

## **BAB I. Pendahuluan**

### **I.1. Latar Belakang**

Program studi Teknik Informatika Institut Teknologi Sumatera (Prodi IF ITERA) mendapatkan umpan balik dari mahasiswa melalui formulir daring. Formulir itu berisi pertanyaan yang ditujukan untuk mahasiswa mengenai layanan, kondisi perkuliahan, pengajaran dosen dan pertanyaan opini lainnya. Opini yang didapat bisa digunakan untuk memperoleh gambaran umum mengenai persepsi mahasiswa selama mahasiswa tersebut berkuliah di Prodi IF ITERA. Data yang diperoleh merupakan data berbentuk teks dan pengolahan data diproses secara manual oleh manusia langsung. Setiap ulasan itu akan dibaca dan dinilai sentimennya satu persatu.

Dari tahun 2013 sampai 2019, jumlah mahasiswa teknik informatika terus mengalami peningkatan. Dimulai dari tahun 2013 dengan total 3 mahasiswa, diikuti tahun berikutnya dengan 4 mahasiswa. Peningkatan drastis dimulai dari mahasiswa angkatan 2015 hingga 2019 dengan jumlah mahasiswa secara berturut-turut 35, 138, 182, 199 dan 233. Pola peningkatan setiap tahun itu, memungkinkan juga terjadinya peningkatan data opini mahasiswa yang disampaikan terhadap Prodi IF. Faktor tersebut mendorong perlunya dilakukan pengolahan data ulasan mahasiswa secara otomatis menggunakan komputer.

Proses penilaian opini akan lebih efektif bila dilakukan menggunakan komputer dibandingkan secara manual. Pengolahan opini secara otomatis ini dikenal secara umum sebagai analisis sentimen. “Analisis sentimen adalah bidang ilmu yang menganalisis opini, sentimen dan emosi yang disampaikan di dalam ulasan terhadap produk, layanan, organisasi, individu, masalah, peristiwa, topik dan atributnya” [1].

Meskipun akurasi analisis sentimen belum mampu menyamai manusia, tetapi pengaplikasiannya sangat berguna untuk data dengan jumlah yang besar.

Secara umum terdapat tiga tingkatan dalam sentimen analisis [1]. Pertama, Level dokumen, tugas utamanya adalah untuk mengklasifikasikan apakah keseluruhan dokumen berisi sentimen positif atau negatif. Pada level ini, Setiap dokumen atau kalimat diasumsikan memiliki satu target opini. Kedua, Level kalimat, mengklasifikasikan sentimen positif atau negatif pada setiap kalimat di dalam dokumen. Terakhir, Level aspek, mengklasifikasikan sentimen positif atau negatif pada aspek atau entitas yang terdapat di dalam dokumen.

Level dokumen dan level kalimat tidak mampu menemukan secara rinci apa yang sebenarnya disukai dan tidak disukai oleh penyampai opini [1]. Dalam kasus ini, jika level dokumen dan level kalimat diterapkan pada opini mahasiswa, hasil yang didapat tidak mampu menangkap entitas yang dijadikan sebagai target opini. Padahal, Opini yang disampaikan mahasiswa bisa membahas entitas secara spesifik. Misalkan terdapat opini mahasiswa di dalam satu kalimat yang berisi penilaian terhadap ruang kelas, cara dosen mengajar dan kondisi laboratorium. Oleh sebab itu, demi memperoleh informasi yang lebih mendetail, Level aspek akan diterapkan di tugas akhir ini untuk mendapatkan target opini dan pasangan polaritas sentimennya. Level aspek dikenal juga sebagai Analisis Sentimen Berbasis Aspek (ABSA).

Tahun 2014 sampai 2016 terdapat lokakarya Semantic Evaluation (SemEval) yang berkaitan dengan ABSA. Penelitian ABSA di SemEval berfokus pada tiga tugas [2], yaitu ekstraksi target opini, bertujuan untuk mengidentifikasi target opini yang disebutkan di dalam kalimat. Selanjutnya, klasifikasi kategori aspek, bertujuan untuk mengidentifikasi pasangan entitas dan atribut yang terdapat pada kalimat opini. Terakhir klasifikasi sentimen, bertujuan untuk melabeli target opini atau aspek kategori dengan polaritas positif, negatif atau netral.

Salah satu metode yang digunakan pada penelitian terkait untuk menyelesaikan permasalahan ABSA ini adalah pembelajaran mesin menggunakan Support Vector Machine (SVM). Pada komparasi metode pembelajaran mesin yang dilakukan pada penelitian [3, 4, 5, 6, 7, 8, 24] untuk mengklasifikasikan sentimen. Hasilnya menunjukkan bahwa SVM mengungguli metode pembelajaran mesin lainnya seperti Naive Bayes (NB), Maximum Entropy (ME), Decision Tree C45 (DT) dan K-Nearest Neighbor (K-NN). Sedangkan untuk klasifikasi kategori aspek, Jihan dkk [9] menunjukkan bahwa metode SVM yang ia gunakan, berhasil mengungguli peringkat 1 pada lokakarya SemEval 2016, yaitu NLANGP [10], yang menggunakan Neural Network dan fitur *deep learning* pada sistemnya dengan selisih f1 skor 1.15% (Jihan dkk : 74.180%, NLANGP : 73.031%).

