

Prediksi Prestasi Belajar Mahasiswa Vokasi Menggunakan Algoritma C4.5

Ning Ratwastuti, Luthfi Atikah, Radix Rascalia dan Suhendra

Politeknik Astra

E-mail: {ning.ratwastuti, luthfi.atika, radix.rascalia, suhendra}@polytechnic.astra.ac.id

Abstrak

Prestasi belajar mahasiswa pada umumnya dilihat melalui IPK (Indeks Prestasi Kumulatif) yang diprediksi selama perkuliahan. Namun, banyak ditemui mahasiswa yang kesulitan selama mengikuti proses pembelajaran, baik mahasiswa yang berasal dari jalur beasiswa maupun jalur reguler. Berdasarkan kondisi tersebut perlu dilakukan pemetaan data pendaftaran calon mahasiswa dari PMB Politeknik ABC terhadap nilai IPK yang menggambarkan prestasi mahasiswa. Prediksi pada penelitian ini memanfaatkan algoritma C4.5 untuk melihat pengaruh data pendaftaran PMB terhadap prestasi belajar mahasiswa, baik jalur beasiswa maupun keseluruhan mahasiswa. Selain itu, untuk melihat pengaruh tiap data atribut dilakukan dengan seleksi menggunakan metode Gain Ratio. Hasil klasifikasi menunjukkan bahwa data pendaftaran PMB berpengaruh terhadap prestasi belajar mahasiswa, dan klasifikasi menggunakan atribut hasil seleksi mendapatkan akurasi yang lebih tinggi dibandingkan menggunakan keseluruhan atribut. Pada mahasiswa beasiswa, atribut kondisi sosial ekonomi mempengaruhi prestasi belajar. Namun pada keseluruhan data mahasiswa Politeknik ABC menunjukkan atribut penguasaan materi di sekolah asal membantu mahasiswa meningkatkan prestasi belajarnya.

Kata kunci : Klasifikasi, Algoritma C4.5, Seleksi Atribut, Gain Ratio, Prestasi Belajar

Pendahuluan

Mengukur kinerja mahasiswa merupakan isu yang menjadi perhatian di dunia pendidikan. Pada penelitian [1] hasil IPK mahasiswa menjadi ukuran keberhasilan mahasiswa pada masa belajarnya dan menjadi tolak ukur kualitas dari suatu institusi pendidikan. Umumnya suatu institusi pendidikan mengukur dan mengevaluasi mahasiswa mereka berdasarkan pada IPK. Kondisi saat ini, nilai IPK diukur berdasarkan nilai kelulusan suatu mata kuliah. Untuk menunjang dan memotivasi mahasiswa suatu institusi melakukan berbagai tindakan seperti bimbingan dosen wali dan bagian kesiswaan. Banyak faktor yang mempengaruhi tingkat dari IPK mahasiswa.

Terdapat berbagai pendekatan yang dilakukan untuk mengukur tingkat IPK mahasiswa. Pada penelitian sebelumnya, disimpulkan bahwa faktor sosial demografi dan lingkungan berpengaruh pada prestasi akademik [2]. Sedangkan pada penelitian [3] mahasiswa dari latar belakang orang tua yang berpisah cenderung memiliki nilai akademik yang rendah dibandingkan dengan mahasiswa yang memiliki kehidupan sosial yang lebih baik. Penelitian lain [4] faktor gender mempengaruhi tingkat keberhasilan dari mahasiswa. Hal ini menunjukkan

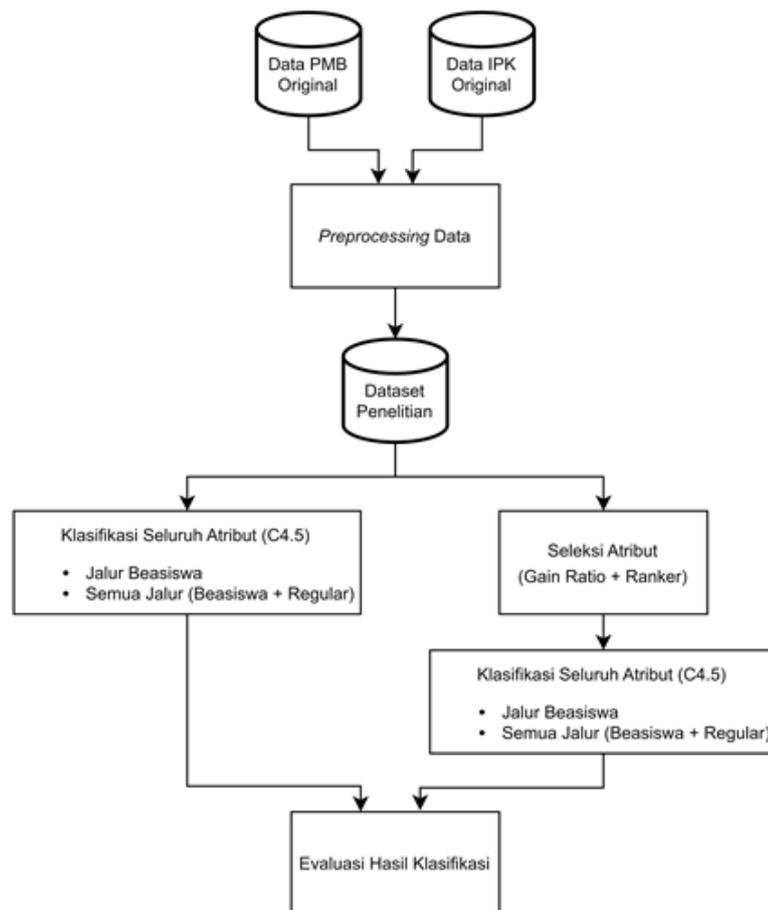
bahwa terdapat banyak faktor yang mempengaruhi prestasi belajar siswa.

