

1. Pendahuluan

1.1 Latar belakang

E-mail spam merupakan pesan yang tidak diminta yang dikirim ke sejumlah penerima melalui *e-mail* dalam jumlah yang besar. Contoh kiriman *e-mail* yang merupakan *spam* seperti iklan, advertensi, tawaran untuk bergabung ke MLM, undian, informasi palsu, *phishing*, dan penipuan [8]. Dengan adanya *e-mail spam* ini mengakibatkan pengguna *e-mail* membutuhkan waktu yang lebih banyak untuk membaca dan memutuskan apakah *e-mail* yang diterima tersebut adalah *spam* atau bukan. Seperti pada laporan hasil studi lembaga riset ICF International yang menyatakan bahwa pada tahun 2008, 80% energi dihabiskan oleh pengguna untuk menghapus *spam* dan mencari *e-mail* bukan *spam* [9]. Untuk menghadapi masalah ini banyak pendekatan yang diterapkan diantaranya *extensions of e-mail protocols*, *certification of e-mail server*, *e-mail spam filtering*, dan *legislation*, diantara pendekatan-pendekatan tersebut, *e-mail spam filtering* adalah solusi yang lebih realistis dalam implementasinya [12].

E-mail spam filtering adalah sebuah mekanisme yang digunakan untuk memisahkan *e-mail spam* dengan *e-mail* bukan *spam* secara otomatis [11]. Salah satu teknik yang digunakan dalam *E-mail spam filtering* yaitu teknik klasifikasi [5]. Klasifikasi adalah salah satu teknik dalam data *mining* yang digunakan untuk memprediksi kelompok keanggotaan(*class*) dari setiap *instance* data. Salah satu teknik *machine learning* untuk klasifikasi yang paling sering digunakan untuk membangun *e-mail spam filtering* adalah SVM.

Support Vector Machines (SVM) adalah salah satu teknik *machine learning* yang digunakan untuk klasifikasi data dan umumnya diimplementasikan untuk menangani *dataset* yang hanya memiliki dua kelas [17][15][2]. SVM banyak digunakan dalam pembangunan sistem *e-mail spam filtering* karena telah terbukti efisien dan akurasi yang tinggi [20][13]. Akan tetapi performansi dari SVM akan menurun ketika *dataset* yang digunakan mengandung banyak atribut sehingga ketika dipetakan ke dalam ruang vektor menimbulkan *curse of dimensionality*. Kondisi lain yang mengakibatkan performansi SVM menurun yaitu jika suatu *dataset* jumlah *class* yang satu dengan yang lainnya sangat berbeda jauh. Kondisi tersebut menyebabkan *class* yang sedikit tersebut dianggap *outlier*. Untuk mengatasi hal tersebut, maka banyak dilakukan modifikasi terhadap SVM yang bertujuan untuk meningkatkan efektifitas dan efisiensi SVM[5].

Salah satu modifikasi pada SVM yaitu penggunaan paradigma *granular computing* dan teori statistik, yang kemudian penggabungannya disebut *Granular Support Vector Machines* (GSVM) [18]. Dalam pembangunan GSVM sendiri, banyak algoritma-algoritma yang diterapkan sesuai dengan tujuan pengimplementasian SVM itu sendiri. Salah satunya adalah algoritma *Recursive Feature Elimination* (RFE). Tahap awal algoritma ini akan menghilangkan *feature-feature* yang *irrelevant* dan *redundant* secara iteratif sehingga mencapai jumlah tertentu. Dan dari hasil tersebut akan dilakukan pemilihan *feature* yang paling merepresentasikan *email spam*. [18].

Pada tugas akhir ini akan dibangun sebuah sistem *e-mail spam filtering* menggunakan *Granular Support Vector Machine* dengan *Recursive Feature*

Elimination (RFE) karena algoritma ini dapat menghasilkan *feature-feature* yang lebih akurat atau informative untuk pengklasifikasian *email spam* [18].

1.2 Perumusan masalah

Berdasarkan latar belakang tersebut, maka dapat dirumuskan permasalahan sebagai berikut:

1. Bagaimana membangun sebuah sistem *e-mail spam filtering* dengan *Granular Support Vector Machines – Recursive Feature Elimination (RFE)*.
2. Bagaimana akurasi dan efisiensi waktu *Support Vector Machines* dimodifikasi dengan *granular computing* dengan algoritma *Recursive Feature Elimination* terhadap *e-mail spam filtering* yang akan dibangun.

Terdapat beberapa batasan masalah dalam penelitian tugas akhir ini, antara lain :

1. Data yang digunakan adalah data yang tidak mengandung gambar dan bahasa yang digunakan adalah bahasa inggris.
2. Sistem yang dibangun adalah aplikasi yang berdiri sendiri (*stand alone application*) dan tidak diimplementasikan dalam *e-mail server*.

1.3 Tujuan

Tujuan dari penelitian tugas akhir ini adalah:

1. Mengimplementasikan teknik *Granular Support Vector Machine with Recursive Feature Elimination (RFE)* untuk mengklasifikasikan *email* berdasarkan kriteria *spam* atau bukan *spam*.
2. Melakukan analisa akurasi pada sistem yang telah dibangun dengan parameter *sensitivity, specificity*.

1.4 Metodologi penyelesaian masalah

Metode yang digunakan untuk menyelesaikan tugas akhir ini adalah :

1. Identifikasi masalah, yakni dengan melakukan identifikasi terhadap permasalahan yang ada.
2. Studi Literatur, yakni mempelajari referensi dan literatur, baik berupa makalah, jurnal, maupun buku yang relevan yang membahas tentang *Granular Support Vector Machine with Recursive Feature Elimination (RFE)*.
3. Mempersiapkan data set yang akan digunakan untuk testing dan training dengan melakukan preprocessing terhadap data set tersebut.
4. Pembuatan desain sistem *e-mail spam filtering* dengan *Granular Support Vector Machine with Recursive Feature Elimination (RFE)*.
5. Implementasi (*Coding*), yaitu mengimplementasikan perancangan menjadi sistem *email spam filtering* dengan menerapkan *Granular Support Vector Machine with Recursive Feature Elimination (RFE)*.
6. Training dan testing sistem, melakukan pelatihan dan pengujian pada sistem dengan menggunakan data training dan data testing.
7. Analisa hasil, melakukan analisa hasil dari sistem dengan cara membandingkan hasil klasifikasi data testing dengan data jawaban sebenarnya.

8. Pembuatan laporan, mendokumentasikan semua tahap metodologi penyelesaian masalah menjadi suatu laporan yang nantinya dapat dikembangkan sesuai perkembangan jaman dan dapat dimanfaatkan sebagaimana mestinya.