

# Prediksi Tingkat Kelulusan Mahasiswa Menggunakan Metode Hybrid

Agus Bahtiar<sup>1\*</sup>, Fadhil Muhamad Basysyar<sup>2</sup>, Gifthera Dwilestari<sup>3</sup>

<sup>1</sup>Program Studi Teknik Informatika, STMIK IKMI, Cirebon

<sup>2</sup>Program Studi Teknik Informatika, STMIK IKMI, Cirebon

<sup>3</sup>Program Studi Teknik Informatika, STMIK IKMI, Cirebon

<sup>1,2,3</sup>Jln. Perjuangan No 10B Majasem Kota Cirebon, 45143, Indonesia

email: <sup>1</sup> [agusbahtiar038@gmail.com](mailto:agusbahtiar038@gmail.com), <sup>2</sup> [fadhilbasyayar@yahoo.co.id](mailto:fadhilbasyayar@yahoo.co.id), <sup>3</sup> [gifteradwilestarsi@gmail.com](mailto:gifteradwilestarsi@gmail.com)

**Abstract** – Student graduation is one of the fields of internal quality assurance standards (SPMI) of a tertiary institution. The number of students who graduate on time is an indicator of the success of a tertiary institution, both public and private. Higher education has a very important role in creating the best graduates for the needs of the workforce. Research in predicting student graduation has been widely carried out. Based on these conditions, this research is carried out the application of one of the data mining classification techniques, namely C4.5. In this study comparing the modeling of C4.5 algorithm with C4.5 based on Particle Swarm Optimization (PSO) which is applied to student graduation data. From the test results used cross validation, confusion matrix and ROC curves, it is known that C4.5 has an accuracy value of 78.16 % And the performance indicated by the AUC value is 5.62% whereas C4.5 based on PSO has an accuracy value of 93.76 % and AUC value 3.25% From this research, it is proven that PSO can improve the accuracy and performance value of AUC

**Keywords:** Graduation prediction, support vector mashine algorithm and particle swarm optimization

**Abstrak** – Kelulusan mahasiswa merupakan salah satu bidang standar penjaminan mutu internal (SPMI) suatu perguruan tinggi. Jumlah mahasiswa yang lulus tepat waktu menjadi indikator keberhasilan dari sebuah perguruan tinggi baik negeri dan swasta. Perguruan tinggi mempunyai peranan yang sangat penting didalam menciptakan lulusan yang terbaik untuk kebutuhan dunia kerja. Penelitian dalam hal prediksi kelulusan mahasiswa telah banyak dilakukan. Berdasarkan kondisi tersebut maka penelitian ini dilakukan penerapan salah satu teknik klasifikasi data mining yaitu C4.5. Dalam penelitian ini membandingkan pemodelan algoritma C4.5 dengan C4.5 berbasis Particle Swarm Optimization (PSO) yang diterapkan pada data kelulusan mahasiswa. Dari hasil pengujian digunakan cross validation, confusion matrix dan kurva ROC, diketahui bahwa C4.5 memiliki nilai akurasi sebesar 78.16 % dan performa yang ditunjukkan nilai AUC adalah 5.62% sedangkan C4.5 berbasis PSO memiliki nilai akurasi sebesar 93.79% dan nilai AUC 3.25% dari penelitian ini, terbukti bahwa PSO dapat meningkatkan nilai akurasi dan performa AUC.

**Keywords:** Student graduation, data mining, data mining techniques, PSO-based C4.5 and C4.5 algorithms..

## I. PENDAHULUAN

Kelulusan mahasiswa merupakan salah satu bidang standar penjaminan mutu internal (SPMI) suatu perguruan tinggi. salah satu standar yang ditetapkan oleh program studi Teknik Informatika STMIK IKMI Cirebon, untuk lulusan teknik informatika yaitu maksimal delapan semester dengan IPK minimal 2,76. Lulus tepat waktu adalah keinginan seluruh mahasiswa. Tidak hanya itu, lulus tepat waktu adalah keuntungan bagi dua pihak, yaitu mahasiswa dan instansi pendidikan. Dalam mengukur tingkat produktifitas kelulusan dapat di peroleh dari jumlah lulusan dengan jumlah penerimaan, jika jumlah lulusan lebih kecil dari jumlah penerimaan maka jumlah penerimaan kecil. [1].

Perguruan tinggi merupakan satuan pendidikan yang menjadi terminal terakhir bagi seseorang yang berpeluang belajar setinggi-tingginya melalui jalur Pendidikan sekolah [2]. Pemanfaatan sumber daya yang dimiliki oleh sebuah perguruan tinggi, selain sumberdaya sarana,

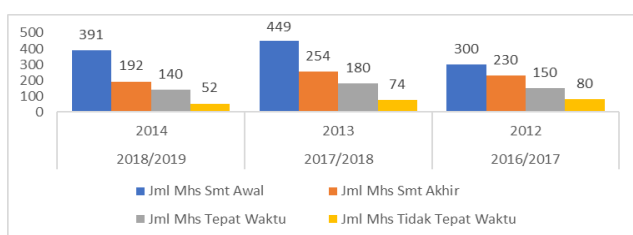
prasarana, dan manusia system informasi merupakan salah satu sumber daya yang dapat digunakan untuk meningkatkan keunggulan.

