

Perbandingan 4 Algoritma Berbasis Particle Swarm Optimization (ps) Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa

Moh. Zainuddin

Program Studi Informatika, STMIK ASIA, Malang
mzein.asia@gmail.com

ABSTRAK. Tujuan dari penelitian ini adalah untuk mencari algoritma terbaik dalam membuat prediksi kelulusan siswa dari 4 algoritma: *Algoritma Naive Bayes*, *Decision Tree (C4.5)*, *k-Nearest Neighbor (k-NN)*, *Neural Network* berbasis *Particle Swarm Optimization (PSO)* sebagai referensi untuk membuat kebijakan dan tindakan bidang akademik (BAAK) dalam mengurangi mahasiswa yang lulus terlambat dan tidak lulus. Hasil menunjukkan *Algoritma k-Nearest Neighbor (k-NN) berbasis PSO* pada k -optimum=19 mempunyai performa terbaik dari 4 algoritma yang ada, dengan nilai **Accuracy = 74,08%** dan nilai **Area Under the Curve (AUC) = 0,788**. Penambahan fitur *Particle Swarm Optimization (PSO)* selalu meningkatkan nilai akurasi, dimana peningkatan nilai akurasi tertinggi terletak pada **Algoritma Decision Tree (C4.5)** sebesar **5,21%**, terendah pada **Algoritma Naive Bayes** sebesar **2,13%**.

Kata kunci: *Algoritma Naive Bayes*, *Decision Tree (C4.5)*, *k-Nearest Neighbor (k-NN)*, *Neural Network*, *Particle Swarm Optimization (PSO)*, *Accuracy*, *Area Under the Curve (AUC)*.

ABSTRACT. The purpose of this study was to find the best algorithm in making predictions of students' graduation from 4 algorithms: *Naive Bayes Algorithm*, *Decision Tree (C4.5)*, *k-Nearest Neighbor (kNN)*, *Neural Network based Particle Swarm Optimization (PSO)* as references to make policies and academic acts (BAAK) in reducing students who graduated late and did not pass. The results show that *PSO-k-Nearest Neighbor (k-NN) algorithm based on k-optimum = 19* has the best performance of 4 algorithms, with **Accuracy = 74,08%** and **Area Under the Curve (AUC) = 0,788**. The addition of the *Particle Swarm Optimization (PSO)* feature always increases the accuracy value, where the highest accuracy value lies in the *Decision Tree Algorithm (C4.5)* of 5.21%, the lowest on the *Naive Bayes Algorithm* of 2.13%.

Keywords: *Naive Bayes Algorithm*, *Decision Tree (C4.5)*, *k-Nearest Neighbor (k-NN)*, *Neural Network*, *Particle Swarm Optimization (PSO)*, *Accuracy*, *Area Under the Curve (AUC)*.

1. PENDAHULUAN

Pendidikan adalah suatu aktivitas sosial yang memungkinkan masyarakat tetap ada dan berkembang (Prabowo, 2012). Jenjang pendidikan perguruan tinggi menjadi salah satu persyaratan dasar dalam mencari pekerjaan, dimana perguruan tinggi akan mempersiapkan calon-calon sarjana yang berkualitas dan mempunyai keterampilan dibidangnya. Pencapaian gelar kesarjana tersebut membutuhkan waktu normal selama 3,5 sampai 4,5 tahun, akan tetapi dalam praktiknya banyak mahasiswa tidak selalu dapat menuntaskan studinya selama waktu normal yang telah ditentukan. Banyak faktor yang menyebabkan ketidaktepatan waktu kelulusan mahasiswa tersebut, faktor-faktor tersebut dapat bersumber dari faktor internal dan faktor eksternal.

Kelulusan tepat waktu bagi mahasiswa menjadi salah satu butir akreditasi pada sebuah perguruan tinggi. Adanya informasi kelulusan tepat waktu tentu akan menjadikan suatu pengambilan keputusan yang tepat bagi manajemen Perguruan Tinggi dalam mengambil langkah berikutnya. Selama ini STMIK ASIA MALANG belum memiliki pola-pola prediksi kelulusan tepat waktu sebagai acuan untuk memprediksi jumlah lulus tepat waktu. Prediksi kelulusan tepat waktu yang dilakukan saat ini hanya berdasarkan ramalan dari data IPK (Indeks Prestasi Kumulatif) dan IMK (Indeks Mutualisme Kumulatif) semester sebelumnya. Prediksi hampir sama dengan klasifikasi dan estimasi, hanya saja prediksi digunakan untuk menduga nilai-nilai tertentu yang akan terjadi dimasa mendatang (Prabowo, 2012)

Sementara itu STMIK ASIA MALANG mempunyai data SIS (*Student Information Services*) yang selama ini belum dimanfaatkan secara maksimal adalah hal yang sangat disayangkan jika *dataset* yang begitu besar tidak dimanfaatkan untuk digali informasi apa yang terdapat di dalamnya. Selain itu, selama ini ada anggapan dari Pembantu Ketua 1 (Bidang Akademik) STMIK ASIA MALANG bahwa untuk memprediksi tingkat kelulusan tepat waktu cukup dengan melihat data IPK dan IMK sebelumnya. Permasalahan yang lebih penting atau segera diantisipasi adalah adanya suara dari para mahasiswa yang mengatakan kepada adik kelas di sekolah masing-masing bahwa jangankan kuliah di STMIK ASIA Malang karena kelulusannya sangat sulit. Ini sangat

membahayakan bagi Perguruan Tinggi, karena bisa berakibat banyak calon mahasiswa yang tidak mau kuliah di STMIK ASIA Malang, berakibat fatal hilangnya pendapatan bagi lembaga. Berangkat dari permasalahan di atas maka dilakukanlah penelitian ini yaitu untuk melakukan *data mining* terhadap *dataset* SIS (*Student Information Services*) sehingga didapatkan informasi mengenai kelulusan tepat waktu dari mahasiswa STMIK ASIA MALANG.

STMIK ASIA MALANG merupakan salah satu organisasi pendidikan yang bergerak dibidang multimedia dan teknologi. Informasi tingkat kelulusan dari mahasiswa sangat penting untuk meningkatkan pelayanan yang dapat membuat mahasiswa nyaman dan bisa lulus tepat waktu. Penggunaan data mining dapat digunakan sebagai pertimbangan dalam mengambil keputusan lebih lanjut tentang faktor yang mempengaruhi kelulusan khususnya faktor dalam data induk mahasiswa (Nuqson, 2010).

