

Analisis Prediksi Kelulusan Mahasiswa Menggunakan Decision Tree Berbasis Particle Swarm Optimization

Hendra^{[1]*}, Mochammad Abdul Azis^[2], Suhardjono^[3]

Program Studi Ilmu Komputer STMIK Nusa Mandiri^[1],

Program Studi Sistem Informasi Universitas Bina Sarana Informatika^{[2][3]}

14002267@nusamandiri.ac.id^[1], mohammad.mmz@bsi.ac.id^[2], Suhardjono@bsi.ac.id^[3]

Abstract—Good accreditation results are the goal of the college. With good accreditation, prospective students can glance at and enter the tertiary institution. To achieve this, there are several aspects that affect good accreditation results, one of which is graduate students who play an important role in determining accreditation. Timely graduate students can benefit the college or a student. Graduates can be predicted before the final semester using a method one of which is the decision tree. Decision tree is a method that is simple and easy to understand by producing rules in the form of a decision tree, but using a decision tree model alone is not enough to produce optimal results. So we need a method for optimization that is particle swarm optimization with advantages can improve accuracy by eliminating unused features. From the results of research with primary data of 2000-2003 graduate students in Amik PPMI Tangerang explained that the particle swarm optimization method can increase accuracy by 87.56% and increase by 01.01% from the decision tree method with a value of 86.55%. From the particle swarm optimization method can also find out which unused attributes have no weight, so that way can improve accuracy. From the results of the increase, it can be used by the Amik University of Tangerang to prevent students from graduating on time.

Keywords— Decision Tree, Student Graduation, Particle Swarm Optimization.

Abstrak—Hasil akreditasi yang baik merupakan tujuan dari perguruan tinggi. Dengan adanya akreditasi yang baik membuat para calon mahasiswa dapat melirik dan masuk perguruan tinggi tersebut. Untuk mencapai tersebut ada beberapa aspek yang mempengaruhi hasil akreditasi yang baik, salah satunya mahasiswa lulusan yang merupakan peran penting dalam menentukan akreditasi. Mahasiswa lulusan yang tepat waktu dapat menguntungkan bagi perguruan tinggi tersebut ataupun bagi seorang mahasiswa. Mahasiswa lulusan dapat diprediksi sebelum semester akhir dengan menggunakan sebuah metode salah satunya yaitu decision tree. Decision tree merupakan metode yang sederhana dan mudah dipahami dengan menghasilkan *rule-rule* berupa pohon keputusan, akan tetapi menggunakan model *decision tree* saja tidak cukup untuk menghasilkan hasil yang optimal. Maka dibutuhkan metode untuk pengoptimalan yaitu *particle swarm optimization* dengan kelebihan dapat meningkatkan akurasi dengan cara menghilangkan *fitur* yang tidak terpakai. Dari hasil penelitian dengan data primer mahasiswa lulusan tahun 2000-2003 pada amik ppmi tangerang menerangkan bahwa metode *particle*

swarm optimization dapat meningkatkan akurasi sebesar 87.56% dan meningkatkan sebesar 01.01% dari metode *decision tree* dengan nilai 86.55%. Dari metode *particle swarm optimization* juga dapat mengetahui mana atribut yang tidak terpakai yang tidak memiliki bobot, sehingga cara tersebut dapat meningkatkan akurasi. Dari hasil kenaikan tersebut dapat dipakai oleh perguruan tinggi amik tangerang untuk menghindari mahasiswa lulus tidak tepat waktu.