Organisasi pendidikan tinggi berada dalam lingkungan kompetitif yang sangat tinggi dan bertujuan untuk mendapatkan keuntungan yang lebih kompetitif melalui kompetisi bisnis lainnya. Organisasi-organisasi ini harus meningkatkan kualitas layanan mereka dan memuaskan pelanggan mereka. Mereka menganggap mahasiswa dan dosen sebagai aset utama mereka dan mereka ingin terus meningkatkan indikator indikator kunci Mereka dengan menggunakan aset secara efektif dan efisien [3]. Mahasiswa bagi sebuah institusi pendidikan merupakan aset penting untuk itu perlu diperhatikan tingkat kelulusannya agar lulus tepat waktu.

Wisuda, tepat waktu merupakan isu kebijakan yang semakin penting. Tingkat kelulusan dianggap sebagai salah satu efektivitas kelembagaan [2], untuk itu perlu diperhatikan

jumlah mahasiswa pada suatu perguruan tinggi. Tingkat penurunan kelulusan mahasiswa yang ada pada perguruan tinggi akan menjadi masalah dan akan berpengaruh terhadap akreditasi perguruan tinggi tersebut. Oleh karena itu perlu adanya pemantauan dan evaluasi terhadap kecenderungan mahasiswa lulus tidak tepat waktu.

STMik IKMI Cirebon merupakan Sekolah Tinggi swasta yang bergerak di bidang pendidikan informatika memiliki Visi menjadi sekolah tinggi unggulan dan berbeda di bidang Ilmu Komputer di wilayah Cirebon, Indramayu, Majalengka, dan Kuningan Tahun 2020. Dalam mewujudkan hal tersebut STMik IKMI Cirebon menerima calon mahasiswa baru sesuai daya tampung. Data mahasiswa diperoleh dari bagian penerimaan mahasiswa baru dan akademik, selama 3 (tiga) tahun terakhir yaitu mahasiswa tahun angkatan 2012 sampai dengan 2014, seperti terlihat dalam Gambar 1 sebagai berikut



Gambar 1.1 Penerimaan Mahasiswa Baru, Mahasiswa Semester Akhir, Mahasiswa tepat waktu dan tidak tepat waktu Prodi Teknik Informatika STMik IKMI Cirebon tahun angkatan 2012 sampai 2014

Tabel 1 Data Mahasiswa Prodi Teknik Informatika(S

No	Tahun Masuk	Mahasiswa Baru	Mahasiswa Semester 8	Lulus Tepat Waktu	Tidak Tepat	Tingkat produktivitas	Penurunan Tingkat Produktivitas
1	2012	300	230	150	80	77%	23%
2	2013	449	254	180	74	57%	43%
3	2014	391	192	140	52	49%	51%

Dari table 1 prodi Teknik Informatika tahun Angkatan 2012 sampai dengan 2014 tersebut diketahui bahwa pada tahun 2012 mahasiswa awal yang terdaftar sejumlah 300 orang, sedangkan jumlah mahasiswa di semester akhir (8) sejumlah 230 mahasiswa, ada penurunan tingkat produktivitas

sebanyak 23% dari jumlah mahasiswa awal sampai dengan mahasiswa semester akhir (8), sedangkan 150 mahasiswa lulus tepat waktu (65%) dan 80 mahasiswa tidak tepat waktu (35%). Pada tahun 2013 mahasiswa awal terdaftar 449 mahasiswa, sedangkan 254 di semester akhir, ada penurunan tingkat produktivitas sebanyak 57% dari jumlah mahasiswa awal sampai mahasiswa semester akhir (8), 180 lulus tepat waktu (71%) dan 74 mahasiswa tidak tepat waktu (29%). dan pada tahun 2014 mahasiswa awal terdaftar 391 mahasiswa, 192 di semester akhir, ada penurunan tingkat produktivitas sebanyak 49% dari jumlah mahasiswa awal sampai mahasiswa semester akhir (8). 140 lulus tepat waktu (73%) dan 52 mahasiswa tidak tepat waktu (27%). Dengan demikian jumlah mahasiswa lulus tidak tepat waktu jika di rata-rata sebanyak 69 mahasiswa lulus tidak tepat waktu. Tingginya tidak tepat waktu kelulusan mahasiswa tersebut disebabkan terjadinya penurunan produktivitas mahasiswa dan penurunan indeks prestasi mahasiswa. Akibat dari permasalahan diatas dalam jangka Panjang dapat menyebabkan penumpukan mahasiswa yang tinggi di setiap periode [1] sehingga sangat berpengaruh terhadap penilaian akreditasi yang ditetapkan oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) adalah prosentase mahasiswa yang lulus tepat waktu 50%[4]. Salah satu solusi permasalahan tersebut, STMik IKMI Cirebon harus dapat meningkatkan kualitas Pendidikan yaitu dengan cara mengantisipasi keterlambatan kelulusan mahasiswa.