2. METODE PENELITIAN

2.1 Penelitian Terkait

Tabel 1 di bawah ini menjelaskan secara ringkas dari beberapa penelitian yang ada sebelumnya:

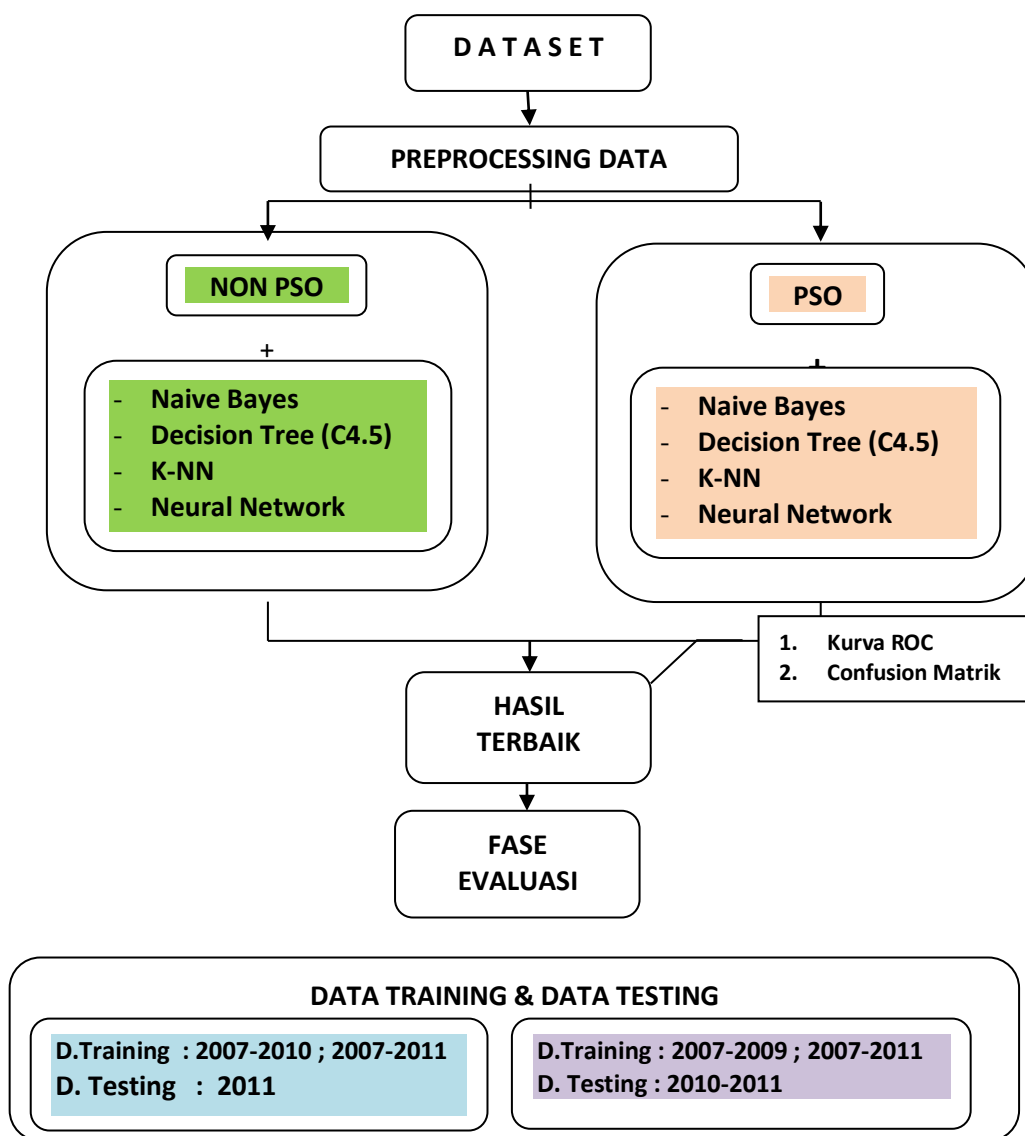
Tabel 1 Perbandingan Hasil Penelitian

N O	PENELITI	BIDANG KAJIAN	METODE & JUMLAH ATRIBUT	HASIL
1.	1. Raheela Asif, 2. Agathe Merceron, 3. Mahmood K. Pathan	Prediksi Kinerja Bidang Akademik	- Algoritma: 1. <i>Decision Tree</i> 2. <i>Naive Bayes</i> 3. <i>Neural Network</i> - Jumlah Atribut:16	Akurasi terbaik sebesar 83,65% pada Dataset II diperoleh dengan menggunakan <i>Algoritma Naive Bayes</i> .
2.	Ade Ricky Rozzaqi	Prediksi Ketepatan Kelulusan Mahasiswa	- Algoritma : 1. <i>Naive Bayes</i> 2. <i>Algoritma Information Gain dan Naive Bayes</i> - Jumlah Atribut:15	Nilai akurasi tertinggi dengan metode yang menggabungkan antara <i>Algoritma Naive Bayes</i> dan <i>Fitur Selection Information Gain</i> .
3	1. David Hartanto Kamagi 2. Seng Hansun	Prediksi Tingkat Kelulusan Mahasiswa	- Algoritma: C4.5 - Jumlah Atribut: 6	Aplikasi desktop berhasil memprediksi kelulusan mahasiswa dengan presentase 87.5% dari enam puluh data training dan empat puluh data testing.
4	1. Mutiara Ayu Banjarsari 2. H.Irwan Budiman 3. Andi Farmadi	Prediksi Kelulusan Tepat Waktu Mahasiswa.	- Algoritma : <i>K - Nearest Neighbor</i> - Jumlah Atribut: 6	Nilai k-Optimal adalah k=5, dengan tingkat akurasi sebesar 80,00%.
5	Lillyan Hadjaratie	Prediksi Dan Pemetaan Kelulusan Mahasiswa	- Algoritma: 1. <i>Decision Tree (C4.5)</i> 2. <i>Artificial Neural Network</i> 3. <i>K-Nearest Neighbour</i> - Jumlah Atribut: 8	Hasil klasifikasi data mahasiswa aktif dan lulusan dapat diperoleh dengan menggunakan 3 metode tersebut.
6.	1. Nursalim,	Klasifikasi Bidang Kerja	- Algoritma:	Algoritma data

