

APLIKASI ALGORITMA *DECISION TREE* C4.5 SEBAGAI MODEL PENDUKUNG KEPUTUSAN SELEKSI ANGGOTA RESIMEN MAHASISWA UNIVERSITAS UDAYANA

Alexander Hiro Wibisono^{1§}, I Putu Eka Nila Kencana², I Wayan Sumarjaya³

¹Program Studi Matematika, Fakultas MIPA-Universitas Udayana [Email: alex.kutsuna@gmail.com]

²Program Studi Matematika, Fakultas MIPA-Universitas Udayana [Email: i.putu.enk@unud.ac.id]

³Program Studi Matematika, Fakultas MIPA-Universitas Udayana [Email: sumarjaya@unud.ac.id]

[§]*Corresponding Author*

ABSTRACT

The Reserve Officer Training Corps (ROTC) of the 901st Mayurajana Battalion. The 2000 Agreement caused the ROTC to be excluded from the Indonesian Military command structure, causing the ROTC in each campus to have a diverse standard of members due to the recruitment process which are done by students. C4.5 algorithm from chefboost was used to find the accuracy as well as assist the construction of an ideal C4.5 algorithm. The result shows the C4.5 algorithm by chefboost to have a 79.16% accuracy with the tendency to underestimate the data. During the modeling of the algorithm it is discovered that the source of the algorithm's tendency to underestimate the data appears to be affected by the recruits administration and medical history.

Keywords: *C4.5, decision support model, decision tree, ROTC, recruitment*

1. PENDAHULUAN

Resimen mahasiswa (menwa) adalah sekelompok mahasiswa atau mahasiswa yang telah menjalani pelatihan dasar militer dan diharapkan mengabdikan pada bangsa dengan cara apapun sebagai komponen cadangan untuk Tentara Nasional Indonesia (Notosusanto, 1984). Pada tahun 1994 Menteri Dalam Negeri, Menteri Pertahanan, dan Menteri Pendidikan mengeluarkan suatu keputusan yang disebut Surat Keputusan Bersama (SKB) Tiga Menteri 1994 yang secara resmi menyatakan keberadaan menwa di bawah komando langsung Angkatan Bersenjata Republik Indonesia. Namun, pada tahun 2000 perjanjian baru yang ditandatangani oleh tiga menteri yang sama telah memberlakukan peraturan baru yang menyatakan bahwa menwa tidak lagi di bawah komando Tentara Nasional Indonesia. Oleh karena itu, proses penerimaan anggota baru tidak terstruktur dan terorganisasi seperti pada tahun 1994.

Proses penerimaan anggota di resimen mahasiswa Universitas Udayana sejak terbitnya SKB Tiga Menteri 2000 menjadi kurang efektif karena tidak lagi diawasi oleh pejabat militer yang berkualifikasi. Oleh karena itu,

berdasarkan data selama proses penerimaan anggota tidak dimanfaatkan dengan baik dan hanya menjadi arsip yang disimpan dengan sia-sia. Data yang dikumpulkan masih memiliki pengaruh terhadap keputusan dalam proses penerimaan anggota.

Agar dapat mengolah dan memperoleh informasi yang berharga serta membuat data tersebut lebih mudah dipahami dan bermanfaat, diperlukan suatu metode yang disebut *data mining* agar dapat diproses. *Data mining* telah berkembang pesat seiring berjalannya waktu, banyak peneliti mulai mengembangkan menggunakan *data mining* untuk menemukan lebih banyak pengetahuan dan dukungan keputusan, karena *data mining* didorong oleh penemuan dan bukan didorong oleh asumsi. (Larose dan Larose, 2014). Pah dan Utama (2020) melakukan sebuah penelitian yang membandingkan antara empat macam *data mining* dalam proses penerimaan karyawan dengan *data mining* ternama pohon keputusan C4.5, *naive Bayes*, *random forest*, dan *support vector machine*. Hasil penelitian menyimpulkan bahwa C4.5 yang memiliki paling akurat dan disarankan untuk digunakan dalam proses seleksi anggota baru.

