

## PREDIKSI KETEPATAN KELULUSAN MAHASISWA MENGUNAKAN METODE DECISION TREE BERBASIS *PARTICLE SWARM OPTIMATION (PSO)*

**Arif Rakhman**

Email : arif@limamedia.net

D III Teknik Komputer Politeknik Harapan Bersama Tegal

Jalan Mataram No. 9 Kota Tegal Telp (0283) 352000

### Abstrak

Tingkat penurunan kelulusan mahasiswa yang signifikan dan terus berkembang merupakan masalah yang ada pada perguruan tinggi. Maka dari itu pemantauan atau evaluasi terhadap kecenderungan mahasiswa lulus tepat waktu atau tidak menjadi sangat vital dan hal ini menjadi tugas bagi semua elemen yang ada di institusi perguruan tinggi, sehingga pemaksimalan kinerja harus dilakukan. Banyak faktor yang menyebabkan ketidaktepatan waktu kelulusan mahasiswa tersebut, faktor-faktor tersebut dapat bersumber dari faktor internal dan faktor eksternal. Banyak penelitian yang menggunakan metode *decision tree* maupun *classification tree* dalam memprediksi tentang kelulusan tetapi nilai akurasi yang dihasilkan akurasinya masih kurang tinggi. Setelah dilakukan pengujian dengan dua model yaitu Algoritma Decision Tree dan Algoritma Decision Tree berbasis *Particle Swarm Optimization (PSO)* maka hasil yang didapat adalah Algoritma Decision Tree menghasilkan nilai akurasi sebesar 96,57 % dan nilai AUC sebesar 0.942 dengan tingkat diagnosa *Excellent Classification*, namun setelah dilakukan penambahan yaitu Algoritma Decision Tree berbasis *Particle Swarm Optimization* nilai akurasi sebesar 97.19 % dan nilai AUC sebesar 0.969 dengan tingkat diagnosa *Excellent Classification*. Sehingga kedua metode tersebut memiliki perbedaan tingkat akurasi yaitu sebesar 0.62 dan perbedaan nilai AUC sebesar 0,027.

**Kata kunci:** *Kelulusan, Decision tree, Particle Swarm Optimization*

### 1. Pendahuluan

Perkembangan pendidikan dewasa ini semakin maju dan semakin bertambah pesat jumlahnya. Pendidikan adalah usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual, keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya, masyarakat, bangsa, dan Negara [1]. Pendidikan merupakan sarana untuk mendapatkan suatu ilmu atau pengalaman yang berguna untuk mengembangkan otak bagi orang yang menempuhnya.

Salah satu jenjang pendidikan yang menjadi persyaratan dasar dalam mencari pekerjaan adalah perguruan tinggi, yang mana perguruan tinggi akan mempersiapkan calon-calon sarjana yang handal dan mempunyai keterampilan dibidangnya. Banyak faktor yang menyebabkan ketidaktepatan waktu kelulusan mahasiswa tersebut, faktor-faktor tersebut dapat

bersumber dari faktor internal dan faktor eksternal [2].

Untuk membantu dalam menemukan informasi-informasi berharga itu diperlukan teknik *data mining*. *Data mining* adalah teknik untuk menemukan dan mendeskripsikan pola-pola yang ada dalam data sebagai sebuah alat untuk membantu menjelaskan data tersebut dan membuat prakiraan dari data itu [3].

Perguruan tinggi saat ini dituntut untuk memiliki keunggulan bersaing dengan memanfaatkan semua sumber daya yang dimiliki. Selain sumber daya sarana, prasarana, dan manusia, sistem informasi adalah salah satu sumber daya yang dapat digunakan untuk meningkatkan keunggulan bersaing [4].

Tingkat penurunan kelulusan mahasiswa yang signifikan dan terus berkembang merupakan sebuah masalah yang ada pada perguruan tinggi. Maka dari itu pemantauan atau evaluasi

terhadap kecenderungan mahasiswa lulus tepat waktu atau tidak menjadi sangat vital dan hal ini menjadi tugas bagi semua elemen yang ada di institusi perguruan tinggi, sehingga pemaksimalan kinerja harus dilakukan. Salah satu pemaksimalan kinerja yang dilakukan adalah pemantauan kinerja yang melibatkan penilaian yang melayani peran penting dalam menyediakan informasi yang diarahkan untuk membantu mahasiswa, guru atau dosen, administrator, dan pembuat kebijakan keputusan [9].

Dalam melakukan penelitian ini menggunakan Metode Decision Tree dan Decision Tree PSO untuk melakukan prediksi terhadap tingkat kelulusan tepat waktu berdasarkan pada parameter input yaitu; Jenis Kelamin, Status Perkawinan, Status Pekerjaan, Indeks Prestasi Semester 1, Indeks Prestasi semester 2, Indeks Prestasi semester 3, Indeks Prestasi semester 4.