Beberapa penelitian sebelumnya memanfaatkan berbagai metode untuk menganalisa latar belakang mahasiswa dan kinerja mahasiswa mendatang. Klasifikasi data menjadi topik yang saat ini populer untuk melakukan prediksi data. Klasifikasi sendiri dilakukan untuk membedakan suatu data dimana kelas dari label data tidak diketahui [5]. Mengaplikasikan *data mining* menjadi perhatian pada penelitian [6] yang melakukan klasifikasi untuk membandingkan kinerja dari algoritma ID3, CART, dan C4.5. Dari hasil perbandingan didapatkan algoritma C4.5 mendapatkan hasil akurasi yang lebih baik. Hasil dari penelitian [7] melakukan klasifikasi kinerja mahasiswa dengan menggunakan algoritma ID3 dan C4.5 untuk meningkatkan kinerja mereka. Hasil dari penelitian didapatkan kedua algoritma menghasilkan akurasi yang sama dengan data yang lebih kecil, sedangkan pada data besar algoritma C4.5 lebih unggul dibandingkan dengan algoritma ID3. Penelitian [8] memanfaatkan algoritma klasifikasi ID3 dan C4.5 untuk menganalisa kinerja mahasiswa. Namun pada penelitian tersebut hanya memprediksi berdasarkan latar belakang nilai kuantitatif dari mahasiswa.

Algoritma C4.5 adalah perbaikan dari standart algoritma *decision tree* yang dapat menangani nilai-nilai dari atribut yang hilang [9]. Penelitian lain [10] melakukan analisa terhadap mahasiswa lulus tepat waktu dengan memanfaatkan algoritma C4.5. Klasifikasi untuk memprediksi kualitas mahasiswa berdasarkan 7 atribut, yaitu nilai akhir Sekolah Menengah Atas (SMA), pekerjaan orang tua, penghasilan orang tua, mata kuliah yang mengulang, IPK mahasiswa pada semester 4, Nilai yang didapatkan setiap mata kuliah, dan banyak semester yang diambil. Penelitian ini sudah mengkombinasikan data kuantitatif dan kondisi sosial ekonomi siswa. Namun atribut-atribut tersebut tidak dapat diaplikasikan pada mahasiswa yang memiliki kurikulum belajar yang berbeda, khususnya institusi pendidikan vokasi seperti Politeknik ABC.

Oleh karena itu, penelitian ini bertujuan untuk menganalisa berbagai faktor sosial ekonomi dari data pendaftaran mahasiswa baru dan pengaruhnya

terhadap prestasi belajar mahasiswa vokasi, yang ditunjukkan dengan nilai IPK. Penelitian ini dilakukan dengan memanfaatkan algoritma klasifikasi C4.5, dan seleksi atribut dengan metode *Gain Ratio* untuk mendapatkan atribut yang berpengaruh signifikan terhadap prestasi belajar mahasiswa. Data yang digunakan pada penelitian ini adalah data mahasiswa Politeknik ABC angkatan masuk tahun 2019. Klasifikasi dilakukan untuk mengetahui pengaruh 16 atribut dari faktor sosial dan ekonomi mahasiswa. Atribut-atribut tersebut seperti Rata-Rata Nilai Raport Pelajaran Matematika, Anak Ke-, Tahun Lulus dari SMA/ sederajat, Jumlah Saudara, Jenis Kelamin, Pekerjaan Ayah, Usia, Asal Provinsi Rumah, Pendidikan Ayah, Rata-Rata Nilai Raport Pelajaran Bahasa Inggris, Asal Provinsi Sekolah, Rata-Rata Nilai Raport Pelajaran Fisika, Rata-Rata Penghasilan Orang Tua, Pekerjaan Ibu, Pendidikan Ibu, dan Kota Pendaftaran.



