

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Dokumen adalah sekumpulan tulisan yang memuat informasi, berbentuk cetak atau elektronik [1]. Klasifikasi dokumen sudah banyak dilakukan untuk mempermudah dalam pencarian dokumen berdasarkan kategori yang ditentukan. Dokumen yang memiliki kata pembahasan yang sama akan dikelompokkan ke dalam satu kategori. Dalam melakukan pencarian informasi, seseorang dapat dengan mudah melewati kategori yang tidak diinginkan atau yang tidak menarik karena adanya proses klasifikasi dokumen tersebut.

Penelitian yang dilakukan oleh Ahmad Yusuf, dkk membahas tentang *Support Vector Machine* yang didukung *K-Means Clustering* dalam klasifikasi dokumen menghasilkan nilai akurasi sebesar 88.1% dengan parameter jumlah kelompok sebesar 5. Dalam penelitian ini, SVM memiliki kinerja yang baik didukung oleh *K-Means Clustering* [2]. Sedangkan penelitian yang dilakukan oleh Amanda Nurul Amalia yang membahas tentang Implementasi SVM pada Klasifikasi Laporan Skripsi menghasilkan akurasi sebesar 34%. Hasil pengujian penelitian tersebut memiliki akurasi cenderung kecil. Pada penelitian ini akurasi yang cenderung kecil dapat disebabkan karena fitur kata yang digunakan pada proses pembobotan kata tidak relevan dalam menentukan kelas yang akan dikelompokkan pada proses klasifikasi [3].

Untuk dapat menghasilkan akurasi yang lebih besar dalam proses klasifikasi dokumen abstrak maka digunakan seleksi fitur *Mutual Information* (MI) dan *Support Vector Machine* (SVM). Penelitian yang dilakukan oleh Julius Gigih Dimastyo membahas tentang perbandingan metode seleksi fitur pada *spam filter* menggunakan klasifikasi *Multinomial Naïve Bayes* dengan menggunakan seleksi fitur *Idf*, *Mutual Information* (MI) dan *Chi Square* dalam meningkatkan akurasi klasifikasi. Dari penelitian yang dilakukan menghasilkan penggunaan metode seleksi fitur yang terbaik berdasarkan tingkat akurasi adalah *mutual information*

dengan akurasi 93.77% dengan waktu komputasi 429.7 detik dan jumlah *vocabulary* sebanyak 9507. *Mutual Information* (MI) juga lebih bagus dalam melakukan klasifikasi dokumen *spam* dibandingkan dengan seleksi fitur yang lainnya jika menggunakan jumlah *vocabulary* yang sedikit (sebanyak 3390). Walaupun dalam seleksi fitur mengalami sedikit penurunan akurasi namun efisiensi dari *term* yang dihasilkan sebagai penciri sangat terlihat. Tanpa menggunakan seleksi fitur membutuhkan waktu komputasi selama 35.8 menit dengan akurasi 94.54% dengan jumlah *vocabulary* 50419. Dibandingkan dengan hasil seleksi fitur *Mutual Information* (MI) mampu menghasilkan akurasi 93.77% [4]. Nilai yang digunakan dalam *Mutual Information* (MI) tidak ada batasan nyata tetapi menggunakan batasan sesuai yang diinginkan. Dalam penelitian ini batasan yang diambil disamakan atau mendekati dengan jumlah *vocabulary* yang dihasilkan dari seleksi fitur agar bisa terlihat perbedaan dari masing-masing seleksi fitur untuk dibandingkan [4]. Maka dalam penelitian ini akan digunakan *Mutual Information* untuk dapat melakukan pemilihan fitur yang relevan dari suatu kategori dalam suatu kelompok dengan *Support Vector Machine* dalam mengklasifikasi dokumen.

1.2 Identifikasi Masalah

Berdasarkan latar belakang diatas, identifikasi masalah yang akan dijadikan bahan penelitian adalah rendahnya akurasi pada proses klasifikasi dokumen abstrak dengan menggunakan SVM.

1.3 Maksud dan Tujuan

Maksud penelitian ini adalah melakukan klasifikasi dokumen menggunakan metode *Support Vector Machine* dengan seleksi fitur *Mutual Information*.

Tujuan yang ingin dicapai dari penelitian ini adalah meningkatkan akurasi dalam melakukan klasifikasi dokumen menggunakan metode *Support Vector Machine* dengan seleksi fitur *Mutual Information*.

