

1. Pendahuluan

1.1 Latar Belakang

Teknologi semakin berkembang seiring dengan berjalannya waktu. Disadari atau tidak, sebagian besar kehidupan ini dibantu oleh teknologi dan banyak sekali manfaat yang bisa diperoleh dengan teknologi yang ada masa kini. Salah satu teknologi yang sering dan bahkan setiap hari digunakan adalah internet. Internet bisa memberikan banyak informasi dan manfaat. Mulai dari mencari data yang dibutuhkan untuk perkuliahan, jejaring sosial, dan mengirim *e-mail* ke orang lain.

E-mail (Electronic Mail) adalah media komunikasi melalui internet yang dapat digunakan untuk saling bertukar data, untuk berdiskusi melalui *mailist*, menginformasikan sesuatu atau mengiklankan produk tertentu dari perusahaan kepada penerima *e-mail*. Mengirim *e-mail* ke orang lain bisa dilakukan dengan berbagai layanan yang disediakan baik yang gratis maupun berbayar. Dalam sebuah *e-mail*, seringkali didapatkan *e-mail* yang tidak diinginkan atau yang bernama *spam*.

Spam-mail dapat didefinisikan sebagai “*unsolicited bulk email*” yaitu *e-mail* yang dikirimkan kepada ribuan penerima (*recipient*)[11]. *Spam-mail* biasanya dikirimkan oleh suatu perusahaan untuk mengiklankan suatu produk, hal ini menyebabkan semakin padatnya *queue* atau antrian dari *mail server* yang telah digunakan.

E-mail spam adalah salah satu jenis *spam* berupa pesan-pesan identik yang dikirim secara massal ke sejumlah penerima dengan *e-mail*. *Spam* cukup mengganggu kenyamanan dalam menggunakan *e-mail*. Biasanya *spam* dikirim oleh suatu perusahaan yang bertujuan untuk mengiklankan produknya dan penerima *spam* berjumlah sangat banyak karena pengiriman dilakukan dengan memasukkan banyak penerima sekaligus. Bahkan ada juga *spam* yang masuk ke *inbox e-mail*, walaupun sudah ada bagian *spam* sendiri, namun masih saja hal itu terjadi dan ini cukup mengganggu pengguna *e-mail*. Beberapa kerugian *spam* antara lain adalah semakin berkurangnya *bandwidth* yang digunakan, pemborosan

waktu dan biaya bagi *user* yang menggunakan koneksi *dial up* untuk berinternet, serta kemungkinan untuk mengakses situs-situs berbahaya yang tidak diinginkan akan semakin besar.

Dalam tugas akhir ini digunakan metode SVM (*Support Vector Machine*) dan *Naïve Bayes*. Berdasarkan penelitian yang telah dilakukan, dua metode ini dipilih karena kinerja yang dihasilkan bagus, tingkat keakuratan yang dinilai cukup tinggi untuk pengklasifikasian data, dan *error rate* yang dihasilkan minimum[8][15]. Diharapkan akan dengan mudah mengklasifikasikan mana *e-mail* yang termasuk *spam* dan bukan *spam*.

Kelebihan dari metode SVM adalah SVM dapat menentukan *hyperplane* atau bidang pemisah dengan memilih bidang dengan optimal *margin* maka generalisasi pada SVM dapat terjaga dengan sendirinya, tingkat generalisasi pada SVM tidak dipengaruhi oleh jumlah data latih, dengan menentukan parameter *soft margin*, *noise* dapat dikontrol sehingga makin besar parameter *soft margin*, makin besar pula pinalti yang dikenakan pada kesalahan pada klasifikasi sehingga proses pelatihan semakin ketat[16]. Kekurangan SVM adalah SVM menggunakan fungsi pemisah yang memisahkan data ke dalam dua kelas, jika kelas yang ingin dipisahkan lebih dari dua maka dibutuhkan modifikasi, ketika data latih yang digunakan sangat besar, mempengaruhi waktu pelatihan dan ukuran *memory* yang dibutuhkan[16].

