, dan topik lainnya yang mendukung penyusunan Tugas Akhir ini. Sumber referensi berupa buku, Tugas Akhir, jurnal, dan *e-book*. Referensi ini akan digunakan sebagai dasar teori penyusunan Tugas Akhir.

1.5.2 Pengumpulan Data

Data yang dibutuhkan dalam penyusunan Tugas Akhir ini adalah *email* yang akan digunakan dalam proses training. Data tersebut diperoleh dari *email* yang terdapat pada *email server* tertentu. Dari data yang diperoleh maka akan didesain Perangkat Lunak yang nantinya akan digunakan untuk memproses data.

1.5.3 Pengembangan Perangkat Lunak:

1. Analisis Kebutuhan Sistem

Pembuatan perancangan perangkat lunak

2. Pengkodean

Implementasi perancangan menjadi perangkat lunak

3. Pengujian Perangkat Lunak

Perangkat lunak diuji dengan menggunakan data dengan beberapa kasus uji

4. Evaluasi Perangkat Lunak

Jika terdapat kekurangan pada perangkat lunak yang dihasilkan, maka akan dievaluasi untuk memperbaiki kinerja dari aplikasi *spam filtering*.

5. Analisis hasil pengujian

Hasil pengujian dianalisis untuk proses *debugging*.