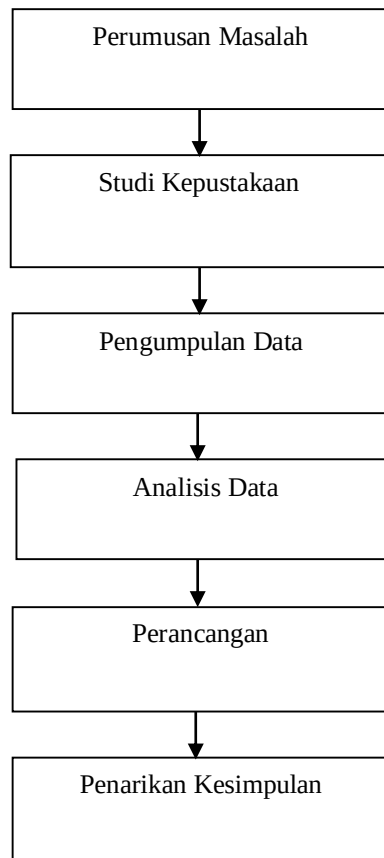
1.4 Batasan Masalah

Adapun batasan dari penulisan tugas akhir ini adalah sebagai berikut:

1. Dokumen masukan yang digunakan adalah file-file abstrak laporan tugas akhir Prodi Teknik Informatika UNIKOM yang diambil dari *elib.unikom.ac.id*.
2. Data yang digunakan berjumlah 200 abstrak laporan tugas akhir dimana 150 abstrak laporan tugas akhir menjadi data latih dan 50 abstrak laporan tugas akhir menjadi data uji dengan format .csv.
3. Teks masukan dalam Bahasa Indonesia dan tidak memperhatikan kesalahan penulisan.
4. Kategori yang ditentukan berdasarkan kelompok keilmuan Prodi Teknik Informatika UNIKOM, yaitu: Kelompok Keilmuan A (Sistem Informasi), Kelompok Keilmuan B (Rekayasa Perangkat Lunak), Kelompok Keilmuan C (Jaringan Komputer dan Multimedia), Kelompok Keilmuan D (Teknologi Internet dan Mobile), dan Kelompok Keilmuan E (Kecerdasan Buatan).
5. Penelitian dibatasi pada tahap seleksi fitur menggunakan *Mutual Information*, klasifikasi dokumen dan penilaian akurasi menggunakan *Support Vector Machine*.

1.5 Metode Penelitian

Metode penelitian adalah cara ilmiah untuk mendapatkan data dengan tujuan dan kegunaan tertentu. Metode penelitian yang digunakan dalam penelitian ini adalah penelitian kuantitatif. Penelitian kuantitatif memiliki tujuan untuk mencari hubungan dan menjelaskan sebab perubahan dalam fakta sosial terukur dengan serangkaian langkah atau prosedur yang baku yang menjadi pegangan peneliti. Dalam melakukan penelitian ini melalui beberapa tahap yang dilakukan seperti pada Gambar 1.1. Langkah-langkah yang terdapat dalam metode penelitian menurut Suryo Guritno [5], diantaranya:



Gambar 1.1 Metode Penelitian

Berdasarkan langkah-langkah penelitian yang terdapat pada Gambar 1.1. dapat di jelaskan sebagai berikut:

1. Perumusan Masalah

Tahapan pertama dalam penelitian adalah perumusan masalah. Peneliti merumuskan masalah apa saja yang terdapat dalam klasifikasi dokumen abstrak.

2. Studi Kepustakaan

Tahapan kedua dalam penelitian adalah melakukan studi kepustakaan. Peneliti melakukan studi kepustakaan untuk mengkaji teori-teori yang mendasari penelitian, yang bersumber dari penelitian-penelitian terdahulu baik melalui buku maupun jurnal.

3. Pengumpulan Data

Tahapan ketiga dalam penelitian adalah mengumpulkan data. Peneliti mengumpulkan data yang akan digunakan untuk proses penelitian dalam hal ini adalah data latih dan data uji yang akan digunakan untuk proses klasifikasi dokumen.

4. Analisis Data

Tahapan keempat dalam penelitian adalah menganalisis data. Peneliti menganalisis data dan langkah yang akan digunakan dalam mengolah data menggunakan teknik analisis statistik deskriptif berupa tabel, analisis langkah-langkah yang akan dijalankan.

5. Perancangan

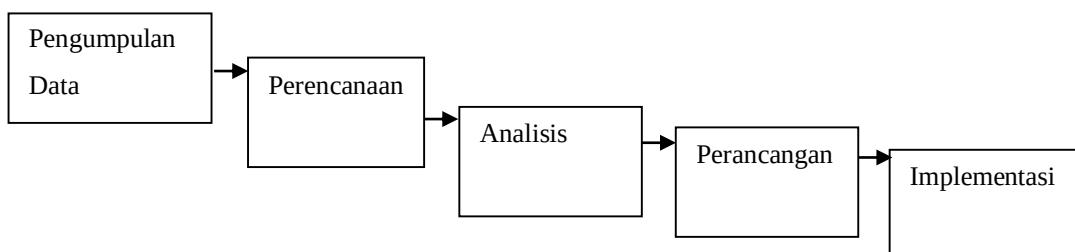
Tahapan kelima dalam penelitian adalah melakukan perancangan. Peneliti melakukan perancangan untuk klasifikasi dokumen menggunakan *Support Vector Machine* dan seleksi fitur *Mutual Information*.