	2. Suprapedi, 3. H. Himawan	Lulusan	<i>K-Nearest Neighbor (K-NN)</i> - Jumlah Atribut: 12	mining <i>K-Nearest Neighbor</i> memiliki kinerja terbaik untuk klasifikasi bidang kerja lulusan dengan nilai accuracy yaitu 83,33% dan nilai <i>Area Under The Curve (AUC)</i> adalah 90%
7	1. C. Anuradha 2. T.Velmurugan	Prediksi Kinerja Mahasiswa	- Algoritma: 1. <i>C4.5 (J48)</i> 2. <i>Naive Bayes</i> 3. <i>K-NN</i> - Jumlah Atribut: 19	Tingkat prediksi yang tidak seragam antara algoritma. Kisaran prediksi bervariasi 61% - 75%.
8	1. Muhamad Hanief Meinanda 2. Metri Annisa 3. Narendi Muhandri 4. Kadarsyah Suryadi	Prediksi Masa Studi Sarjana	- Algoritma: <i>Neural Network</i> - Jumlah Atribut:5	<i>Artificial Neural Network</i> dengan arsitektur <i>Multilayer Perceptron</i> merupakan model terbaik untuk prediksi masa studi.
9	Khoirul Mu'aris	Prediksi Kelulusan Mahasiswa.	Algoritma : 1. <i>C4.5</i> 2. <i>C4.5 dg PSO</i> Jumlah Atribut: 8	<i>Algoritma C4.5</i> berbasis <i>PSO</i> terbukti dapat meningkatkan nilai akurasi dan <i>AUC</i> .

Pada tabel di atas rata-rata hanya menjelaskan tentang prediksi kelulusan mahasiswa dengan menggunakan salah satu algoritma dengan atau tanpa menggunakan fitur seleksi. Pada point 5, 3 algoritma yang diterapkan hanya menceritakan kalau 3 metode tersebut dapat menghasilkan klasifikasi data mahasiswa aktif dan lulusan. Pada point 7, 3 algoritma yang diterapkan hanya menceritakan kalau 3 metode tersebut mempunyai tingkat prediksi yang tidak seragam antar algoritma. Sehingga belum pernah diketahui sebenarnya algoritma mana yang terbaik dari beberapa algoritma yang ada, khususnya tentang prediksi kelulusan mahasiswa. Dari beberapa jurnal (± 35) yang dibaca penulis banyak sekali peneliti sebelumnya menggunakan *Algoritma Naive Bayes* sebagai urutan pertama, *Algoritma Decision Tree (C4.5)* sebagai urutan kedua disusul *Algoritma K-NN* sebagai urutan ketiga serta *Algoritma Neural Network* sebagai urutan terakhir dalam melakukan penelitian prediksi kelulusan tepat waktu mahasiswa. Hal tersebut diatas yang menjadi alasan untuk melakukan penelitian guna mencari mana diantara 4 jenis algoritma tersebut yang menghasilkan nilai akurasi dan *AUC* terbaik. Selanjutnya ditambahkan fitur seleksi *Particle Swarm Optimization (PSO)* untuk diterapkan pada 4 algoritma guna mengetahui kenaikan akurasi dan *AUC*, serta ingin mengetahui algoritma mana yang terbaik.

Metodologi atau tahapan penelitian diperlukan sebagai kerangka dan panduan proses penelitian, sehingga rangkaian proses penelitian dapat dilakukan secara terarah, teratur dan sistematis. Berikut metode yang diusulkan dalam penelitian ini ditunjukkan pada Gambar 1.



Gambar 1 Metode Yang Diusulkan

2.2 Pengumpulan Data

Penelitian ini menggunakan data Alumni Mahasiswa Jurusan Teknik Informatika dan Sistem Komputer STMIK ASIA MALANG mulai Angkatan 2007 sampai 2011 sebanyak 1.064 record yang diambil dari BAAK. Dari data tersebut hanya data dari Lulusan SLTA saja sebanyak 845 yang digunakan penelitian dengan alasan kuliahnya mulai dari semester 1 (0 sks). Sedangkan data transfer dari D1/D2/D3 tidak dimasukkan dengan alasan sudah ada beberapa SKS yang sudah ditempuh di Perguruan Tinggi sebelumnya.

2.3 Preprocessing Data

Ada 12 atribut yang dilakukan dalam penelitian ini, poin 1 sampai 11 sebagai variabel penentu prediksi kelulusan dan poin 12 sebagai target atau atribut tujuan yaitu kelulusan tepat waktu, seperti tercantum pada Tabel 2 dibawah ini:

Tabel 2 Atribut Dan Prediksi Kelulusan Tepat Waktu

NO.	ATRIBUT	KETERANGAN
1	Jenis Kelamin	Laki-laki / Perempuan
2	Umur	17 – 37
3-9	IPS (Sem 1 – 7)	0,00 – 4,00
10	Status Pekerjaan Mahasiswa	Belum Bekerja
		Sudah Bekerja
11	Status Pernikahan Mahasiswa	Belum Menikah
		Sudah Menikah
12	Prediksi Kelulusan	Lulus Tepat Waktu
		Lulus Tidak Tepat Waktu

Untuk selanjutnya dilakukan teknik *preprocessing* agar kualitas data yang diperoleh lebih baik dengan cara:

1. **Data Validation**, untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*).
2. **Data Dcretization**, dalam data *training* yang digunakan dalam penelitian ini, dilakukan seleksi atribut prediksi kelulusan dengan mengikuti aturan sebagai berikut:
 - a. Jika Lama Studi 3,5 - 4,5 tahun, dikatakan Lulus Tepat Waktu
 - b. Jika Lama Studi > 4,5 tahun, dikatakan Lulus Tidak Tepat Waktu

2.4 Modeling Phase (Fase Pemodelan)

Pada tahapan ini merupakan tahapan pemrosesan data training yang diklasifikasikan oleh model untuk menghasilkan nilai akurasi dan *AUC* yang tertinggi. Pada fase ini akan dilakukan tahapan-tahapan:

1. Menghitung nilai akurasi dan *AUC* dari masing-masing algoritma (*Naïve Bayes*, *Decision Tree (C4.5)*, *k-NN* dan *Neural Network*), dan dicari algoritma mana yang menghasilkan nilai akurasi dan *AUC* tertinggi.
2. Menghitung nilai akurasi dan *AUC* dari masing-masing algoritma (*Naïve Bayes*, *Decision Tree (C4.5)*, *k-NN* dan *Neural Network*) setelah penambahan fitur seleksi *Particle Swarm Optimization (PSO)*, dan dicari algoritma mana yang menghasilkan nilai akurasi dan *AUC* tertinggi.

