

BAB I

PENDAHULUAN

1.1. Latar Belakang

Setiap perguruan tinggi memiliki penumpukan data akademik yang terjadi dari tahun ke tahun. Penumpukan data ini dapat diolah dan dimanfaatkan menjadi sebuah strategi dalam memprediksi kelulusan mahasiswa. Prediksi (*prediction*) diartikan sebagai teknik untuk memperkirakan nilai pada masa yang akan datang dengan memperhatikan data dan informasi yang relevan, termasuk data dan informasi masa lalu atau informasi pada saat ini (Nachrowi, D. N., dan Usman, H., 2004). Melakukan prediksi kelulusan mahasiswa merupakan hal yang penting diperhatikan, karena persentase jumlah kelulusan mahasiswa dari tahun ke tahun mempengaruhi penilaian pemerintah serta status akreditasi program studi. Oleh karena itu, diperlukan suatu teknik pengolahan data untuk mendapatkan hasil prediksi yang tepat serta memiliki tingkat akurasi yang tinggi (David H., dan Seng H., 2014). Teknik pengolahan data yang digunakan pada prediksi ini adalah algoritma C4.5 dan *Support Vector Machine* (SVM).

Algoritma C4.5 bekerja dengan menghasilkan pohon keputusan yang mudah diinterpretasikan serta tingkat akurasi yang dapat diterima, mampu mengatasi atribut yang bersifat kontinu, mengatasi nilai yang hilang, melakukan pemangkasan pohon yang kompleks dalam menghasilkan prediksi, yang menjadi kelebihan dibandingkan algoritma lainnya (Han., et al., 2001). Dalam pembentukan pohon keputusan sering terjadi *overlap* terutama ketika kelas-kelas dan kriteria yang digunakan jumlahnya sangat banyak, hal itu yang menyebabkan meningkatnya waktu pengambilan keputusan dan jumlah memori yang diperlukan. Hasil keputusan yang didapatkan dari algoritma C4.5 bergantung pada proses pembentukan pohon keputusan.

Salah satu metode statistik yang dapat diterapkan untuk melakukan klasifikasi dalam prediksi adalah *Support Vector Machine* (SVM). *Support Vector Machine* (SVM) merupakan suatu teknik untuk menemukan *hyperplane* yang bisa

memisahkan dua set data dari dua kelas yang berbeda. *Support Vector Machine* membutuhkan dataset, kemudian di tes untuk mendapatkan akurasi dan hasil tes tertinggi diperoleh dengan menggunakan mesin dukungan *vector*. *Support Vector Machine* (SVM) memiliki kelebihan dalam menentukan jarak dengan menggunakan *support vector* sehingga proses prediksi menjadi cepat dan tingkat akurasi yang lebih tinggi (Wenyu Zhang, et. Al., 2015).

Berdasarkan uraian diatas dilakukan perbandingan tingkat akurasi prediksi, maka judul yang diajukan pada tugas akhir ini adalah “Aplikasi Perbandingan Algoritma Prediksi Kelulusan Mahasiswa Menggunakan C4.5 dan *Support Vector Machine* (SVM)”.

1.2. Rumusan Masalah

Sesuai dengan latar belakang yang sudah diuraikan sebelumnya, maka yang menjadi rumusan masalah dalam penulisan tugas akhir ini adalah belum ditemukannya algoritma terbaik untuk memprediksi kelulusan mahasiswa dengan membandingkan hasil akurasi yang lebih tinggi terhadap algoritma C4.5 dan *Support Vector Machine* (SVM).

1.3. Tujuan dan Manfaat

Tujuan dari penulisan tugas akhir ini adalah mampu menghasilkan prediksi kelulusan mahasiswa berdasarkan tingkat akurasi yang dihasilkan dengan membandingkan algoritma C4.5 dan *Support Vector Machine* (SVM).

Adapun manfaat yang diharapkan adalah sebagai berikut :

1. Dapat digunakan sebagai aplikasi penunjang keputusan bagi suatu perguruan tinggi dalam mengambil keputusan memprediksi lulus tepat waktu atau lulus terlambat seorang mahasiswa.
2. Hasil penulisan tugas akhir ini dapat dijadikan sebagai referensi maupun pedoman bagi mahasiswa lain yang mengambil penelitian sejenis.

