PENERAPAN METODE DECISION TREE ALGORITMA C4.5 UNTUK SELEKSI CALON PENERIMA BEASISWA TINGKAT UNIVERSITAS

Nadiya Hijriana[1)] dan Muhammad Rasyidan[2)]

[1,2]Program Studi Teknik Informatika Universitas Islam Kalimantan, Banjarmasin
E-mail: nadyahijriana@yahoo.com

ABSTRACT

Scholarship is an award of financial aid for a student to further their education where it usually provided by goverment or any private institution. One of the problems that arise in distribution of scholarship is how to select student who is really approriate to get scholarship. This research using Decision Tree Method from Algorithm C4.5 and RapidMiner 5.3 as software for making a model that it can be used as a tool to help making decision in selection's process of scholarship recipients.

Keywords: Scholarship, Selection, Decision Tree Method, C4.5 Algorithm

PENDAHULUAN

Mendapatkan pendidikan adalah merupakan hak asasi sekaligus kewajiban bagi setiap warga negara Indonesia sebagaimana yang tercantum dalam UUD 1945 Pasal 31 ayat 1-2. Sektor pendidikan merupakan salah satu hal yang menjadi prioritas atau perhatian utama pemerintah dikarenakan kemajuan suatu bangsa salah satunya ditentukan oleh bagaimana kualitas pendidikan warga negaranya. Oleh karena itu sesuai UUD 1945 pemerintah mengusahakan dan menyelenggarakan satu sistem pendidikan nasional serta memprioritaskan anggaran pendidikan sekurang-kurangnya 20% dari anggaran pendapatan dan belanja negara serta dari anggaran pendapatan dan belanja daerah untuk memenuhi kebutuhan penyelenggaraan pendidikan nasional.

Namun tak dapat dipungkiri bahwa masih terdapat warga negara yang mempunyai tingkat penghasilan yang rendah sehingga mengakibatkan terhambatnya kesempatan untuk mendapatkan pendidikan yang layak dan sesuai serta berkelanjutan. Untuk itu pemerintah memberikan bantuan dana (beasiswa) sebagaimana yang tercantum dalam UU No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional khususnya Pasal 12 ayat 1 butir c dan d, dimana setiap peserta didik pada setiap satuan pendidikan berhak mendapatkan beasiswa bagi yang berprestasi yang orangtuanya tidak mampu membiayai pendidikannya dan mendapatkan biaya pendidikan bagi mereka yang orangtuanya tidak mampu membiayai pendidikannya.

Salah satu permasalahan dalam penyaluran dana beasiswa adalah masih adanya pemberian beasiswa yang kurang tepat sasaran, sehingga menyebabkan tidak maksimalnya pendayagunaan dana beasiswa tersebut. Selain itu proses seleksi calon penerima beasiswa yang dilakukan secara manual terkadang membutuhkan waktu cukup lama sehingga kurang efisien.

Beberapa penelitian telah dilakukan untuk membantu pengambilan keputusan dalam proses seleksi calon penerima beasiswa, dengan harapan agar beasiswa dapat diberikan kepada mahasiswa yang memang pantas mendapatkannya atau tepat sasaran, serta dapat memberikan efisiensi waktu dalam proses pengambilan keputusan. Salah satu metode yang sering digunakan adalah metode Decision Tree Algoritma C4.5 seperti penelitian yang dilakukan oleh Muh. Safri Juliardi, dkk. (2017) dengan kesimpulan bahwa Algoritma C4.5 dapat digunakan sebagai metode alternatif untuk membantu proses seleksi penerima beasiswa Bidikmisi.

Selain itu, Jose Augusto Duarte Guterres (2015) juga melakukan perbandingan metode untuk klasifikasi calon penerima beasiswa. Hasilnya adalah bahwa Algoritma C45 layak dijadikan sebagai pendukung keputusan dalam proses pengajuan penerima beasiswa pada Stikom Artah

Buana Kupang karena hasil persentasi dari Algoritma C45 yang lebih tinggi daripada metode klasifikasi yang lainnya. Sedangkan Jonathan Rinaldy (2014) membuat sebuah sistem penunjang keputusan berbasis web dengan menggunakan rule yang terbentuk dari pengujian metode Decision Tree Algoritma C4.5. Dengan adanya sistem tersebut mempermudah pendaftaran pengajuan beasiswa dan calon penerima beasiswa dapat dipilih secara adil.