2.5 Evaluation Phase (Fase Evaluasi)

Pada fase ini dilakukan pengujian terhadap algoritma berbasis *PSO* terbaik dengan membagi jumlah data training dan data testing sebagai berikut:

1. Data Training (Mahasiswa Angkatan 2007-2010) dan Data Testing (Mahasiswa Angkatan 2011).
2. Data Training (Mahasiswa Angkatan 2007-2011) dan Data Testing (Mahasiswa Angkatan 2011).
3. Data Training (Mahasiswa Angkatan 2007-2009) dan Data Testing (Mahasiswa Angkatan 2010-2011).
4. Data Training (Mahasiswa Angkatan 2007-2011) dan Data Testing (Mahasiswa Angkatan 2010-2011).

Dari 4 pengujian data training dan testing di atas dipilih yang mempunyai nilai akurasi atau *AUC* tertinggi untuk selanjutnya ditentukan sebagai model.

3. HASIL DAN PEMBAHASAN

3.1 Pengolahan Data Awal

Pada tahap pengumpulan dataset yang digunakan pada penelitian ini berasal dari data Alumni Mahasiswa Jurusan Teknik Informatika dan Sistem Komputer Angkatan 2007–2011 STMIK ASIA Malang sebanyak 845 dari lulusan SLTA murni dengan 19 atribut. Data asal tersebut dilakukan *preprocessing* dengan mengisi data yang kosong (Indek Prestasi Semester 1 sampai 7) akibat tidak aktif kuliah (cuti kuliah) atau tidak kuliah selama semester itu dan hanya membayar SPP saja, sehingga datanya blank (kosong). Data yang kosong tersebut diisi dengan nilai minimal dari masing-masing atribut yang sama. Tabel 3 merupakan deskripsi dan pemetaan dari dataset Alumni STMIK ASIA Malang Angkatan 2007-2011.

Tabel 3 Dataset Alumni STMIK ASIA Malang Angkatan 2007-2011

NO.	JK	UMUR	IP														STATUS PEKERJAAN	STATUS PERNIKAHAN	LAMA STUDI	PREDIKSI KELULUSAN
			S.1	S.2	S.3	S.4	S.5	S.6	S.7	S.8	S.9	S.10	S.11	S.12	S.13	S.14				
1	L	22	2,15	2,56	2,88	2,43	2,63	2,35	2,69	0,00	1,80	0,00	2,50				Belum Kerja	Belum Nikah	5,42	Tidak Tepat Waktu
2	L	20	2,28	1,78	2,30	1,50	1,50	0,69	0,31	0,00	2,68	2,83	1,81	2,00	2,00	3,00	Belum Kerja	Belum Nikah	6,92	Tidak Tepat Waktu
3	L	23	2,38	2,69	2,57	2,62	2,81	2,14	2,00	2,85	2,39	3,50					Belum Kerja	Belum Nikah	4,67	Tidak Tepat Waktu
4	L	20	2,40	1,94	2,17	1,33	2,61	1,76	1,25	1,40	1,23	2,75	2,12	0,00	2,96		Belum Kerja	Belum Nikah	6,58	Tidak Tepat Waktu
5	L	19	2,48	2,22	2,67	2,71	2,52	2,21	3,00	1,70	0,60	2,71	0,00	3,00			Belum Kerja	Belum Nikah	5,58	Tidak Tepat Waktu
6	L	20	2,53	2,62	2,55	2,63	2,47	2,50	3,22	2,88	2,96	0,00	3,00				Belum Kerja	Belum Nikah	5,42	Tidak Tepat Waktu
7	F	20	2,55	2,74	2,88	2,76	2,65	2,95	2,93	2,85	0,00						Belum Kerja	Belum Nikah	4,17	Tepat Waktu
8	L	20	2,55	2,70	2,90	2,48	0,53	1,97	1,89	1,54	1,14	0,00	2,90	2,14	2,00	3,00	Belum Kerja	Belum Nikah	6,67	Tidak Tepat Waktu
9	L	20	2,60	2,75	2,66	2,90	3,17	3,29	3,50	3,36							Belum Kerja	Belum Nikah	3,92	Tepat Waktu
10	F	20	2,60	2,81	2,71	2,88	2,69	2,86	2,88	3,00	0,83						Belum Kerja	Belum Nikah	4,67	Tidak Tepat Waktu
.
.
.
843	L	18	3,65	3,50	3,71	3,50	3,43	3,25	3,06	3,50							Belum Kerja	Belum Nikah	4,00	Tepat Waktu
844	L	18	3,65	3,23	3,65	3,39	3,61	3,59	2,91	4,00							Belum Kerja	Belum Nikah	4,00	Tepat Waktu
845	L	34	3,65	3,70	3,39	3,68	3,30	3,25	3,50	3,80	1,75						Sudah Kerja	Sudah Nikah	4,08	Tepat Waktu

Dataset di atas merupakan data awal yang belum dipilih atribut, selanjutnya dilakukan pemilihan atribut-atribut yang diperlukan, khusus untuk pengisian kolom atribut tujuan (prediksi kelulusan), digunakan acuan sebagai berikut:

- a. Jika Lama Studi 3,5 - 4,5 tahun, dikatakan Lulus Tepat Waktu
- b. Jika Lama Studi > 4,5 tahun, dikatakan Lulus Tidak Tepat Waktu,

akhirnya diperoleh dataset baru seperti pada Tabel 5 dan Tabel 6.

Alasan digunakan kategori Lama Studi 3,5 - 4,5 tahun dinyatakan sebagai Lulus Tepat Waktu adalah sebagai berikut. Berdasarkan Permenristekdikti no 44 tahun 2015 tentang Standar Nasional Pendidikan Tinggi Pasal 16d: paling lama 7 (tujuh) tahun akademik untuk program sarjana, program diploma empat/sarjana terapan, dengan beban belajar mahasiswa paling sedikit 144 (seratus empat puluh empat) sks. Informasi yang diperoleh dari Pembantu Ketua 1 Bidang Akademik Perti ASIA Malang, mahasiswa dikatakan lulus tepat waktu jika mahasiswa mampu menempuh lama studi 3,5 sampai 4 tahun untuk STIE ASIA dan 4 sampai 4,5 tahun untuk STMIK ASIA. Alasan lain, rata-rata jumlah Total SKS yang ditempuh di STMIK ASIA Malang sebanyak 153 SKS, secara perhitungan normal maka membutuhkan waktu kuliah 8 Semester (4 tahun) untuk menyelesaikan perkuliahan teori saja jika per semester mengambil 20 SKS (rata-rata yang diambil mahasiswa), sehingga ada waktu sisa selama 1 semester untuk menyelesaikan Tugas Akhir (Skripsi). Dari data yang ada diperoleh informasi tambahan, rata-rata mahasiswa STMIK ASIA Malang lulus selama kurang lebih 4,74 tahun.

