

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Machine Learning (ML) merupakan teknik yang banyak digunakan untuk mengekstraksi pengetahuan dari data. ML adalah bidang penelitian di persimpangan antara statistik, kecerdasan buatan, dan ilmu komputer. ML juga dikenal sebagai analitik prediktif atau pembelajaran statistik[1]. Pada perkembangannya konsep ML memiliki empat kategori yaitu *Supervised Learning*, *Unsupervised Learning*, *Semi-Unsupervised Learning* dan *Reinforcement Learning*. *Supervised Learning* biasa digunakan untuk mengklasifikasikan data dimana dataset yang digunakan harus memiliki label data terutama pada variable target. Salah satu algoritma klasifikasi adalah *Decision Tree* (DT).

Decision tree merupakan algoritma yang menggabungkan perhitungan entropy dan informasi gini pada pembentukan pohon keputusan. Pohon keputusan kemudian digunakan untuk mengklasifikasikan data baru yang belum memiliki label (*class target*)[2]. Pada konsep klasifikasi kendala yang biasa dihadapi adalah rumit dan mahal nya proses komputasi, terutama untuk jumlah kombinasi variable yang sangat tinggi (dimensi tinggi). Kendala ini juga mempengaruhi akurasi dan kualitas model klasifikasi yang terbentuk. Inilah yang menjadi perhatian bagi banyak peneliti untuk diatasi dengan menawarkan berbagai pendekatan sebagai upaya untuk melakukan optimasi pada teknik klasifikasi seperti DT. Salah satu pendekatan yang bisa digunakan adalah melakukan optimasi algoritma. Ada banyak teknik optimasi algoritma yang bisa digunakan salah satunya adalah *Grid Search Optimization* (GSO).

Grid Search Optimization (GSO) merupakan teknik optimasi algoritma *cross validation* untuk meningkatkan akurasi yang signifikan terhadap model klasifikasi. Metode ini melakukan uji coba kombinasi dan validasi satu persatu kemudian memilih kombinasi yang menghasilkan performa model terbaik untuk prediksi[3]. Beberapa penelitian menggunakan *Grid Search Optimization* telah banyak digunakan, misalnya untuk meningkatkan akurasi pada *Support Vector*

Regression (SVR)[4] serta mengatasi *Hyperparameter Tuning* pada Algoritma Klasifikasi[3].

Terdapat banyak kasus yang sudah diselesaikan menggunakan algoritma DT, salah satunya terkait permasalahan klasifikasi mahasiswa non-aktif. Namun, akurasi hasil klasifikasi belum maksimal padahal batas akurasi bisa mencapai 100%. Oleh karena itu penelitian dilakukan sebagai upaya untuk mengoptimalkan teknik klasifikasi DT menggunakan pendekatan GSO. Tujuannya untuk menaikkan akurasi hasil klasifikasi khususnya pada kasus mahasiswa non-aktif.

Berdasarkan permasalahan yang dipaparkan di atas, *Grid Search Optimization* diterapkan pada algoritma *Decision Tree* sebagai metode penyelesaian masalah yang ada. Sehingga penelitian yang dilakukan mengambil judul “**Optimasi Algoritma *Decision Tree* Menggunakan *Grid Search Optimization* Untuk Klasifikasi Mahasiswa Non-Aktif**”.

1.2. Identifikasi Masalah

Berdasarkan latar belakang permasalahan yang telah diuraikan identifikasi masalah dapat disajikan sebagai berikut:

1. Pengoptimasian akurasi algoritma *Decision Tree* dengan *Grid Search Optimization*.
2. Pengaruh variable dimensi tinggi terhadap klasifikasi.
3. Penyebab adanya mahasiswa non-aktif.

1.3. Rumusan Masalah

Berdasarkan latar belakang di atas maka rumusan masalah dalam penelitian ini sebagai berikut :

1. Bagaimana melakukan Optimasi Algoritma *Decision Tree* Menggunakan *Grid Search Optimization*?
2. Bagaimana mengklasifikasikan Mahasiswa Non-Aktif menggunakan Algoritma *Decision Tree* yang sudah dioptimasi menggunakan *Grid Search Optimization*?

1.4. Batasan Masalah

Batasan masalah adalah membatasi ruang lingkup penelitian agar hasil yang dihasilkan lebih terarah dan fokus. Pembatasan masalah dalam penelitian prediksi mahasiswa non-aktif antara lain adalah:

1. Data yang diuji adalah data dari Universitas Bhayangkara Jakarta Raya dalam kasus ini penulis mengambil sebagai atribut penentu meliputi: NPM, Jenis Kelamin, Lokasi Daerah, SKS yang sudah dijalankan, IPK, Semester, Tahun Lahir, Gender, Status, Pendidikan, Pekerjaan Orangtua, Jenis Kelas, Non Aktif / Aktif.
2. Algoritma yang dipakai adalah *Decision Tree* (C4.5), yang dipakai lalu diinterpretasikan menjadi pohon keputusan.
3. Optimasi yang digunakan adalah *Grid Search Optimization*, yang menentukan penambahan akurasi dalam penelitian ini.

1.5. Tujuan dan Manfaat Penelitian

Sesuai dengan topik atau permasalahan yang ada, maka tujuan dari penulisan ini sebagai berikut:

1. Melakukan optimasi algoritma *Decision Tree* menggunakan *Grid Search Optimization*.
2. Mengklasifikasikan mahasiswa non-aktif menggunakan algoritma *Decision Tree* yang sudah dioptimasi menggunakan *Grid Search Optimization*.

Diharapkan penelitian yang penulis lakukan dapat membawa manfaat bagi penulis sendiri khususnya, dan orang lain yang membaca penelitian ini. Berikut ini kegunaan atau manfaat yang diharapkan:

1. Dapat membantu pihak Universitas lainnya untuk mencari informasi dalam mendapatkan informasi untuk penyebab terjadinya mahasiswa non-aktif.
2. Menjadi tambahan ilmu pengetahuan dan referensi terkait penggunaan algoritma *Decision Tree* dengan *Grid Search Optimization*.

1.6. Sistematika Penulisan

Sistematika penulisan skripsi terbagi menjadi beberapa bab, sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini menguraikan tentang pendahuluan yang meliputi latar belakang masalah, identifikasi masalah, rumusan masalah, tujuan penulisan, manfaat penulisan, batasan masalah dan metode penulisan yang digunakan dalam pengumpulan informasi atau data dan sistematika penulisan.

BAB 2 LANDASAN TEORI

Bab ini memaparkan teori-teori yang berkaitan dengan topik guna menjadi acuan dalam pengembangan analisa prediksi yang dibuat.

BAB 3 METODE PENELITIAN

Bab ini menguraikan secara spesifik mengenai prediksi, alur penelitian, perencanaan penelitian, keterangan data, dan analisa prediksi.

BAB 4 HASIL DAN PEMBAHASAN

Bab ini berisi hasil penelitian berserta pembahasannya.

BAB 5 KESIMPULAN DAN SARAN

Bab ini berisi simpulan dari penulisan skripsi dan saran yang diberikan penulis untuk mengembangkan skripsi ini.