Kata Kunci—*Decision Tree, Kelulusan Mahasiswa, Particle Swarm Optimization.*

I. PENDAHULUAN

Perguruan tinggi untuk menentukan hasil akreditasi perguruan tinggi atau program studinya dapat ditentukan dari beberapa aspek salah satunya mahasiswa yang lulus. Kelulusan mahasiswa sangat berpengaruh dalam menjalani proses akademik sebuah perguruan tinggi. Terlebih kelulusan mahasiswa sangat dinantikan bagi mahasiswanya sendiri maupun orang tua atau keluarganya. Mahasiswa dalam kelulusan sangat berharap lulus dengan tepat waktu, dikarenakan dengan lulus tepat waktu dapat membantu mahasiswa dalam mencari kerja dengan cepat dan dapat melanjutkan ke jenjang berikutnya. Kelulusan dengan terlambat dapat mengakibatkan hasil akreditasi atau kualitas perguruan tinggi menurun serta memerlukan dosen yang lebih untuk memenuhi rasio dosen. Kelulusan dapat diprediksi sebelum mahasiswa tersebut mencapai semester akhir maka diperlukan evaluasi dalam proses belajar mengajar, sehingga dapat diperbaiki prestasi mahasiswa. Untuk memperbaiki prestasi mahasiswa sebuah perguruan tinggi tentu saja mempunyai data mahasiswa di server yang mereka simpan sendiri dan untuk menganalisa dan mengambil suatu informasi dari data tersebut tidak mudah, mengingat data yang sangat besar dan tidak memiliki kecerdasan dalam keahlian data yang tersimpan tersebut.

Ada beberapa data yang tersimpan dalam perguruan tinggi tapi tidak mempengaruhi dengan kelulusan mahasiswa. Kelulusan mahasiswa dipengaruhi dari data yang ada seperti jurusan SLTA, asal SLTA, jenis kelamin dan tentu saja IPK [1]. Penelitian tentang kelulusan dengan menganalisis faktor IPK yang diduga dapat mempengaruhi kelulusan mahasiswa ternyata tidak signifikan dengan kelulusan mahasiswa tersebut [2]. Peneliti lain juga telah melakukan dengan ketiga metode

NN, NB dan DT dengan hasil bahwa metode NN lebih tinggi akurasi yang didapat dibandingkan dengan metode yang lain [3]. Ada juga yang membandingkan metode dari ketiga tersebut dengan menambahkan metode *regresi* dengan data yang berbeda akan tetapi dari hasil tersebut bahwa metode NB mendapatkan nilai akurasi tertinggi [4]. Dari atribut yang disebutkan ada beberapa metode untuk memprediksi suatu data tersebut contohnya *Decision Tree* (DT), *Neural Network* (NN), *Support Vector Machine* (SVM) dan *Naïve Bayes* (NB). Dari metode tersebut sudah pernah dilakukan untuk kelulusan mahasiswa dengan ketiga metode kecuali metode naïve bayes, dari hasil penelitian yang telah dilakukan akurasi metode *support vector machine* mendapatkan nilai akurasi tertinggi sebesar 85.18% dan menempati peringkat pertama dan ditempat kedua dengan akurasi 84.96% ditempat metode *decision tree* dan yang ditempat terakhir dengan akurasi 84.68% yaitu metode *neural network* [1].

Dari hasil penelitian prediksi kelulusan di kembangkan oleh peneliti lain dengan *support vector machine* berbasis *particle swarm optimization*, dimana dari permasalahan yang ada terhadap *support vector machine* tentang kelemahan optimasi telah ditutupi oleh metode *particle swarm optimization*, dan dari hasil yang didapat oleh metode PSO sebesar 86.43% naik menjadi 00.62% dari metode SVM sebelumnya [5] dan dari hasil penelitian menunjukkan adanya peningkatan akurasi terhadap mahasiswa lulusan tersebut, dengan hasil akurasi sebesar 86.43% dari hasil sebelumnya menggunakan *support vector machine* sebesar 85.18% dan mengalami kenaikan sebesar 01.29%. PSO juga dapat meningkatkan akurasi dari metode NN sebesar 00.02% [6]. *Decision tree* menghasilkan akurasi terbaik sebesar 84.4% [7]. Namun penelitian ini tidak menggunakan *support vector machine* melainkan menggunakan *decision tree* dimana pada penelitian terdahulu *decision tree* mendirikan peringkat kedua setelah *support vector machine*.

Decision tree merupakan metode yang sering dipakai dan populer dari metode lainnya dengan peringkat yang kesepuluh [8]. *Decision tree* mampu menghasilkan pohon keputusan dengan aturan atau rule-rule untuk membuat klasifikasi baru dari nilai atribut menjadi *class* dengan cara penklasifikasiannya [9]. *Decision tree* mampu menghasilkan keputusan yang sederhana dari hasil keputusan yang rumit dengan cara mengubahnya menjadi sederhana [10]. *Decision tree* sangat

mudah dimengerti dalam mengolah data yang kecil [11]. *Decision tree* merupakan metode yang tanpa mengurangi hasil dari kualitas yang didapat dengan menggunakan kriteria setiap *node* [12].