Kelebihan metode *Naïve Bayes* adalah memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter (rata-rata dan variansi dari fitur) yang dibutuhkan untuk klasifikasi, menangani nilai yang hilang dengan mengabaikan instansi selama perhitungan estimasi peluang, waktu pelatihan cepat dan hanya membutuhkan ukuran *memory* yang kecil, mampu menangani data kuantitatif dan diskrit[17]. Kekurangan dari *Naïve Bayes* adalah metode ini tidak berlaku jika probabilitas kondisionalnya adalah nol, apabila nol maka probabilitas prediksi akan bernilai nol juga, *Naïve Bayes* mengasumsikan independensi fitur[17].

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan sebelumnya, maka rumusan masalah dalam pengerjaan tugas akhir ini adalah :

- a. Bagaimana mengimplementasikan metode SVM dan *Naïve Bayes* untuk pengklasifikasian *spam e-mail*.
- b. Bagaimana mengukur kinerja dan keakuratan metode SVM dan *Naïve Bayes* dalam pengklasifikasian *spam e-mail*.

Adapun beberapa batasan masalah dalam tugas akhir ini adalah :

- a. Pengklasifikasian hanya untuk konten *e-mail* dengan bahasa Inggris dan tidak mencakup *attachment* pada *e-mail* tersebut.
- b. Implementasi dilakukan pada program dan bukan di *e-mail server* dengan data *e-mail* yang diambil dari *repository* penyedia data UCI *Machine Learning Repository*.

1.3 Tujuan

Tujuan dari tugas akhir ini adalah sebagai berikut :

- a. Merancang, mengimplementasikan, dan membandingkan metode SVM dan *Naïve Bayes* dalam pengklasifikasian *spam e-mail*.
- b. Menganalisis kinerja sistem dalam pengklasifikasian *spam e-mail* menggunakan metode SVM dan *Naïve Bayes*.

1.4 Metodologi Penyelesaian Masalah

Metodologi penelitian yang akan digunakan dalam penyelesaian tugas akhir ini adalah:

a. Studi Literatur

Mencari, mengumpulkan, dan mempelajari informasi referensi yang bersumber dari buku, jurnal maupun sumber lain dari internet sebagai landasan teori. Khususnya referensi yang berkaitan dengan metode SVM dan *Naïve Bayes*. Tahap ini dilakukan selama pengerjaan tugas akhir berlangsung.

b. Pengumpulan Data

Mencari dan mengumpulkan data berupa *e-mail* yang akan digunakan dalam pelatihan dan pengujian pengklasifikasian *spam e-mail* pada tahap implementasi.

c. Perancangan Sistem

Merancang alur sistem pengklasifikasian *spam e-mail* menggunakan metode SVM dan *Naïve Bayes*.

d. Implementasi Sistem

Mengimplementasikan rancangan sistem sesuai dengan hasil perancangan yang telah dilakukan dalam proses pengklasifikasian.

e. Analisis Hasil Implementasi Sistem

Menganalisis hasil implementasi berupa parameter-parameter metode SVM dan *Naive Bayes* beserta kinerja sistem dalam penyelesaian tugas akhir ini.

f. Pembuatan Laporan Tugas Akhir

Mendokumentasikan penyelesaian tugas akhir ini kedalam bentuk laporan tertulis.

1.5 Sistematika Penulisan

Sistematika penulisan dalam penyusunan Tugas Akhir ini direncanakan sebagai berikut:

BAB 1 PENDAHULUAN

Pada bab pertama ini akan dijelaskan latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, metodologi penelitian, sistematika penulisan, dan jadwal kegiatan.

BAB 2 DASAR TEORI

Pada bab kedua ini akan diuraikan dasar-dasar teori yang mendukung penyelesaian tugas akhir, khususnya teori yang berkaitan dengan praproses dan metode klasifikasi pada proses pengklasifikasian *spam e-mail*.

BAB 3 ANALISIS PERANCANGAN DAN IMPLEMENTASI

Pada bab ketiga ini akan dijelaskan mengenai proses analisis perancangan dan implementasi yang dibangun secara terperinci.

BAB 4 ANALISIS HASIL PENGUJIAN

Pada bab keempat ini akan dijelaskan hasil dari analisis perancangan dan implementasinya mengenai parameter-parameter dan kinerja sistem pada metode yang digunakan dalam sistem yang telah dibangun pada bab sebelumnya.

BAB 5 PENUTUP

Pada bab terakhir ini akan diuraikan kesimpulan terhadap hasil analisa yang telah dilakukan berikut saran-saran untuk pengembangan tugas akhir selanjutnya.