BAB 1
PENDAHULUAN

1.1. Latar Belakang Masalah

Spam adalah penyalahgunaan sistem pesan elektronik (termasuk sebagian besar media penyiaran, sistem pengiriman digital) untuk mengirim pesan massal yang tidak diminta oleh penerimanya. *Email spam* adalah salah satu jenis *spam* berupa pesan-pesan identik yang dikirim secara massal ke sejumlah penerima dengan menggunakan *email.* Informasi yang tersedia pada *website* Kaspersky Lab terdapat 55.28% *spam* yang terkirim di seluruh dunia pada tahun 2015. Angka tersebut menurun 11.48% dari tahun sebelumnya [1].

Spam yang terkirim kepada pengguna layanan *email* dapat menimbulkan masalah berupa menghabiskan kapasitas penyimpanan dan menghabiskan waktu pengguna untuk menghapus *spam* tersebut. *Spam filter* muncul untuk mengatasi permasalahan tersebut yang akan mendeteksi keberadaan *spam* sehingga tidak dimunculkan pada *inbox mail. Spam filter* telah banyak dikembangkan dengan berbagai macam metode, salah satunya menggunakan fungsi klasifikasi Naïve Bayes. Naïve Bayes banyak digunakan sebagai teknik dalam klasifikasi oleh pengembang *spam filter* baik oleh pengembang *open-source* maupun komersial karena kesederhanaan algoritmanya dan mudah dalam mengimplementasikannya [2]. Namun dengan semakin berkembangnya jenis *email spam*, akurasi dari metode *spam filter* harus dapat dikembangkan dan disesuaikan dengan perkembangan *email spam*.

Kelemahan metode Naïve Bayes yaitu hanya dapat menghitung peluang perkata dari *subject* saja yang berakibat keterkaitan antar kata atau pola pada *subject email* tidak diperhatikan [3]. Kelemahan lain dari metode Naïve Bayes yaitu memunculkan nilai 1 dalam perhitungan peluang *email spam* sehingga dapat menimbulkan keraguan untuk hasil perhitungan peluang yang didapatkan. Kelemahan yang dimiliki Naïve Bayes mengakibatkan metode ini dapat menurunkan akurasi dari klasifikasi *email spam*. Metode untuk menangani masalah perhitungan peluang perkata akan menggunakan metode Teori

Kombinasi karena dalam kasus ini metode digunakan untuk menemukan kombinasi antar kata yang diperkirakan dapat membentuk *subject* dari *email spam*. Metode *Smoothing* dalam Naïve Bayes berfungsi untuk menangani munculnya nilai 1 dalam Naïve Bayes. Penggunaan metode *Smoothing Dirichlet* Naïve Bayes dalam penelitian ini karena memiliki akurasi lebih baik dibandingkan dengan metode *Smoothing Jelinek*, metode *Smoothing Absolite Discounting* dan metode *Smoothing Two-Stage* [3]. Dengan menggabungkan Teori Kombinasi dan Metode *Smoothing Dirichlet* Naïve Bayes diharapkan dapat menangani kelemahan dari Metode Naïve Bayes sehingga dapat meningkatkan akurasi klasifikasi *email spam*.

Penelitian ini akan menganalisis klasifikasi *email spam* menggunakan kombinasikan Teori Kombinasi dan metode *Smoothing Dirichlet* Naïve Bayes dengan tujuan meningkatkan akurasi pengklasifikasian *Email spam*.

1.2. Perumusan Masalah

Berdasarkan latar belakang masalah yang diuraikan, maka dapat dirumuskan masalah dalam penelitian ini yaitu Bagaimana hasil akurasi pengklasifikasian *Email Spam* menggunakan Teori Kombinasi dan *Metode Smoothing Dirichlet* Naïve Bayes.

1.3. Maksud dan Tujuan

Maksud dari penelitian ini adalah untuk menganalisis akurasi pengklasifikasian *email* menggunakan Teori Kombinasi dan Metode *Smoothing Dirichlet* Naïve Bayes.

Tujuan yang akan dicapai dalam penelitian adalah untuk mengetahui akurasi klasifikasi *email spam* menggunakan Teori Kombinasi dan Metode *Smoothing Dirichlet* Naïve Bayes dengan membandingkan akurasi dari klasifikasi *email spam* menggunakan Metode *Smoothing Dirichlet* Naïve Bayes saja.