Dari kelebihan algoritma *decision tree* diperlukan metode untuk mengoptimasi algoritma tersebut, diantara metode optimasi yang ada penelitian ini menggunakan metode optimasi *particle swarm optimization* (PSO). *Particle swarm optimization* merupakan konsep yang sederhana dengan implementasi sangat mudah, dan *konvergensi* cepat [13]. Terutama proses dalam pemilihan *fitur* yang secara optimal dapat dilakukan dengan menghilangkan *fitur-fitur* yang tidak relevan [14]. Dari dilakukan penghilangan *fitur* yang berlebih akan mengakibatkan akurasi meningkat [15].

II. BAHAN DAN METODE

Model *decision tree* dengan sistem pengambilan keputusan dan menghasilkan sebuah pohon keputusan yang terdiri dari simpul internal, tepi dan simpul daun dengan simpul internal yang merupakan keputusan dengan mewakili setiap set pada atribut dan masing label diberikan dengan masing-masing edge [16]. Model *decision tree* dengan alur proses untuk membentuk pohon keputusan dalam data mining sebagai berikut: 1). Membuat simpul akar dengan memilih atribut 2). Untuk tiap-tiap nilai akan dibuat cabang. 3). Bagi kasus dalam tiap-tiap cabang. 4). Ulangi proses pembagian untuk tiap-tiap cabang sampai tiap-tiap kasus pada cabang mendapatkan kelas yang sama [7].

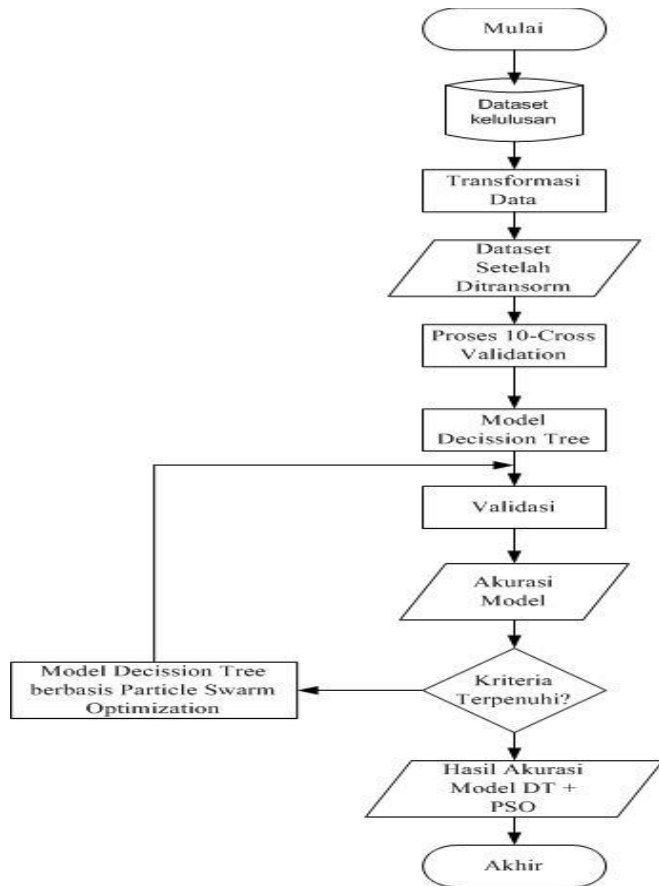
Model *particle swarm optimization* adalah teknik optimisasi global dengan melalui simulasi perilaku social kawanan burung yang sering berkelompok, dengan ide dasarnya adalah pertama menginisialisasi seperangkat partikel acak dan kemudian menemukan solusi optimal dengan menggunakan iterasi [17].

Penelitian ini menggunakan *decision tree* dengan *particle swarm optimization* dimana data yang diambil pada tahun 2000-2003 dari satu perguruan tinggi swasta amik tangerang dengan *parameter* sebanyak 796. Dari data tersebut dilakukan diuji coba menggunakan algoritma DT dan DT berbasis PSO. *Sample* data kelulusan mahasiswa dari perguruan tinggi swasta dapat dilihat ditabel 1.

