

Implementasi Algoritma Data Mining J48 Untuk Klasifikasi Mahasiswa Yang Layak Mendapat Beasiswa PPA

Ananoma Buulolo¹, Efori Bu'ulolo^{2*}

Fakultas Ilmu Komputer dan Teknologi Informasi, Teknik Informatika, Universitas Budi Darma, Medan, Indonesia

Email: ¹ananoma15bll@gmail.com, ^{2*}buuloloefori21@email.com

Email Penulis Korespondensi: buuloloefori21@email.com

Submitted: 06/06/2022; Accepted: 30/06/2022; Published: 30/06/2022

Abstrak—Beasiswa PPA (Peningkatan Prestasi Akademik) merupakan dukungan biaya pendidikan yang diberikan kepada mahasiswa/i aktif yang telah menempuh minimal semester 2 (dua) di Perguruan Tinggi dengan pertimbangan keterbatasan kemampuan ekonomi serta memiliki prestasi akademik maupun non akademik. Dalam mengklasifikasi mahasiswa/i yang layak mendapat beasiswa PPA pihak pengelola mengalami kesulitan karena beasiswa PPA sangatlah terbatas tidak sebanding dengan jumlah mahasiswa. Salah satu solusi dalam mengklasifikasi mahasiswa/i yang layak mendapatkan beasiswa PPA adalah dengan memanfaatkan data-data mahasiswa/i yang telah menerima beasiswa PPA pada tahun sebelumnya. Namun dalam proses pengklasifikasiannya diperlukan teknik penggalian data, salah satu teknik penggalian data adalah data mining. Dimana data mining merupakan ekstraksi pola yang menarik terhadap data dalam jumlah yang besar [1]. Salah satu algoritma dalam data mining yang digunakan dalam mengklasifikasi adalah algoritma J48. Algoritma J48 merupakan implementasi algoritma C4.5 di tools WEKA, selain itu algoritma J48 merupakan implementasi *decision tree* pada tools rapidminer. Algoritma ini, diharapkan dapat membantu pihak pengelola beasiswa PPA di STMIK Budi Darma dalam mengklasifikasi mahasiswa/i yang benar-benar layak menerima beasiswa PPA. Hasilnya, salah satu kriteria seorang mahasiswa diklasifikasikan menerima beasiswa PPA adalah memiliki prestasi.

Kata Kunci : Beasiswa; PPA; Data; Mining; J48

Abstract—The PPA Scholarship is a support for educational costs that is given to students who have taken at least semester 2 in Higher Education with consideration of limited economic capacity and have academic and non-academic achievements. In classifying students who are eligible for PPA scholarships, the management has difficulties because the limited PPA scholarships are not proportional to the number of students. One solution in classifying students who are eligible for PPA scholarships is to utilize the data of students who have received PPA scholarships in the previous year. However, in the classification process, data research is needed, one of the data research techniques is data mining. Where data mining is an interesting pattern extraction of large amounts of data [1]. One of the algorithms in data mining used in classification is the J48 algorithm. The J48 algorithm is the implementation of the C4.5 algorithm in the WEKA tools, besides the J48 algorithm is the decision tree implementation in the rapidminer tools. This algorithm is expected to help the management of the PPA scholarship at STMIK Budi Darma in classifying students who really deserve to receive the PPA scholarship. As a result, one of the criteria for a student to be classified as receiving a PPA scholarship is to have achievement.

Keywords: Scholarship; PPA; Data; Mining; J48

1. PENDAHULUAN

Beasiswa PPA (Peningkatan Prestasi Akademik) merupakan dukungan biaya pendidikan yang diberikan kepada mahasiswa/i aktif yang telah menempuh minimal semester 2 (dua) di Perguruan Tinggi dengan pertimbangan keterbatasan kemampuan ekonomi serta memiliki prestasi baik dibidang akademik maupun non akademik. Dengan adanya program beasiswa PPA ini akan mendorong mahasiswa/i yang telah menerima beasiswa lebih berprestasi lagi dikemudian hari, serta memberikan motivasi bagi mahasiswa/i lain untuk lebih giat dalam belajar, serta dapat mengurangi jumlah mahasiswa yang putus kuliah, karena tidak mampu dalam pembiayaan pendidikan [2].

Selama ini, pengelola beasiswa PPA di STMIK Budi Darma kesulitan dalam hal menentukan mahasiswa yang benar-benar layak menerima beasiswa PPA, karena kuota yang sangat terbatas dan tidak sebanding dengan jumlah mahasiswa sedangkan jumlah mahasiswa yang layak menerima beasiswa PPA dan yang memenuhi kriteria / persyaratan cukup banyak. Berdasarkan permasalahan tersebut, penulis telah menemukan salah satu solusi yaitu memanfaatkan data-data mahasiswa/i yang telah menerima beasiswa PPA pada tahun sebelumnya. Kemudian menggunakan data-data mahasiswa/i tersebut sebagai sampel data untuk mengklasifikasi mahasiswa/i yang benar-benar layak beasiswa PPA dengan menggunakan Algoritma J48 [3][4].