Khusus untuk analisa Algoritma Neural Network, data di atas dirubah menjadi data seperti Tabel 6 di bawah ini dengan mengikuti aturan sebagai berikut:

- a. Untuk Jenis Kelamin (JK),
 - Laki-laki diberi kode 1 (653 mahasiswa)
 - Perempuan diberi kode 2 (192 mahasiswa)
- b. Untuk Status Pekerjaan,
 - Belum Kerja diberi kode 1 (779 mahasiswa)
 - Sudah Kerja diberi kode 2 (66 mahasiswa)
- c. Untuk Status Pernikahan
 - Belum Nikah diberi kode 1 (833 mahasiswa)
 - Sudah Nikah diberi kode 2 (12 mahasiswa)

Tabel 4 Dataset Untuk Analisa Algoritma Naive Bayes, C4.5, k-NN

NO.	JK	UMUR	IP							Stat.Kerja	Stat.Nikah	Prediksi
			S.1	S.2	S.3	S.4	S.5	S.6	S.7			
1	L	22	2,15	2,56	2,88	2,43	2,63	2,35	2,69	Belum Kerja	Belum Nikah	Tidak Tepat Waktu
2	L	20	2,28	1,78	2,30	1,50	1,50	0,69	0,31	Belum Kerja	Belum Nikah	Tidak Tepat Waktu
3	L	23	2,38	2,69	2,57	2,62	2,81	2,14	2,00	Belum Kerja	Belum Nikah	Tidak Tepat Waktu
4	L	20	2,40	1,94	2,17	1,33	2,61	1,76	1,25	Belum Kerja	Belum Nikah	Tidak Tepat Waktu
5	L	19	2,48	2,22	2,67	2,71	2,52	2,21	3,00	Belum Kerja	Belum Nikah	Tidak Tepat Waktu
6	L	20	2,53	2,62	2,55	2,63	2,47	2,50	3,22	Belum Kerja	Belum Nikah	Tidak Tepat Waktu
7	F	20	2,55	2,74	2,88	2,76	2,65	2,95	2,93	Belum Kerja	Belum Nikah	Tepat Waktu
8	L	20	2,55	2,70	2,90	2,48	0,53	1,97	1,89	Belum Kerja	Belum Nikah	Tidak Tepat Waktu
9	L	20	2,60	2,75	2,66	2,90	3,17	3,29	3,50	Belum Kerja	Belum Nikah	Tepat Waktu
10	F	20	2,60	2,81	2,71	2,88	2,69	2,86	2,88	Belum Kerja	Belum Nikah	Tidak Tepat Waktu
.
.
.
843	L	18	3,65	3,50	3,71	3,50	3,43	3,25	3,06	Belum Kerja	Belum Nikah	Tepat Waktu
844	L	18	3,65	3,23	3,65	3,39	3,61	3,59	2,91	Belum Kerja	Belum Nikah	Tepat Waktu
845	L	34	3,65	3,70	3,39	3,68	3,30	3,25	3,50	Sudah Kerja	Sudah Nikah	Tepat Waktu

Tabel 5 Dataset Untuk Analisa Algoritma NN

NO.	JK	UMUR	IP							Stat.Kerja	Stat.Nikah	Prediksi
			S.1	S.2	S.3	S.4	S.5	S.6	S.7			
1	1	22	2,15	2,56	2,88	2,43	2,63	2,35	2,69	1	1	Tidak Tepat Waktu
2	1	20	2,28	1,78	2,30	1,50	1,50	0,69	0,31	1	1	Tidak Tepat Waktu
3	1	23	2,38	2,69	2,57	2,62	2,81	2,14	2,00	1	1	Tidak Tepat Waktu
4	1	20	2,40	1,94	2,17	1,33	2,61	1,76	1,25	1	1	Tidak Tepat Waktu
5	1	19	2,48	2,22	2,67	2,71	2,52	2,21	3,00	1	1	Tidak Tepat Waktu
6	1	20	2,53	2,62	2,55	2,63	2,47	2,50	3,22	1	1	Tidak Tepat Waktu
7	2	20	2,55	2,74	2,88	2,76	2,65	2,95	2,93	1	1	Tepat Waktu
8	1	20	2,55	2,70	2,90	2,48	0,53	1,97	1,89	1	1	Tidak Tepat Waktu
9	1	20	2,60	2,75	2,66	2,90	3,17	3,29	3,50	1	1	Tepat Waktu
10	2	20	2,60	2,81	2,71	2,88	2,69	2,86	2,88	1	1	Tidak Tepat Waktu
.
.
.
843	L	18	3,65	3,50	3,71	3,50	3,43	3,25	3,06	1	1	Tepat Waktu
844	L	18	3,65	3,23	3,65	3,39	3,61	3,59	2,91	1	1	Tepat Waktu
845	L	34	3,65	3,70	3,39	3,68	3,30	3,25	3,50	2	2	Tepat Waktu

3.2 EKSPERIMEN DAN PENGUJIAN MODEL

3.2.1 Perhitungan 4 Algoritma Non Berbasis PSO

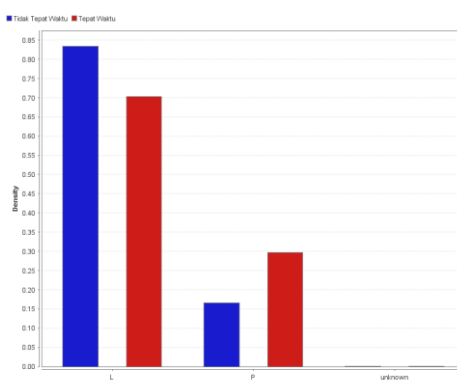
Dari 4 Algoritma Berbasis Non PSO diketahui bahwa Algoritma k-Nearest Neighbor (k=15) menghasilkan pengujian terbaik dengan nilai accuracy **69,81%** dan Area Under the Curve (AUC) **0,764**, seperti terlihat pada Tabel 6 di bawah ini.