Tabel 1. *Sample* Data Kelulusan Mahasiswa

JK	Jurusan SLTA	Asal SLTA	IPK1	IPK2	IPK3	IPK4	IPK 5	IPK 6	HASIL
L	IPA	SMA N 1 Lais	2,55	2,61	2,64	2,74	2,97	3	YA
L	Teknik Komputer Jaringan	SMK TI YPML	3,23	3,15	3,07	3,17	3,25	3,17	YA
P	Administrasi Perkantoran	SMK Nusantara 1	2,73	2,73	2,62	2,64	2,89	3	YA
L	IPA	SMA Yuppentek 1	2,73	3,1	2,97	3,16	3,3	3,33	YA
L	IPS	Hidayatul Ikhwan	2,18	2,44	2,11	2,36	2,66	2,57	TIDAK
P	TKJ	SMK Pustek	2,32	2,39	2,43	2,74	2,67	2,95	TIDAK
P	Penjualan	SMK Negeri 1	3,64	3,56	3,59	3,53	3,53	3,51	YA
L	Otomotif	SMK	2,64	2,22	2,3	2,11	2,49	2,74	YA
L	Otomotif	SMK Negeri 2	2,5	2,71	2,1	2,57	2,6	2,58	TIDAK
P	IPA	SMA Negeri 1 Jiwan	2,91	3,07	2,35	2,84	3,3	2,95	TIDAK

Data pada tabel 1 telah dilakukan penghilangan data yang tidak valid, setelah itu dilakukan normalisasi data dan mentransformasi data dengan *decision tree* dan *particle swarm optimization*. Dari data tersebut diuji coba dengan model yang diusulkan yang dapat dilihat pada gambar 1.



Gambar 1. Model Penelitian yang diusulkan

Setelah data set di *transform* dan dari hasil pembersihan data, data yang sudah bersih maka dilakukan uji *validasi* dengan menggunakan *10-cross validation* dan model *decision tree*, setelah itu di *validasi* dan menghasilkan akurasi dari model *decision tree*. Jika kriteria akurasi tidak terpenuhi maka dilakukan model dengan *particle swarm optimization* untuk memperbaiki fitur yang berlebih dan tidak relevan. Jika PSO diterapkan maka divalidasi kembali sehingga mendapatkan akurasi model yang memenuhi kriteria.

III. HASIL DAN PEMBAHASAN

Penelitian ini dengan melakukan pengolahan data melalui *software* rapid miner 9.3 Data yang disudah melalui tahap model penelitian menghasilkan *confusion matrix* dari metode *decision tree* yang tertera pada table 2.

Tabel 2. Confusion Matrix DT

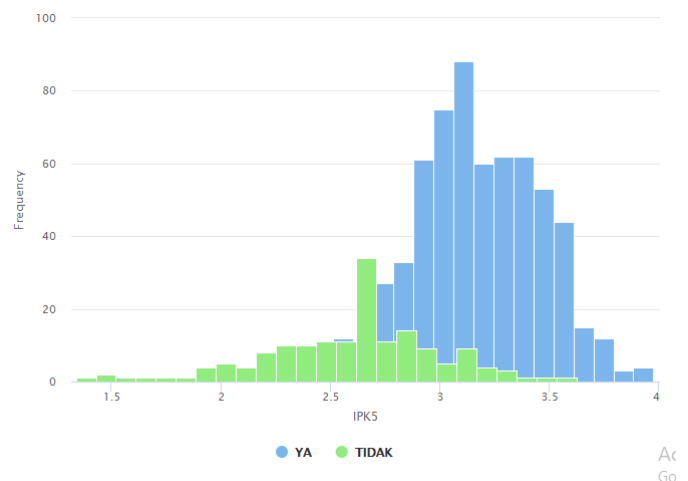
	Ya	Tidak
Prediksi Lulus	582	55
Prediksi Tidak Lulus	52	107