Untuk itu melalui penelitian ini penulis mencoba menguji kegunaan dan akurasi metode Decision Tree Algoritma C4.5 sebagai salah satu alternatif untuk mendukung keputusan dalam menyeleksi calon penerima beasiswa pada universitas XYZ sehingga pemberian dana beasiswa dapat tepat sasaran, diberikan kepada mahasiswa yang memang layak dan pantas mendapatkannya. Selain itu penulis juga ingin mengetahui bagaimana pengaruh suatu atribut jika data pada atribut tersebut diubah ke dalam bentuk kelas atau klasifikasi, apakah lebih mempengaruhi tingkat akurasi nilai dari model yang diuji atau malah sebaliknya.

METODE PENELITIAN

Penelitian ini menggunakan metode eksperimen, yang terdiri dari: (1) Pengumpulan data, (2) Pengolahan data awal, (3) Model yang diusulkan, (4) Pengujian model dan (5) Evaluasi dan validasi model.

Pengumpulan Data

Data primer adalah data yang diperoleh dari hasil penelitian. Data primer dalam penelitian ini adalah data hasil uji dengan menggunakan metode *Decision Tree* Algoritma C4.5. Sedangkan Data sekunder adalah data yang diperoleh secara tidak langsung bersumber dari dokumentasi, literatur, buku, jurnal dan informasi lainnya yang ada hubungannya dengan masalah yang diteliti. Data sekunder pada penelitian ini antara lain buku dan jurnal tentang metode *Decision Tree* Algoritma C4.5 serta data mahasiswa Angkatan 2012 di Universitas XYZ untuk Tahun Akademik 2014/2015 (Semester 6) dan Tahun Akademik 2015/2016 (Semester 7) sebanyak 444 orang.

Pengolahan Data Awal

Data awal yang didapatkan berupa data mahasiswa seluruh angkatan pada Program Studi Teknik Informatika di Universitas XYZ, kemudian dilakukan pengolahan data lagi termasuk pengolahan klasifikasi data sehingga didapatkan data mahasiswa Angkatan 2012 dan atribut yang akan dipakai dalam pengujian antara lain NPM, Nama, Status, Beasiswa (Sudah pernah mendapatkan beasiswa atau tidak), Indeks Prestasi Semester 6 atau IP-6, Indeks Prestasi Semester 7 atau IP-7, Rerata, Kelas (Klasifikasi IPK), Penghasilan Orangtua, Klasifikasi Penghasilan, Jumlah Tanggungan, dan Klasifikasi Tanggungan.

Model Yang Diusulkan

Model yang diusulkan untuk seleksi calon penerima beasiswa dalam penelitian ini adalah *Decision Tree* Algoritma C4.5 dengan menggunakan *software* RapidMiner 5.3 untuk pembuatan analisa dan pengujian model.

Pengujian Model

Model yang telah diusulkan akan diuji dengan sampel data pada Tahun Akademik 2014/2015 dan Tahun Akademik 2015/2016 melalui suatu simulasi menggunakan *software* RapidMiner 5.3. Data Sampel terdiri dari atribut NPM, Nama, Status, Beasiswa (Sudah pernah mendapatkan beasiswa atau tidak), Indeks Prestasi Semester 6 atau IP-6, Indeks Prestasi Semester 7 atau IP-7, Rerata, Kelas (Klasifikasi IPK), Penghasilan Orangtua, Klasifikasi Penghasilan, Jumlah Tanggungan, dan Klasifikasi Tanggungan. Sebanyak 90 % data akan digunakan sebagai *data training* untuk membangun struktur pohon keputusan melalui metode *Decision Tree* Algoritma C4.5 dan 10 % lainnya akan digunakan sebagai data uji atau *data testing*.