Gambar 1: Tahapan Penelitian

Metode Penelitian

Penelitian ini terbagi menjadi beberapa tahapan, seperti ditunjukkan pada Gambar 1. Penelitian menggunakan keseluruhan data PMB Politeknik

ABC tahun 2019. Terdapat total 4610 calon mahasiswa yang mendaftar di Politeknik ABC melalui semua jalur masuk dan gelombang masuk. Pada setiap data mahasiswa, terdapat beberapa infor-

masi seputar data diri, kondisi sosial ekonomi keluarga, kondisi sekolah asal termasuk beberapa nilai dari sekolah asal. Data prestasi belajar mahasiswa yang ditunjukkan dengan nilai IPK (Indeks Prestasi Kumulatif) saat mahasiswa lulus juga digunakan dalam penelitian ini. Data ini kemudian disebut sebagai data mentah/*raw data* dalam penelitian.

Preprocessing data

Preprocessing data dijalankan untuk mempersiapkan data mentah menjadi dataset matang yang siap diolah. *Preprocessing data* perlu dilakukan karena data mentah sangat rentan dengan adanya *noise data*, *missing data*, dan data yang tidak konsisten karena besarnya ukuran database atau data tersebut berasal dari gabungan beberapa *dataset* yang berbeda. Rendahnya kualitas data akan menyebabkan hasil pengolahan data juga memiliki kualitas yang rendah. Teknik *preprocessing* yang tepat akan meningkatkan kualitas data.

Terdapat beberapa teknik *preprocessing data* dalam *data mining*, diantaranya adalah pembersihan data (*data cleaning*), integrasi data (*data integration*), reduksi data (*data reduction*), dan transformasi data (*data transformation*). Data yang memiliki banyak *noise* dan tidak konsisten dapat dilakukan *preprocessing* dengan teknik pembersihan data. Sedangkan untuk menggabungkan beberapa sumber data ke dalam suatu penyimpanan data digunakan teknik integrasi data. Teknik reduksi data dapat digunakan untuk mengurangi ukuran data dan menghilangkan duplikasi data. Sementara untuk mengubah atau mengkonversi skala pada atribut data digunakan teknik transformasi data [11], [12].

Dataset Penelitian

Pada penelitian ini, *dataset* penelitian adalah dataset yang dihasilkan dari serangkaian *preprocessing* data yang dilakukan sebelumnya. *Dataset* terdiri 192 *instances* mahasiswa aktif Politeknik ABC angkatan tahun 2019 dan lulus tahun 2022. Masing-masing *instance* terdiri dari 16 atribut kondisi mahasiswa yang didapatkan saat pendaftaran mahasiswa. Atribut kondisi mahasiswa tersebut dijadikan sebagai atribut masukan. Sementara itu, atribut keluaran (*class attribute*) adalah nilai IPK mahasiswa selama 6 semester. Nilai IPK ini menunjukkan prestasi belajar mahasiswa selama proses perkuliahan. Keterangan mengenai atribut masukan, keluaran, dan rentang nilainya ditunjukkan pada Tabel 1. Selain atribut keluaran dan masukan, terdapat keterangan tambahan pada masing-masing *instance*, yaitu apakah mahasiswa tersebut masuk melalui jalur beasiswa atau jalur reguler.

Klasifikasi

Tahapan selanjutnya adalah melakukan pemetaan atribut masukan terhadap atribut keluaran, dalam hal ini adalah atribut data pendaftaran terhadap prestasi belajar mahasiswa. Proses pemetaan dilakukan dengan mekanisme klasifikasi. Klasifikasi yang berasal dari bahasa Latin *classis*, merupakan suatu proses pengelompokan. Pengelompokan tersebut bertujuan untuk mengumpulkan benda/entitas yang sama dan memisahkan benda/entitas yang berbeda/tidak sama. Pada konteks umum, klasifikasi adalah usaha menata alam pengetahuan ke dalam tata urutan sistematis [13].

Dalam ranah *data mining*, klasifikasi adalah proses menemukan model (atau fungsi) yang menjabarkan dan membedakan kelas data atau konsep. Terdapat dua fase dalam proses klasifikasi, yaitu fase pelatihan dan fase pengujian. Fase pelatihan disebut juga fase pembentukan model. Model dibentuk didasarkan pada analisis dari set data pelatihan pada obyek data yang kelasnya dikenal. Pembentukan model ini bertujuan untuk memprediksi obyek yang kelas labelnya tidak/belum diketahui. Fase prediksi ini disebut juga sebagai fase pengujian, yang melibatkan data pengujian. Pada umumnya terdapat perbedaan data yang digunakan dalam proses pengujian dan pelatihan, sehingga hasil pengujian yang didapatkan dapat digunakan sebagai evaluasi hasil klasifikasi [11].