1.4. Batasan Masalah

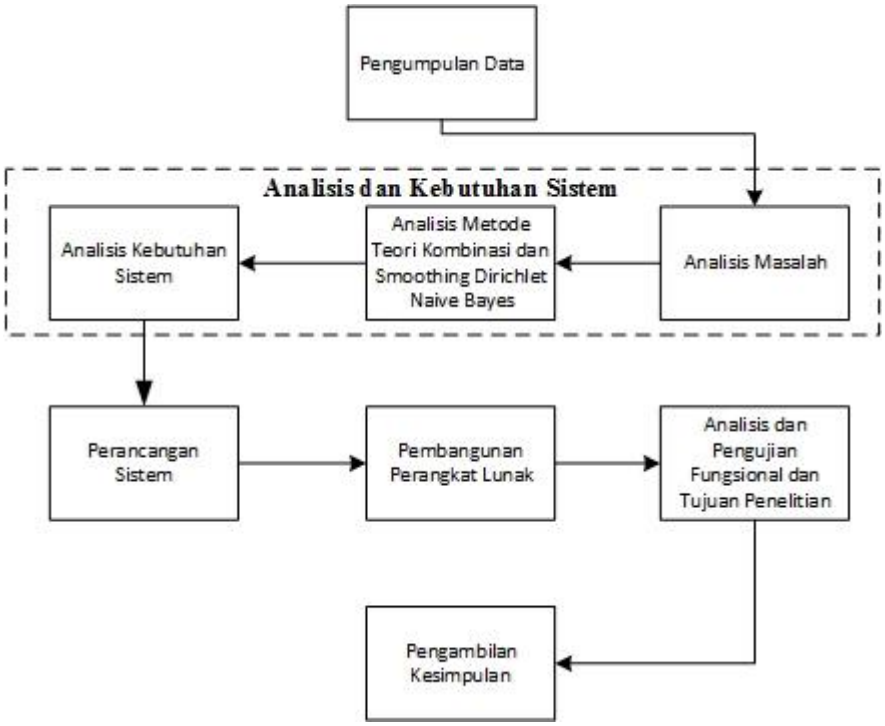
Pembahasan permasalahan diharapkan tidak menyimpang dari pokok permasalahan, sehingga diperlukan batasan masalah. Adapun batasan dari penulisan tugas akhir ini adalah sebagai berikut:

a. Data yang digunakan dalam penelitian adalah korpus *email* publik yang tersedia pada halaman *web Spam*assassin dengan kode prefix "20030228" berbahasa Inggris.
b. Korpus yang digunakan adalah dokumen *email* dengan standar MIME dalam format *raw*.
c. Tidak dilakukan proses *stemming* karena akan mengubah makna kata dalam *subject email*.
d. Bahasa untuk sistem menggunakan Bahasa Indonesia.

1.5. Metodologi Penelitian

Metodologi yang digunakan adalah kuantitatif. Metode kuantitatif adalah metode penelitian yang digunakan untuk meneliti pada populasi atau sampel tertentu, pengumpulan data menggunakan instrumen penelitian, analisis data bersifat kuantitatif/statistik, dengan tujuan untuk menguji hipotesis yang telah ditetapkan.

Alur dari penelitian yang akan dilakukan terdapat di *Block Diagram* pada Gambar 1.1.



Gambar 1. 1 Block Diagram Alur Penelitian

1.5.1. Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah Studi Literatur. Studi Literatur merupakan pengumpulan data dengan cara mengumpulkan literatur, jurnal, *paper* dan bacaan-bacaan yang ada kaitannya dengan Teori Kombinasi dan *Smoothing Dirichlet* Naïve Bayes.

1.5.2. Metode Analisis dan Kebutuhan Sistem

Analisis dan kebutuhan sistem yaitu kegiatan melakukan analisis pada metode yang digunakan yaitu Teori Kombinasi dan Metode *Smoothing Dirichlet* Naïve Bayes untuk melakukan klasifiskasi *email spam* dan melakukan penyusunan tahapan proses yang akan dilakukan.

1.5.3. Metode Perancangan Sistem

Metode perancangan sistem yang digunakan dalam penelitian dan sistem yang akan dibangun adalah perancangan *Object Oriented Analys Design*.

Perancangan *Object Oriented Analys Design* merupakan perancangan yang berorientasi obyek, sehingga perancangan sistem akan menggunakan *Use Case Diagram*, *Use Case Scenario*, *Activity Diagram*, *Class Diagram* dan *Sequence Diagram*.

1.5.4. Metode Pembangunan Perangkat Lunak

Metode pembangunan perangkat lunak yang digunakan adalah model *Prototype.* Berikut merupakan tahapan-tahapan proses pengembangan yang terdapat dalam model *prototype*:

1. Analysis

Tahap ini merupakan tahap dimana pengembang mengumpulkan berbagai macam informasi serta kebutuhan-kebutuhan sistem yang nantinya akan membantu pengembang dalam menentukan garis besar dari sistem yang akan dibuat.

2. Build/ Revise Mock-Up

Tahap ini merupakan tahap pembuatan atau tahap pengimplementasian informasi dan kebutuhan sistem yang sudah didapatkan sebelumnya. Pada tahap ini pengembang mencoba menerjemahkan seluruh informasi dan kebutuhan kedalam logika-logika program.

Pada tahap ini juga dilakukan perbaikan atau revisi dari pengimplementasian yang tidak sesuai atau kurang sesuai dengan informasi dan kebutuhan yang didapat.