Dewasa ini pendekatan data mining berkembang untuk mengatasi berbagai permasalahan menyangkut tentang pengolahan data. Beberapa peneliti menggunakan Teknik data mining untuk menyelesaikan permasalahan prediksi kelulusan mahasiswa (hilda, 2012), (suharniti & Ernastuti, 2010).

Data mining adalah suatu cara yang bertujuan dalam penemuan pola secara otomatis atau semi otomatis dari data lain yang dimanfaatkan untuk menyelesaikan suatu masalah melalui berbagai atura proses (witten, I.H, 2011). Data mining memiliki beberapa teknik, diantaranya klasifikasi dan clustering. Teknik klasifikasi adalah teknik pembelajaran yang digunakan untuk memprediksi nilai dari atribut kategori target(Vercellis, 2009). Klasifikasi tujuan untuk membagi objek yang ditugaskan ke salah satu nomor kategori yang disebut kelas(Max Bramer, 2007). Clustering merupakan pengelompokan objek atau data berdasarkan kemiripan antar data, sehingga anggota dalam satu kelompok memiliki banyak kemiripan dibandingkan dengan kelompok lain(Gorunescu, 2011). Untuk memprediksi kelulusan mahasiswa, maka hasil pengolahan data akan diklasifikasikan menjadi dua kelas, tepat dan terlambat. Metode yang paling populer yang biasa digunakan untuk teknik klasifikasi adalah Decision trees, Naive Bayes Classifier(NBC), Statistis Analysis, dan lain-lain(Gorunescu, 2011).

Beberapa penelitian terdahulu yang telah dilakukan mengenai analisis prediksi kelulusan mahasiswa dengan metode klasifikasi data mining, diantaranya adalah penelitian yang dilakukan oleh Hilda Amalia pada tahun 2012 dengan tema mengkomparasi algoritma C4.5, Naive Bayes, dan Neural Network. Dalam penelitian tersebut diketahui bahwa algoritma nilai akurasi dan

AUC tertinggi adalah neural network dan ternyata algoritma C4.5 masih tergolong fair classification.

Decision tree sering digunakan dalam klasifikasi karena memiliki hasil yang cukup baik jika dibandingkan dengan algoritma yang lainnya. C4.5 suatu model pembelajaran dari data yang tergolong cepat, selain itu model yang digambarkan dalam diagram pohon dan mudah dipahami. Namun jika ada data yang tidak relevan dapat menurunkan akurasi C4.5 (Tai & Chen, 2009). Di C4.5 seluruh atribut diseleksi untuk kemudian dibagi menjadi himpunan yang lebih kecil (Wu, 2009). Data yang akan diolah sebaiknya dilakukan proses pre processing data.

Beberapa kelebihan dari pohon keputusan (Gorunescu, 2011) adalah sebagai berikut :

- a. Hasil dari analisa berupa diagram pohon keputusan yang sangat mudah dimengerti
- b. Mudah untuk dibangun, serta membutuhkan data percobaan yang lebih sedikit dibandingkan algoritma klasifikasi lainnya.
- c. Mampu mengolah data nominal dan kontinyu.
- d. Model yang dihasilkan dapat dengan mudah dimengerti, berbeda dengan teknik klasifikasi yang lain seperti neural network yang menyajikan model dengan informasi logis yang tersirat.
- e. Menggunakan metode statistik sehingga dapat divalidasi.
- f. Waktu komputasi relatif lebih cepat dibandingkan dengan teknik klasifikasi yang lain.
- g. Akurasi yang dihasilkan mampu menandingi teknik klasifikasi yang lainnya.

*Particle swarm optimization* (PSO) adalah algoritma pencarian stokastik berbasis populasi yang dimodelkan setelah perilaku social kawanan burung dalam konteks *particle swarm optimization* (PSO), segerombolan mengacu pada sejumlah solusi potensial untuk masalah optimasi, di mana masing-masing solusi potensial disebut sebagai partikel. Tujuan *particle swarm optimization* (PSO) adalah untuk menemukan posisi partikel yang menghasilkan evaluasi terbaik dari fungsi kebugaran (objektif) yang diberikan [5][6]. Particle swarm optimization memiliki perbandingan lebih untuk pemilihan fitur dan memiliki kinerja yang lebih unggul untuk banyak masalah optimasi dengan lebih cepat dan tingkat konvergensi yang lebih stabil (Ling, Nguyen, & Chan, 2009). Karakteristik particle swarm optimization adalah interaksi social yang mempromosikan pembagian informasi antara partikel yang akan membantu dalam pencarian solusi yang optimal (Park, Lee, & Choi, 2009). Parameter yang terdapat pada particle swarm optimization diantaranya posisi, kecepatan, kecepatan maksimum, percepatan konstanta dan berat inersia. Ada beberapa cara atau teknik untuk pengoptimalan diantaranya : meningkatkan bobot atribut weight terhadap semua atribut variable yang dipakai, attribute selection, dan feature selection./

Untuk melakukan prediksi terhadap informasi yang belum diketahui sebelumnya maka digunakan teknik klasifikasi. Metode klasifikasi adalah merupakan pendekatan untuk menjalankan klasifikasi dalam data mining.