Terdapat berbagai macam metode yang dapat digunakan untuk proses klasifikasi. Penelitian ini menggunakan algoritma klasifikasi C4.5. Algoritma C4.5 adalah sebuah metode yang termasuk dalam kategori *decision tree* (pohon pengambilan keputusan). Algoritma ini melakukan pemilihan atribut berdasarkan *Gain Ratio* yang dimilikinya. Atribut yang memiliki *Gain Ratio* tertinggi akan menjadi root [11]. *Decision tree* kemudian dibangun dengan cara membagi data secara rekursif. Pembagian ini nantinya akan menghasilkan tiap bagian terdiri dari data yang berasal dari kelas yang sama, sehingga terciptalah klasifikasi tiap kategori. Langkah-langkah pembentukan *tree* dapat dideskripsikan sebagai berikut, dimana D dianggap sebagai data pelatihan yang dipartisi secara rekursif dan C adalah kelas yang ada pada data [11]. Langkah-langkah tersebut adalah:

1. Menghitung nilai informasi dari atribut kelas, yang dinyatakan pada Persamaan 1.

$$Info(D) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (1)$$

dimana p_i merupakan proporsi jumlah kasus atau *instance* dari kelas i terhadap jumlah semua kasus atau *instance* di D atau peluang bahwa sebuah kasus atau *instance* di D memiliki kelas C_i .

- Menghitung nilai informasi dari setiap atribut, nilai informasi dari atribut A dapat dinyatakan dalam Persamaan 2.

$$Info_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times Info(D_j) \quad (2)$$

dimana D_j merupakan jumlah *instance* dengan nilai j untuk atribut A dan $Info(D_v)$ adalah nilai relatif informasi untuk D_j .

- Menghitung informasi *gain* dari setiap atribut, dapat dinyatakan pada Persamaan 3.

$$Gain(A) = Info(D) - Info_A(D) \quad (3)$$

- Menghitung informasi *split*, sesuai dengan Persamaan 4.

$$SplitInfo_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times \log_2 \left(\frac{|D_j|}{|D|} \right) \quad (4)$$

- Menghitung *Gain Ratio* dari setiap atribut
Gain Ratio adalah rasio antara informasi *gain* dan info *split* dari atribut, yang dapat dinyatakan dalam Persamaan 5.

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)} \quad (5)$$

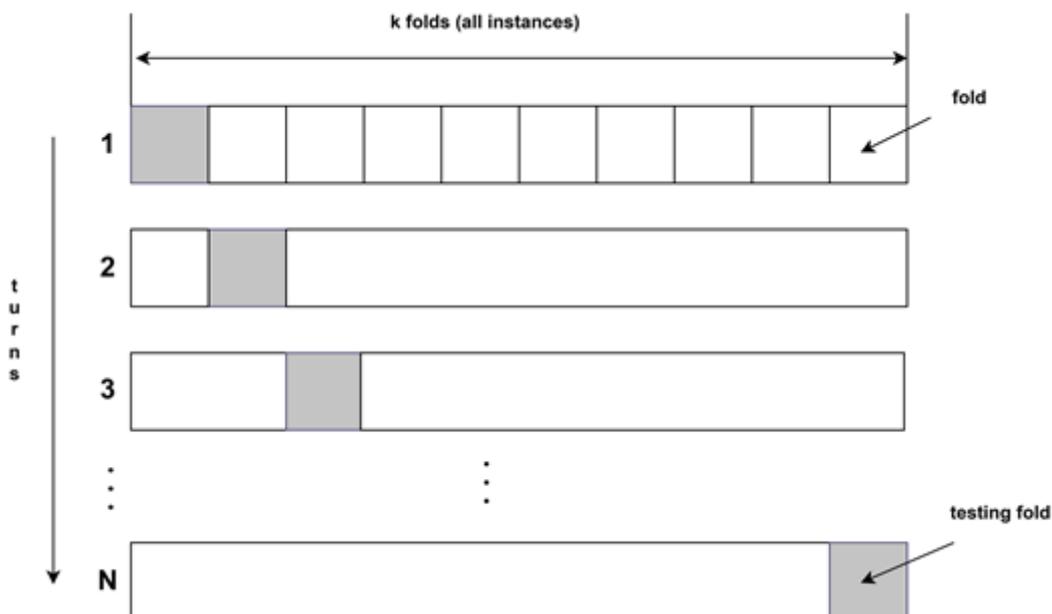
Atribut yang memiliki *Gain Ratio* tertinggi akan menjadi *root*.

Seleksi Atribut

Proses selanjutnya adalah reduksi 16 atribut masukan menjadi *subset-subset* dengan metode seleksi atribut. Metode seleksi atribut yang digunakan adalah *Gain Ratio*, dengan sistem evaluasi *Ranker*. *Gain Ratio* merupakan suatu metode seleksi yang melihat sejauh mana suatu atribut memberikan pengaruh terhadap atribut keluaran (*class attribute*). Metode *Gain Ratio* menormalkan nilai ‘informasi’ yang dihasilkan oleh masing-masing atribut, sehingga dapat digunakan untuk mengembangkan pohon yang tidak merata. Hasil pengolahan atribut masukan dengan *Gain Ratio* kemudian diurutkan dengan metode *Ranker*, untuk mengurutkan atribut masukan dari yang paling berpengaruh hingga yang tidak berpengaruh terhadap atribut keluaran [14]. Hasil dari proses seleksi atribut kemudian diolah kembali dengan metode klasifikasi, terutama untuk 4 atribut masukan yang paling berpengaruh.