Tujuan dari penelitian ini adalah untuk mengetahui akurasi algoritma pohon keputusan C4.5 menggunakan data dari anggota dan calon tahun sebelumnya Resimen Mahasiswa Ugraçena Batalyon A-901/Mayurajana Universitas Udayana.

Pohon keputusan C4.5 merupakan *data mining* yang menggunakan *entropy* lebih tepatnya *gain ratio* untuk menentukan kriteria pemisah pohon (Quinlan, 1993). *Entropy* merupakan metode untuk kuantifikasi informasi dengan rumus yang ditulis sebagai berikut (Gray, 2011; Shannon, 1948),

$$Entropy(S) = - \sum_{i=1}^n P_i \log P_i \quad (1)$$

Information Gain adalah kriteria berbasis ketidakmurnian yang digunakan oleh ID3 dan C4.5 yang dapat ditulis sebagai berikut (Rokach and Maimon, 2014),

$$Info\ Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} Entropy(S_i) \quad (2)$$

Gain ratio merupakan normalisasi dari *information gain* yang dapat ditulis sebagai berikut,

$$Gain\ Ratio(S, A) = \frac{Info\ Gain(S, A)}{SplitInfo(S, A)} \quad (3)$$

Entropy, *information gain*, dan *gain ratio* telah dibuktikan menjadi hal penting dalam klasifikasi terutama di bidang seleksi anggota (Rozi, 2015).

2. METODE PENELITIAN

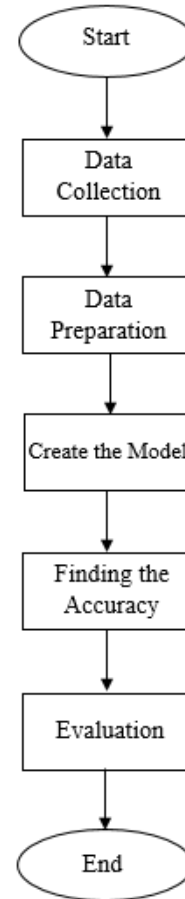
Penelitian ini menggunakan data sekunder. Data tersebut meliputi data administrasi, kesehatan jasmani, ujian fisik yang mencakup lari 12 menit (*cooper test*), *push up*, *sit up*, *pull up*, dan lari angka delapan (*shuttle run*), serta ujian mental ideologi anggota Resimen Mahasiswa Ugraçena Batalyon A-901/Mayurajana Universitas Udayana tahun 2016-2018.

Metode penelitian terdiri dari:

1. Pengumpulan data (*data collection*),
2. Persiapan data (*data preparation*),
3. Pembentukan model pohon keputusan yang dibentuk oleh paket *chefboost* dari data sebanyak 32 anggota tahun 2017 dan 2018 (*create the model*),
4. Mencari akurasi model yang akan memeriksa keakuratan model menggunakan data sebanyak 24 anggota tahun 2016 yang akan dimasukkan ke dalam *confusion matrix*

yang ditunjukkan pada Tabel 1 (*finding the accuracy*),

5. Mengevaluasi hasil yang diperoleh (*evaluation*). Gambar 1 menunjukkan alur penelitian.



Gambar 1. Alur Penelitian

Tabel 1. *Confusion Matrix*

<i>Confusion Matrix</i>		Hasil Nyata	
		<i>Passed</i>	<i>Failed</i>
Prediksi	<i>Passed</i>	True <i>Passed</i> (TP)	False <i>Passed</i> (FP)
	<i>Failed</i>	False <i>Failed</i> (FF)	True <i>Failed</i> (TF)