Dari table 2 *confusion matrix* yang dihasilkan *True Negatif* (TN) sejumlah 582, *False Negatif* (FN) sejumlah 52, *True Positif* (TP) sejumlah 107 dan *False Positif* (FP) sejumlah 55. Dari data tersebut menerangkan bahwa mahasiswa yang diprediksi lulus tepat sejumlah waktu sejumlah 582 mahasiswa dan prediksi tersebut benar dan sesuai dengan penelitian, sedangkan dengan jumlah mahasiswa yang diprediksi lulus senilai 55 ternyata setelah dilakukan ujicoba dengan metode *decision tree* mahasiswa tersebut tidak lulus. Untuk mahasiswa yang diprediksi tidak lulus sejumlah 107 dan hasil uji coba model *decision tree* benar bahwa mahasiswa tersebut tidak lulus, sedangkan dengan jumlah mahasiswa yang diprediksi tidak lulus sejumlah 52 ternyata hasil membuktikan mahasiswa tersebut lulus tepat waktu. Dari tabel tersebut dapat dihitung rumus akurasi dibawah ini.

$$\text{Akurasi} = \frac{(TN + TP)}{(TN + FN + TP + FP)}$$

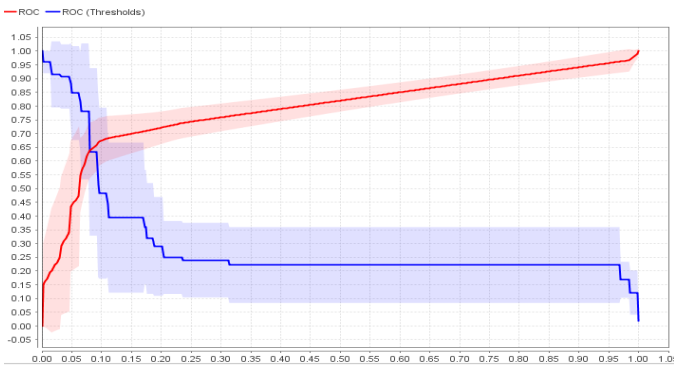
$$\text{Akurasi} = \frac{(582 + 107)}{(582 + 52 + 107 + 55)} = 0,865578$$

Dengan rumus akurasi *decision tree* tersebut dapat dihasilkan akurasi dengan model *decision tree* sebesar 0.865578 atau diubah dengan persen senilai 86.5578%. Dari hasil prediksi kelulusan mahasiswa dapat dilihat dengan diagram batang di bawah ini.



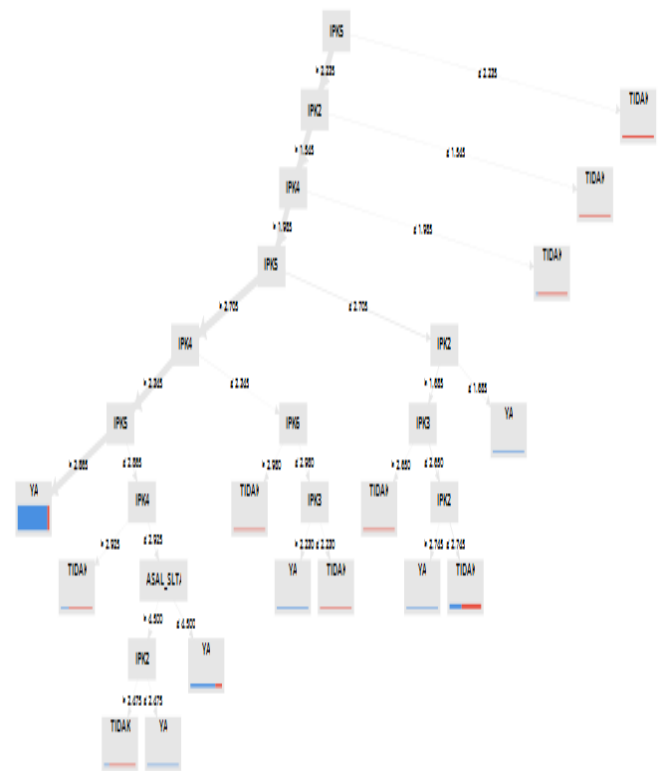
Gambar 2. Digram Batang Model Decision Tree

Dari diagram batang tersebut dapat juga dilihat kurva dengan model *decision tree*, kurva tersebut merupakan *area under curve* (AUC) yang dapat dilihat dari gambar 3.



Gambar 3. Diagram AUC Kinerja Assiten

Dari hasil uji coba dengan model *decision tree* maka didapatkan *rule* algoritma kelulusan mahasiswa seperti gambar berikut.



