STUDI KOMPARASI ALGORITMA DECESION TREE (C4.5) DENGAN ALGORITMA K-NN DALAM MEMPREDIKSI KELULUSAN TEPAT WAKTU MAHASISWA

Moh. Zainuddin[1], Achmad Noercholis[2]
[1,2]STMIK ASIA Malang
[1]mzein@asia.ac.id, [2]anoercholis@asia.ac.id

ABSTRACT

The purpose of this study is to study which is better between the Decision Tree (C4.5) Algorithm and the k-Nearest Neighbor (k-NN) algorithm based on Particle Swarm Optimization (PSO) based on an assessment of its accuracy on passing the right delivery time. The results showed the closest k-Neighbor (k-NN) algorithm based on PSO at k-optimum = 19 had better results than the Decesion Tree (C4.5) algorithm, with an Accuracy value of 76.69% for the k- algorithm NN and 69.82% for the Decesion Tree Algorithm (C4.5). Increasing the accuracy value of the k-NN algorithm (6.88%) is greater than the Decesion Tree Algorithm (5.80%). Semester Achievement Index Attribute (IPS) 6 has the highest weighting value with the Decesion Tree Algorithm (C4.5) and the k-NN Algorithm. The Age attribute has a weighting value of 0,000 meaning it does not make a real contribution to the right graduation of student time.

Keywords: *Accuracy, Algorithm, Decision Tree (C4.5), k-Nearest Neighbor (k-NN), Particle Swarm Optimization (PSO).*

1. PENDAHULUAN

Pendidikan adalah suatu aktivitas sosial yang memungkinkan masyarakat tetap ada dan berkembang[1]. Jenjang pendidikan perguruan tinggi merupakan salah satu persyaratan dasar dalam mencari pekerjaan, yang akan mempersiapkan calon-calon sarjana yang berkualitas dan mempunyai keterampilan dibidangnya. Pencapaian gelar kesarjanaan tersebut membutuhkan waktu normal selama 3,5 sampai 4,5 tahun, akan tetapi dalam praktiknya banyak mahasiswa yang lulusnya melebihi 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.

Prediksi kelulusan tepat waktu yang dilakukan saat ini hanya berdasarkan ramalan dari data IPK (Indeks Prestasi Komulatif) dan IMK (Indeks Mutualisme Komulatif) semester sebelumnya. Prediksi hampir sama dengan klasifikasi dan estimasi, hanya saja prediksi digunakan untuk menduga nilai-nilai tertentu yang akan terjadi dimasa mendatang[1].

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. Berangkat dar 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[2].

2. KAJIAN LITERATUR

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

Tabel 1: Perbandingan Hasil Penelitian

NO	PENELITI	BIDANG KAJIAN	METODE & JUMLAH ATRIBUT	HASIL
1	1. Raheela Asif,[3] 2. Agathe Merceron, 3. Mahmood K. Pathan	Prediksi Kinerja Bidang Akademik	- Algoritma: 1. *Decision Tree* 2. *Naive Bayes* 3. *Neural Network* - Jumlah Atribut:16	Akurasi terbaik sebesar 83,65% pada Dataset II diperoleh dengan menggunakan *Algoritma Naive Bayes.*
2	1. Mutiara Ayu Banjarsari[4] 2. H.Irwan Budiman 3. Andi Farmadi	Prediksi Kelulusan Tepat Waktu Mahasiswa.	- Algoritma : *K-Nearest Neighbor* - Jumlah Atribut: 6	Nilai k-Optimal adalah k=5, dengan tingkat akurasi sebesar 80,00%.
3	Lilyan Hadjaratie[5]	Prediksi Dan Penetaan Kelulusan Mahasiswa	- Algoritma: 1. *Decision Tree (C4.5)* 2. *Artificial Neural Network* 3. *K-Nearest Neighbour* - Jumlah Atribut: 8	Hasil klasifikasi data mahasiswa aktif dan lulusan dapat diperoleh dengan menggunakan 3 metode tersebut.
4	1. Nuraslim, [6] 2. Suprapedi, 3. H. Himawan	Klasifikasi Bidang Kerja Lulusan	- Algoritma: *K-Nearest Neighbor (K-NN)* - Jumlah Atribut: 12	Algoritma data mining *K-Nearest Neighbor* memiliki kinerja terbaik untuk klasifikasi bidang kerja lulusan dengan nilai accuracy yaitu 83,33% dan nilai *Area Under The Curve (AUC)* adalah 90%
5	1. C. Anuradha,[7] 2. T.Velmurugan	Prediksi Kinerja Mahasiswa	- Algoritma: 1. *C4.5 (J48)* 2. *Naive Bayes* 3. *K-NN* - Jumlah Atribut: 19	Tingkat prediksi yang tidak seragam antara algoritma. Kisaran prediksi bervariasi 61% - 75%.
6	1. David Hartanto Kamagi,[8] 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.