Evaluasi Hasil Klasifikasi

Langkah penelitian selanjutnya menuju ke tahap evaluasi hasil klasifikasi. Evaluasi performa dari metode klasifikasi dapat dilakukan menggunakan data pelatihan dan pengujian yang didapatkan selama proses klasifikasi. Untuk memisahkan data pada proses pelatihan dan pengujian dapat dilakukan dengan berbagai macam metode, salah satunya dengan metode *k-fold cross validation*. Pada metode *k-fold cross validation*, *dataset* dibagi menjadi sebanyak k bagian (*fold*), seperti ditunjukkan pada Gambar 2 [11].



Gambar 2: Skema *K-fold Cross Validation*

Nilai k yang umum digunakan adalah 10. Pada *10-fold cross validation*, dataset dibagi secara acak menjadi 10 bagian. Tiap bagian memiliki proporsi jumlah *instance* yang sama. Pada setiap iterasi, satu bagian digunakan sebagai data pengujian dan sisa 9/10 bagian digunakan sebagai data pelatihan. Iterasi dilakukan sebanyak 10 kali, dan nilai *error* akan dihasilkan pada setiap proses iterasinya. Setelah selesai iterasi, estimasi *error* dari 10 bagian tersebut dijumlahkan dan dirata-rata untuk mendapatkan estimasi *error* keseluruhan [13].

Proses evaluasi klasifikasi menghasilkan parameter tingkat akurasi. Akurasi merupakan persentase perbandingan *instance* yang berhasil terklasifikasi benar dengan jumlah total seluruh *instance* dalam *dataset*, seperti ditunjukkan pada Persamaan 6.

$$Accuracy = \frac{Correctlyclassified}{Totalinstance} \times 100\% \quad (6)$$

Semakin tinggi persentase akurasi yang dihasilkan dari fase pelatihan dan pengujian menunjukkan semakin baiknya klasifikasi yang dilakukan [17]. Pada penelitian ini, data akurasi masing-masing klasifikasi kemudian dibandingkan untuk mengetahui apakah pemilihan atribut dengan proses seleksi atribut berpengaruh terhadap akurasi hasil klasifikasi.

Hasil dan Pembahasan

Tahap awal penelitian adalah *preprocessing* data pada data mentah yang sudah dikumpulkan. Teknik pembersihan dan reduksi data dilakukan pada 4610 data calon mahasiswa pada PMB Politeknik ABC tahun 2019. Data-data calon mahasiswa yang terduplikasi, memiliki *missing value* maupun memiliki noise dihilangkan dari list data penelitian. Begitupun atribut atau parameter yang tidak diperlukan dalam penelitian juga dihilangkan. Data calon mahasiswa kemudian diintegrasikan dengan data IPK kelulusan tahun 2022, sehingga tersisa data 192 mahasiswa untuk diolah ke proses selanjutnya. Dari total mahasiswa tersebut, 65 orang diantaranya adalah mahasiswa dengan jalur beasiswa, sementara sisanya masuk melalui jalur reguler.

Preprocessing data selanjutnya menggunakan teknik transformasi data. Dari keseluruhan 16 atribut kondisi mahasiswa yang diperoleh pada saat proses PMB merupakan data kualitatif. Misalnya data pendidikan orang tua yang ditunjukkan dalam keterangan lulus SD, SMP, dan seterusnya sampai S3. Teknik transformasi data akan mengkonversi nilai dari data kualitatif menjadi kategori angka. Pada data pendidikan orang tua, kondisi tidak sekolah dikonversi menjadi angka 0, kemudian lulus SD menjadi angka 1, SMP menjadi angka 2, SMA menjadi angka 3, Diploma menjadi angka 4, S1/Sarjana menjadi angka 5, S2/Master menjadi angka 6, dan S3/Doktor dikonversikan menjadi

angka 7. Proses tersebut juga kemudian dilakukan terhadap 15 atribut masukan dan 1 atribut keluaran yang digunakan dalam penelitian ini.

Hasil dari serangkaian teknik *preprocessing* data adalah *dataset* penelitian yang sudah siap diolah. *Dataset* terdiri dari 192 *instances*, dengan 16 atribut masukan dan 1 atribut keluaran (*class attribute*). Atribut masukan didapatkan dari calon mahasiswa saat proses PMB, sementara atribut keluaran adalah prestasi belajar yang ditunjukkan dengan nilai IPK kelulusan. Detail atribut yang digunakan dalam penelitian ini termasuk rentang nilai yang digunakan ditunjukkan pada Tabel 1. Terlihat pada Tabel 1 bahwa atribut masukan terdiri dari data mengenai diri calon mahasiswa, kondisi sosial ekonomi keluarga, hingga nilai raport untuk 3 mata pelajaran utama.