Evaluasi dan Validasi Model

Evaluasi diperlukan untuk menganalisa dan mengukur sejauh mana keakuratan hasil yang telah dicapai oleh model dengan menggunakan teknik atau tools yang terdapat pada *software* RapidMiner yaitu *Confusion Matrix*. *Confusion Matrix* adalah *tool* yang digunakan untuk evaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah (F. Gorunescu, 2011). Sedangkan untuk validasi model digunakan teknik *Split Validation* yaitu teknik validasi yang membagi data menjadi dua bagian secara acak, sebagian sebagai *data*

training dan sebagian lainnya sebagai *data testing* Dwi Untari (2014).

HASIL DAN PEMBAHASAN

Pada penelitian ini telah dilakukan pengujian dengan menggunakan *software* RapidMiner 5.3 pada data mahasiswa Angkatan 2012 sebanyak 444 orang, dimana atribut yang didapat adalah NPM, Nama, Status, Beasiswa (Sudah pernah mendapatkan beasiswa atau tidak), Indeks Prestasi Semester 6 atau IP-6, Indeks Prestasi Semester 7 atau IP-7, Penghasilan Orangtua, dan Jumlah Tanggungan. Untuk IP-6 dan IP-7 dibuat rerata agar memudahkan proses klasifikasi nilai IP sehingga didapat tabel Klasifikasi Nilai IPK sebagai berikut:

Tabel 1. Klasifikasi Nilai IPK

No.	Nilai IPK	Klasifikasi
1	≥ 3,51	Dengan Pujian
2	≥ 2,76	Sangat Memuaskan
3	≥ 2,0	Memuaskan
4	< 2,0	Kurang

Kemudian untuk atribut Penghasilan Orangtua dan Jumlah Tanggungan juga dibuat klasifikasi seperti tampak pada tabel dibawah ini:

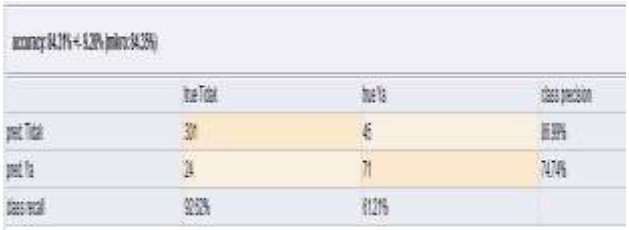
Tabel 2. Klasifikasi Penghasilan

No.	Penghasilan	Klasifikasi
1	≥ 3.500.000	Sangat Tinggi
2	≥ 2.500.000	Tinggi
3	≥ 1.500.000	Sedang
4	< 1.500.0000	Rendah

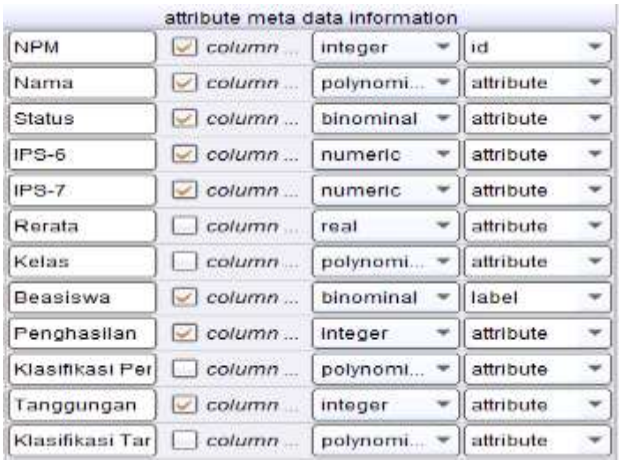
Tabel 3. Klasifikasi Tanggungan

No.	Jumlah Tanggungan	Klasifikasi
1	1-2 orang	Sedikit
2	3-4 orang	Sedang
3	≥ 5 orang	Banyak

Selanjutnya dilakukan pengujian data menggunakan *software* RapidMiner sebanyak 10 kali dengan teknik validasi *Split Validation*. Dari hasil eksperimen ke-3 didapatkan hasil nilai akurasi sebesar 84.31% (Gambar 1) tanpa melibatkan atribut Rerata, Kelas (Klasifikasi IPK), Klasifikasi Penghasilan, dan Klasifikasi Tanggungan (Gambar 2). Eksperimen ini menghasilkan nilai *class precision* berturut-turut sebesar 86.99% dan 74.74% dan nilai *class recall* sebesar 92.62% dan 61.21%.