Tabel 6. Rekapitulasi Hasil Pengujian 4 Algoritma Non PSO

Jenis Pengujian	Naive Bayes	Decision Tree (C4.5)	k-NN (k=15)	Neural Network
Accuracy	63,79 %	64,02 %	69,81 %	67,44 %
Precision	58,18 %	61,64 %	68,09 %	66,55 %
Recall	81,22 %	62,16 %	66,97 %	63,22 %
AUC	0,736	0,633	0,764	0,745

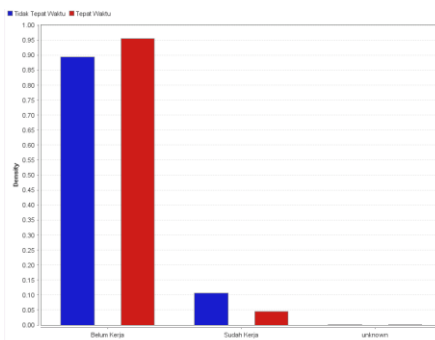
3.2.2 Perhitungan 4 Algoritma Berbasis PSO

Perhitungan dengan menggunakan *Algoritma Naive Bayes Berbasis PSO* diperoleh nilai akurasi sebesar **65,92%** dan nilai *AUC* sebesar **0,721**, juga diperoleh informasi seperti pada Gambar 2 sampai 4 di bawah ini.



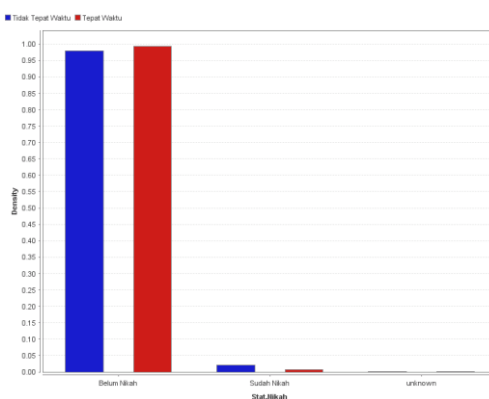
Gambar 2 Grafik Kelulusan Mahasiswa Berdasarkan Jenis Kelamin

Dari Gambar 2 di atas dapat dijelaskan, bahwa untuk jenis kelamin laki-laki, jumlah lulus tepat waktu lebih sedikit dibanding lulus tidak tepat waktu. Untuk jenis kelamin perempuan, jumlah lulus tepat waktu lebih banyak daripada lulus tidak tepat waktu.



Gambar 3 Grafik Kelulusan Mahasiswa Berdasarkan Status Kerja

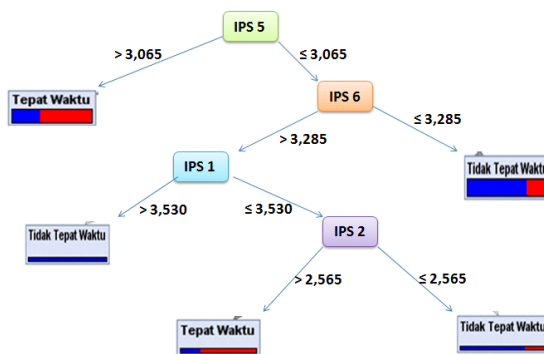
Dari Gambar 3 di atas dapat dijelaskan, bahwa untuk status belum kerja, jumlah lulus tepat waktu lebih banyak dibanding lulus tidak tepat waktu. Untuk status sudah kerja, jumlah lulus tepat waktu lebih sedikit daripada lulus tidak tepat waktu.



Gambar 4 Grafik Kelulusan Mahasiswa Berdasarkan Status Nikah

Dari Gambar 4 di atas dapat dijelaskan, bahwa untuk status belum menikah, jumlah lulus tepat waktu lebih banyak dibanding lulus tidak tepat waktu. Untuk status sudah menikah, jumlah lulus tepat waktu lebih sedikit daripada lulus tidak tepat waktu.

Perhitungan dengan menggunakan *Algoritma Decision Tree (C4.5) Berbasis PSO* diperoleh nilai akurasi tertinggi sebesar **69,23%** dan nilai AUC sebesar **0,687**. Rule Algoritma klasifikasi *Decision Tree (C4.5)* berbasis *PSO*, bisa dipelajari dari Gambar 5 di bawah ini:

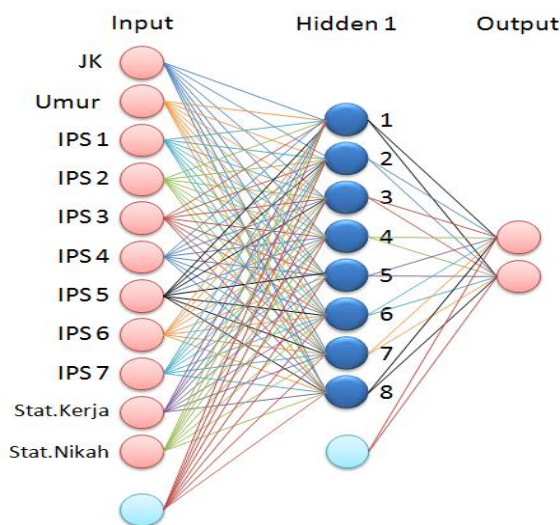


Gambar 5 Pohon Keputusan Kelulusan Mahasiswa Menggunakan Algoritma C4.5 + PSO

Dari Gambar 5 di atas dapat diuraikan: jika Indeks Prestasi Semester (IPS) 5 > 3,065 maka akan lulus tepat waktu, jika IPS 5 ≤ 3,065 akan dilihat IPS 6, jika ≤ 3,285 maka lulus tidak tepat waktu, jika IPS 6 > 3,285 akan dilihat IPS 1, jika > 3,530 maka lulus tidak tepat waktu, jika IPS 1 ≤ 3,530, akan dilihat IPS 2, jika > 2,565, maka lulus tepat waktu, jika IPS 2 ≤ 2,565, maka lulus tidak tepat waktu.

Perhitungan dengan menggunakan *Algoritma k-NN Berbasis PSO* diperoleh nilai akurasi tertinggi sebesar **74,08%** dan nilai *AUC* sebesar **0,788** pada nilai **k-Optimal=19**.