Penelitian yang dilakukan oleh Fergie Joanda Kaunang pada tahun 2018 meneliti tentang tingkat kemiskinan di Indonesia dengan algoritma J48. Hasil dari penelitian tersebut adalah tingginya jumlah penduduk miskin di Indonesia diprediksi terjadi pada provinsi dengan angka harapan hidup lebih besar dari 69,95, rata-rata lama sekolah lebih kecil dari 7,71, dan pengeluaran perkapita lebih besar dari Rp. 8.957,00 [5]. Penelitian yang dilakukan oleh Kristin Sinaga dkk tahun 2019 memprediksikan resiko kredit pada koperasi simpan pinjam dengan algoritma J48. Hasil dari penelitian tersebut yaitu tinggi atau rendahnya resiko kredit dipengaruhi oleh kriteria salah satunya pendapatannya lancar atau tidak [6].

Algoritma J48 merupakan salah satu teknik dalam data mining yang dapat digunakan untuk mengidentifikasi informasi atau pengetahuan dari data yang besar [7][8][9]. Adapun syarat sebagai pertimbangan yang digunakan untuk kelayakan mendapatkan beasiswa PPA adalah mahasiswa aktif pada STMIK Budi Darma, mahasiswa yang memiliki IPK paling tinggi, mahasiswa yang memiliki prestasi, mahasiswa yang memiliki keterbatasan kemampuan ekonomi

dan lain sebagainya. Dengan menggunakan algoritma J48 dalam klasifikasi mahasiswa yang layak mendapatkan beasiswa PPA, memudahkan pengelola beasiswa PPA serta memperoleh calon mahasiswa/I penerima beasiswa PPA terpilih yang benar-benar layak menerima beasiswa PPA.

2. METODOLOGI PENELITIAN

Berikut ini merupakan metodologi yang digunakan dalam penelitian ini adalah, sebagai berikut:

- a. Studi pendahuluan
Langkah pertama yang harus dilakukan adalah mempelajari masalah yang akan diteliti, dan agar fokus pada bidang yang akan diteliti maka ditentukan ruang lingkup masalah. Adapun ruang lingkup masalah yaitu data yang digunakan sebagai sampel data yaitu data seleksi PPA 2017 dan PPA 2018, hasil dalam bentuk pohon keputusan dan pengujian menggunakan aplikasi rapidminer
- b. Studi Pustaka
Studi pustaka dilakukan untuk mencari referensi-referensi yang berkaitan dengan pembahasan serta tema yang akan dibuat dari berbagai sumber, salah satunya adalah buku-buku pendukung, jurnal pendukung, serta hasil browsing dari internet.
- c. Studi Lapangan
Pada tahap ini, penulis melakukan wawancara langsung kepada pengelola beasiswa PPA di STMIK Budi Darma untuk mengetahui permasalahan yang dihadapi tentang pengelolaan beasiswa PPA serta dokumentasi berupa pengambilan sampel data yang akan digunakan[10]
- d. Analisa dan implementasi
Setelah penulis menganalisa masalah yang berkaitan dengan pengelolaan beasiswa PPA dan solusi dari permasalahan tersebut. Penulis menemukan masalah yaitu pengelola kesulitan dalam menentukan mahasiswa yang benar-benar layak menerima beasiswa PPA karena jumlah mahasiswa yang banyak dengan kuota yang sedikit. Solusinya adalah mengimplementasikan algoritma J48 dengan sampel data dan menguji data dengan aplikasi rapidminer yaitu data seleksi beasiswa PPA 2017 dan 2018. Hasil dan implementasi algoritma J48 adalah pohon keputusan yang menggambarkan kriteria mahasiswa yang benar-benar layak menerima beasiswa PPA.
- e. Laporan dan Publikasi
Kegiatan penelitian ini, dari awal hingga selesai dan proses yang dilakukan dibuat dalam bentuk laporan penelitian dan dipublikasi sebagai karya ilmiah[11].

3. HASIL DAN PEMBAHASAN

Sampel data yang menjadi pokok pembahasan dalam penelitian ini adalah data-data mahasiswa yang menerima dan yang tidak menerima beasiswa PPA di STMIK Budi Darma pada tahun sebelumnya yaitu pada tahun 2017 dan 2018. Data tersebut diolah dengan algoritma J48 untuk mendapatkan pola atau pohon keputusan. Berikut ini merupakan data mahasiswa yang menerima dan tidak menerima beasiswa PPA pada STMIK Budi Darma pada tabel 1:

Tabel 1. Data Mahasiswa yang Menerima dan Tidak Menerima Beasiswa PPA di STMIK Budi Darma