Pada tabel di atas rata-rata hanya menjelaskan tentang prediksi kelulusaan mahasiswa dengan menggunakan salah satu algoritma dengan atau tanpa menggunakan fitur seleksi. Pada nomer 3, 3 algoritma yang diterapkan hanya menceritakan kalau 3 metode tersebut dapat menghasilkan klasifikasi data mahasiswa aktif dan lulusan. Pada nomer 5, 3 algoritma yang diterapkan hanya menceritakan kalau 3 metode tersebut mempunyai tingkat prediksi yang tidak seragam antar algoritma. Sehingga belum pernah dilakukan penelitian sebenarnya algoritma mana yang lebih baik dari beberapa algoritma yang ada, khususnya tentang prediksi kelulusan mahasiswa. Dari beberapa jurnal yang dibaca penulis banyak sekali peneliti sebelumnya menggunakan *Algoritma Decision Tree (C4.5)* dan *Algoritma K-NN* dalam melakukan penelitian prediksi kelulusan tepat waktu mahasiswa. Hal tersebut diatas yang menjadi alasan untuk melakukan penelitian guna mencari mana diantara 2 jenis algoritma tersebut yang menghasilkan nilai akurasi yang lebih baik. Selanjutnya ditambahkan fitur seleksi *Particle Swarm Optimization (PSO)* untuk diterapkan pada 2 algoritma untuk mengetahui kenaikan akurasi, serta ingin mengetahui algoritma mana yang lebih baik.

Landasan Teori

Mahasiswa merupakan masyarakat kalangan elite dimana mahasiswa mempunyai ciri intelektualitas yang lebih komplek dibandingkan kelompok seusia mereka yang bukan mahasiswa, ataupun kelompok usia dibawah dan diatas mereka. “Ciri intelektualitas tersebut adalah kemampuan mahasiswa menghadapi, memahami dan mencari cara pemecahan masalah secara lebih sistematis”[9].

Ketepatan masa studi mahasiswa adalah hal yang sangat penting untuk diperhatikan, hal ini dikarenakan penurunan jumlah kelulusan akan menghilangkan jumlah pendapatan institusi dan mempengaruhi penilaian pemerintah serta mempengaruhi status akreditasi institusi. Menurut Suhartinah & Ernastuti ada beberapa faktor yang dapat mempengaruhi kelulusan mahasiswa antara lain adalah nilai akhir SMA, Indeks Prestasi Semester (IPS), gaji orang tua dan pekerjaan orang tua[10]. Suatu perguruan tinggi biasanya menggunakan indeks prestasi sebagai penilaian akademik, banyak universitas memberi standar minimum yang sulit diperoleh mahasiswa. Banyak variabel yang dapat digunakan dalam prediksi kelulusan mahasiswa seperti umur, status pernikahan, jumlah saudara dan lain-lain.

Peraturan Menteri Pendidikan dan Kebudayaan No. 49/2014 yang mengatur masa kuliah sarjana maksimal lima tahun dianggap terlalu berat oleh mahasiswa. Untuk itu, pemerintah akan mengembalikan masa kuliah sarjana menjadi tujuh tahun. Peraturan Menteri tersebut mengatur beban belajar minimal mahasiswa S-1/D-4 adalah 144 SKS (satuan kredit semester). Untuk menuntaskan seluruh beban SKS tadi, mahasiswa S-1/D-4 diberi batas waktu 4 hingga 5 tahun (8 sampai 10 semester).