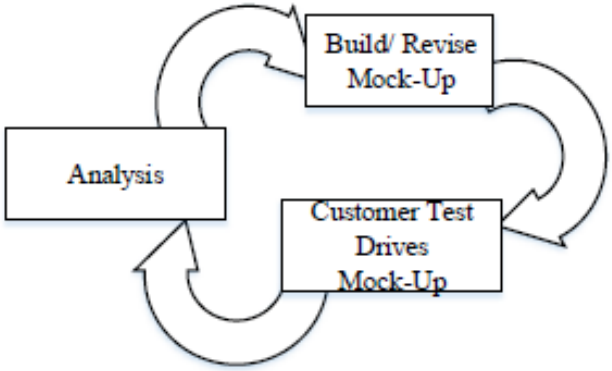
3. Customer Test Drives Mock-Up

Tahap ini merupakan tahap uji coba, dimana uji coba dilakukan untuk mengetahui apakah informasi dan kebutuhan yang didapat sudah seluruhnya diimplementasikan kedalam sistem atau belum. Dan pada tahap ini juga diuji apakah keluaran dari sistem sudah seperti yang diharapkan atau tidak.

Jika keseluruhan hasil uji coba pada tahap ini masih memiliki kekurangan, maka *prototype* direvisi atau diperbaiki dengan tahapan-tahapan yang sebelumnya,

proses ini terus berulang sampai *prototype* mengeluarkan hasil yang diinginkan atau *prototype* sudah terbangun dengan minim kesalahan.

Secara garis besar model *Prototype* dapat dilihat pada Gambar 1.2 [4]



Gambar 1. 2 Prototype Model

1.5.5. Metode Pengujian Perangkat Lunak dan Tujuan Penelitian

Metode pengujian perangkat lunak yang digunakan yaitu metode *Black box*. *Black box* adalah pengujian perangkat lunak yang berfokus pada fungsionalitas dari perangkat lunak yang dibangun. Pengujian *black box* memungkinkan perekayasa perangkat lunak mendapatkan serangkaian input yang sepenuhnya menggunakan semua persyaratan fungsional untuk suatu program. Pengujian yang dilakukan pada metode *black box* yaitu:

a. Fungsi-fungsi yang tidak benar
b. Kesalahan *interface*
c. Kesalahan dalam struktur data atau akses *database*
d. Kesalahan inisialisasi dan tujuan akhir

Metode pengujian tujuan penelitian yang akan digunakan yaitu metode *Confusion Matrix*. *Confusion Matrix* adalah pengujian tujuan penelitian yang akan melihat keakurasian status yang sistem berikan dari hasil pemrosesan dengan membandingkan status asli yang dilakukan secara manual.

1.5.6. Metode Pengambilan Kesimpulan

Kesimpulan adalah intisari dari hasil penelitian dan pernyataan mengenai hubungan hasil penelitian dengan hipotesis, termasuk alasan-alasan yang menyebabkan hasil penelitian berbeda dengan hipotesis. Metode pengambilan kesimpulan pada penelitian ini yaitu dilihat dari hasil penelitian dan pengujian yang telah dilakukan sebelumnya. Pengambilan keputusan ini bertujuan memberikan ringkasan hasil akhir penelitian yang dilakukan sesuai dengan maksud dan tujuan yang telah ditentukan.

1.6. Sistematika Penulisan

Sistematika penulisan laporan akhir penelitian ini disusun untuk memberikan gambaran umum tentang penelitian yang dilakukan. Sistematika penulisan tugas akhir ini adalah sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini berisi penjelasan mengenai latar belakang masalah, perumusan masalah, maksud dan tujuan, batasan masalah, metodologi penelitian serta sistematika penulisan yang dimaksdukan agar dapat memberikan gambaran tentang urutan pemahaman dalam menyajikan laporan ini.

BAB 2 LANDASAN TEORI

Bab ini membahas mengenai landasan teori yang digunakan untuk menganalisis masalah dan teori yang dipakai dalam mengolah data penelitian yaitu teori mengenai *email*, *email spam*, Teori Kombinasi, Metode *Smoothing Dirichlet*, Naïve Bayes, *Web Server*, dan lain-lain.

BAB 3 ANALISIS DAN PERANCANGAN

Bab ini menguraikan penjelasan mengenai analisis dan perancangan yang akan dibangun mulai dari analisis masalah ,analisis data masukan, analisis metode, analisis fungsional dan non fungsional, analisis kebutuhan perangkat lunak, analisis kebutuhan perangkat keras, perancangan antarmuka, dan perancangan pesan.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi tentang implementasi metode *Smoothing Dirichlet* Naïve Bayes dan Teori Kombinasi kedalam perangkat lunak berdasarkan perancangan yang sudah dibuat, serta pengujian simulator untuk mendapatkan akurasi untuk metode tersebut.

BAB 5 KESIMPULAN DAN SARAN

Bab ini menjelaskan tentang kesimpulan yang diperoleh dari hasil analisis performansi Teori Kombinasi dan metode *Smoothing Dirichlet* Naïve Bayes dalam klasifikasi *email spam*, serta saran-saran untuk pengembangan kedepannya.