Gambar 4. Rule Algoritma Decision Tree

Tabel 3. Bobot Atribut

Atribut	Bobot
Jurusan SLTA	0.996
Asal SLTA	1
IPK1	0
IPK2	0.679
IPK3	0
IPK4	0.735
IPK5	1
IPK6	0.230

Setelah di uji coba dengan menggunakan *decision tree* maka di optimasikan dengan *particle swarm optimization*. Setelah dilakukan uji coba menggunakan optimasi PSO maka didapatkan bobot atribut yang dapat dilihat pada tabel 3.

Dari hasil bobot tersebut, terdapat bobot yang tidak mempengaruhi secara signifikan terhadap kelulusan mahasiswa diantaranya atribut IPK1 (ipk semester 1) dan IPK3 (ipk semester 3). Untuk atribut yang lain seperti Jurusan dan Asal SLTA, IPK2, IPK4, IPK5, IPK6 merupakan atribut yang mempengaruhi secara signifikan. Setelah mendapatkan bobot atribut, terdapat juga *confusion matrix*.

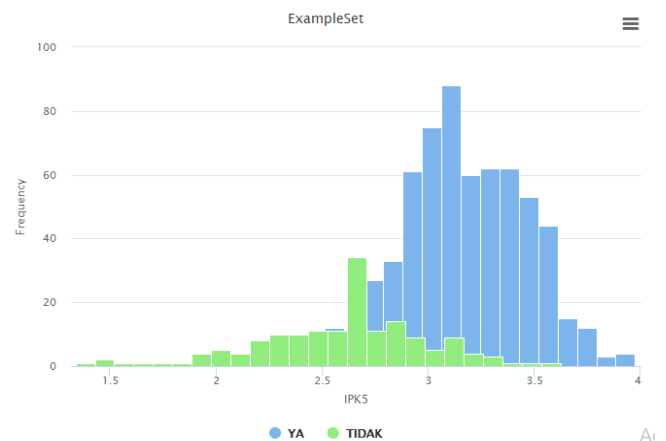
Tabel 4. Confusion Matrix DT+PSO

	Ya	Tidak
Prediksi Lulus	589	54
Prediksi Tidak Lulus	45	108

Dari table 4 *confusion matrix* yang dihasilkan *True Negatif* (TN) sejumlah 589, *False Negatif* (FN) sejumlah 45, *True Positif* (TP) sejumlah 108 dan *False Positif* (FP) sejumlah 54. Dari data tersebut menerangkan bahwa mahasiswa yang diprediksi lulus tepat waktu sejumlah 589 mahasiswa dan prediksi tersebut benar dan sesuai dengan penelitian, sedangkan dengan jumlah mahasiswa yang diprediksi lulus senilai 54 ternyata setelah dilakukan ujicoba dengan metode *decision tree* mahasiswa tersebut tidak lulus. Untuk mahasiswa yang diprediksi tidak lulus sejumlah 108 dan hasil uji coba model *decision tree* benar bahwa mahasiswa tersebut tidak lulus, sedangkan dengan jumlah mahasiswa yang diprediksi tidak lulus sejumlah 45 ternyata hasil membuktikan mahasiswa tersebut lulus tepat waktu. Dari tabel tersebut dapat dihitung rumus akurasi dibawah ini.

$$\text{Akurasi} = \frac{(TN+TP)}{(TN+FN+TP+FP)} \dots\dots\dots(1)$$

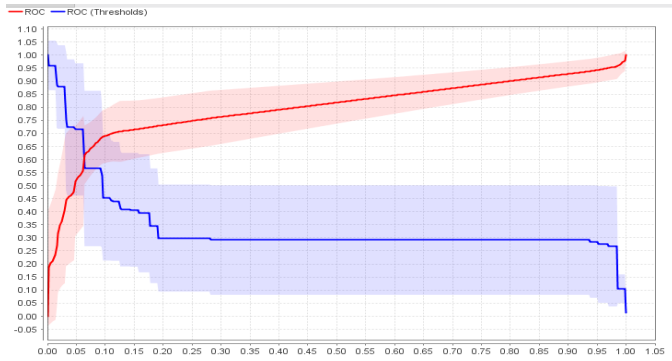
$$\text{Akurasi} = \frac{(589 + 108)}{(589 + 45 + 108 + 54)} = 0,875628$$