Prediksi merupakan pemikiran terhadap satu atau beberapa produk untuk memperoleh suatu pengetahuan yang secara sistematis berdasarkan bukti fisik [1], [7].

Beberapa algoritma yang dapat digunakan diantaranya adalah algoritma *Decision Tree C4.5*, *Artificial Neural Network* (ANN), Algoritma Naïve Bayes, K-NN (*K-Nearest Neighbors*), Algoritma genetik, dan *Support Vector Machine* [8].

*Artificial neural network* (ANN) adalah salah satu metode machine learning yang sering digunakan untuk pemodelan atau prediksi. Pemodelan yang sering menggunakan ANN adalah pemodelan data kualitatif atau analisis data biner. ANN ini mampu memecahkan kasus nonlinier pada klasifikasi dan dapat memberikan akurasi yang tinggi [9]. Pada metode *Artificial neural network* mempunyai kekurangan dalam menentukan nilai parameter yang digunakan sehingga penggunaan dataset dapat mempengaruhi tingkat akurasi yang dihasilkan [10].

Pada penelitian ini penerapan algoritma optimasi Particle Swarm Optimization digunakan untuk memaksimalkan kinerja dari model algoritma C4.5 dalam memprediksi kelulusan mahasiswa sehingga hasil prediksi kelulusan mahasiswa lebih tinggi.

## II. PENELITIAN YANG TERKAIT

Penelitian yang dilakukan oleh Mongan Winny Amelia, Dkk dengan topik "Prediksi masa studi mahasiswa dengan menggunakan algoritma *Naïve Bayes*". Penelitian tersebut membahas tentang penerapan metode algoritma *Naïve Bayes* dalam memprediksi masa studi mahasiswa, dengan menggunakan data mining untuk klasifikasinya, untuk prediksi dengan algoritma *Naïve Bayes* dapat dilakukan prediksi terhadap ketepatan waktu studi dari mahasiswa berdasarkan data training. Pengujian yang digunakan yaitu *k-fold cross validation 10-fold*. Hasil pengujian tersebut didapat nilai akurasi rata-rata sebesar 85.17% sedangkan nilai akurasi sebesar 88.96% [11].

Supardi Salmu, Achmad Solichin, dalam penelitiannya yang berjudul "Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan *Naïve Bayes*". Dalam penelitian tersebut, peneliti melakukan penelitian menggunakan algoritma *Naïve Bayes* dengan metode probalistik dan statistik dengan memprediksi peluang yang akan datang dimasa depan. Dengan menggunakan algoritma *Naïve Bayes* didapatkan prediksi yang akurat dengan tingkat akurasi sebesar 80.72% pada tingkat prediksi kelulusannya.[1]

Agus Romadhona, supraedi, H.Himawan, dalam penelitiannya yang berjudul "Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Usia, Jenis Kelamin, dan Indeks Prestasi Menggunakan Algoritma *Decision Tree*". Dalam penelitian tersebut, peneliti melakukan penelitian menggunakan algoritma *Decision Tree 4.5*. Dari penelitian tersebut ditemukan bahwa prediksi masa studi siswa dapat dipengaruhi oleh mahasiswa yang masuk usia, jenis kelamin,

IPK semester 1 sampai 4 mahasiswa pascasarjana tepat waktu dengan nilai gain 0,340 dari semua atribut. Dengan algoritma *Decision Tree C4.5* mencapai akurasi tertinggi pada jumlah data 389 dengan nilai akurasi 91,51% untuk nilai *k-fold*= 3, dan 90,75 untuk *k-fold*=5, dan 90.77 untuk *k-fold*=10.[12]

Eko Prasetyo Rohmawan, dalam penelitiannya yang berjudul “Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode *Decision Tree* Dan *Artificial Neural Network*”. Metode yang digunakan dalam penelitian *Decision Tree* dan *Artificial Neural Network*. Pada penelitian tersebut menggunakan data mahasiswa yang sudah dinyatakan lulus sebagai data training dan mahasiswa yang belum lulus sebagai data testing dengan hasil prediksi menggunakan *Artificial Neural Network* menunjukkan dengan nilai 79,74% dengan rincian prediksi telat sebesar 91,11% dan prediksi tepat waktu sebesar 63,49%. Dari kedua metode tersebut dapat dihasilkan perbandingan antara Metode *Decision Tree* dan *Artificial Neural Network* dengan hasil yang tertinggi adalah 79,74% dan *Decision Tree* 74,51% [8].