Maka diperlukan sebuah prediksi kelulusan mahasiswa dengan teknik Data mining menggunakan algoritma Decision Tree dengan Particle Swarm Optimization (PSO).

A. Pengertian Data Mining

Istilah data mining memiliki beberapa padanan, seperti *knowledge discovery* ataupun *pattern recognition*. Kedua istilah tersebut sebenarnya memiliki ketepatannya masing-masing, istilah *knowledge discovery* atau penemuan pengetahuan tepat digunakan karena tujuan utama dari data mining memang untuk mendapatkan pengetahuan yang tersembunyi di dalam bongkahan data. Istilah *pattern recognizing* atau pengenalan pola pun tepat digunakan karena pengetahuan yang hendak digali memang berbentuk pola-pola yang mungkin juga masih perlu digali dari dalam bongkahan data yang tengah dihadapi. Bila dalam penulisan ini digunakan istilah data mining hal ini lebih didasarkan pada lebih populernya istilah tersebut dalam kegiatan penggalian pengetahuan data. Istilah data mining memiliki hakikat ( notion ) sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau

menambang pengetahuan dari data atau informasi yang kita miliki. Kegiatan inilah yang menjadi bidang garapan dari disiplin ilmu data mining [14].

B. Algoritma Decision Tree

Ada beberapa algoritma yang sering digunakan dalam pembentukan pohon keputusan, antara lain: ID3, CART, dan C 4.5. Algoritma C 4.5 ini adalah algoritma pengembangan dari algoritma ID3. Data yang digunakan dalam pohon keputusan dinyatakan dalam bentuk tabel dengan atribut, dimana atribut menyatakan parameter sebagai kriteria dalam pembentukan pohon keputusan. Proses pohon keputusan yaitu mengubah bentuk tabel menjadi model pohon selanjutnya model pohon menjadi rule, dan terakhir menyederhanakan rule.

Pada umumnya algoritma Decision Tree dalam membangun pohon keputusan adalah sebagai berikut:

- 1) Memilih atribut sebagai akar
- 2) Membuat cabang untuk tiap – tiap nilai
- 3) Membagi kasus dalam cabang
- 4) Mengulangi proses untuk tiap cabang semua kasus pada cabang memiliki kelas yang sama.

Dalam pemilihan atribut sebagai akar berdasarkan nilai gain tertinggi dari atribut yang ada. Untuk menghitung nilai gain dapat menggunakan rumus persamaan:

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

Keterangan:

- S : himpunan kasus
- A : atribut
- n : jumlah partisi atribut A
- |S<sub>i</sub>| : jumlah kasus pada partisi ke-i
- |S| : jumlah kasus dalam S

Untuk nilai entropi, dengan persamaan:

$$Entropy(S) = \sum_{i=1}^n -p_i * log_2 p_i$$

Keterangan:

- S : himpunan kasus
- A : fitur
- n : jumlah partisi S
- p<sub>i</sub> : proporsi dari S<sub>i</sub> terhadap S

C. Algoritma PSO

*Particle Swarm Optimimization* (PSO) adalah algoritma pencarian berbasis populasi dan diinisialisasi dengan populasi solusi acak yang disebut partikel [20]. PSO merupakan metode pencarian yang berasal dari penelitian untuk gerakan sekelompok burung atau ikan. Serupa dengan algoritma genetik (GA), PSO melakukan pencarian menggunakan populasi (*swarm*) dari individu (partikel) yang akan diperbaharui dari iterasi ke iterasi.

Untuk PSO dapat diasumsikan sebagai kelompok burung secara acak mencari makanan disuatu daerah. Burung tersebut tidak tahu dimana makanan tersebut berada, tapi mereka tahu seberapa jauh makanan itu berada, jadi strategi terbaik untuk menemukan makanan tersebut adalah dengan mengikuti burung yang terdekat dari makanan tersebut [20]. PSO digunakan untuk memecahkan masalah optimasi.