Gambar 5. Digram Batang Model Decision Tree Berbasis Particle Swarm Optimization

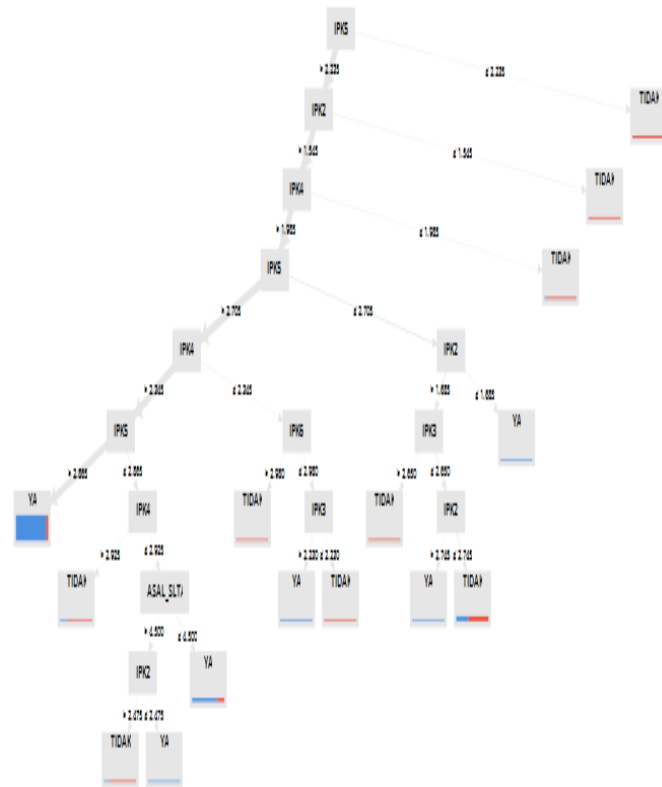
Dengan rumus akurasi *decision tree* dengan *particle swarm optimization* tersebut dapat dihasilkan akurasi dengan model *decision tree* berbasis *particle swarm optimization* sebesar 0.875628 atau diubah dengan persen senilai 87.56%. Dari hasil prediksi kelulusan mahasiswa dapat dilihat dengan diagram batang pada Gambar 5.

Dari diagram batang tersebut dapat juga dilihat kurva dengan model *decision tree* berbasis PSO, kurva tersebut merupakan *area under curve* (AUC) yang dapat dilihat dari gambar 6.



Gambar 6. Digram AUC Model *Decision Tree* Berbasis *Particle Swarm Optimization*

Dari hasil uji coba dengan model *decision tree* berbasis PSO maka didapatkan rule algoritma kelulusan mahasiswa seperti gambar berikut.



Gambar 7. Rule Algoritma *Decision Tree* Berbasis *Particle Swarm Optimization*

IV. KESIMPULAN

Particle swarm optimization dengan meningkatkan kelemahan dari *decision tree* yang diusulkan dalam makalah ini yang mempunyai tujuan dalam menghilangkan parameter untuk mendapatkan satu set data kelulusan mahasiswa yang efektif. Dari data kelulusan mahasiswa yang di uji dengan menggunakan metode *decision tree* menunjukkan bahwa *particle swarm optimization* berbasis *decision tree* yang berkinerja lebih baik dari pada metode *decision tree* yang tanpa optimasi.

Decision tree mendapatkan hasil akurasi sebesar 86.55% dengan prediksi kesalahan untuk mahasiswa yang lulus sebesar 107 data mahasiswa dimana data tersebut menyatakan bahwa yang seharusnya lulus tepat waktu dinyatakan tidak tepat waktu, dan sebaliknya yang tidak tepat waktu dinyatakan tepat waktu.

Particle swarm optimization mendapatkan hasil akurasi sebesar 87.56% dengan prediksi kesalahan untuk mahasiswa yang lulus sebesar 99 data mahasiswa dimana data tersebut menyatakan bahwa yang seharusnya lulus tepat waktu dinyatakan tidak tepat waktu, dan sebaliknya yang tidak tepat waktu dinyatakan tepat waktu.