Algoritma Decesion Tree (C4.5)

Secara umum *Algoritma Decesion Tree (C4.5)* untuk membangun pohon keputusan adalah sebagai berikut[11]:

a. Pilih atribut sebagai akar
b. Buat cabang untuk masing-masing nilai
c. Bagi kasus dalam cabang
d. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti yang tertera berikut:

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

Keterangan:
S : Himpunan kasus
A : Atribut
n : Jumlah partisi atribut A
|Si| : Jumlah kasus pada partisi ke i
|S| : Jumlah kasus dalam S

Sebelum mendapatkan nilai Gain adalah dengan mencari nilai Entropi. Entropi digunakan untuk menentukan seberapa informatif sebuah masukan atribut untuk menghasilkan sebuah atribut. Rumus dasar dari Entropi adalah sebagai berikut:

$$\text{Entropy (S)} = \sum_{i=1}^{n} -pi * \log_2 pi \quad \text{.....} \quad (2)$$

Keterangan:
S : Himpunan Kasus
n : Jumlah partisi S
pi : Proporsi dari Si terhadap S

Algoritma k-Nearest Neighbor

Algoritma k-Nearest Neighbor (k-NN) merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. K-NN termasuk algoritma ***supervised learning*** dimana hasil dari query instance yang baru diklasifikan berdasarkan mayoritas dari kategori pada k-NN. Kelas yang paling banyak muncul itu yang akan menjadi kelas hasil klasifikasi. Tujuan dari algoritma ini adalah mengklasifikasikan objek baru berdasarkan atribut dan *training sample.*

Algoritma k-Nearest Neighbor (k-NN) menggunakan klasifikasi ketetanggaan *(neighbor)* sebagai nilai prediksi dari query instance yang baru. Algoritma ini sederhana, bekerja berdasarkan jarak terpendek dari *query instance* ke *training sample* untuk menentukan ketetanggaannya[9].

Langkah-langkah untuk menghitung metode ***k-Nearest Neighbor*** dengan jarak terdekat (euclidian) antara lain:

a. Menentukan parameter k
b. Menghitung jarak antara data yang akan dievaluasi dengan semua pelatihan
c. Mengurutkan jarak yang terbentuk
d. Menentukan jarak terdekat sampai urutan k
e. Memasangkan kelas yang bersesuaian
f. Mencari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi

$$d_i = \sqrt{\sum_{i=1}^{p} (x_{2i} - x_{1i})^2} \quad \text{.....} \quad (3)$$

Keterangan:

x_1 = Sampel data

x_2 = Data uji atau data testing
i = Variabel data
d = Jarak
p = Dimensi data

Seleksi Fitur Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) dirumuskan oleh Dr. Eberhart dan Dr. Kennedy pada tahun 1995. PSO adalah metode pencarian populasi, yang berasal dari penelitian untuk pergerakan berkelompok burung dan ikan dalam mencari makan [12]. Seperti *Algoritma Genetika*, PSO melakukan pencarian menggunakan populasi (disebut swarm) dari individu-individu (disebut partikel) yang diperbarui dari setiap iterasi yang dilakukan. Untuk menemukan solusi yang optimal, setiap partikel bergerak ke arah posisi terbaik yang sebelumnya (pbest) dan posisi global terbaik (gbest).

Pada penelitian ini, PSO digunakan untuk menyeleksi fitur *(feature selection)* atau atribut tersebut, maka PSO menggunakan bobot atribut yang telah dihitung dan atribut yang telah diseleksi akan diprediksi menggunakan *Algoritma Decesion Tree (C4.5)* dan *Algoritma k-Nearest Neighbor (k-NN)*

k-Fold Cross Validation

Menurut Fu[13], *k-Fold Cross Validation* mengulang k-kali untuk membagi sebuah himpunan contoh secara acak menjadi k subset yang saling bebas, setiap ulangan disisakan satu subset untuk pengujian dan subset lainnya untuk pelatihan. Menurut Vercellis[14], dengan k = 5 atau 10 dapat digunakan untuk memperkirakan tingkat kesalahan yang terjadi,

sebab data training pada setiap fold cukup berbeda dengan data training yang asli. Secara keseluruhan, 5 atau 10 *Fold Cross Validation* sama-sama direkomendasikan dan disepakati bersama. Menghitung nilai akurasinya dapat dilakukan dengan menggunakan persamaan :

$$\text{Akurasi} = \frac{jumlah.klasifikasi.benar}{jumlah.data.uji} x100\% \ldots\ldots \quad (4)$$

Confusion Matrix

Untuk melakukan evaluasi terhadap model klasifikasi berdasarkan perhitungan objek testing mana yang diprediksi benar dan tidak benar. Perhitungan ini ditabulasikan kedalam tabel yang disebut *confusion matrix*[15]. C*onfusion matrix* merupakan data set yang hanya memiliki dua kelas, kelas yang satu sebagai positif dan kelas yang lain sebagai negatif. Terdiri dari empat sel yaitu *True Positives (TP), False Positives (FP), True Negatives (TN)* dan *False Negatives (FN)*.

