

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi saat ini telah banyak menunjukkan kemajuan yang luar biasa. Berkembang secara drastis dan terus berevolusi hingga sekarang semakin canggih dan mendunia. Banyak hal dari sektor kehidupan yang telah menggunakan keberadaan dari teknologi itu sendiri. Kehadirannya telah memberikan dampak yang cukup besar terhadap kehidupan umat manusia dalam berbagai aspek dan dimensi. Salah satunya dalam dunia pendidikan. Perkembangan teknologi informasi dirasa mulai memberikan dampak positif didalam proses belajar-mengajar. Banyak hal yang dirasa beda dan berubah dibandingkan dengan cara yang berkembang sebelumnya, di mana sekarang ini jarak dan waktu bukanlah sebagai masalah yang berarti untuk mendapatkan ilmu, berbagai media tercipta untuk memfasilitasinya [1].

Saat ini berbagai institusi ataupun organisasi yang berhubungan dengan akademis memiliki *repository* sendiri didalam menyimpan naskah publikasi sebuah penelitian yang telah ada. Akan tetapi dengan meningkatnya aktivitas penulisan dokumen akademis dari kegiatan riset, konferensi dan pertemuan-pertemuan ilmiah saat ini juga menyebabkan tingginya *volume* dokumen teks yang harus disimpan. Dokumen teks yang disimpan tersebut jarang digunakan secara optimal karena manusia seringkali tidak memiliki waktu dan kemampuan yang cukup untuk mengelolanya [2].

Data bervolume besar seperti dokumen teks, jauh melampaui kapasitas pengolahan manusia yang sangat terbatas [3]. Kondisi ini menimbulkan kesulitan bagi manusia dalam mencerna dan memilah sebuah informasi misalnya dalam pengklasifikasian atau pengkategorian sebuah naskah publikasi penelitian.

Data mining lebih terkhusus pada bidang *text mining* adalah sebuah bidang ilmu yang dianggap mampu memberi solusi didalam analisis informasi dari *unstructured data*. Terlebih, saat ini sudah banyak aplikasi *text mining* yang tersedia dalam menganalisa dan mencari *interesting patterns* termasuk *trend* dan *outliers* suatu dokumen [4]. *Support Vector Machine* atau SVM adalah salah satu teknik pembelajaran mesin yang terkenal dalam menyelesaikan permasalahan klasifikasi. Konsep pembelajaran pada SVM adalah dengan mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah class pada input space dengan cara mengukur margin *hyperplane* tersebut dan mencari titik maksimal nya. Salah satu keuntungan dari SVM adalah tidak membutuhkan data latih yang banyak. Maka dari itu diangkatlah sebuah penelitian skripsi yang berjudul "Klasifikasi Naskah Publikasi Menggunakan Metode *Support Vector Machine Linear*".

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka masalah yang muncul adalah:

1. Bagaimana sebuah program atau sistem dapat mengkategorikan sebuah dokumen dengan implementasi bidang ilmu *text mining*?
2. Bagaimana penerapan SVM dalam pengklasifikasian dokumen teks?

1.3 Batasan Masalah

Adapun yang menjadi ruang lingkup batasan masalah dalam penelitian ini meliputi:

1. Prototipe sistem dibuat berbasis Desktop menggunakan bahasa pemrograman C# .Net Framework.
2. Metode Pengklasifikasi yang diterapkan pada sistem ini adalah *SVM Linear*.
3. Pembobotan pada tahap *Preprocessing* sendiri menggunakan TF-IDF
4. Pengklasifikasian naskah publikasi berdasarkan Pengkategorian didalam *Prosiding Seminar Nasional Teknik Industri UGM*.
5. Penelitian ini menggunakan Weka dan Rapid Miner sebagai *tools* untuk menemukan model persamaan terbaik serta melakukan validasi.

1.4 Maksud dan Tujuan Penelitian

Berdasar latar belakang dan rumusan masalah yang diangkat dalam penelitian ini, maka maksud dan tujuan penelitian yang ingin dicapai adalah:

1. Menemukan model terbaik dengan metode SVM yang dapat digunakan untuk klasifikasi dokumen.
2. Membangun sebuah *prototipe* sistem menggunakan model tersebut agar dapat mengkategorikan naskah publikasi

1.5 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah:

1. Secara umum diharapkan dapat memberikan sumbangan bagi ilmu pengetahuan di bidang Data mining terkhusus pada *Text Mining*.