6. Penarikan Kesimpulan

Tahapan keenam dalam penelitian adalah membuat kesimpulan. Peneliti membuat kesimpulan dalam proses klasifikasi dokumen dan hasil akurasi yang diperoleh dalam proses klasifikasi dokumen tersebut.

1.5.1 Metode Pembangunan Perangkat Lunak

Metode pembangunan perangkat lunak yang digunakan menggunakan model *waterfall* pada Gambar 1.2. Langkah-langkah yang terdapat dalam model *waterfall* menurut Roger S. Pressman [6], diantaranya:



Gambar 1.2 Metode Waterfall

Berdasarkan langkah-langkah yang terdapat pada Gambar 1.2. dapat di jelaskan sebagai berikut:

1. Pengumpulan Data

Tahap mengumpulkan data digunakan untuk mengumpulkan data yang menjadi data masukan dalam penelitian ini. Data masukan yang digunakan untuk menjadi data latih dan data uji dalam melakukan klasifikasi dokumen menggunakan *Support Vector Machine* dan seleksi fitur *Mutual Information* dengan jumlah data sebanyak 200 data.

2. Perencanaan

Tahap perencanaan dilakukan setelah tahap mengumpulkan data untuk proses awal selesai dilaksanakan, tahap perencanaan ini meliputi menyiapkan data yang akan diolah dalam tahap perencanaan juga harus dapat memahami proses perhitungan yang akan dilakukan. Data masukan yang akan digunakan adalah abstrak akademik.

3. Analisis

Tahap Analisis dilakukan setelah melakukan tahap perencanaan. Tahap analisis meliputi analisis sistem yang akan dibuat atau dibangun, analisis data masukan yang akan digunakan dan analisis proses yang akan dilakukan. Tahap analisis harus dilakukan langkah perlangkah sesuai dengan apa yang direncanakan untuk melakukan klasifikasi dokumen menggunakan *Support Vector Machine* dan seleksi fitur *Mutual Information*.

4. Perancangan

Tahap perancangan dilakukan setelah tahap analisis. Tahap perancangan dilakukan untuk dapat merancang aplikasi yang akan dibangun. Aplikasi yang akan dibangun berisi menu atau petunjuk dari cara penggunaannya. Dari mulai struktur menu dan perancangan antarmuka. Perancangan dilakukan dengan mendesain tampilan antarmuka yang diinginkan dan akan dilakukan penyesuaian didalam aplikasi yang akan dibuat untuk mengolah klasifikasi dokumen abstrak yang dilakukan.

5. Implementasi

Tahap implementasi dilakukan setelah tahap perancangan. Tahap implementasi dibuat untuk dapat menerjemahkan hasil perancangan antarmuka kedalam aplikasi yang akan dibuat. Aplikasi yang akan dibuat ini harus sesuai dengan perancangan yang sudah digambarkan. Sehingga implementasi aplikasi merupakan penggambaran dari apa yang telah dirancang selain itu menyesuaikan dengan data yang akan diolah dan dianalisis.

1.6 Sistematika Penulisan

Sistematika penulisan laporan yang akan disusun untuk memberikan gambaran secara umum tentang Klasifikasi Dokumen Menggunakan *Support Vector Machine* dan *Mutual Information* adalah sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini membahas mengenai latar belakang masalah, identifikasi masalah, maksud dan tujuan, batasan masalah, metode penelitian serta sistematika penulisan.

BAB 2 TINJAUAN PUSTAKA

Bab ini membahas mengenai buku, jurnal dan sumber yang digunakan untuk menganalisis masalah dan teori yang dipakai dalam mengolah data penelitian yaitu teori mengenai dokumen, pengelompokan, Seleksi fitur *Mutual Information*, Metode *Support Vector Machine* dan teori mengenai software pembangunan sistem.

BAB 3 ANALISIS DAN PERANCANGAN SISTEM

Bab ini menguraikan penjelasan analisis kebutuhan 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 dan perancangan antarmuka.

BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM

Bab ini membahas tentang implementasi dan pengujian seleksi fitur *Mutual Information* dan *Support Vector Machine* dalam klasifikasi dokumen dari hasil analisis dan perancangan sistem.

BAB 5 KESIMPULAN DAN SARAN

Bab ini membahas tentang kesimpulan dari proses penggunaan *Mutual Information* dan *Support Vector Machine* dalam klasifikasi dokumen serta saran untuk pengembangan kedepannya.