Perhitungan dengan menggunakan *Algoritma Neural Network Berbasis PSO* diperoleh nilai akurasi tertinggi sebesar **72,79%** dan nilai *AUC=0,780* pada *Training Cycles* 100, dengan menghasilkan arsitektur Neural Network seperti pada Gambar 6 di bawah ini:



Gambar 6 Arsitektur *Neural Network* Berbasis *PSO*

Dari 4 *Algoritma Berbasis PSO* akhirnya diperoleh data rekapitulasi hasil pengujian seperti Tabel 7, dimana *Algoritma k-Nearest Neighbor (k-NN) Berbasis PSO* menghasilkan pengujian terbaik dengan nilai accuracy **74,08%** dan Area Under the Curve (*AUC*) = **0,788**, seperti terlihat pada tabel di bawah ini.

Tabel 7 Rekapitulasi Hasil Pengujian 4 *Algoritma Berbasis PSO*

Jenis Pengujian	<i>Naive Bayes</i>	<i>Decision Tree (C4.5)</i>	<i>k-NN (k=19)</i>	<i>Neural Network</i>
Accuracy	65,92 %	69,23 %	74,08 %	72,55 %
Precision	59,77 %	71,56 %	73,12 %	74,62 %
Recall	83,25 %	57,87 %	70,30 %	63,96 %
AUC	0,721	0,687	0,788	0,780

3.3 Pembahasan

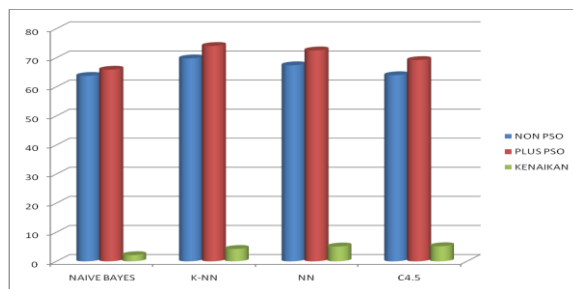
3.3.1 Peningkatan Nilai Akurasi

Berdasarkan Tabel 6 dan Tabel 7, akhirnya diperoleh hitungan peningkatan nilai akurasi seperti pada Tabel 8.

Tabel 8 Peningkatan Nilai Akurasi

ALGORITMA	Non PSO	+ PSO	Peningkatan Nilai Akurasi	Keterangan
<i>Naive Bayes</i>	63,79 %	65,92 %	2,13 %	Terendah
<i>k-NN</i>	69,81 %	74,08 %	4,27 %	Tertinggi ke-3
<i>Neural Network</i>	67,44 %	72,55 %	5,11 %	Tertinggi ke-2
<i>Decision Tree (C4.5)</i>	64,02 %	69,23 %	5,21 %	Tertinggi ke-1

Untuk lebih mempermudah pembacaan kenaikan nilai akurasi, berikut ini tampilan grafik peningkatan nilai akurasi dari Tabel 8 di atas:



Gambar 7 Grafik Kenaikan Nilai Akurasi

Dari grafik Gambar 7 di atas, diketahui urutan kenaikan nilai akurasi dari terendah menuju tertinggi adalah Algoritma Naive Bayes, k-NN, Neural Network (NN) dan Decision Tree (C4.5).

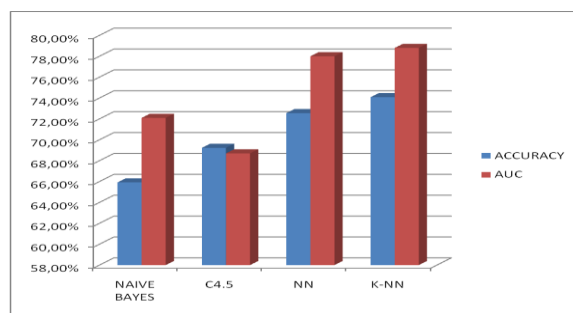
3.3.2 Penentuan Algoritma Terbaik

Rekapitulasi hasil pengujian dari 4 *Algoritma berbasis PSO* dapat dilihat pada Tabel 9 di bawah ini:

Tabel 9. Rekapitulasi Hasil Pengujian 4 Algoritma + PSO

Jenis Pengujian	Naive Bayes	Decision Tree (C4.5)	k-NN (k=19)	Neural Network
Accuracy	65,92 %	69,23 %	74,08 %	72,55 %
Precision	59,77 %	71,56 %	73,12 %	74,62 %
Recall	83,25 %	57,87 %	70,30 %	63,96 %
AUC	0,721	0,687	0,788	0,780

Dari Tabel 9 di atas, dapat ditarik kesimpulan bahwa Algoritma k-NN Berbasis PSO menghasilkan pengujian terbaik dengan nilai **accuracy** tertinggi sebesar **74,08%** (pada k-optimum=19) dan nilai **Area Under the Curve (AUC)=0,788**. Untuk lebih mempermudah pembacaan data di atas, berikut ini tampilan grafik nilai accuracy dan AUC dari Tabel 9:



Gambar 8 Grafik Nilai Akurasi Dan AUC

Dari grafik Gambar 8 di atas, diketahui urutan nilai akurasi dari terendah menuju tertinggi adalah Algoritma Naive Bayes, Decision Tree (C4.5), Neural Network dan k-NN.

Untuk mengetahui pembobotan masing-masing atribut, maka diterapkan pada *Algoritma k-NN+ PSO (k=19)* sebagai algoritma yang memberikan nilai akurasi terbaik. Tabel 11 menunjukkan nilai pembobotan masing-masing atribut, dimana atribut: Umur, IPS 3, IPS 5 dan Status Pernikahan tidak memberikan kontribusi yang nyata terhadap kelulusan tepat waktu. Atribut: Jenis Kelamin, IPS 1, IPS 2, IPS 6, IPS 7 dan Status Pekerjaan memberikan kontribusi yang nyata terhadap kelulusan tepat waktu. Untuk atribut IPS 4 tetap memberikan kontribusi terhadap kelulusan tepat waktu meskipun hanya sebesar **0,682**.