Tabel 2: Confusion Matrix Untuk 2 Model Kelas

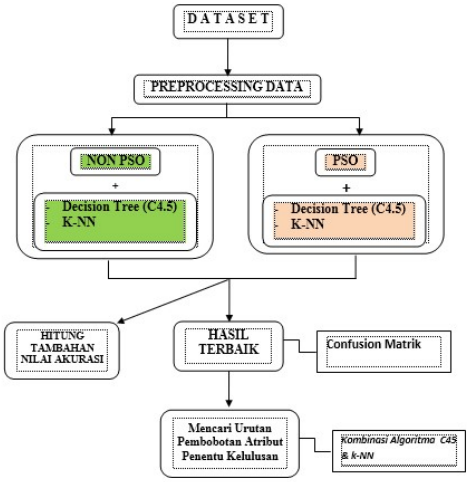
CLASSIFICATION	PREDICTED CLASS		
		Class = YES	Class = NO
OBSERVED CLASS	Class = YES	*A* (True Positives – TP)	*b* (False Negatives – FN)
	Class = NO	*c* (False Possitives – FP)	*d* (True Negatives – TN)

Untuk menghitung nilai *accuracy* menggunakan rumus[13]:

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

3. METODE PENELITIAN

Metodologi 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

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.

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 3 dibawah ini:

Tabel 3: Attribut 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 Menikahp
		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 Dicretization***, 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

Fase Pemodelan

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

1. Menghitung nilai akurasi dari masing-masing algoritma (*Decesion Tree dan k-NN)*, dan dicari algoritma mana yang menghasilkan nilai akurasi tertinggi.
2. Menghitung nilai akurasi dari masing-masing algoritma (*Decesion Tree dan k-NN)* setelah penambahan fitur seleksi *Particle Swarm Optimization (PSO)*, dan dicari algoritma mana yang menghasilkan nilai akurasi tertinggi.
3. Menghitung nilai penambahan akurasi sebelum dan sesudah penambahan fitur seleksi *Particle Swarm Optimization (PSO)*
4. Mencari urutan pembobotan atribut penentu kelulusan berdasarkan kombinasi 2 algoritma (C4.5 dan k-NN).

4. HASIL DAN PEMBAHASAN

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 11 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.

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, dinyatakan Lulus Tepat Waktu
b. Jika Lama Studi > 4,5 tahun, dinyatakan Lulus Tidak Tepat Waktu, akhirnya diperoleh dataset baru seperti pada Tabel 4.

Tabel 4: Dataset Untuk Analisa *Algoritma Decesion Tree dan 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	P	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	P	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

Eksperimen Dan Pengujian Model

Perhitungan 2 Algoritma Non Berbasis PSO

Perhitungan dengan menggunakan ***Algoritma Decesion Tree*** diperoleh nilai akurasi tertinggi sebesar **64,02%** seperti pada Tabel 5 di bawah ini.

Tabel 5: Perhitungan *Algoritma Decesion Tree*

Accuracy : 64,02% +/-5,13% (mikro:63,79%)			
	True Tepat Waktu	**True Tidak Tepat Waktu**	**Class Precision**
Pred.Tepat Waktu	**245**	**155**	**61,25%**
Pred.Tidak Tepat Waktu	**149**	**296**	**66,52%**
Class Recall	**62,18%**	**65,63%**	

Perhitungan dengan menggunakan ***Algoritma k-NN (k=15)*** diperoleh nilai akurasi tertinggi sebesar **69,81%** seperti pada Tabel 6 di bawah ini.