Tabel 1: Atribut Penelitian dan Rentang Nilai

No	Nama Atribut	Jenis Atribut	Rentang Nilai
1.	Kota Pendaftaran	Masukan	1-7
2.	Usia	Masukan	1-6
3.	Jenis Kelamin	Masukan	1-2
4.	Anak Ke	Masukan	1-5
5.	Jumlah Saudara	Masukan	0-5
6.	Asal Provinsi Rumah	Masukan	1-7
7.	Pekerjaan Ayah	Masukan	0-5
8.	Pendidikan Ayah	Masukan	0-7
9.	Pekerjaan Ibu	Masukan	0-5
10.	Pendidikan Ibu	Masukan	0-7
11.	Rata-rata Penghasilan Orang Tua	Masukan	1-5
12.	Asal Provinsi Sekolah	Masukan	1-7
13.	Rata-Rata Nilai Raport Pelajaran Matematika	Masukan	1-5
14.	Rata-Rata Nilai Raport Pelajaran Fisika	Masukan	1-5
15.	Rata-Rata Nilai Raport Pelajaran Bahasa Inggris	Masukan	1-5
16.	Tahun Lulus dari SMA/ sederajat	Masukan	1-5
17.	IPK (Indeks Prestasi Kumulatif)	Keluaran	1-5

Proses klasifikasi dengan metode C4.5 dilakukan terhadap *dataset* penelitian. Klasifikasi dilakukan 2 kali, pertama terhadap data mahasiswa yang mendapatkan beasiswa (65 *instances*), kemudian proses kedua dilakukan terhadap data keseluruhan mahasiswa (192 *instances*). Evaluasi klasifikasi dengan *10-fold cross validation* menggunakan *WEKA Data Mining Tools* menghasilkan persentase nilai akurasi yang ditunjukkan pada Tabel 2. Nilai akurasi merupakan persentase perbandingan *instance* yang berhasil terklasifikasi benar dengan jumlah total seluruh *instance* dalam *dataset*. Semakin tinggi nilai akurasi menunjukkan semakin baik hasil klasifikasi. Nilai maksimal akurasi adalah 100%, apabila proses klasifikasi berhasil mengklasifikasikan semua *instance* dengan benar.

Hasil akurasi pada Tabel 2 menunjukkan bahwa atribut masukan data pendaftaran berpengaruh

terhadap prestasi belajar mahasiswa, walaupun persentase yang ditunjukkan belum terlalu tinggi. Pada pengolahan klasifikasi *dataset* total mahasiswa, nilai akurasi yang dihasilkan adalah 69.79%, yang berarti bahwa sebanyak 134 dari total 192 instances berhasil terklasifikasi dengan benar. Pada Tabel 2 juga terlihat bahwa hasil akurasi dengan *dataset* yang hanya berisi mahasiswa beasiswa memiliki akurasi yang lebih rendah dibandingkan apabila menggunakan data total mahasiswa. Kondisi ini menunjukkan bahwa penambahan *instances* berpengaruh terhadap peningkatan akurasi.

Tabel 2: Akurasi Hasil Klasifikasi C4.5 dengan Seluruh Atribut

<u>Jumlah Atribut</u>	Akurasi <i>Dataset</i> Mahasiswa Beasiswa - 65 instances (dalam %)	Akurasi <i>Dataset</i> Total Mahasiswa - 192 instances (dalam %)
Total (16 atribut)	67.19	69.79

Proses seleksi atribut dilakukan untuk mereduksi atribut masukan yang tidak memiliki pengaruh signifikan terhadap class attribute. Dengan kata lain, proses ini akan memilih atribut yang berpengaruh signifikan terhadap class attribute. Pada penelitian ini, atribut yang terpilih dari hasil seleksi atribut menunjukkan atribut masukan kondisi mahasiswa yang berpengaruh signifikan terhadap prestasi belajar mahasiswa selama perkuliahan. Pada penelitian ini, digunakan seleksi atribut dengan metode Gain Ratio, dan sistem Ranker untuk mengurutkan atribut dari yang paling signifikan. Hasil proses seleksi atribut dilakukan 2 kali, yaitu pada data mahasiswa jalur beasiswa dan data mahasiswa total, seperti ditunjukkan pada Tabel 3.

Pada urutan signifikansi atribut masukan yang ditunjukkan di Tabel 3 terlihat bahwa, 4 hal paling berpengaruh terhadap prestasi belajar mahasiswa jalur beasiswa adalah rata-rata nilai raport pelajaran Matematika, anak ke berapa dalam keluarga, tahun lulus dari jenjang SMA/ sederajat, dan jumlah saudara. Ada 1 atribut nilai kuantitatif dan 3 atribut kondisi sosial ekonomi mahasiswa. Sementara itu, pada *dataset* total mahasiswa secara keseluruhan, 4 hal yang paling berpengaruh terhadap prestasi mahasiswa adalah rata-rata nilai raport pelajaran Matematika, kota pendaftaran, rata-rata nilai raport pelajaran Bahasa Inggris dan rata-rata nilai raport pelajaran Fisika. Ada 3 atribut nilai kuantitatif dan 1 atribut kondisi sosial ekonomi mahasiswa.