No	Nama	IPK	Pekerjaan Orang Tua (POT)	Prestasi	Keaktifan Dalam Organisasi Kampus (KDOK)	Keterangan
1	Adam Malik	Rendah	Petani	Tidak Ada	Tidak Aktif	Tidak Menerima
2	Agus Adi Pramadi	Tinggi	Petani	Ada	Aktif	Menerima
3	Ahmad Safitra	Rendah	Wiraswasta	Ada	Aktif	Menerima
4	Ayulia Sari	Tinggi	Ibu Rumah Tangga	Tidak Ada	Tidak Aktif	Menerima
5	Dani Iqbal	Tinggi	Petani	Ada	Aktif	Menerima
6	Firman Telaumbanua	Tinggi	Petani	Tidak Ada	Tidak Aktif	Menerima
7	Hendri Cahaya Putra	Tinggi	Wiraswasta	Ada	Aktif	Menerima
8	Herlina Sari	Tinggi	Petani	Tidak Ada	Aktif	Tidak Menerima
9	Ihsan Maulana	Rendah	Petani	Tidak Ada	Tidak Aktif	Tidak Menerima
10	Indra Lesmana	Rendah	Petani	Tidak Ada	Tidak Aktif	Tidak Menerima
11	Indra Williamsyah Sinaga	Tinggi	Pegawai Swasta	Ada	Aktif	Menerima
12	Irma Solin	Tinggi	Petani	Tidak Ada	Tidak Aktif	Tidak Menerima
13	Jepri Tarigan	Sedang	PNS	Ada	Aktif	Menerima
14	Johansyah	Tinggi	Petani	Tidak Ada	Tidak Aktif	Menerima
15	Juni Ade Nawer Purba	Tinggi	Petani	Tidak Ada	Aktif	Tidak Menerima
16	Lia Angriani	Rendah	Petani	Tidak Ada	Tidak Aktif	Tidak Menerima
17	Lukas Sarumaha	Tinggi	Wiraswasta	Tidak Ada	Tidak Aktif	Menerima

18	Mardiyah Lubis	Tinggi	Petani	Tidak Ada	Tidak Aktif	Tidak Menerima
19	May Sarah Margolang	Sedang	Polisi	Ada	Aktif	Menerima
20	Paulina Tanjung	Tinggi	Petani	Tidak Ada	Tidak Aktif	Menerima
21	Raymond Advendi G Manurung	Rendah	Petani	Tidak Ada	Tidak Aktif	Tidak Menerima
22	Riyo Oktaviany Finola	Sedang	Wiraswasta	Ada	Aktif	Menerima
23	Rizky Saptiansyah Nasution	Tinggi	Petani	Ada	Aktif	Menerima
24	Saima Ronita Purba	Tinggi	Petani	Ada	Aktif	Menerima
25	Sonya Malinda Harahap	Tinggi	Petani	Tidak Ada	Aktif	Menerima
26	Suftri Yono Hutagalung	Sedang	Petani	Ada	Aktif	Menerima

Berdasarkan tabel 1 diatas, menunjukkan bahwa mahasiswa/i yang menerima beasiswa PPA pada STMIK Budi Darma adalah sebanyak 17 orang, sedangkan yang tidak menerima adalah sebanyak 9 orang.

3.1 Tahapan Penerapan Algoritma J48

Berikut ini merupakan langkah-langkah dalam penyelesaian algoritma data mining J48 yaitu:

- Menetapkan *training dataset*
- Menentukan akar pohon keputusan/ *root*, dengan cara:
 - Menghitung nilai *entropy* menggunakan rumus sebagai berikut:

$$Entropy (S) = \sum_{i=1}^n -p_i * \log_2 p_i \tag{1}$$
 - Menghitung nilai *gain* menggunakan rumus sebagai berikut:

$$Gain (S,A) = S - \sum_{i=1}^n \frac{|S_i|}{|S|} * S_i \tag{2}$$
 - Nilai gain tertinggi merupakan pohon keputusan/ *root*
- Ulangi langkah ke-2 sampai semua tupel pada masing-masing atribut terbagi
- Proses pembagian akan berhenti jika semua tupel/ kriteria dalam titik N telah memperoleh kelas yang sama tidak ada atribut dalam tupel yang dibagi lagi atau tidak ada tupel dalam cabang yang kosong[12][13].

Adapun proses perhitungan dengan algoritma J48 adalah sebagai berikut:

Diketahui:

Jumlah kasus (S) = 26 orang

Jumlah kasus yang menerima beasiswa PPA(S₁) = 17 orang

Jumlah kasus yang tidak menerima beasiswa PPA (S₂) = 9 orang

$$\begin{aligned} \text{Maka Entropy Total} &= \left(-\frac{S_1}{S} * \log_2 \left(\frac{S_1}{S}\right)\right) + \left(-\frac{S_2}{S} * \log_2 \left(\frac{S_2}{S}\right)\right) = \left(-\frac{17}{26} * \log_2 \left(\frac{17}{26}\right)\right) + \left(-\frac{9}{26} * \log_2 \left(\frac{9}{26}\right)\right) \\ &= (-0,653846154 * \log_2(0,653846154)) + (-0,346153846 * \log_2(0,346153846)) \\ &= (-0,653846154 * (-0,6129788768)) + (-0,346153846 * (-1,5305147166)) \\ &= 0,400792573 + 0,529793556 = 0,930586129 \end{aligned}$$

a. Menghitung nilai Entropy Atribut IPK

Menghitung nilai entropy atribut IPK (Tinggi)

Diketahui:

Jumlah kasus IPK “Tinggi” (S) = 16 orang

Jumlah kasus IPK “Tinggi” yang menerima (S₁) = 12 orang

Jumlah kasus IPK “Tinggi” yang tidak menerima (S₂) = 4 orang

$$\begin{aligned} \text{Entropy atribut IPK (Tinggi)} &= \left(-\frac{S_1}{S} * \log_2 \left(\frac{S_1}{S}\right)\right) + \left(-\frac{S_2}{S} * \log_2 \left(\frac{S_2}{S}\right)\right) \\ &= \left(-\frac{12}{16} * \log_2 \left(\frac{12}{16}\right)\right) + \left(-\frac{4}{16} * \log_2 \left(\frac{4}{16}\right)\right) = (-0,75 * (-0,4150374992)) + (-0,25 * (-2)) \\ &= 0,811278124 \end{aligned}$$

Menghitung nilai entropy atribut IPK (Sedang)

Diketahui:

Jumlah kasus IPK “Sedang” (S) = 4 orang

Jumlah kasus IPK “Sedang” yang menerima (S₁) = 4 orang

Jumlah kasus IPK “Sedang” yang tidak menerima (S₂) = 0 orang

$$\begin{aligned} \text{Entropy atribut IPK (Tinggi)} &= \left(-\frac{S_1}{S} * \log_2 \left(\frac{S_1}{S}\right)\right) + \left(-\frac{S_2}{S} * \log_2 \frac{S_2}{S}\right) \\ &= \left(-\frac{4}{4} * \log_2 \left(\frac{4}{4}\right)\right) + \left(-\frac{0}{4} * \log_2 \frac{0}{4}\right) = ((-1 * 0) + (0 * 0)) \\ &= 0 \end{aligned}$$

Menghitung nilai entropy atribut IPK (Rendah)

Diketahui:

Jumlah kasus IPK “Rendah” (S) = 6 orang

Jumlah kasus IPK “Rendah” yang menerima (S₁) = 1 orang

Jumlah kasus IPK “Rendah” yang tidak menerima (S₂) = 5 orang

$$\begin{aligned} \text{Entropy atribut IPK (Tinggi)} &= \left(-\frac{S_1}{S} * \log_2 \left(\frac{S_1}{S}\right)\right) + \left(-\frac{S_2}{S} * \log_2 \frac{S_2}{S}\right) \\ &= \left(-\frac{1}{6} * \log_2 \left(\frac{1}{6}\right)\right) + \left(-\frac{5}{6} * \log_2 \frac{5}{6}\right) \\ &= (-0,166666667 * (-2,584962501)) + (-0,833333333 * (-0,263034405)) \\ &= 0,918295834 \end{aligned}$$

Perhitungan nilai gain atribut IPK

Gain (Total, IPK)

$$\begin{aligned} &= \text{Entropy(Total)} - \sum_{i=1}^n \frac{\text{IPK}}{\text{Total}} * \text{Entropy(IPK)} \text{ Gain (Total, IPK)} \\ &= \text{Entropy Total} - \left(\left(\frac{\text{Jumlah IPK Tinggi}}{\text{Total kasus}} * \text{Entropy IPK Tinggi}\right) + \right. \\ &\quad \left.\left(\frac{\text{Jumlah IPK Sedang}}{\text{Total Kasus}} * \text{Entropy IPK Sedang}\right) + \left(\frac{\text{Jumlah IPK Sedang}}{\text{Total Kasus}} * \right. \right. \\ &\quad \left. \left. \text{Entropy IPK Rendah}\right)\right) \\ &= (0,930586129) - \left(\left(\frac{16}{26} * 0,811278124\right) + \left(\frac{4}{26} * 0\right) + \left(\frac{6}{26} * 0,650022422\right)\right) \\ &= (0,930586129) - ((0,499248077) + (0) + (0,150005174)) \\ &= 0,281332878 \end{aligned}$$

b. Menghitung nilai entropy atribut pekerjaan orang tua (POT)

Menghitung nilai entropy atribut Pekerjaan orang tua (Petani)

Diketahui:

Jumlah kasus pekerjaan orang tua ”petani” (S) = 18 orang

Jumlah kasus pekerjaan orang tua “petani” yang menerima (S₁) = 9 orang

Jumlah kasus pekerjaan orang tua “petani” yang tidak menerima (S₂) = 9 orang

$$\begin{aligned} \text{Entropy atribut POT (Petani)} &= \left(-\frac{S_1}{S} * \log_2 \left(\frac{S_1}{S}\right)\right) + \left(-\frac{S_2}{S} * \log_2 \frac{S_2}{S}\right) \\ &= \left(-\frac{9}{18} * \log_2 \left(\frac{9}{18}\right)\right) + \left(-\frac{9}{18} * \log_2 \frac{9}{18}\right) = (-0,5 * (-1)) + (-0,5 * (-1)) \\ &= 1 \end{aligned}$$

Menghitung nilai entropy atribut pekerjaan orang tua (Wiraswasta)

Diketahui:

Jumlah kasus pekerjaan orang tua “wiraswasta” (S) = 4 orang

Jumlah kasus pekerjaan orang tua “wiraswasta” menerima (S₁) = 0 orang

Jumlah kasus pekerjaan orang tua “wiraswasta” yang tidak menerima (S₂) = 4 orang

$$\begin{aligned} \text{Entropy atribut POT (Wiraswasta)} &= \left(-\frac{S_1}{S} * \log_2 \left(\frac{S_1}{S}\right)\right) + \left(-\frac{S_2}{S} * \log_2 \frac{S_2}{S}\right) \\ &= \left(-\frac{4}{4} * \log_2 \left(\frac{4}{4}\right)\right) + \left(-\frac{0}{4} * \log_2 \frac{0}{4}\right) = (-1 * (0)) + (0 * (0)) \\ &= 0 \end{aligned}$$

Menghitung nilai entropy atribut pekerjaan orang tua (Ibu Rumah Tangga)

Diketahui:

Jumlah kasus pekerjaan orang tua “IRT” (S) = 1 orang

Jumlah kasus pekerjaan orang tua “IRT” yang menerima (S₁) = 1 orang

Jumlah kasus pekerjaan orang tua “IRT” yang tidak menerima (S₂) = 0 orang

$$\text{Entropy atribut POT (IRT)} = \left(-\frac{S_1}{S} * \log_2 \left(\frac{S_1}{S}\right)\right) + \left(-\frac{S_2}{S} * \log_2 \frac{S_2}{S}\right)$$

$$= \left(-\frac{1}{1} * \log_2 \left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} * \log_2 \frac{0}{1}\right) = (-1 * (0)) + (0 * (0))$$

$$= 0$$

Menghitung nilai entropy atribut pekerjaan orang tua (Pegawai Swasta)

Diketahui:

Jumlah kasus pekerjaan orang tua ”PS” (S) = 1 orang

Jumlah kasus pekerjaan orang tua “PS” yang menerima (S₁) = 1 orang

Jumlah kasus pekerjaan orang tua “PS” yang tidak menerima (S₂) = 0 orang

$$\text{Entropy atribut POT (PS)} = \left(-\frac{S_1}{S} * \log_2 \left(\frac{S_1}{S}\right)\right) + \left(-\frac{S_2}{S} * \log_2 \frac{S_2}{S}\right)$$

$$= \left(-\frac{1}{1} * \log_2 \left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} * \log_2 \frac{0}{1}\right) = (-1 * (0)) + (0 * (0))$$

$$= 0$$

Menghitung nilai entropy atribut pekerjaan orang tua (PNS)

Diketahui:

Jumlah kasus pekerjaan orang tua “PNS” (S) = 1 orang

Jumlah kasus pekerjaan orang tua “PNS” menerima (S₁) = 1 orang

Jumlah kasus pekerjaan orang tua “PNS” yang tidak menerima (S₂) = 0 orang

$$\text{Entropy atribut POT (PNS)} = \left(-\frac{S_1}{S} * \log_2 \left(\frac{S_1}{S}\right)\right) + \left(-\frac{S_2}{S} * \log_2 \frac{S_2}{S}\right)$$

$$= \left(-\frac{1}{1} * \log_2 \left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} * \log_2 \frac{0}{1}\right)$$

$$= (-1 * (0)) + (0 * (0))$$

$$= 0$$

Menghitung nilai entropy atribut pekerjaan orang tua (Polisi)

Diketahui:

Jumlah kasus pekerjaan orang tua “Polisi” (S) = 1 orang

Jumlah kasus pekerjaan orang tua “Polisi” yang menerima (S₁) = 1 orang

Jumlah kasus pekerjaan orang tua “Polisi” yang tidak menerima (S₂) = 0 orang

$$\text{Entropy atribut POT (Polisi)} = \left(-\frac{S_1}{S} * \log_2 \left(\frac{S_1}{S}\right)\right) + \left(-\frac{S_2}{S} * \log_2 \frac{S_2}{S}\right)$$

$$= \left(-\frac{1}{1} * \log_2 \left(\frac{1}{1}\right)\right) + \left(\frac{0}{1} * \log_2 \frac{0}{1}\right) = (-1 * (0)) + (0 * (0))$$

$$= 0$$

Perhitungan nilai gain atribut POT:

$$\text{Gain (Total, POT)} = \text{Entropy(Total)} - \sum_{i=1}^n \frac{\text{POT}}{\text{Total}} * \text{Entropy(POT)}$$

Gain (Total, POT)

$$= 0,930586129 - \left(\left(\frac{18}{26} * 1\right) + \left(\frac{4}{26} * 0\right) + \left(\frac{1}{26} * (0)\right) + \left(\frac{1}{26} * 0\right) + \left(\frac{1}{26} * 0\right) + \left(\frac{1}{26} * 0\right)\right)$$

$$= 0,1290013 - ((0,692307692) + (0) + (0) + (0) + (0) + (0))$$

$$= 0,238278437$$

c. Menghitung nilai entropy atribut prestasi

Menghitung nilai entropy atribut prestasi (Ada)

Diketahui:

Jumlah kasus prestasi ”ada” (S) = 11 orang

Jumlah kasus prestasi “ada” yang menerima (S_1) = 11 orang

Jumlah kasus prestasi “ada” yang tidak menerima (S_2) = 0 orang

$$\begin{aligned} \text{Entropy atribut Prestasi (Ada)} &= \left(-\frac{S_1}{S} * \log_2 \left(\frac{S_1}{S}\right)\right) + \left(-\frac{S_2}{S} * \log_2 \frac{S_2}{S}\right) \\ &= \left(-\frac{11}{11} * \log_2 \left(\frac{11}{11}\right)\right) + \left(-\frac{0}{11} * \log_2 \frac{0}{11}\right) \\ &= (-1 * (0)) + (0 * (0)) \\ &= 0 \end{aligned}$$

Menghitung nilai entropy atribut prestasi (Tidak Ada)

Diketahui:

Jumlah kasus prestasi “Tidak Ada” (S) = 15 orang

Jumlah kasus prestasi “Tidak Ada” menerima (S_1) = 6 orang

Jumlah kasus prestasi “Tidak Ada” yang tidak menerima (S_2) = 9 orang

$$\begin{aligned} \text{Entropy atribut Prestasi (Tidak Ada)} &= \left(-\frac{S_1}{S} * \log_2 \left(\frac{S_1}{S}\right)\right) + \left(-\frac{S_2}{S} * \log_2 \frac{S_2}{S}\right) \\ &= \left(-\frac{6}{15} * \log_2 \left(\frac{6}{15}\right)\right) + \left(-\frac{9}{15} * \log_2 \frac{9}{15}\right) = (-0,4 * (-1,321928094)) + (-0,6 * (-0,736965594)) \\ &= 0,970950594 \end{aligned}$$

Perhitungan nilai gain atribut Prestasi:

Gain (Total, Prestasi)

$$\begin{aligned} &= \text{Entropy(Total)} - \sum_{i=1}^n \frac{\text{Prestasi}}{\text{Total}} * \text{Entropy(Prestasi)} \\ &= 0,930586129 - \left(\left(\frac{11}{26} * 0\right) + \left(\frac{15}{26} * 0,970950594\right)\right) \\ &= 0,930586129 - ((0) + (0,560163804)) \\ &= 0,930586129 - 0,560163804 \\ &= 0,370422325 \end{aligned}$$

d. Menghitung nilai entropy atribut keaktifan dalam organisasi kampus (KDOK)

Menghitung nilai entropy atribut keaktifan dalam organisasi kampus (Aktif)

Diketahui:

Jumlah kasus keaktifan dalam organisasi kampus ”Aktif” (S) = 14 orang

Jumlah kasus keaktifan dalam organisasi kampus ”Aktif” yang menerima (S_1) = 12 orang

Jumlah kasus keaktifan dalam organisasi kampus ”Aktif” yang tidak menerima (S_2) = 2 orang

$$\begin{aligned} \text{Entropy atribut KDOK (Aktif)} &= \left(-\frac{S_1}{S} * \log_2 \left(\frac{S_1}{S}\right)\right) + \left(-\frac{S_2}{S} * \log_2 \frac{S_2}{S}\right) \\ &= \left(-\frac{12}{14} * \log_2 \left(\frac{12}{14}\right)\right) + \left(-\frac{2}{14} * \log_2 \frac{2}{14}\right) \\ &= (-0,857142857 * (-0,22239242)) + (-0,142857143 * (-2,80735492)) \\ &= 0,591672779 \end{aligned}$$

Menghitung nilai entropy atribut keaktifan dalam organisasi kampus (Tidak Aktif)

Diketahui:

Jumlah kasus keaktifan dalam organisasi kampus “Tidak Aktif” (S) = 12 orang

Jumlah kasus keaktifan dalam organisasi kampus “Tidak Aktif” menerima (S_1) = 5 orang

Jumlah kasus keaktifan dalam organisasi kampus “Tidak Aktif” yang menerima (S_2) = 7 orang

$$\begin{aligned} \text{Entropy atribut KDOK (Tinggi)} &= \left(-\frac{S_1}{S} * \log_2 \left(\frac{S_1}{S}\right)\right) + \left(-\frac{S_2}{S} * \log_2 \frac{S_2}{S}\right) \\ &= \left(-\frac{5}{12} * \log_2 \left(\frac{5}{12}\right)\right) + \left(-\frac{7}{12} * \log_2 \left(\frac{7}{12}\right)\right) \\ &= (-0,41666667 * (-1,263034405)) + (-0,58333333 * (-1,2630343)) = 0,526264336 + 0,453604421 \end{aligned}$$



$$= 0,979868757$$

Perhitungan nilai gain atribut KDOK:

Gain (Total, KDOK)

$$= \text{Entropy}(\text{Total}) - \sum_{i=1}^n \frac{\text{KDOK}}{\text{Total}} * \text{Entropy}(\text{KDOK})$$

$$= 0,930586129 - \left(\left(\frac{14}{26} * 0,591672779 \right) + \left(\frac{12}{26} * 0,9798688757 \right) \right)$$

$$= 0,930586129 - \left((0,318593035) + (0,452247118) \right) = 0,930586129 - 0,770840153$$