Menemukan solusi yang optimal, masing-masing partikel bergerak kearah posisi yang terbaik sebelumnya dan posisi terbaik secara global. Sebagai contoh, partikel ke-i dinyatakan sebagai:  $x_i = (x_{i1}, x_{i2}, \dots, x_{id})$  dalam ruang d-dimensi. Posisi terbaik sebelumnya dari partikel ke-i disimpan dan dinyatakan sebagai  $pbest_i = (pbest_{i,1}, pbest_{i,2}, \dots, pbest_{i,d})$ . Indeks partikel terbaik diantara semua partikel dalam kawanan group dinyatakan sebagai  $gbest_d$ . Kecepatan partikel ke-i dinyatakan sebagai:  $v_i = (v_{i,1}, v_{i,2}, \dots, v_{i,d})$ . Modifikasi kecepatan dan posisi tiap partikel dapat dihitung menggunakan kecepatan saat ini dan jarak  $pbest_i, d$  ke  $gbest_d$  seperti ditunjukkan persamaan berikut:

$$v_{i,d} = w * v_{i,d} + c1 * R * (pbest_{i,d} - x_{i,d}) + c2 * R * (gbest_d - x_{i,d})$$

$$x_{id} = x_{i,d} + v_{i,d}$$

Dimana:

- $V_{i,d}$  = Kecepatan partikel ke-i pada iterasi ke-i
- $w$  = Faktor bobot inersia
- $c1, c2$  = Konstanta akselerasi (learning rate)
- $R$  = Bilangan random (0-1)
- $x_{i,d}$  = Posisi saat ini dari partikel ke-i pada iterasi ke-i
- $pbest_i$  = Posisi terbaik sebelumnya dari partikel ke-i

- $gbesti$  = Partikel terbaik diantara semua partikel dalam satu kelompok atau populasi
- $n$  = Jumlah partikel dalam kelompok
- $d$  = Dimensi

Persamaan (2.3) menghitung kecepatan baru untuk tiap partikel (solusipotensial) berdasarkan pada kecepatan sebelumnya ( $V_{i,m}$ ), lokasi partikel dimana nilai fitness terbaik telah dicapai ( $pbest$ ), dan lokasi populasi global ( $gbest$  untuk versi global,  $lbest$  untuk versi local) atau local neighborhood pada algoritma versi local dimana nilai fitness terbaik telah dicapai. Persamaan (2.4) memperbaharui posisi tiap partikel pada ruang solusi. Dua bilangan acak  $c1$  dan  $c2$  dibangkitkan sendiri. Penggunaan berat inersia  $w$  telah memberikan performa yang meningkat pada sejumlah aplikasi. Hasil dari perhitungan partikel yaitu kecepatan partikel diantara interval  $[0,1]$ .

2. Metode Penelitian

Penelitian ini memakai data lulusan mahasiswa program DIII Politeknik Harapan Bersama Tegal, data yang dibutuhkan dalam penelitian ini antara lain jumlah data eksperimen 300 data dengan atribut seperti jenis kelamin, status pernikahan, status pekerjaan, IPS\_1, IPS\_2, IPS\_3, IPS\_4. Atribut-atribut tersebut diambil dari data sistem akademik mahasiswa Politeknik Harapan Bersama Tegal.

Adapun tahapan proses penelitian yang akan dilakukan adalah sebagai berikut:

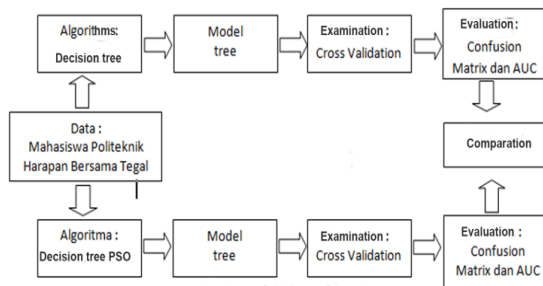


Figure 1. Phase Experiment

Gambar 1. Proses Penelitian

### 3. Hasil dan Pembahasan

#### A. Hasil Pengujian Model Algoritma Decision Tree

**Tabel 1** Confusion Matrix

Observed class	Predicted Class	
	True Tepat waktu	True Tidak Tepat waktu
True Tepat waktu	224	9
True Tidak Tepat waktu	1	66

**Tabel 2** Nilai accuracy, recall dan precision

	Nilai (%)
Accuracy	96.67
Recall	96.13
Precision	99.55

#### B. Hasil Pengujian Model Algoritma Decision Tree PSO

**Tabel 3** Confusion Matrix

Observed class	Predicted Class	
	True Tepat waktu	True Tidak Tepat waktu
True Tepat waktu	224	9
True Tidak Tepat waktu	1	66

**Tabel 4** Nilai accuracy, recall dan precision

	Nilai (%)
Accuracy	97.67
Recall	97.39
Precision	99.55

#### C. Analisa Hasil Pengujian

Dari hasil pengujian diatas, baik evaluasi menggunakan *counfusion matrix* maupun *ROC curve* terbukti bahwa hasil pengujian algoritma decision tree dengan PSO memiliki nilai akurasi yang lebih tinggi dibandingkan dengan algoritma decision tree. Nilai akurasi untuk model algoritma decision tree sebesar 96.67 % dan nilai akurasi untuk model algoritma Decision Tree dengan PSO sebesar 97.67 % dengan selisih akurasi 1 %, dapat dilihat pada Tabel dibawah ini :

**Tabel 5** selisih akurasi

	Accuracy	Step-up	AUC	Step-up
Decision tree	96.67 %		0.951	
Decision tree PSO	97.67 %	1 %	0.962	0.011

Untuk evaluasi menggunakan *ROC curve* sehingga menghasilkan nilai *AUC (Area Under Curve)* untuk model algoritma Decision Tree mengasilkan nilai 0.951 dengan nilai diagnosa *Excellent Classification*, sedangkan untuk algoritma Decision Tree dengan *PSO (Particle Swarm Optimization)* menghasilkan nilai 0.962 dengan nilai diagnose *Excellent Classification*.