U. Fayyad dkk [11] menjelaskan terdapat 5 tahap di dalam kerangka kerja menambang data, yaitu data seleksi, data *preprocessing*, data transformasi, menambang data dan interpretasi atau evaluasi. Tugas Akhir ini akan mengotomasi ABSA dengan menerapkan kerangka kerja penambangan data. ABSA yang akan diimplementasikan adalah Klasifikasi Kategori Aspek dan Klasifikasi Sentimen dengan menggunakan SVM.

## **I.1 Rumusan Masalah**

Berdasarkan latar belakang di atas maka diperoleh rumusan masalah, yaitu:

1. Bagaimana cara menerapkan analisis sentimen berbasis aspek untuk ulasan teks prodi teknik informatika itera untuk tugas klasifikasi kategori aspek dan klasifikasi sentimen menggunakan SVM?
2. Bagaimana menerapkan kerangka kerja Penambangan data yang meliputi pemilihan data, *preprocessing*, transformasi data, penambangan data hingga evaluasi pada sentimen analisis berbasis aspek.

## **I.2 Tujuan**

Tujuan dari tugas akhir ini adalah sebagai berikut:

1. Membangun Sistem ABSA menggunakan kasus data teks ulasan Program studi teknik informatika Itera.
2. Mengkaji Penelitian ABSA yang telah ada sebagai acuan pembangunan sistem.
3. Mengkaji penerapan SVM pada ABSA dengan menggunakan data bahasa indonesia.
4. Menerapkan 5 kerangka kerja menambang data dalam proses pembangunan sistem.
5. Menganalisis hasil sistem yang telah dibuat.

## **I.3 Batasan Masalah**

Guna memperjelas penelitian ini maka akan definisikan batasan-batasan masalah sebagai berikut:

1. Tidak mengimplementasikan ABSA kategori tugas ekstraksi target opini
2. Penelitian menggunakan data milik program studi Teknik Informatika ITERA
3. Penelitian hanya menggunakan data teks berbahasa Indonesia baik kalimat baku maupun tidak baku.
4. Data ulasan yang diolah meliputi kalimat lengkap dan tidak lengkap.
5. Pelabelan data dilakukan secara manual.

## **I.4 Metodologi**

### 1. Studi literatur

Mencari dan mempelajari literatur yang meliputi buku, artikel dan jurnal terkait topik yang bersesuaian.

### 2. Pengolahan data

Melakukan pengumpulan dan pengolahan data sebagai tahap awal pengkajian. pengolahan data dapat berupa melabeli, mendefinisikan aspek, dan pra proses data.

### 3. Analisis Dan Perancangan Sistem

Merancang sistem dan membangun strategi untuk menyelesaikan masalah pada data.

### 4. Eksperimen

Menerapkan dan melakukan analisis mengenai penerapan SVM termasuk pengaruh fitur terhadap hasil.

## **I.5 Sistematika Penulisan**

### **BAB 1 Pendahuluan**

Bab 1 merupakan bab pendahuluan. Bab ini membahas latar belakang masalah yang diangkat, Rumusan masalah, tujuan penelitian, batasan masalah, metodologi penelitian dan sistematika penulisan. Batasan masalah merupakan bab yang berisi ruang lingkup tugas akhir mengenai apa yang tidak dikerjakan atau batasan-batasan terhadap ketersediaan data.

### **BAB 2 Landasan Teori**

Bab 2 adalah bab yang berisi ilmu pengetahuan yang digunakan sebagai referensi untuk pembuatan tugas akhir ini. Sumber referensi berasal dari buku dan jurnal.

### **BAB 3 Perancangan Sistem**

Bab 3 merupakan rencana rancangan sistem. Rencana memuat tahapan-tahapan dan contoh untuk membuat bagian dari sistem. Bab ini juga berisi analisis dalam menyelesaikan masalah.

### **BAB 4 Implementasi dan Pengujian**

Bab 4 merupakan bab yang membahas implementasi dari perancangan sistem yang telah dibuat pada Bab 3. Bab ini juga berisi pengujian terhadap hasil eksperimentasi yang akan dilakukan.

### **BAB 5 Kesimpulan dan Saran**

Bab 5 membahas kesimpulan hasil implementasi. Di bab ini juga akan dibahas mengenai saran terhadap penelitian dan kelemahan dari sistem