Keempat atribut yang signifikan terhadap keluaran kemudian dilakukan proses klasifikasi dan evaluasi performa. Klasifikasi pertama melibatkan 4 atribut masukan yang dipetakan ke atribut keluaran. Proses dilanjutkan dengan menghilangkan 1 atribut masukan sehingga hanya 3 atribut masukan yang paling signifikan yang digunakan. Hal tersebut berulang untuk 2 atribut masukan paling sig-

nifikan, dan kemudian 1 atribut masukan saja yang digunakan.

Tabel 3: Urutan Hasil Seleksi Atribut *Gain Ration* + *Ranker*

Rank	<i>Dataset</i> Mahasiswa Beasiswa (65 instances)	<i>Dataset</i> Total Mahasiswa (192 instances)
	1.	Rata-Rata Nilai Raport Pelajaran Matematika
2.	<u>Anak Ke</u>	<u>Kota Pendaftaran</u>
3.	Tahun Lulus dari SMA/ sederajat	Rata-Rata Nilai Raport Pelajaran B. Inggris
4.	<u>Jumlah Saudara</u>	Rata-Rata Nilai Raport Pelajaran Fisika
5.	<u>Jenis Kelamin</u>	<u>Asal Provinsi Rumah</u>
6.	<u>Pekerjaan Ayah</u>	<u>Jumlah Saudara</u>
7.	<u>Usia</u>	<u>Anak Ke</u>
8.	<u>Asal Provinsi Rumah</u>	<u>Jenis Kelamin</u>
9.	<u>Pendidikan Ayah</u>	<u>Usia</u>
10.	Rata-Rata Nilai Raport Pelajaran B. Inggris	Tahun Lulus dari SMA/ sederajat
11.	<u>Asal Provinsi Sekolah</u>	<u>Pekerjaan Ayah</u>
12.	Rata-Rata Nilai Raport Pelajaran Fisika	<u>Pekerjaan Ibu</u>
13.	Rata-rata Penghasilan Orang Tua	<u>Asal Provinsi Sekolah</u>
14.	<u>Pekerjaan Ibu</u>	Rata-rata Penghasilan Orang Tua
15.	<u>Pendidikan Ibu</u>	<u>Pendidikan Ibu</u>
16.	<u>Kota Pendaftaran</u>	<u>Pendidikan Ayah</u>

Hasil evaluasi ditunjukkan pada Tabel 4, memperlihatkan bahwa proses klasifikasi dengan seleksi atribut menghasilkan nilai akurasi yang lebih tinggi dibandingkan dengan penggunaan seluruh atribut seperti ditunjukkan pada Tabel 2. Dengan akurasi tertinggi adalah 77.60% dari pengolahan dataset mahasiswa beasiswa, dapat diartikan bahwa 50 dari 65 *instances* dapat terprediksi benar.

Tabel 4: Akurasi Hasil Klasifikasi C4.5 dengan Atribut Hasil Seleksi

<u>Jumlah Atribut</u>	Akurasi <i>Dataset</i> Mahasiswa Beasiswa - 65 instances (dalam %)	Akurasi <i>Dataset</i> Total Mahasiswa - 192 instances (dalam %)
<u>4 atribut</u>	77.08	75.00
<u>3 atribut</u>	77.60	75.00
<u>2 atribut</u>	77.60	75.00
<u>1 atribut</u>	72.40	75.00

Penutup

Dari hasil penelitian yang dilakukan, didapatkan kesimpulan bahwa faktor sosial ekonomi yang didapatkan dari data pendaftaran mahasiswa baru berpengaruh terhadap prestasi belajar mahasiswa. Hal ini ditunjukkan dengan persentase akurasi hasil klasifikasi dengan algoritma C4.5. Faktor sosial ekonomi digunakan sebagai atribut masukan, sementara IPK mahasiswa digunakan sebagai atribut keluaran. Proses klasifikasi dengan subset hasil seleksi atribut menghasilkan nilai akurasi yang

lebih baik dibandingkan klasifikasi dengan seluruh atribut. Proses seleksi atribut membuat proses komputasi lebih efisien, mendapatkan atribut masukan yang berpengaruh signifikan terhadap atribut keluaran, dengan tidak mengorbankan kualitas performa, bahkan meningkatkannya.

Hasil seleksi atribut menunjukkan bahwa rata-rata nilai raport pelajaran Matematika adalah atribut yang paling berpengaruh pada prestasi belajar mahasiswa, baik pada mahasiswa jalur beasiswa maupun jalur reguler. Hal ini menunjukkan bahwa penguasaan terhadap pelajaran Matematika secara umum berpengaruh signifikan terhadap prestasi belajar mahasiswa. Pada mahasiswa beasiswa, kondisi sosial ekonomi cukup signifikan mempengaruhi prestasi belajar. Namun kondisi total mahasiswa Politeknik ABC pada umumnya menunjukkan bahwa penguasaan materi Matematika, Fisika, dan Bahasa Inggris membantu mahasiswa dalam meningkatkan prestasi belajarnya. Hasil ini sekaligus menjadi bahan pertimbangan bagi institusi dalam proses penerimaan mahasiswa baru, baik melalui jalur beasiswa maupun reguler.