1.4. Ruang Lingkup

Adapun batasan masalah dari tugas akhir ini, antara lain :

1. *Dataset* yang digunakan adalah nilai akademik mahasiswa Mikroskil.

2. *File input* yang didukung oleh sistem yang dibuat dengan berekstensi “.csv”.
3. *Dataset* yang digunakan untuk perhitungan algoritma C4.5 dan *Support Vector Machine* (SVM) memiliki atribut Indeks Prestasi Sementara (IPS) dari semester 1 sampai dengan Indeks Prestasi Sementara (IPS) semester 6.

1.5. Metodologi Penelitian

Langkah penyusunan dan penyelesaian tugas akhir ini menggunakan langkah-langkah sebagai berikut:

1. Pengumpulan Data

- a. Melakukan kajian literatur yang berhubungan dengan prediksi kelulusan mahasiswa dan metode yang digunakan.
- b. Melakukan pengumpulan data yang akan digunakan sebagai bahan penelitian yaitu Data Mahasiswa dari STMIK-Mikroskil Medan.

2. Membuat aplikasi dengan metode *Waterfall*, yaitu:

a. Identifikasi

Melakukan identifikasi terhadap masalah, dan menyesuaikan dengan batasan masalah serta mencari solusi untuk masalah.

b. Analisis

Menganalisis kebutuhan yang diperlukan sistem, baik secara fungsional dengan menggunakan *Use Case* maupun non-fungsional dengan menggunakan PIESCES.

c. Perancangan sistem

Merancang tampilan antarmuka sistem prediksi dengan menggunakan *tools form visual studio 2012*.

d. Implementasi

Pada langkah ini, mengimplementasikan rancangan tampilan antarmuka kedalam pemrograman C#.

e. Pengujian

Pengujian dilakukan dengan metode *k-fold* dan pengujian *Incremental*. Pengujian *k-fold* dilakukan dengan membagi *dataset* kedalam 10 *subset* data mahasiswa. Pembagian *dataset* dapat dilihat pada Tabel 1.1.

Tabel 1.1 Pengujian dengan *k-fold*

<i>Fold</i> (<i>Pengujian</i>)	<i>Data Training</i>	<i>Data Testing</i>
<i>Fold 1</i>	B, C, D (75%)	A (25%)
<i>Fold 2</i>	A, C, D (75%)	B (25%)
<i>Fold 3</i>	A, B, D (75%)	C (25%)
<i>Fold 4</i>	A, B, C (75%)	D (25%)
<i>Fold 5</i>	A, B (50%)	C, D (50%)
<i>Fold 6</i>	B, C (50%)	A, D 50%
<i>Fold 7</i>	C, D (50%)	A, B (50%)
<i>Fold 8</i>	A, D (50%)	B, C (50%)

UNIVERSITAS
MIKROSKIL

Untuk pengujian secara *incremental* dilakukan dengan membagi data *training* sebanyak 5 bagian dan data *testing* sebanyak 2 bagian. Dapat ditunjukkan pada tabel 1.2 dibawah ini.

Tabel 1.2 Pengujian secara *incremental*

<i>Fold</i> (<i>Pengujian</i>)	Data Training	Data Testing
<i>Fold 1</i>	B (15%)	A (25%)
<i>Fold 2</i>	B, C (30%)	A (25%)
<i>Fold 3</i>	B, C, D (45%)	A (25%)
<i>Fold 4</i>	B, C, D, E (60%)	A (25%)
<i>Fold 5</i>	B, C, D, E, F (75%)	A (25%)
<i>Fold 6</i>	B (10%)	A (50%)
<i>Fold 7</i>	B, C (20%)	A (50%)
<i>Fold 8</i>	B, C, D (30%)	A (50%)
<i>Fold 9</i>	B, C, D, E (40%)	A (50%)
<i>Fold 10</i>	B, C, D, E, F (50%)	A (50%)