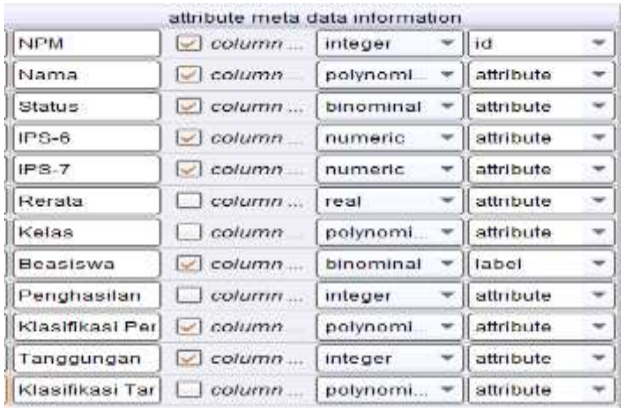


Gambar 1. Performance Vector C4.5 pada Eksperimen ke-3



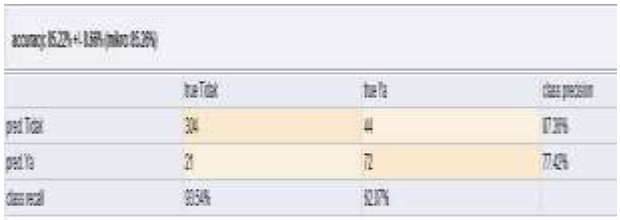
Gambar 2.Atribut yang terlibat dalam Eksperimen ke-3

Pada eksperimen ke-4, atribut Penghasilan tidak dilibatkan dan diganti dengan atribut Klasifikasi Penghasilan seperti yang terlihat pada Gambar 3:



Gambar 3. Atribut yang terlibat dalam Eksperimen ke-4

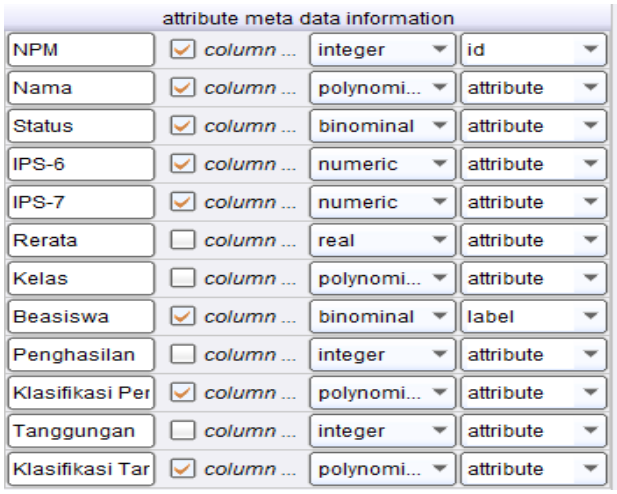
Dari eksperimen tersebut didapatkan hasil nilai akurasi meningkat menjadi 85.22% dengan seperti yang tampak pada gambar 4:



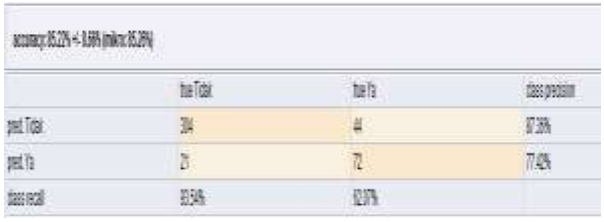
Gambar 4. Performance Vector C4.5 pada Eksperimen ke-4

Dari gambar 4 dapat dilihat bahwa eksperimen ke-4 ini menghasilkan nilai *class precision* berturut-turut sebesar 87.36% dan 77.42% dan nilai *class recall* sebesar 93.54% dan 62.07%.

