

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penjurusan

Data siswa/I salah satu asset terpenting bagi sekolah. Data siswa berisi nilai-nilai mata pelajaran, mulai dari semester awal hingga semester akhir. Dengan adanya nilai-nilai siswa yang tersimpan sehingga dapat dijadikan sebagai penentuan jurusan dalam tingkat SMA. Tidak hanya nilai-nilai mata pelajaran saja, tetapi nilai keahlian siswa juga sangat mendorong dalam penentuan jurusan. Penjurusan adalah proses pemilihan pengelompokkan siswa berdasarkan nilai, minat dan bakat agar siswa mampu mengikuti pelajaran dengan focus (Novianti, Rismawan, & Bahri, 2016).

Dalam melakukan pengolah data nilai siswa untuk penentuan jurusan dapat menggunakan teknik data mining. Data mining merupakan salah satu proses yang menggunakan *statistic*, matematika, kecerdasan buatan dan *mechine learning* untuk mengekstrak dan mengidentifikasi informasi yang bermanfaat serta pengetahuan yang terkait dalam berbagai database (Turban Efraim, 2005). Salah satu teknik data mining sering dilakukan yaitu klasifikasi. Klasifikasi sebuah proses untuk menentukan model yang menjelaskan atau membedakan data dan klasifikasi yang banyak digunakan yaitu pohon keputusan (Novianti et al., 2016). Pada data mining terdapat algoritma yang dilakukan dalam pohon keputusan, yaitu algoritma C4.5 (T, 2005).

Algoritma C4.5 memiliki akurasi yang baik dalam menentukan pemilihan jurusan (Hermawanti, Asriyanik, & Sunarto, 2019). Algoritma C4.5 memiliki kelebihan yaitu mampu menangani nilai data yang dihitung dan nilai data yang diukur (Kustiyahningsih & Rahmanita, 2016). Dari hasil penelitian menggunakan algoritma C4.5 bahwa nilai-nilai ekstra dan psikotes bisa berpengaruh pada penjurusan siswa baru (Nugroho, 2018). Pada algoritma C4.5 terdapat pohon keputusan. Pohon keputusan yang mempresentasikan aturan untuk dimengerti dan

dipahami dengan baik (Novianti et al., 2016). Pohon keputusan (*decision tree*) sering digunakan karena mudah untuk dipahami dan dimengerti walaupun bagi pemula (Mantas, 2014). Pohon keputusan bisa digunakan untuk klasifikasi maupun prediksi (Sammut, 2011).

Penelitian ini akan membahas tentang analisis algoritma C4.5 dalam melakukan penentuan pemilihan jurusan IPA (*Natural Sciences Major*) dan IPS (*Social Sciences Major*) berdasarkan nilai bidang keahlian siswa.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini yaitu untuk melakukan analisis algoritma C4.5 dalam menentukan pemilihan jurusan IPA (*Natural Sciences Major*) dan IPS (*Social Sciences Major*) berdasarkan keahlian siswa.

1.3 Tujuan

Tujuan dari penelitian ini untuk melakukan analisis klasifikasi algoritma C4.5 terhadap nilai mata pelajaran siswa sehingga dapat menghasilkan pohon keputusan yang pemilihan jurusan siswa sesuai bidang keahliannya.

1.4 Manfaat

Manfaat dari penelitian ini untuk memudahkan pihak sekolah dalam menyeleksi siswa/i baru untuk pemilihan jurusan. Pihak sekolah bisa memberikan konsultasi pada siswa/i agar tidak terjadinya pro dan kontra terhadap masing-masing pihak dalam pemilihan jurusan.

1.5 Batasan Masalah

Ada beberapa batasan masalah dalam penelitian ini, yaitu sebagai berikut :

1. Data yang kami gunakan merupakan data dari sekolah yang diambil mulai dari tahun 2017 sampai tahun 2019.
2. Algoritma yang kami gunakan dalam penelitian ini merupakan algoritma C4.5
3. Dalam pengolahan data ini, kami menggunakan aplikasi Rapid Miner versi 9.9

1.6 Keterbaruan

(Sasongko, Linawati, & Parhusip, 2015) melakukan penelitian berjudul penerapan algoritma Fuzzy C-Means Guna Penentuan Jurusan Program Peserta Didik Tingkat SMA untuk menentukan jurusan IPA, IPS dengan hasil yang di peroleh merupakan nilai keanggotaan fuzzy setiap data pada setiap Cluster sebesar 93.79%.

(Sasongko et al., 2015) menggunakan algoritma C4.5 untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal. Algoritma tersebut merupakan metode data mining yang digunakan memprediksi kemampuan siswa. Parameter pemilihan fitur adalah factor-faktor dalam bidang studi matematika. Hasil pada pengujian dan analisis menampilkan jika algoritma C4.5 memiliki tingkatan akurasi 60%.

(Hermawanti et al., 2019) dalam penelitiannya melakukan implementasi algoritma C4.5 untuk prediksi kelulusan tepat waktu. Algoritma ini digunakan untuk membangun sebuah pohon keputusan yang dapat dimengerti, fleksibel dan menarik karena dapat divisualisasikan dalam bentuk gambar. Menggunakan teknik klasifikasi decision tree dengan penerapan algoritma C4.5. Untuk datanya menggunakan atribut dari mahasiswa meliputi IP (Indeks Prestasi) dari semester 1 sampai 7 dan nilai BTQ di semester 6. Data mahasiswa tersebut merupakan contoh data training yang digunakan dalam penyusunan decision tree. Pengujian ini menggunakan data training mahasiswa yang sudah lulus pada tahun 2015 sampai 2018. Selanjutnya dapat di dimanfaatkan oleh pihak Prodi Teknik Informatika UMMI sebagai langkah preventif untuk menghindari penurunan kelulusan mahasiswa setiap tahunnya.

(Algoritma et al., n.d.) dalam penelitiannya yang berjudul Penerapan Data Mining Untuk Rekomendasi Beasiswa Pada SMA Muhammadiyah Gubug Menggunakan Algoritma C4.5 melakukan pengumpulan data yaitu penghasilan orang tua, jumlah saudara kandung, jurusan, kelas dan jumlah nilai. Penelitian ini menggunakan pohon keputusan dengan metode confusion matrix dan akurasi yang dihasilkan sebesar 77%.

(Tasikmalaya, Sambani, & Nuraeni, 2017) melakukan penelitian yang berjudul Penerapan Algoritma C4.5 Untuk Klasifikasi Pola Penjurusan Di Sekolah Menengah Kejuruan dengan menentukan pola penjurusan siswa berdasarkan nilai rapor, tes kesehatan dan olahraga. Pola tersebut dapat membantu pihak sekolah dalam menentukan jurusan yang tepat untuk siswa. Algoritma C4.5 merupakan salah satu bagian dari teknik klasifikasi dengan kemampuan model yang dapat melakukan proses learning dan klasifikasi yang sederhana, cepat dan secara umum mempunyai tingkat akurasi yang tinggi. Dengan menggunakan data set 1104 record data siswa baru, hasil dari pengklasifikasian data pemilihan jurusan SMK untuk jurusan AP, TKJ, RPL, TKR dan TSM memperoleh akurasi sebesar 92,30% dengan tingkat akurasi yang baik.

Sedangkan pada penelitian ini penulis menerapkan algoritma C4.5 dalam menentukan klasifikasi penjurusan IPA/IPS untuk pihak sekolah berdasarkan bidang keahlian siswa.