Tabel 6: Perhitungan *Algoritma k-NN (k=15)*

Accuracy : 69,81% +/-5,83% (mikro:69,82%)			
	True Tepat Waktu	True Tidak Tepat Waktu	Class Precision
Pred.Tepat Waktu	264	125	67,87%
Pred.Tidak Tepat Waktu	130	326	71,49%
Class Recall	67,01%	72,28%	

Dari Tabel 5 dan 6 di atas, diketahui bahwa *Algoritma k-Nearest Neighbor (k=15)* menghasilkan pengujian yang lebih baik dari pada *Algoritma Decesion Tree* dengan nilai accuracy sebesar **69,81%**.

Perhitungan 2 Algoritma Berbasis PSO

Perhitungan dengan menggunakan ***Algoritma Decesion Tree Berbasis PSO*** diperoleh nilai akurasi tertinggi sebesar **69,82%** seperti pada Tabel 7 di bawah ini.

Tabel 7: Perhitungan *Algoritma Decesion Tree+PSO*

Accuracy : 69,82%			
	True Tepat Waktu	True Tidak Tepat Waktu	Class Precision
Pred.Tepat Waktu	283	144	66,28%
Pred.Tidak Tepat Waktu	111	307	73,44%
Class Recall	71,83%	68,07%	

Perhitungan dengan menggunakan ***Algoritma k-NN (k=19) Berbasis PSO*** diperoleh nilai akurasi tertinggi sebesar 76**,69%** seperti pada Tabel 8 di bawah ini.

Tabel 8: Perhitungan *Algoritma k-NN (k=19) + PSO*

Accuracy : 76,69%			
	True Tepat Waktu	True Tidak Tepat Waktu	Class Precision
Pred.Tepat Waktu	290	93	75,72%
Pred.Tidak Tepat Waktu	104	358	77,49%
Class Recall	73,60%	79,38%	

Dari Tabel 7 dan 8, diketahui bahwa *Algoritma k-Nearest Neighbor (k=19)* menghasilkan pengujian yang lebih baik dari pada *Algoritma Decision Tree (C4.5)* dengan nilai accuracy sebesar **76,69%**.

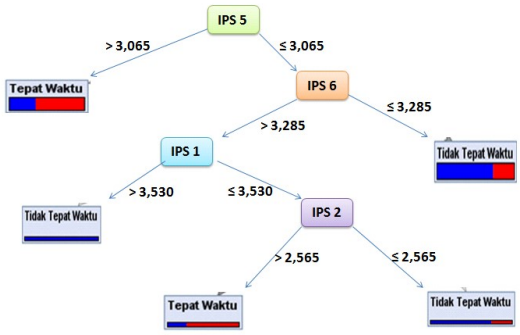
Pembobotan Atribut Algoritma Naïve Bayes Dan k-NN Berbasis PSO

Untuk mengetahui nilai pembobotan dari masing-masing variabel yang berpengaruh terhadap kelulusan tepat waktu mahasiswa, maka dapat dilihat pada Tabel 9 di bawah ini:

Tabel 9: Pembobotan Atribut *Algoritma Decesion Tree (C4.5) dan k-NN*

Atribut	*C4.5 + PSO*	*k-NN + PSO*
Jenis Kelamin (JK)	0,623 (4)	1,000 (1)
Umur	0,863 (3)	0,000 (6)
IP Semester 1	0,565 (5)	0,479 (4)
IP Semester 2	0,069 (10)	1,000 (1)
IP Semester 3	0,221 (8)	1,000 (1)
IP Semester 4	0,485 (7)	1,000 (1)
IP Semester 5	0,182 (9)	0,873 (3)
IP Semester 6	0,968 (1)	1,000 (1)
IP Semester 7	0,562 (6)	0,900 (2)
Status Pekerjaan	0,937 (2)	0,317 (5)
Status Pernikahan	0,062 (11)	1,000 (1)

Dari Tabel 9 di atas dapat diketahui bahwa pembobotan atribut terbesar terletak pada atribut IP Semester 6, baik dengan menggunakan *Algoritma Decesion Tree (C4.5)* maupun *Algoritma k-NN*. Pembobotan terkecil pada *Algoritma Decesion Tree (C4.5)* terletak pada atribut Status Pernikahan (0,062), sedangkan pada *Algoritma k-NN* terletak pada atribut Umur (0,000). Untuk mengetahui atribut yang berpengaruh terhadap kelulusan mahasiswa bisa dilihat melalui Rule *Algoritma Decision Tree (C4.5)*, seperti Gambar 2 di bawah ini:



Gambar 2: Pohon Keputusan Kelulusan Mahasiswa Menggunakan Algoritma C4.5

Dari Gambar 2 di atas dapat diuraikan: IPS 5 mempunyai pengaruh yang kuat (node 1), jika Indeks Prestasi Semester (IPS) 5 > 3,065 maka akan lulus tepat waktu, jika IPS 5 ≤ 3,065 akan dilihat IPS 6, jika IPS 6 ≤ 3,285 maka lulus tidak tepat waktu, jika IPS 6 > 3,285 akan

dilihat IPS 1, jika IPS 1 > 3,530 maka lulus tidak tepat waktu, jika IPS 1 ≤ 3,530 akan dilihat IPS 2, jika IPS 2 > 2,565, maka lulus tepat waktu, jika IPS 2 ≤ 2,565, maka lulus tidak tepat waktu.

Peningkatan Nilai Akurasi

Berdasarkan Tabel 5 sampai 8, akhirnya diperoleh hitungan peningkatan nilai akurasi seperti pada Tabel 10 di bawah ini:

Tabel 10. Peningkatan Nilai Akurasi

ALGORITMA	NON PSO	+ PSO	Peningkatan Nilai Akurasi
Decesion Tree	64,02%	69,82%	5,80%
k-NN	69,81%	76,69%	6,88%

Berdasarkan Tabel 10 di atas, peningkatan nilai akurasi *Algoritma k-NN* lebih tinggi dari pada *Algoritma Decesion Tree (C4.5)* yaitu sebesar 6,88%. Dari Tabel 10 juga diketahui bahwa nilai akurasi sebelum dan sesudah penambahan fitur PSO, pada *Algoritma k-NN* nilai akurasinya lebih tinggi.

Penentuan Algoritma Terbaik

Dari Tabel 10 di atas, dapat ditarik kesimpulan bahwa *Algoritma k-NN Berbasis PSO* menghasilkan pengujian lebih baik dengan nilai ***accuracy*** sebesar **76,69%** (pada k-optimum=19).

Untuk mengetahui pembobotan masing-masing atribut, maka diterapkan pada ***Algoritma k-NN + PSO (k=19)*** sebagai algoritma yang memberikan nilai akurasi lebih baik. Tabel 9 menunjukkan nilai pembobotan masing-masing atribut adalah sebagai berikut:

1. Atribut IPS 6 merupakan atribut yang mempunyai nilai pembobotan tertinggi, termasuk pada *Algoritma Decesion Tree (C4.5)*
2. Nilai pembobotan Atribut Umur (=0,000), artinya tidak memberikan kontribusi yang nyata terhadap kelulusan tepat waktu mahasiswa.
3. Atribut Jenis Kelamin, IPS 2, IPS 3, IPS 4, IPS 6 dan Status Pernikahan memberikan kontribusi yang nyata terhadap kelulusan tepat waktu mahasiswa (=1,000).
4. Atribut IPS 1 dan Status Pekerjaan tetap memberikan kontribusi terhadap kelulusan tepat waktu meskipun nilai pembobotannya kurang dari **0,500**.

Pada *Algoritma k-NN*, ada 6 atribut yang mempunyai nilai pembobotan yang sama (=1,000), untuk mencari atribut mana yang lebih dominan dapat dibantu dengan melihat nilai pembobotan pada *Algoritma Decesion Tree,* dalam hal ini yang tertinggi ada pada atribut IP Semester 6, sehingga bisa disimpulkan nilai pembobotan IP Semester 6 lebih punya pernanan penting dalam kelulusan tepat waktu mahasiswa.

5. KESIMPULAN

Hasil analisis menggunakan *Confusion Matrix* dapat disimpulkan bahwa ***Algoritma k-Nearest Neighbor (k-NN) Berbasis PSO*** memiliki kinerja lebih baik untuk klasifikasi bidang kerja lulusan dibanding ***Algoritma Decesion Tree (C4.5)*** dengan nilai ***Accuracy*** sebesar **76,69%.** Peningkatan nilai akurasi *Algoritma k-NN* (=6,88%) lebih besar daripada *Algoritma Decesion Tree* (=5,80%). Nilai akurasi sebelum dan sesudah penambahan fitur PSO lebih tinggi pada *Algoritma k-NN*. Atribut Indek Prestasi Semester (IPS) 6 mempunyai nilai pembobotan tertinggi baik dengan *Algoritma Decesion Tree (C4.5)* maupun *Algoritma k-NN*. Atribut Umur mempunyai nilai pembobotan sebesar 0,000 artinya tidak memberikan pengaruh yang kuat terhadap kelulusan tepat waktu mahasiswa.