Ryan Dwi Pambudi, Ahmad Afif Supianto, Nanang Yudi Setiawan dalam penelitian yang berjudul “Prediksi Kelulusan Mahasiswa Berdasarkan Kinerja Akademik Menggunakan Pendekatan Data Mining pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya”. Metode penelitian yang digunakannya adalah metode algoritma klasifikasi *Naïve Bayes*. Dengan memanfaatkan metode algoritma klasifikasi *Naïve Bayes* akan dihasilkan pola – pola berdasarkan probabilitas pada setiap atribut mahasiswa yang terkumpul berjumlah 1354, data kemudian dilakukan pre-processing untuk proses *mining* pada sistem yang dikembangkan. Hasil pengujian usability menggunakan *System Usability Scale* (SUS) yang dilakukan kepada Kaprodi SI menghasilkan nilai 57.5. berdasarkan nilai tersebut pengembang sistem dalam penelitian ini termasuk kedalam kategori *adjective rating Good* namun tingkat *acceptability* yang masih tergolong *Low* [13]

Penelitian yang dilakukan oleh Karamouiz dan Vretoz pada tahun 2009 dengan judul *Sentivity Analysis of Neural Network for Identifying the Factors for Collage Students Success*. Permasalahan pada penelitian ini adalah tingkat kelulusan dianggap sebagai indikator efektivitas suatu lembaga institusi, metode yang digunakan adalah NN (*Neural Network*). Dari hasil data training yang dilakukan diperoleh presentasi hasil yang benar untuk kategori yang lulus adalah 86.04% dan training data untuk kategori yang tidak sukses yakni 68.21%, dan error yang diperoleh untuk kedua kategori tersebut adalah 0.18% (Karamouiz & Vrentoz, 2009). Penelitian yang dilakukan oleh Qudri dan Kalanyar pada tahun 2010 dengan judul *Drop Out Feature of Student Data for Academic Performance Using Decision Tree techniques*. Masalah dalam penelitian ini adalah prestasi akademik siswa sangat penting bagi lembaga pendidikan karena program-program strategis dapat direncanakan untuk meningkatkan atau mempertahankan siswa kinerja selama periode mereka studi di lembaga. Kinerja akademik dalam penelitian ini

diukur oleh indek kumulatif rata-rata setelah lulus. Hal ini penting untuk memahami faktor-faktor penentu tingkat penyelesaian yang sukses dan tepat waktu. Metode yang digunakan adalah *Decision Tree*, yakni algoritma J4.8. Hasil penelitian ini adalah sebuah pohon keputusan yang dapat dijadikan rule bagi prediksi siswa yang putus sekolah[3].

Penelitian yang dilakukan oleh Suhartina dan Ernastuti pada tahun 2010 dengan judul *Graduation Prediction of Gunadarma University Students Using Algorithm and Naive Bayes C4.5 Algorithm*. Masalah dalam penelitian ini adalah banyaknya mahasiswa yang tidak lulus sesuai dengan waktu studi yang telah ditetapkan. Algoritma yang digunakan adalah C45 dan naïve bayes.

Hasil dari penelitian ini adalah akurasi dengan dua metode tersebut yakni akurasi untuk metode naïve bayes adalah 80,85% dengan presentasi kesalahan 19,05% Akurasi ketepatan hasil prediksi C4.5 85.7%, dan presentasi kesalahannya adalah 14,3% [14]

### III. METODE PENELITIAN

Langkah-langkah penelitian ini sesuai alur pada model CRISP-DM (*Cross Industry Standard Process for Data Mining*) sebagai berikut :

#### A. Pemahaman Bisnis(Bussines Understanding)

Institusi Perguruan tinggi dituntut untuk memiliki keunggulan dalam berdaya saing yaitu dengan meningkatkan mutu dan pencapaian akreditasi yang baik, diantaranya dipengaruhi oleh jumlah kelulusan mahasiswa pertahun. Mahasiswa yang lulus tidak tepat waktu menjadi catatan penting bagi institusi ketika ingin mengajukan akreditasi.

#### B. Pemahaman data(Data Understanding)

Data yang digunakan dalam penelitian ini bersumber dari data lulusan yang berjumlah 445 record mahasiswa yang terdiri dari 14 atribut 10 atribut bertipe numerik 3 atribut bertipe kategorikal dan 1 atribut bertipe label.

Tabel 3.1 Kedudukan atribut yang akan digunakan

Atribut	Kegunaan	
NIM	√	ID
NAMA	X	No
KELAS	√	regular(nilai model)
JK	√	regular(nilai model)
UMUR	√	regular(nilai model)
IPS1	√	regular(nilai model)
IPS2	√	regular(nilai model)
IPS3	√	regular(nilai model)

		model)
IPS4	√	regular(nilai model)
IPS5	X	No
IPS5	X	No
IPS7	X	No
IPS8	X	No
STATUS	√	Label(hasil)

C. Pengolahan Data(Data Preparation)

Seluruh data yang berhasil dikumpulkan sebanyak 445 data. Data terlebih dahulu dilakukan preprocessing data. Ditahapan ini data dibersihkan.Atribut-atribut yang sekiranya tidak penting untuk digunakan dalam proses analisis, (contoh : atribut nama mahasiswa), maka atribut tersebut akan dihilangkan. Atau jika ada data mahasiswa yang kurang lengkap, maka data tersebut akan dihapus/tidak digunakan. Hasil dari pre processing data ini, jumlah data set nya menjadi 338 data.

Pada tahap ini adalah pemilihan atribut data yang akan digunakan untuk pemodelan.