$$Akurasi = \frac{TP+TF}{TP+FP+FF+TF} \quad (4)$$

3. HASIL DAN PEMBAHASAN

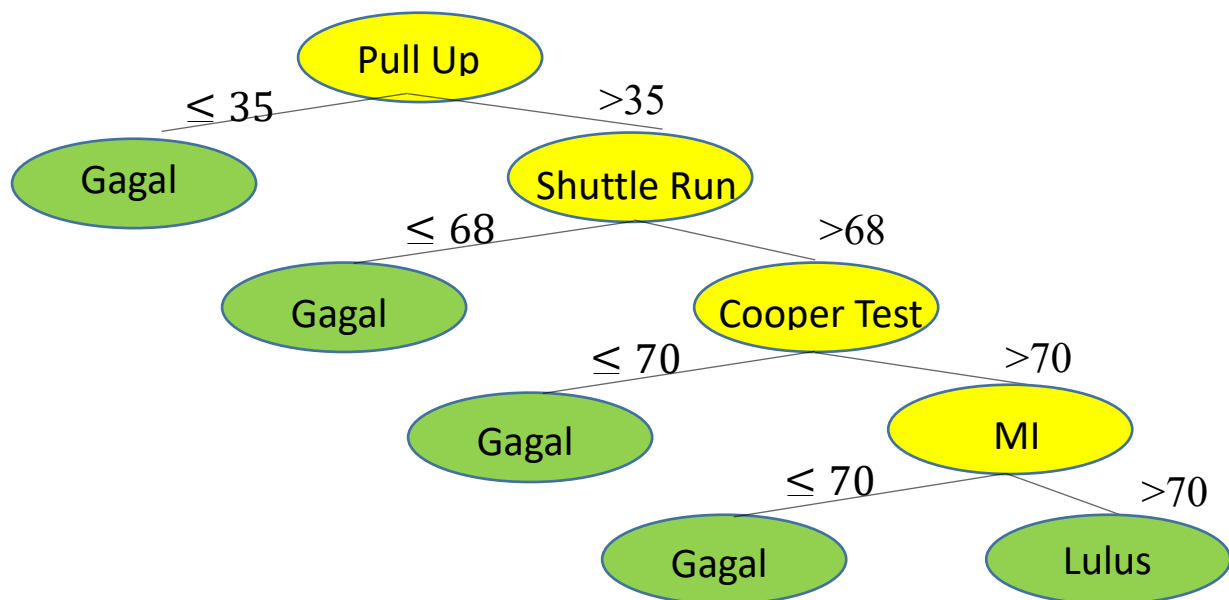
Penelitian ini menggunakan sebuah paket pohon keputusan Python bernama *chefboost* untuk melatih dan membentuk model pohon keputusan C4.5. Proses pembentukan model pohon keputusan oleh *chefboost* dapat dilihat

dengan memanggil fungsi *training* dari paket *chefboost*.

Hasil yang diperoleh pada pemisahan pertama dengan *entropy* induk sebesar 0,9283 adalah atribut *pull up* dengan *gain ratio* 0,5170 diikuti dengan pemisahan kedua dengan *entropy* induk 0,6343 adalah *shuttle run* dengan *gain ratio* 0,6021. Namun, untuk pemisahan ketiga dengan *entropy* induk 0,4262 terdapat tiga atribut dengan *gain ratio* sama sebesar 0,6629 yaitu *cooper test*, *push up*, dan *mental ideologi*. Dalam penelitian ini atribut *cooper test* digunakan sebagai pemisah ketiga. Pemisah keempat dengan *entropy* induk 0,2667 adalah atribut *mental ideologi* dengan *gain ratio* 1, ini mengindikasikan bahwa pada atribut *mental ideologi* sudah terdapat keputusan yang

jas dan pohon keputusannya dapat berakhir di sini. Model pohon keputusan C4.5 dan kriteria pemisah masing-masing atribut dapat dilihat pada Gambar 2.

Model yang telah dirancang oleh *chefboost* yang ditunjukkan pada Gambar 2 berhasil memprediksi dengan tepat sebanyak 16 calon yang lulus, tiga calon yang gagal, dan lima prediksi salah di mana terlihat pada Tabel 2 bahwa lima calon yang seharusnya lulus dinyatakan gagal. Berdasarkan lima data calon yang tidak dapat diprediksi sama modelnya, dapat disimpulkan bahwa model pohon keputusan C4.5 memiliki kecenderungan untuk meremehkan (*underestimate*) data yang diberikan.