Kenaikan akurasi dari metode *particle swarm optimization* sebesar 01.01% dan mengurangi kesalahan prediksi sejumlah 8 data mahasiswa. Dari hasil tersebut dapat digunakan oleh perguruan tinggi untuk menanggulangi mahasiswa yang lulus tidak tepat waktu, sehingga akreditasi pada perguruan tinggi tersebut mendapatkan hasil yang baik.

REFERENCES

- [1] V. Riyanto, A. Hamid, and Ridwansyah, "Prediction of Student Graduation Time Using the Best Algorithm," *Indones. J. Artif. Intell. Data Min.*, vol. 2, no. 2, pp. 1–9, 2019.
- [2] F. A. Freitas and L. J. Leonard, "Maslow's hierarchy of needs and student academic success," *Teach. Learn. Nurs.*, vol. 6, no. 1, pp. 9–13, 2011.
- [3] M. V. Ashok and A. Apoorva, "Data mining approach for predicting student and institution's placement percentage," *2016 Int. Conf. Comput. Syst. Inf. Technol. Sustain. Solut. CSITSS 2016*, pp. 336–340, 2016.
- [4] M. T. Devasia, M. V. T. P., and M. V. Hegde, "Prediction of Students Performance using Educational Data Mining," *Int. J. Cogn. Ther.*, vol. 1, no. 3, pp. 266–279, 2008.
- [5] Suhardjono, G. Wijaya, and A. Hamid, "PREDIKSI WAKTU KELULUSAN MAHASISWA MENGGUNAKAN SVM BERBASIS PSO," *Bianglala Inform.*, vol. 7, no. 2, pp. 97–101, 2019.
- [6] Ridwansyah and E. Purwaningsih, "Particle Swarm Optimization Untuk Meningkatkan Akurasi Prediksi Pemasaran Bank," *J. PILAR Nusa Mandiri*, vol. 14, no. 1, pp. 83–88, 2018.
- [7] H. M. Nawawi, S. Rahayu, J. J. Purnama, and S. I. Komputer, "Algoritma c4.5 untuk memprediksi pengambilan keputusan memilih deposito berjangka," *J. Techno Nuasa Mandiri*, vol. 16, no. 1, pp. 65–72, 2019.
- [8] J. Han and M. Kamber, *Data Mining Concept and Techniques*. India: New Age International Limited, 2006.
- [9] Wu and Kumar, *The Top Ten Algorithms in Data Mining*. USA: CRC Press, 2009.
- [10] C. A. Sugianto, "Penerapan Teknik Data Mining Untuk Menentukan Hasil Seleksi Masuk Sman 1 Cibeer Untuk Siswa Baru Menggunakan Decision Tree," *Tedc*, vol. 9, pp. 39–43, 2015.
- [11] F. Gorunescu, *Data Mining: Concepts, Models and Technique*. 2011.
- [12] D. Setiawati, I. Taufik, J. Jumadi, and W. B. Zulfikar, "Klasifikasi Terjemahan Ayat Al-Quran Tentang Ilmu Sains Menggunakan Algoritma Decision Tree Berbasis Mobile," *J. Online Inform.*, vol. 1, no. 1, p. 24, 2016.

- [13] H. Liu, H. Q. Tian, C. Chen, and Y. F. Li, "An experimental investigation of two Wavelet-MLP hybrid frameworks for wind speed prediction using GA and PSO optimization," *Int. J. Electr. Power Energy Syst.*, vol. 52, no. 1, pp. 161–173, 2013.
- [14] T. Xu, Q. Peng, and Y. Cheng, "Identifying the semantic orientation of terms using S-HAL for sentiment analysis," *Knowledge-Based Syst.*, vol. 35, pp. 279–289, 2012.
- [15] J. Cao, H. Cui, H. Shi, and L. Jiao, "Big data: A parallel particle swarm optimization-back-propagation neural network algorithm based on MapReduce," *PLoS One*, vol. 11, no. 6, pp. 1–17, 2016.
- [16] W. Dai and J. W., "A mapreduce implementation of C4.5 decision tree algorithm," in *International Journal of Database Theory and Application*, 2014, pp. 49–60.
- [17] L. Bin and Y. Min, "Analysis model of drilling tool failure based on PSO-SVM and its application," *Proc. - 4th Int. Conf. Comput. Inf. Sci. ICCIS 2012*, no. 8, pp. 1307–1310, 2012.