### 4. Kesimpulan

Dari hasil analisis optimasi dapat disimpulkan bahwa nilai akurasi yang didapat pada model algoritma Decision Tree dengan *PSO* adalah 97.67 % lebih baik jika dibandingkan dengan model algoritma Decision Tree yaitu 96.67 %. Dari hasil tersebut didapatkan selisih antara kedua model yaitu sebesar 1 %. Sementara untuk evaluasai menggunakan *ROC curve* untuk kedua model yaitu, untuk model algoritma Decision Tree nilai *AUC* adalah 0.951 dengan tingkat diagnosa *Excellent Classification*, dan untuk model algoritma Decision Tree dengan *PSO* nilai *AUC* adalah 0.962 dengan tingkat diagnosa *Excellent Classification*, sehingga didapatkan selisih nilai *AUC* sebesar 0.011. Sehingga dapat disimpulkan bahwa penerapan teknik optimasi *particle swarm optimization* dapat meningkatkan nilai akurasi pada algoritma Decision Tree.

### 5. Daftar Pustaka

- [1] Undang-undang *sisdiknas*. 2011.
- [2] I. A. Sri Padmini, "Analisis Waktu Kelulusan Mahasiswa Dengan Metode CHAID," *e-Jurnal Matematika*, vol. 1, no. 1, pp. 89-93, 2012.
- [3] I. H. Witten and E. Frank, *Data Mining*. 2005.
- [4] N. M. Huda, "Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa," 2010.
- [5] M. N. Quadri and N. V. Kalyankar, "Drop Out Feature of Student Data for Academic Performance Using Decision Tree Techniques," *Global Journal of Computer Science and Technology*, vol. 10, no. 2, p. 2, Apr. 2010.
- [6] Depkumham, "Peraturan pemerintah RI

- no 66 Tentang Perubahan atas peraturan Pemerintah no 17 tahun 2010 tentang pengolahan dan penyelenggaraan pendidikan," 2010.
- [7] H. H. Nawawi and H. M. Martini, *Kebijakan pendidikan di indonesia ditinjau dari sudut hukum*. Gajah Mada University Press, 1994.
- [8] T. S. Karamouzis and A. Vrettos, "An Artificial Neural Network for Predicting Student Graduation Outcomes," *Proceeding of World Congress on Engineering and Computer Science*, pp. 1-8, 2008.
- [9] E. N. Ogor, "Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques," in *fourth Congress of Electronics, Robotics and Automotive Mechanics*, p. 8, 2007.
- [10] S. T. Hijazi and R. Naqvi, "Factors Affecting Students' Performance," *Bangladesh e-Journal of Sociology*, vol. 3, no. 1, pp. 45-51, Jan. 2006.
- [11] D. L. Fithri and D. Eko, "Sistem Pendukung Keputusan Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Metode Naive Bayes," *SNATIF*, 2014.
- [12] M. S. Suhartinah and Ernastuti, "Graduation Prediction Of Gunadarma University Students Using Algorithm And Naive Bayes DECISION TREE Algorithm," 2010.
- [13] E. Supriyadi and D. I. Sensuse, "Optimasi Algoritma Support Vector Machine Dengan Particle Swarm Optimization Dalam Mendeteksi Ketepatan Waktu Kelulusan Mahasiswa," *SNIT*, 2015.
- [14] S. Susanto and D. Suryadi, *Pengantar Data Mining; Menggali Pengetahuan dari Bongkahan Data*. Andi Offset, 2010.
- [15] D. T. Larose, *Discovering Knowledge in data : An Introduction to Data Mining*. 2005.
- [16] M. W. Berry and M. Browne, *Lecture notes in data mining*. Singapore: world scientific, 2006.
- [17] J. Han and M. Kamber, *Data Mining Concepts and Techniques*. San Francisco: Morgan Kaufmann Publishers, 2006.
- [18] O. Maimon and L. Rokach, *Data Mining and Knowledge Discovery Handbook*. New York, USA: Springer, 2005.
- [19] Kusriani and T. Emha, *Algoritma Data Mining*. Yogyakarta: Andi Offset, 2009.