D. Pemodelan(Modelling)

Metode yang digunakan adalah algoritma C4.5 dan C4.5 berbasis PSO untuk membentuk decision treenya menggunakan Rapidminer. Algoritma C4.5 dan C4.5 berbasis PSO akan menghasilkan sebuah pohon keputusan. Cara menentukan atribut mana yang menjadi akar, dilihat dari nilai gain tertinggi dari atribut-atribut yang ada.

E. Evaluasi

Dalam tahap ini akan dilakukan pengukuran keakuratan hasil yang dicapai. Akurasi klasifikasi didapatkan berdasarkan tabel confusion matrix. Confusion matrix dari data testing yang digunakan dengan output Tepat waktu dan Tidak Tepat Waktu dapat dilihat pada Gambar ... Confusion matrix merupakan data set yang hanya memiliki dua kelas, kelas yang satu bersifat positif dan yang satu lagi bersifat negatif. Terdiri dari empat sel yaitu True Positif (TP), False Positif(FP), True Negatif(TN) dan False Negatif(FN) (Max Bramer, 2007).

CLASSIFICATION	PREDIKTED CLASS		
	Class=YE S	Class=N O	
OBSERVED CLASS	Class=YE S	a (true positive-TP)	b (false negative-FN)
	Class=N O	c (false positive-FP)	d (true negative-TN)

Gambar 3.1 Confusion matrix untuk 2 model kelas(Gorunescu, 2011)

Untuk menghitung akurasi menggunakan rumus (Gorunescu, 2011) :

$$Accuracy = \frac{a + d}{a + b + c + d} = \frac{TP + TN}{TP + FP + TN + FN}$$

Kurva ROC

Kurva ROC menunjukkan validasi dari akurasi model dan perbandingan perbedaan antar model klasifikasi. ROC mengekspresikan *confusion matrix* (Vercellis, 2009). ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horizontal dan *true positive* untuk mengukur perbedaan performance metode yang digunakan. Kurva ROC adalah teknik untuk memvisualisasikan dan menguji kinerja pengklasifikasian berdasarkan performanya(Gonrunescu, 2011). Model klasifikasi yang lebih baik adalah yang mempunyai kurva ROC lebih besar(Vercellis, 2009). Performa keakuratan AUC dapat diklasifikasikan menjadi lima kelompok yaitu(Gonrunescu, 2011):

F. Penyebaran(Deployment)

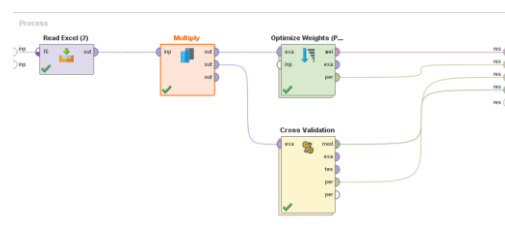
Hasil dari penelitian ini adalah berupa Analisa yang mengarah kepada pengambilan keputusan. Hasil dari Analisa diharapkan dapat dimanfaatkan oleh institusi perguruan tinggi sebagai bahan pertimbangan dalam menentukan langkah guna mengatasi permasalahan mahasiswa yang lulus tidak tepat waktu dimasa yang akan datang.

Selanjutnya dari data tersebut diatas maka langkah selanjutnya lakukan *replace missing value* berdasarkan pada *average* untuk menggantikan nilai yang kosong dengan nilai rata-rata yang akan muncul pada data training.

IV. HASIL DAN PEMBAHASAN

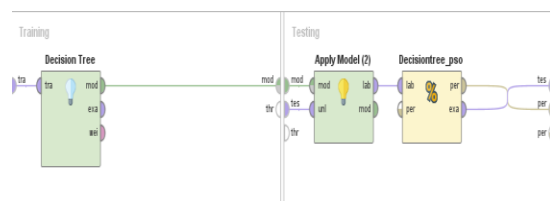
Pembahasan terhadap hasil penelitian dan pengujian yang diperoleh disajikan dalam bentuk uraian sebagai berikut :

Pada tahap awal peneliti melakukan eksperimen dengan menguji dataset pada aplikasi rapidminer menggunakan algoritma C4.5 dan C4.5 berbasis PSO, hasilnya terlihat seperti padagambar dari proses yang terdapat dalam frame main proses.



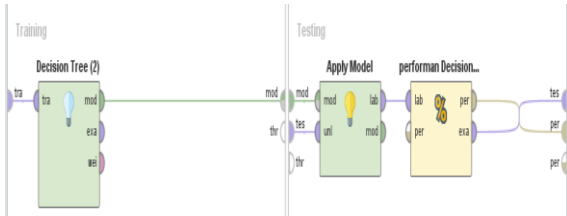
Sumber hasil penelitian 2019  
Gambar 4.1 Proses:

Didalam validasi terbagi menjadi dua frame yaitu Training dan Testing. Pada frame training peneliti menggunakan algoritma Decision tree C4.5 berbasis PSO yang kemudian dihungkan dengan application model dan performance pada frame testing. Gambar 2. merupakan gambar pengujian algoritma Decision tree C4.5 berbasis PSO menggunakan rapidminer pada frame training dan testing