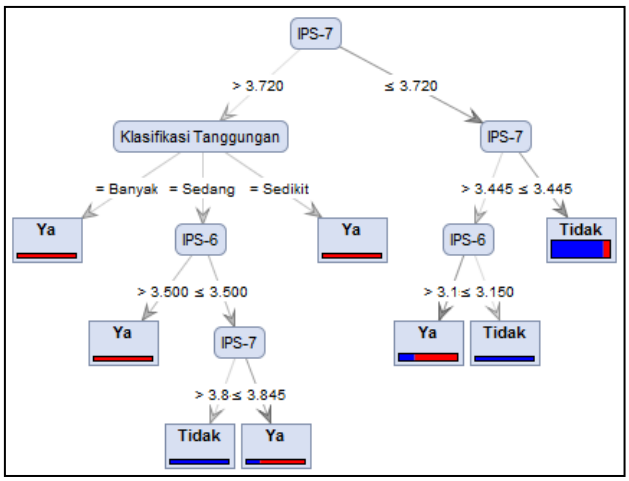
Kemudian pada eksperimen ke-9, atribut Penghasilan dan Tanggungan tidak dilibatkan dan diganti dengan Klasifikasi Penghasilan dan Klasifikasi Tanggungan (Gambar 5). Ternyata hasil nilai akurasi tetap tinggi yakni sebesar 85.22% (Gambar 6) dengan nilai *class precision* berturut-turut sebesar 87.36% dan 77.42% dan nilai *class recall* sebesar 93.54% dan 62.07%. dan menghasilkan rule yang terbentuk dari pohon keputusan yang dapat dilihat pada Gambar 7.



Gambar 5. Atribut yang terlibat dalam Eksperimen ke-9

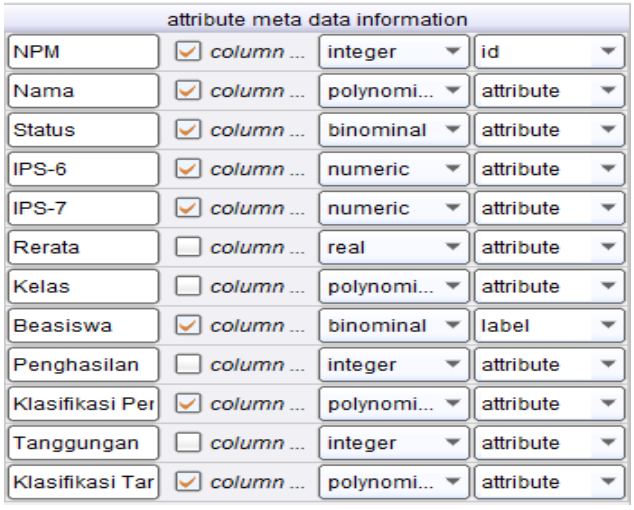


Gambar 6. Performance Vector C4.5 pada Eksperimen ke-9

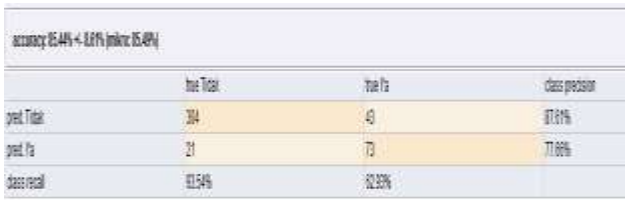


Gambar 7. Struktur pohon keputusan yang terbentuk pada Eksperimen ke-9

Selanjutnya pada eksperimen ke-10 dilakukan perubahan pada atribut Klasifikasi Tanggungan menjadi dua kategori saja yakni jika jumlah anak yang ditanggung kurang dari 4 orang maka dikategorikan "Sedikit" dan sebaliknya jika jumlah anak yang ditanggung lebih dari sama dengan 4 orang maka dikategorikan "Banyak". Setelah dilakukan pengujian kembali dengan *software* RapidMiner, dengan pemilihan atribut yang digunakan seperti pada Gambar 8, didapatkan hasil nilai akurasi meningkat menjadi 85.44%, diiringi dengan peningkatan nilai *class precision* berturut-turut sebesar 87.61% dan 77.66% dan nilai *class recall* sebesar 93.54% dan 62.93% (Gambar 9).