Tabel 10. Pembobotan Atribut *Algoritma k-NN + PSO*

Atribut	Pembobotan
JK (Jenis Kelamin)	1
Umur	0
IP Semester 1	1
IP Semester 2	1
IP Semester 3	0
IP Semester 4	0,682
IP Semester 5	0
IP Semester 6	1
IP Semester 7	1
Status Pekerjaan	1
Status Pernikahan	0

3.3.3 Confusion Matrix

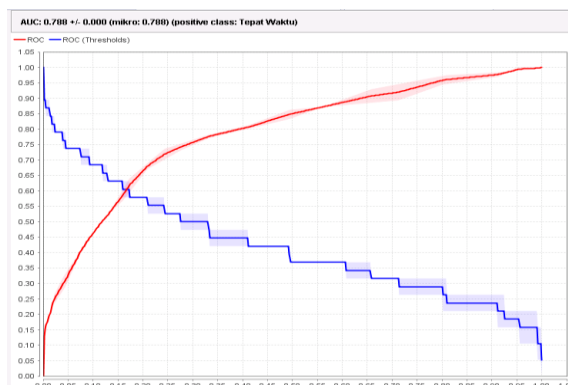
Penilaian kinerja klasifikasi *Algoritma k-NN Berbasis Particle Swarm Optimization(PSO)* menggunakan data privat STMIK ASIA Malang yang terdiri dari 845 record seperti pada Tabel 11. Dimana sebanyak 277 record sesuai prediksi *true (tepat waktu)* sedangkan 102 record tidak sesuai dengan *prediksi true (tepat waktu)*. Untuk kelompok “*tidak tepat waktu*” terdapat 349 record yang sesuai dengan hasil prediksi dan 117 record yang tidak sesuai dengan hasil prediksi.

Tabel 11. *Confusion Matrix*

Accuracy : 74,08% +/- 0,62% (Mikro: 74,08%)			
	True Tepat Waktu	True Tidak Tepat Waktu	Class Precision
Pred.Tepat Waktu	277	102	73,09 %
Pred.Tidak Tepat Waktu	117	349	74,89 %
Class Recall	70,30 %	77,38 %	

3.3.4 Kurva ROC

Gambar 9 merupakan kurva *Receiver Operating Characteristic (ROC)* yang digunakan untuk memvisualisasikan kinerja *Algoritma k-NN* pada *Particle Swarm Optimization (PSO)*. Grafik *Area Under Curve (AUC)* dibawah ini adalah bagian dari daerah unit persegi yang nilainya antara 0 hingga 1,00. Pada gambar, garis vertikal adalah *true positives* dan garis horizontal merupakan *false positives*. Sedangkan grafik berwarna merah merupakan kurva ROC dan grafik berwarna biru adalah kurva ambang (*thresholds*). Diperoleh nilai *AUC* sebesar 0,788, untuk kategori akurasi yaitu klasifikasi cukup atau *fair classification* untuk *Algoritma k-NN* berbasis *Particle Swarm Optimization* pada dataset Kelulusan Mahasiswa STMIK ASIA Malang.



Gambar 9 Kurva Receiver Operating Characteristic

4. KESIMPULAN DAN SARAN

Hasil analisis menggunakan *Confusion Matrix* dan *ROC Curve* dapat disimpulkan bahwa **Algoritma k-Nearest Neighbor Berbasis PSO** memiliki kinerja terbaik untuk klasifikasi bidang kerja lulusan dengan nilai **Accuracy** yaitu **74,08%** dan nilai **Area Under The Curve (AUC) = 0,788**. Penambahan fitur *PSO* selalu meningkatkan nilai akurasi, dimana peningkatan nilai akurasi tertinggi terletak pada **Algoritma Decision Tree (C4.5)** sebesar **5,21%**, terendah pada **Algoritma Naive Bayes** sebesar **2,13%**.

Berkaitan untuk penelitian selanjutnya perlu dicoba dengan penambahan kuantitas (jumlah dataset) dan kualitas (jumlah atribut) yang akan digunakan untuk klasifikasi data mining serta dapat pula ditambahkan beberapa model klasifikasi yang baru misalnya *ID3*, *Support Vector Machine* untuk diuji coba untuk mengetahui tingkat *accuracy* dan *Area Under the Curve (AUC)* untuk klasifikasi bidang kerja lulusan mahasiswa.

DAFTAR PUSTAKA

- Anuradha,C., T.Velmurugan. (July 2015). *A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Students Performance*. *Indian Journal of Science and Technology*, Vol 8(15), DOI: 10.17485/ijst/2015/v8i15/74555, ISSN (Print) : 0974-6846. ISSN (Online) : 0974-5645,.
- Asif R., Agathe M., Mahmood KP. (2015). *Predicting Student Academic Performance at Degree Level: A Case Study*. *I.J. Intelligent Systems and Applications*, 01, 49-61. Published Online December 2014 in MECS (<http://www.mecs-press.org/>). DOI: 10.5815/ijisa.2015.01.05.
- Ayu, Mutiara B.,H.Irwan Budiman and Andi Farmadi. (September 2015). Penerapan K-Optimal Pada Algoritma k-NN untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer FMIPA UNLAM Berdasarkan IP Sampai Dengan Semester 4. *Kumpulan jurnaL Ilmu Komputer (KLIK)* ISSN: 2406-7857. Volume 02, No.02.
- Handjaratie,Lillyan. (2015). *Prediction And Data Mapping of Students Of Engineering Faculty*, Universitas Negeri Gorontalo Using Data Mining.
- Hanief Muhamad M., Metri Annisa, Narendi Muhandri and Kadarsyah Suryadi. (2009). *Prediksi Masa Studi Sarjana Dengan Artificial Neural Network*. *Internetworking Indonesia Journal*. Vol.1/No.2.
- Hartanto,David H.,Seng Hansun. (Juni 2014). Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa. *ULTIMATICS*, Vol. VI, No. 1 | ISSN 2085-4552.
- Mu'aris Khoirul. (2015). *Komparasi Pemodelan Data Menggunakan C4.5 Dan C4.5 Berbasis Particle Swarm Optimization Untuk Memprediksi Kelulusan Mahasiswa*.
- Nursalim, Suprapedi and H.Himawan. (April 2014). *Klasifikasi Bidang Kerja Lulusan Menggunakan Algoritma K-Nearest Neighbor*. *Jurnal Teknologi Informasi*, ISSN 1414-9999. Volume 10 Nomor 1.
- Nuqson Masykur Huda. (2010). "Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa", Semarang.
- Prabowo. (2012). *Aneka Teknik, Piranti dan Penerapan Data Mining : Studi Kasus Peramalan Harga Saham Industri Telekomunikasi Berbasis Jaringan Saraf Tiruan*. Modul Perkuliahan Universitas Budi Luhur.
- Ricky, Ade Rozzaqi. (Juni 2015). *Naïve Bayes dan Filtering Feature Selection Information Gain untuk Prediksi Ketepatan Kelulusan Mahasiswa*. *Jurnal Informatika UPGRIS* Volume 1.