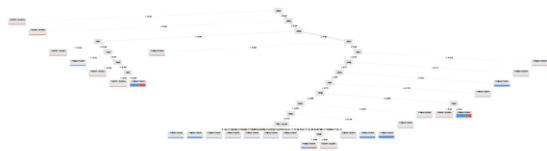


$$\text{Accurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{248 + 69}{248 + 69 + 11 + 10} = 93.79\%$$

Sumber hasil penelitian 2019  
Gambar 4.2 Pengujian Algoritma Decision Tree C4.5 berbasis PSO



Sumber hasil penelitian 2019  
Gambar 4.3 Pengujian Algoritma Decision Tree C4.5



Sumber hasil penelitian 2019  
Gambar 4.3. Pohon Keputusan Menggunakan Rapidminer

Hasil Pengujian metode C4.5 berbasis PSO

A. Confusion matrix

Pada Tabel 4.1 menunjukkan hasil dari confusion matrix metode C4.5 berbasis PSO.

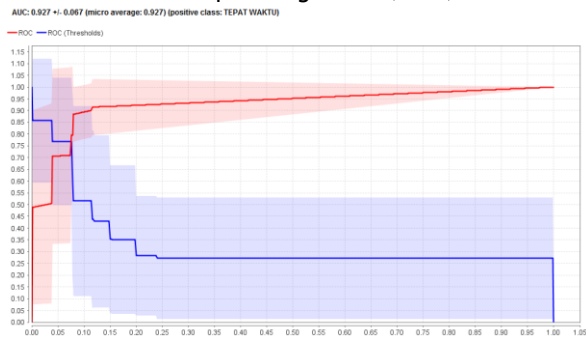
accuracy: 93.79% +/- 3.25% (micro average: 93.7)

	true TIDAK TEPAT	true TEPAT WAKTU	class precision
pred. TIDAK TEPAT	248	10	96.12%
pred. TEPAT WAKTU	11	69	86.25%
class recall	95.75%	87.34%	

Sumber hasil penelitian 2019  
Tabel 4.1 Hasil Confusion matrix metode C4.5 berbasis PSO

Keterangan : Berdasarkan hasil dari confusion matrix untuk metode C4.5 berbasis PSO dapat dilihat dari tabel 1 diatas tingkat akurasi yang dihasilkan menggunakan algoritma C4.5 berbasis PSO adalah sebesar 93.79% dan dapat dihitung untuk mencari nilai accuracy, sensitivity, specificity, PPV dan NPV pada persamaai dibawah ini :

B. Receiver Operating Curve(ROC)



Sumber hasil penelitian 2019  
Gambar 4.4 Kurva ROCmetode C4.5 berbasis PSO

Hasil pengujian metode C4.5

A. Confusion Matrix

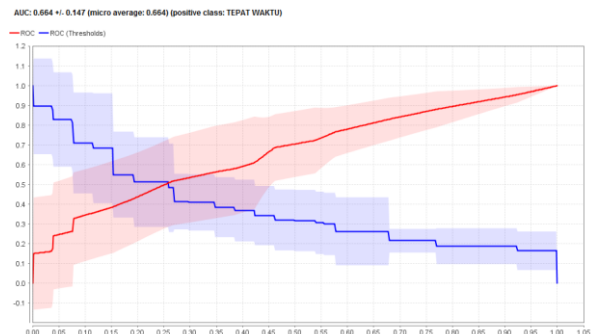
Pada tabel 4.2 menunjukkan hasil confusion matrix dengan menggunakan metode C4.5

accuracy: 78.16% +/- 5.62% (micro average: 78.16%)

	true TIDAK TEPAT	true TEPAT WAKTU	class precision
pred. TIDAK TEPAT	245	60	80.33%
pred. TEPAT WAKTU	14	19	57.58%
class recall	94.59%	24.05%	

Sumber hasil penelitian 2019  
Tabel 4.2 Hasil Confusion matrix metode C4.5

B. Receiver Operating Curve(ROC)



Sumber hasil penelitian 2019  
Gambar 4.5 Kurva ROCmetode C4.5



Tabel 4. 3 hasil pengujian Data Mahasiswa

Confidence	minimal Gain	Pop size	C4.5		C4.5 +PSO [2]	
			Acc%	AUC	Acc%	AUC
0.5	0.05	5	81.32%	7.24%	84.31%	7.75%
0.5	0.06	5	81.91%	6.76%	83.74%	5.94%
0.25	0.05	5	81.61%	6.94%	84.61%	8.21%
0.25	0.05	3	79.33%	4.24%	92.91%	3.46%
0.5	0.08	5	80.45%	5.99%	83.74%	5.39%
0.25	0.09	3	80.21%	3.79%	83.44%	8.44%
<b>0.25</b>	<b>0.025</b>	<b>3</b>	<b>78.16%</b>	<b>5.62%</b>	<b>93.79%</b>	<b>3.25%</b>
0.25	0.025	5	81.36%	6.82%	92.90%	6.10%

Dari hasil pengujian / eksperimen yang telah dilakukan terhadap data mahasiswa dapat disimpulkan bahwa data dengan nilai Confidence = 0.25 , Minimal Gain = 0.025 dan Populasi Size = 3 memiliki akurasi tertinggi sebesar 93.79% dan AUC sebesar 3.25%