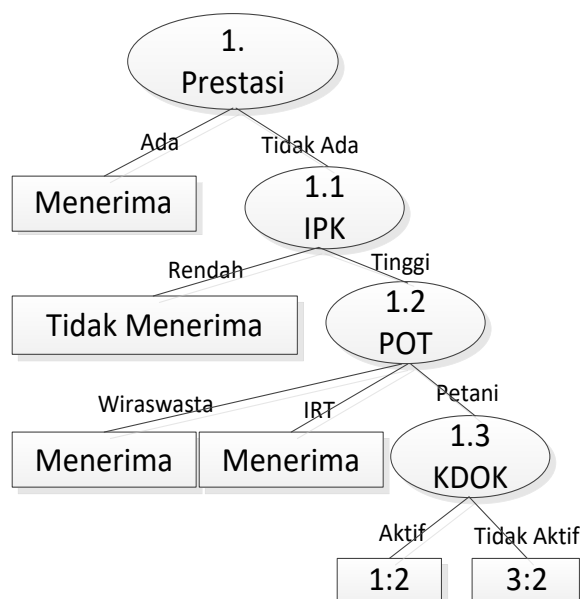
$$= 0,159745976$$

Lakukan hal yang sama untuk mengitung node ke-2 dan 3, sehingga diperoleh nilai *entropy* dan *gain* sebagai berikut:

Tabel 2. Nilai *Entropy* dan *Gain*

Atribut	Value	Jumlah kasus (S)	Menerima (S1)	Tidak Menerima (S2)	Entropy	Gain
Jumlah Kasus		26	17	9	0,930586129	
			IPK			0,281332878
	Tinggi	16	12	4	0,811278124	
	Sedang	4	4	0	0	
	Rendah	6	1	5	0,650022422	
			Pekerjaan Orang Tua (POT)			0,238278437
	Petani	18	9	9	1	
	Wiraswasta	4	4	0	0	
	Ibu Rumah Tangga	1	1	0	0	
	Pegawai Swasta	1	1	0	0	
	PNS	1	1	0	0	
	Polisi	1	1	0	0	
			Prestasi			0,370422325
	Ada	11	11	0	0	
	Tidak Ada	15	6	9	0,970950594	
			Keaktifan Dalam Organisasi Kampus (KDOK)			0,159745976
	Aktif	14	12	2	0,591672779	
	Tidak Aktif	12	5	7	0,979868757	

Sehingga diperoleh role/ pohon keputusan sebagai berikut:



Gambar 1. Role/ Pohon Keputusan

Setelah mengetahui *role* secara manual, maka dilakukan implementasi atau pengujian dengan menggunakan *tools rapidminer*, hasil pengujian pada gambar 1. Algoritma merupakan kumpulan perintah yang tertulis secara sistematis guna menyelesaikan permasalahan logika dari matematika. Algoritma J48 adalah klasifikasi algoritma *data mining* yang mengimplementasikan *Decision Tree* pada *tools rapidminer*[7].

Adapun *Input*, *Process*, dan *Output* pengujian data pada algoritma J48 adalah sebagai berikut:

Input : S = Himpunan/ Jumlah Kasus

S,A = Nilai *Gain* berdasarkan atribut terhadap jumlah kasus

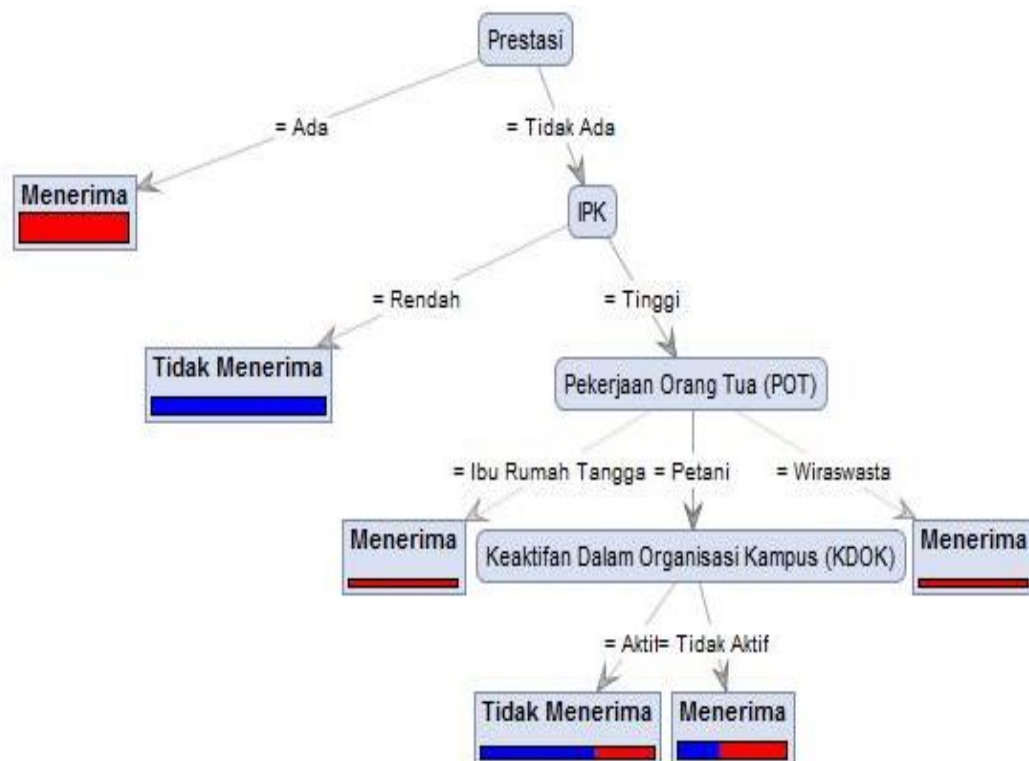
Process : For S = 1 To 26

Decision Tree (S)

$$Gain(S,A) = S - \sum_{i=1}^n \frac{|s_i|}{|s|} * s_i$$

Minimal leaf size = Jumlah pohon yang ditampilkan

Output : *Decision Tree*/ pohon keputusan



Gambar 2. Role/ Pohon Keputusan

4. KESIMPULAN

Berdasarkan hasil dan pembahasan serta pengujian dengan aplikasi *tools Rapiminer* pada gambar 2, dapat ditarik kesimpulannya yaitu algoritma J48 dapat digunakan untuk mengklasifikasi mahasiswa yang benar-benar layak menerima beasiswa PPA berdasarkan data-data tahun sebelumnya, dan kriteria yang paling menentukan seorang calon mahasiswa penerima beasiswa PPA adalah prestasi.

REFERENCES

- [1] E. Buulolo, "ALGORITMA APRIORI PADA DATA PENJUALAN DI SUPERMARKET," in *Seminar Nasional Inovasi dan Teknologi Informasi 2015 (SNITI)*, 2015, no. September 2015, pp. 4–7.
- [2] D. P. dan K. Jenderal, *Panduan Beasiswa Peningkatan Prestasi Akademik*. Jakarta: Jenderal, Direktorat Pembelajaran dan Kemahasiswaan, 2018.
- [3] F. Nasari and S. Darma, "PENERAPAN K-MEANS CLUSTERING PADA DATA PENERIMAAN MAHASISWA BARU," in *Seminar Nasional Teknologi Informasi dan Multimedia 2015*, 2015, pp. 6–8.
- [4] E. Buulolo and R. Syahputra, "Implementasi Algoritma Clustering K-Means Untuk Mengelompokkan Mahasiswa Baru Yang Berpotensi (Studi Kasus : Stmik Budi Darma)," vol. 2, no. September, pp. 17–24, 2019.
- [5] F. Joanda Kaunang, "Penerapan Algoritma J48 Decision Tree Untuk Analisis Tingkat Kemiskinan di Indonesia," *Cogito Smart J.*, vol. 4, no. 2, 2018.
- [6] K. Sinaga, E. Buulolo, and B. Nadeak, "Implementasi Algoritma Decision Tree J48 untuk Memprediksi Resiko Kredit pada Koperasi Simpan Pinjam (Studi Kasus : Kofipindo Lubuk Pakam)," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 20–24, 2019, doi: 10.30865/komik.v3i1.1561.



- [7] G. K. Amit Chhabra, “Improved J48 Classification Algorithm for the Prediction of Diabetes,” *Int. J. Comput. Appl.*, vol. 98, no. 0975 – 8887, p. 15, 2014.
- [8] E. Buulolo, “Implementasi Algoritma Apriori Pada Sistem Persediaan Obat (Studi Kasus : Apotik Rumah Sakit Estomihi Medan),” *Pelita Inform. Budi Darma*, vol. 4, no. Agustus 2013, pp. 71–83, 2013, [Online]. Available: <https://www.researchgate.net/publication/312495968>.
- [9] A. K. Wardhani, “IMPLEMENTASI ALGORITMA K-MEANS UNTUK PENGELOMPOKKAN PENYAKIT PASIEN PADA PUSKESMAS KAJEN PEKALONGAN,” *J. Transform.*, vol. 14, pp. 30–37, 2016.
- [10] N. Martono, *METODE PENELITIAN KUANTITATIF: Analisis Isi dan Analisis Data Sekunder (sampel halaman gratis)*. RajaGrafindo Persada, 2010.
- [11] J. Noor, *Metodologi Penelitian: Skripsi, Tesis, Disertasi & Karya Ilmiah*. Prenada Media, 2015.
- [12] I. P. D. Lesmana, “Pengembangan Decision Tree J48 Untuk Diagnosis Penyakit Diabetes Mellitus,” in *Seminar Nasional Teknologi Informasi & Komunikasi Terapan*, 2012, no. Semantik, pp. 189–193.
- [13] S. Dian Handy Permana and C. Maya, “ANALISISPERBANDINGAN ALGORITMA NAIVE BAYES, J48,DAN RANDOM FOREST TREE DALAMPENINGKATAN LOYALITASPELANGGAN UMKM DENGAN VOUCHERBELANJA,” *J. Integr.*, vol. 11, no. 2, 2019.