Berkaitan untuk penelitian selanjutnya perlu dicoba dengan menggunakan Regresi Linier Berganda guna mengetahui pembobotan masing-masing atribut dengan melihat koefisien regresi masing-masing atribut tersebut.

6. REFERENSI

[1] Prabowo, *Aneka Teknik, Piranti dan Penerapan Data Mining : Studi Kasus Peramalan Harga Saham Industri Telekomunikasi Berbasis Jaringan Saraf Tiruan*. Modul Perkuliahan Universitas Budi Luhur, 2012.

[2] N. M. Huda, "APLIKASI DATA MINING UNTUK MENAMPILKAN INFORMASI TINGKAT

KELULUSAN MAHASISWA (Studi Kasus di Fakultas MIPA Universitas Diponegoro) Nuqson Masykur Huda PROGRAM STUDI TEKNIK INFORMATIKA," *Tek. Inform. MIPA Univ. Diponegoro*, 2010.

[3] R. Asif, A. Merceron, dan M. K. Pathan, "Predicting Student Academic Performance at Degree Level : A Case Study," no. December 2014, hal. 49–61, 2015.

[4] M. A. Banjarsari, H. I. Budiman, dan A. Farmadi, "Penerapan K-Optimal Pada Algoritma Knn untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan IP Sampai Dengan Semester 4," vol. 02, no. 02, hal. 50–64, 2015.

[5] L. Hadjaratie, "PREDIKSI DAN PEMETAAN DATA MAHASISWA FAKULTAS TEKNIK UNIVERSITAS NEGERI GORONTALO MENGGUNAKAN PENDEKATAN DATA MINING," *Tek. Inform. Univ. Negeri Gorontalo*.

[6] Nursalim, Suprapedi, dan H.Himawan, "KLASIFIKASI BIDANG KERJA LULUSAN MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR," *J. Teknol. Inf.*, vol. 10, no. April, hal. 31–43, 2014.

[7] C. Anuradha dan T. Velmurugan, "A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Students Performance," vol. 8, no. July, hal. 1–12, 2015.

[8] D. H. Kamagi dan S. Hansun, "Implementasi Data Mining dengan Algoritma C4 . 5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," vol. VI, no. 1, hal. 15–20, 2014.

[9] A. R. Alfarisi, H. Tjandrasa, dan I. Arieshanti, "Perbandingan Performa antara Imputasi Metode Konvensional dan Imputasi dengan Algoritma Mutual Nearest Neighbor," vol. 2, no. 1, hal. 1–4, 2013.

[10] dan K. S. Muhamad Hanief Meinanda, Metri Annisa, Narendi Muhandri, "Prediksi Masa Studi Sarjana dengan Artificial Neural Network," *INTERNETWORKING Indones. J.*, vol. 1, no. 2, hal. 31–35, 2009.

[11] Jefri, "IMPLEMENTASI ALGORITMA C4.5 DALAM APLIKASI UNTUK MEMPREDIKSI JUMLAH MAHASISWA YANG MENGULANG MATA KULIAH DI STMIK AMIKOM YOGYAKARTA," *STMIK AMIKOM Yogyakarta*, 2013.

[12] C. Agustina, "Optimasi Naive Bayes Menggunakan Particle Swarm Optimization Untuk Meningkatkan Akurasi Deteksi Autisme Spectrum Disorder," *J. Speed (Sentra Penelit. Eng. dan Edukasi)*, vol. 10, no. 2, hal. 10–14, 2018.

[13] M. Bramer, *Principles of Data Mining*, 2 ed. London, 2013.

[14] C. Vercellis, *Business Intelligence: Data Mining and Optimization for Decision Making*, 1 ed. United Kingdom: John Wiley & Sons Ltd, 2009.

[15] F. Gorunescu, *Data Mining Concepts, Models and Techniques*, 12 ed. Verlag Berlin Heidelberg: Springer, 2011.