#### V. KESIMPULAN

##### Kesimpulan

Dari penelitian yang telah dilakukan bahwa optimasi terhadap metode C4.5 dalam memprediksi kelulusan mahasiswa dapat meningkatkan nilai akurasi sebesar 93.79 % dan nilai AUC 3.325 sedangkan hasil nilai akurasi metode C4.5 adalah 78.16 % dan AUC 5.62%

##### Saran

Agar penelitian ini dapat ditingkatkan, berikut adalah saran-saran yang diusulkan:

1. Beberapa atribut, nim, nama, jenis kelamin, ipk dan status kelulusan, sebaiknya gunakan metode penyeleksian untuk menyeleksi data.
2. Membandingkan kelulusan sebelumnya dan akan datang
3. Untuk penelitian lainnya hasil penelitian ini dapat dikembangkan dengan metode klasifikasi data mining lainnya seperti, naïve bayes, svm, dan lainnya untuk pengembangan.
4. Memperbanyak jumlah atribut dan dataset

#### UCAPAN TERIMA KASIH

Terima kasih kami sampaikan kepada STMIK IKMI Cirebon yang telah membantu baik dalam hal sarana dan prasarana dalam penelitian ini.

#### DAFTAR PUSTAKA

[1] S. Salmu and A. Solichin, "Prediksi Tingkat Kelulusan Tepat Waktu Menggunakan Naïve Bayes : Studi Kasus UIN SyMahasiswaarif Hidayatullah Jakarta Prediction of Timeliness Graduation of Students Using Naïve Bayes: A Case Study at Islamic State University Syarif Hidayatullah Jakarta," *Pros. Semin. Nas. Multidisiplin Ilmu*, no. April, pp.

701–709, 2017.

S. A. Prakoso and E. T. Tutik, "Komparasi algoritma C4.5 dengan naive bayes untuk klasifikasi kelulusan mahasiswa tepat waktu di PTS 'KZX,'" *Komputaki*, vol. 3, no. 1, pp. 103–132, 2017.

M. Quadri and D. Kalyankar, "Drop out feature of student data for academic performance using decision tree techniques," *Glob. J. Comput.*, vol. 10, no. 2, pp. 2–5, 2010.

Badan Akreditasi Nasional, "Akreditasi Perguruan Tinggi," 2018.

P. Studi and M. Informatika, "PENERAPAN PARTICLE SWARM OPTIMIZATION ( PSO ) UNTUK SELEKSI ATRIBUT DALAM MENINGKATKAN AKURASI PREDIKSI DIAGNOSIS PENYAKIT HEPATITIS DENGAN METODE ALGORITMA C4 . 5 Lis Saumi Ramdhani b . Knowledge Discovery in Database ( KDD )," vol. IV, no. 1, pp. 1–15, 2016.

[6] J. Techno, N. Mandiri, S. Nurajizah, M. Informatika, and J. Selatan, "Penerapan Metode Suppor Vector Machine Berbasis Particle," *J. Techno Nusa Mandiri*, vol. d, no. 1, pp. 216–226, 2013.

[7] R. Naufal, "Sistem Informasi Inventory Berdasarkan Prediksi Data Penjualan Barang Menggunakan Metode Single Moving Average Pada CV . Agung Youanda," *Protekinfo*, vol. 4, pp. 29–33, 2017.

[8] E. P. Rohmawan, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree Dan Artificial Neural Network," *J. Ilm. Matrik*, vol. 20, no. 1, pp. 21–30, 2018.

[9] J. Jiang, J. Zhang, G. Yang, D. Zhang, and L. Zhang, "Application of back propagation neural network in the classification of high resolution remote sensing image: Take remote sensing image of Beijing for instance," *2010 18th Int. Conf. Geoinformatics, Geoinformatics 2010*, no. 20091108110003, 2010.

[10] O. Somantri and C. Supriyanto, "Algoritme Genetika untuk Peningkatan Prediksi Kebutuhan Permintaan Energi Listrik," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 5, no. 2, pp. 108–114, 2016.

[11] M. W. Amelia, A. S. M. Lumenta, and A. Jacobus, "Prediksi masa studi mahasiswa dengan menggunakan algoritma naive bayes," *E-Journal Tek. Inform.*, vol. 11, no. 1, 2017.

[12] H. Romadhona, Agus; suprapedi; himawan, "Prediksi Kelulusan Tepat Waktu Mahasiswa Stmik-Ymi," vol. 13, p. 917, 2017.

[13] R. D. Pambudi, A. A. Supianto, and N. Y. Setiawan, "Prediksi Kelulusan Mahasiswa Berdasarkan Kinerja Akademik Menggunakan Pendekatan Data Mining Pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2194–2200, 2019.

- [14] E. Marselina Silvia Suhartinah, "Graduation Prediction Of Gunadarma University Students Using Algorithm And Naive Bayes C4.5 Algorithm," *Fac. Ind. Technol. Gunadarma Univ.*, 2010.