2. Hasil penelitian ini dapat menambah referensi penelitian mengenai pengklasifikasian naskah publikasi menggunakan metode *Support Vector Machine*.
3. Untuk Memudahkan didalam pengkategorian naskah publikasi yang ada didalam *repository* institusi atau organisasi akademik

1.6 Metode Penelitian

1.6.1 Metode Pengumpulan Data

1.6.1.1 Metode Observasi

Pada pengumpulan data melalui metode observasi ini, penulis mengambil dari *Prosiding* Seminar Nasional Teknik Industri UGM tahun 2015 sampai tahun 2017.

1.6.1.2 Studi Literatur

Pada tahapan ini penulis mempelajari dan mengambil kajian dari buku, artikel ilmiah, ataupun jurnal penelitian terkait yang mampu menjadi bahan pendukung penelitian ini.

1.6.2 Metode Analisis

Pada penelitian ini, data naskah publikasi yang berupa judul serta abstrak akan melalui tahap *preprocessing* pemberian bobot dengan metode TF-IDF, setelah itu digunakan metode *Support Vector Machine* untuk klasifikasi.

1.6.3 Metode Perancangan

Prototipe sistem dirancang dengan UML. UML (*Unified Modeling Language*) merupakan Bahasa visual untuk pemodelan dan komunikasi mengenai

sebuah sistem dengan menggunakan diagram dan teks-teks pendukung (Rosa dan Shalahuddin, 2013).

1.6.3 Metode Pengembangan

Tahap-tahap yang peneliti terapkan dalam metode pengembangan adalah sebagai berikut:

1. Pengumpulan data.
2. Preprocessing text.
3. Pembobotan kata tiap kategori.
4. Pembuatan dataset.
5. Pengujian akurasi tiap kernel.
6. Menggunakan model pada prototipe program.

1.6.4 Metode Testing

Pengujian untuk algoritma menggunakan metode pengujian cross validation dengan bantuan tools yaitu weka dan rapidminer. Beberapa hasil yang dilihat saat pengujian yaitu akurasi, presisi dan recall.

1.7 Sistematika Penulisan

Penyusunan skripsi ini perlu adanya sistematika penulisan yang terdiri dari bagian-bagian yang saling berhubungan sehingga dapat digunakan sebagai acuan pokok untuk penyusunan laporan skripsi antara lain:

BAB I PENDAHULUAN

Bab ini menguraikan tentang latar belakang masalah, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, serta sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisis tentang teori-teori yang mendasari pembahasan secara rinci, dapat berupa definisi atau model matematis yang langsung berkaitan dengan ilmu atau masalah yang diteliti. Landasan teori dalam bab ini dikelompokkan menjadi empat. Yaitu teori dasar tentang *data mining*, teori dasar mengenai sistem, teori analisis dan perancangan serta tujuan umum yang meliputi tinjauan pustaka dan tinjauan perangkat lunak.

BAB III ANALISIS DATA DAN PERANCANGAN

Bab ini membahas tentang analisis sistem dan perancangan sistem. Analisis sistem dimulai dari melakukan studi pendahuluan, identifikasi masalah, memahami kerja sistem yang ada, dan hasil analisis. Perancangan sistem meliputi perancangan struktur menu, perancangan proses, serta perancangan *interface*.

BAB IV IMPLEMENTASI DAN PEMBAHASAN

Bab ini membahas tentang implementasi dari aplikasi yang akan dirancang yang meliputi cara instalasi aplikasi dan pengoprasian aplikasi. Bab ini juga akan dibahas mengenai hasil dari uji coba aplikasi dan algoritma yang digunakan.

BAB V PENUTUP

Bab ini berisi kesimpulan yang didapat dari penelitian pembuatan prototipe sistem klasifikasi naskah publikasi dan tingkat akurasi SVM dalam melakukan proses klasifikasi serta kekurangan dari aplikasi ini.