Gambar 2. Model Pohon Keputusan C4.5

Table 2. Hasil *Confusion Matrix*

<i>Confusion Matrix</i>		Hasil Nyata		Akurasi
		<i>Passed</i>	<i>Failed</i>	
Prediksi	<i>Passed</i>	16	0	79,16%
	<i>Failed</i>	5	3	

Hasil penelitian skripsi menemukan bahwa model algoritma pohon keputusan C4.5 yang dirancang oleh *chefboost* menggunakan 32 data anggota tahun 2017 dan 2018 berhasil memprediksi kelulusan calon anggota resimen mahasiswa dengan akurasi sebesar 79,16% yang terlihat pada tabel 2. Jelas dengan model pohon keputusan C4.5 pada Gambar 2 bahwa

Resimen Mahasiswa Ugraçena Batalyon A-901/Mayurajana tampaknya lebih fokus pada kemampuan fisik dan mental para calon dan tidak mempertimbangkan kondisi medis calon. Berdasarkan algoritma, dapat disimpulkan bahwa proses penerimaan anggota resimen mahasiswa Ugraçena Batalyon A-

901/Mayurajana tidak mempertimbangkan kesehatan dan administrasi calon anggota.

4. KESIMPULAN DAN SARAN

Pohon keputusan C4.5 memiliki akurasi 79,16% dalam memprediksi penerimaan anggota resimen mahasiswa Ugraçena Batalyon A-901/Mayurajana, ini menunjukkan kemampuan pohon keputusan C4.5 dalam pengambilan keputusan untuk seleksi anggota baru masih perlu penelitian lebih lanjut.

Algoritma pohon keputusan C4.5 ideal yang dibangun dalam penelitian ini masih memiliki ruang untuk perbaikan. Seperti pada pemisahan ketiga, atribut *cooper test* dipilih sebagai pengganti ujian mental ideologi dan push up. Meskipun sudah jelas bahwa jika ujian mental ideologi dipilih daripada *cooper test*, maka hasilnya akan tetap sama dengan hanya dua atribut untuk bertukar posisi, masih belum jelas hasil yang akan terjadi apabila atribut *push up* dipilih sebagai pemisah ketiga.

DAFTAR PUSTAKA

- Gray, R. M. (2011). *Entropy and Information Theory* (2nd ed.). Springer Science+Business Media. <https://doi.org/10.1007/978-1-4419-7970-4>
- Larose, D. T., & Larose, C. D. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition*. In D. T. Larose (Ed.), *John Wiley & Sons, Inc.* (2nd ed.). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118874059>
- Notosusanto, Nugroho. (1984). *Pejuang dan Prajurit*. Jakarta: Sinar Harapan
- Pah, C. E. A., & Utama, D. N. (2020). Decision support model for employee recruitment using data mining classification. *International Journal of Emerging Trends in Engineering Research*, 8(5), 1511–1516. <https://doi.org/10.30534/ijeter/2020/06852020>
- Quinlan, J. R. (1993). *C4.5 programs for machine learning by J. Ross Quinlan* (P. Langley (ed.)). Morgan Kauffman Publishers, Inc.
- Rokach, L., & Maimon, O. (2014). *Data Mining with Decision Trees: Theory and Applications 2nd Edition* (H. Bunke & P. S. P. Wang (eds.); 2nd ed., Vol. 81). World Scientific Publishing Co. Pte. Ltd.
- Rozi, A. F. (2015). Sistem Pendukung Keputusan Seleksi Penerimaan Calon Siswa/i Baru Menggunakan Algoritma C4.5. *Teknoin*, 21(1), 1–11. <https://doi.org/10.20885/teknoin.vol21.iss1.art2>
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(4), 379–423, 623–656. <https://doi.org/10.1002/j.1538-7305.1948.tb00917.x>