Saran yang dapat diberikan untuk pengembangan penelitian ini adalah penambahan atribut lain yang mempengaruhi prestasi belajar mahasiswa, sehingga dapat meningkatkan nilai akurasi. Atribut lain yang ditambahkan misalnya nilai-nilai dan data yang diperoleh selama proses seleksi mahasiswa, seperti nilai seleksi masuk dan hasil wawancara. Selain itu, penambahan data penelitian menjadi multi years juga diharapkan dapat meningkatkan performa model klasifikasi. Apabila model klasifikasi yang dibuat sudah cukup stabil, dapat pula ditingkatkan untuk penerapannya ke dalam sistem penerimaan mahasiswa baru.

Daftar Pustaka

- [1] Usamah bin Mat, Norlida Buniyamin, Pauziah Mohd Arsad, Rosni Abu Kassim and Usamah bin Mat , “An Overview of Using Academic Analytics to Predict”, IEEE 5th Conference on Engineering Education (ICEED), 2013.
- [2] U. A. Mirza Irfan Ali, “Attribution to success and failure among medical students”, Pak ArmedForcesMed J , vol. 6, p. 67, 2017.
- [3] S. Lumley, P. Ward, L. Roberts and J. P. Mann, “Self-reported extracurricular activity, academic success, and quality of life in UK medical students”, Int J Med Educ, vol. 6, pp. 111–117, Sep. 2015, doi: 10.5116/ijme.55f8.5f04.
- [4] J. R. Lechien, C. Kempnaers, M. Dramaix and P. Linkowski, “Influence of gender and selection procedures on the academic performance of undergraduate medical students”, Acta Med Acad, vol. 45, no. 2, pp. 145–151, doi: 10.5644/ama2006-124.170, Nov. 2016.
- [5] M. Nowak, “Defining Project Approach using Decision Tree and Quasi-hierarchical Multiple Criteria Method”, in Procedia Engineering, vol. 172, pp. 791–799. doi: 10.1016/j.proeng.2017.02.125, 2017.
- [6] R. Singh and S. Pal, “Application of Machine Learning Algorithms to Predict Students Performance”, Article in International Journal of Advanced Computer Research, vol. 29, no. 5, pp. 7249–7261, 2020.
- [7] H. Pallathadka, A. Wenda, E. Ramirez-Asís, M. Asís-López, J. Flores-Albornoz and K. Phasinam”, Classification and prediction of student performance data using various machine learning algorithms,” Mater Today Proc, doi: 10.1016/j.matpr.2021.07.382, Jul. 2021.
- [8] K. Adhatrao, A. Gaykar, A. Dhawan, R. Jha and V. Honrao, “Predicting Students’ Performance Using ID3 and C4.5 Classification Algorithms”, International Journal of Data Mining & Knowledge Management Process, vol. 3, no. 5, pp. 39–52, doi: 10.5121/ijdkp.2013.3504, Sep. 2013.
- [9] X. Wang, C. Zhou, and X. Xu, “Application of C4.5 decision tree for scholarship evaluations”, in Procedia Computer Science, vol. 151, pp. 179–184. doi: 10.1016/j.procs.2019.04.027, 2019.
- [10] N. A. Prahastiwi, R. Andreswari, and R. Fauzi, “Students Graduation Prediction Based On Academic Data Record Using The Decision Tree Algorithm C4.5 Method”, JURTEKSI (Jurnal Teknologi dan Sistem Informasi), vol. 8, no. 3, pp. 295–304, doi: 10.33330/jurteksi.v8i3.1680, Aug. 2022.
- [11] J. Han, M. Kamber, and J. Pei, “Data Mining. Concepts and Techniques, 3rd Edition”, The Morgan Kaufmann Series in Data Management Systems, 2011.
- [12] Ian H. Witten, Eibe Frank and Mark A. Hall , “Data Mining: Practical Machine Learning Tools and Techniques”, The Morgan Kaufmann Series in Data Management Systems, 2011.
- [13] Ahmad Saikhu, Joko Lianto dan Umi Hanik, “Fuzzy Decision Tree dengan Algoritma C4.5 pada Data Diabetes IndianPima”, Konferensi Nasional Sistem dan Informatika, Bali, 2011.
- [14] E. Harris, “Information Gain Versus Gain Ratio: A Study of Split Method Biases”, Ann Math Artif Intell, 2002.
- [15] C. Cassisi, P. Montalto, M. Aliotta, A. Cannata, and A. Pulvirenti, “Similarity Measures and Dimensionality Reduction Techniques for Time Series Data Mining”, in Advances in Data Mining Knowledge Discovery and Applications, InTech, doi: 10.5772/49941, 2012.