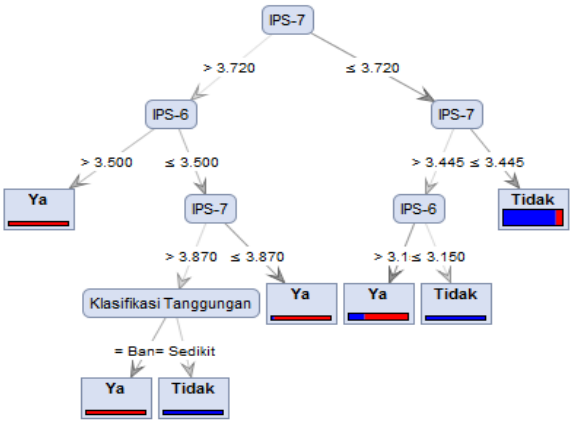


Gambar 8. Atribut yang terlibat dalam Eksperimen ke-4



Gambar 9. Performance Vector C4.5 pada Eksperimen ke-10

Adapun rule yang terbentuk dari pohon keputusan pada eksperimen ke-10 terbilang bagus seperti yang tampak pada gambar berikut ini:



Gambar 10. Struktur pohon keputusan yang terbentuk pada Eksperimen ke-10

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil eksperimen yang sudah dilakukan maka dapat disimpulkan bahwa:

1. Hasil evaluasi dengan *tool Confussion Matrix* yang ditunjukkan melalui tabel Performance Vector C4.5 dan validasi dengan teknik *Split Validation* menunjukkan tingkat akurasi tertinggi didapat pada eksperimen ke-10 yakni sebesar 85.44%, dengan nilai *class precision* berturut-turut sebesar 87.61% dan 77.66% dan nilai *class recall* sebesar 93.54% dan 62.93%.
2. Adanya tingkat akurasi yang bagus menjadikan metode Decision Tree Algoritma C4.5 layak digunakan sebagai alternatif pendukung dalam proses seleksi calon penerima beasiswa sehingga penyaluran dana beasiswa dapat tepat sasaran.
3. Atribut yang berpengaruh terhadap proses seleksi calon penerima beasiswa adalah Indeks Prestasi Semester, Penghasilan yang dibuat menjadi 3 atau 4 kategori kelas (Klasifikasi Penghasilan) dan Tanggungan, baik dengan 2 kategori kelas maupun dengan 3 kategori kelas (Klasifikasi Tanggungan). Sehingga data atribut Penghasilan dan Tanggungan akan lebih berpengaruh dalam penentuan nilai akurasi jika diubah ke dalam bentuk klasifikasi terlebih dahulu dibandingkan jika hanya dengan data asli berupa angka.

Saran

Penelitian ini dapat dikembangkan lebih lanjut dengan menambahkan atribut Pengeluaran, misalnya berupa tagihan rekening listrik dan rekening air. Selain itu juga dapat dikembangkan dengan membuat aplikasi Sistem Pendukung Keputusan (SPK) agar proses seleksi calon penerima beasiswa menjadi lebih efektif dan efisien.

DAFTAR PUSTAKA

Direktorat Jenderal Pembelajaran dan Kemahasiswaan. 2015 *Pedoman Umum Beasiswa Dan Bantuan Biaya Pendidikan Peningkatan Prestasi Akademik (PPA).*, Kementerian Riset, Teknologi, Dan Pendidikan Tinggi. Jakarta.

Dwi Untari. 2014. Data Mining Untuk Menganalisa Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Metode Decision Tree C4.5. Semarang. Universitas Dian Nuswantoro

F. Gorunescu. 2011. *Data Mining Concept, Models and Techniques*. Verlag Berlin Heidelberg: Springer.

Jonathan Rinaldy. 2014. *Sistem Penunjang Keputusan Penerima Beasiswa Dengan Metode Decision Tree C4.5 pada SMAK Yos Sudarso Batu*. Malang. Sekolah Tinggi Informatika & Komputer Indonesia (STIKI)

Jose Augusto Duarte Guterres. 2015. *Kelayakan Algoritma C4.5 Sebagai Pendukung Keputusan Dalam Pengajuan Penerima Beasiswa.* Seminar Nasional Teknologi Informasi.

Muh. Safri Juliardi, Ristu saptono, dan Denis Eka Cahyani. 2017 *Universitas Sebelas Maret Bidikmisi Applicant's Classification Using C4.5 Algorithm*. ITSMART: Jurnal Ilmiah Teknologi dan Informasi Vol. 6, No. 1.