ISSN 2502-1613/E-ISSN 2541-3740

PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN *K-NEAREST NEIGHBOR* BERBASIS *PARTICLE SWARM OPTIMIZATION*

Nursetia Wati
e-mail: nursetiawati@umgo.ac.id

[1] Prodi Sistem Informasi Univesitas Muhammadiyah Gorontalo

Abstract— Completion of studies of students in a timely manner is one measure of the quality of higher education, as well as in finding a job. Anticipation that can be done is by predicting graduation, with these predictions do evaluation efforts in organizing lectures at the faculty or study program. The data in this study is the student data Gorontalo State University Faculty of Education and Faculty of Engineering from 2008 until 2012. From 5104 the total number of records is done with attribute data sorting empty, so the existing data into 2312 record. The algorithm used in this study is a K-Nearest Neighbor which will then be optimized using Particle Swarm Optimization. By using the technique Fold Cross Validation on K-Nearest Neighbor algorithm produces the highest accuracy 88.58 on the value of k = 14. The next test using particle swarm optimization algorithm to get the highest accuracy on the population size = 10 with accuracy of 89.14%.

Intisari— Penyelesaian studi mahasiswa tepat waktu merupakan salah satu tolak ukur kualitas pendidikan tinggi, juga dalam mencari pekerjaan. Antisipasi yang dapat dilakukan adalah dengan memprediksi kelulusan, dengan prediksi tersebut dilakukan upaya evaluasi dalam penyelenggaraan perkuliahan di fakultas atau program studi. Data dalam penelitian ini adalah data mahasiswa Fakultas Ilmu Pendidikan Universitas Negeri Gorontalo dan Fakultas Teknik Universitas Negeri Gorontalo dari tahun 2008 sampai dengan tahun 2012. Dari 5104 jumlah record dilakukan pemilahan data atribut kosong, sehingga data yang ada menjadi 2.312 record. Algoritma yang digunakan dalam penelitian ini adalah K-Nearest Neighbor yang selanjutnya akan dioptimasi menggunakan Particle Swarm Optimization. Dengan menggunakan teknik Fold Cross Validation pada algoritma K-Nearest Neighbor menghasilkan akurasi tertinggi 88,58 pada nilai k = 14. Pengujian selanjutnya menggunakan algoritma particle swarm optimization untuk mendapatkan akurasi tertinggi pada populasi size = 10 dengan akurasi 89,14%.
Kata Kunci— K- Nerest Neighbor, Particle Optimazation, Prediksi Kelulusan

I. PENDAHULUAN

Berdasarkan data pada Fakultas Ilmu Pendidikan Univeristas Negeri Gorontalo strata D2 lulusan tahun 2006 sampai dengan tahun 2012 menunjukan rata-rata presentasi kelulusan tepat waktu atau menempuh studi dalam kurun waktu dua tahun yaitu sebesar 89% dan untuk presentasi mahasiswa yang melebihi waktu studi atau lulus tidak tepat waktu sebesar 10%, sedangkan untuk S1 yang ditempuh dalam kurun waktu 4 tahun dengan lulusan sejak tahun 2005 sampai tahun 2009 sebesar 77% dan untuk mahasiswa yang melebihi jangka waktu studi atau lulus tidak tepat waktu sebesar 28%. Untuk data mahasiswa pada Fakultas Teknik strata D3 lulusan tahun 2006 sampai tahun 2012 yang ditempuh dalam kurun waktu 3 tahun rata-rata presentasi yaitu sebesar 20% dan mahasiswa lulus tidak tepat waktu atau melebihi jangka waktu studi yang telah

ISSN 2502-1613/E-ISSN 2541-3740

ditetapkan sebesar 71%, sedangkan pada strata S1 lulusan pada tahun 2007 sampai dengan tahun 2012 menunjukan presentasi mahasiswa yang lulus tepat waktu yaitu sebesar 73% dan presentasi kelulusan untuk mahasiswa tidak tepat waktu yaitu sebesar 18% [1]. Dengan data yang ada terlihat presentasi kelulusan tidak tepat waktu terutama pada Fakultas Teknik sangatlah tinggi, kondisi tersebut mendorong setiap program studi terus meningkatkan kelulusan agar saran mutu bisa tercapai.

Dalam memprediksi kelulusan mahasiswa sangatlah penting untuk menjadi bahan pertimbangan bagi fakultas agar dapat melakukan tindak lanjut dalam menghadapi mahasiswa yang nantinya berpotensi lulus tidak tepat waktu. Dan berdasarkan latar belakang permasalahan yang ada, terlihat dalam penelitian sebelumnya yang menggunakan algoritma *K-Nearest Neighbor* hasil akurasi yang didapatkan masih terbilang kurang. Untuk itu diperlu ditambahkan algoritma *Particle Swarm Optimization* untuk lebih meningkatkan akurasi dalam memprediksi kelulusan mahasiswa yang nantinya dengan nilai yang didapatkan dapat menjadi parameter tersendiri untuk fakultas agar bisa mengatasi mahasiswa yang berpotensi lulus tidak tepat waktu.

II. KAJIAN PUSTAKA

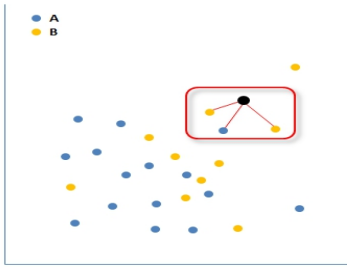
A. Data Mining

Alasan utama mengapa data mining diperlukan adalah karena adanya sejumlah besar data yang dapat digunakan untuk menghasilkan informasi dan *knowledge* yang berguna. Informasi dan *knowledge* yang didapat tersebut dapat digunakan pada banyak bidang, mulai manajemen bisnis, control produksi, kesehatan, dan lain-lain. Data mining adalah suatu proses pengerukan atau pengumpulan informasi penting dari suatu data yang besar. Proses data mining seringkali menggunakan metode statistika, matematika, hingga memanfaatkan teknologi artificial intelligence. Nama alternatifnya yaitu Knowledge discovery (mining) in databases (KDD), knowledge extraction, data/pattern analysis, data archeology, data dredging, information harvesting, business intelligence, dan lain-lain.

B. K-Nearest Neighbor (KNN)

Dalam *K-Nearest Neighbor*, kita mengelompokkan suatu data baru bedasarkan jarak data baru ke beberapa data/tetangga (*Neighbor)* terdekat. Dalam hal ini jumlah data/tetangga terdekat di tentukan oleh *user* yang diyatakan dengan k [4] . *K-Nearest Neighbor* merupakan salah satu model yang digunakan dalam algoritma *supervised* yang hasil instance diklasifikasikan berdasarkan mayoritas dari kategori pada *K-Nearest Neighbor. K-Nearest Neighbor* bekerja bersadarkan jarak minimum dari data baru ke tetangga terdekat yang sudah ditetapkan. Tujuan dari algoritma ini adalah mengklasifikasikan obyek baru berdasarkan atribut dan training sample. Jauh dekatnya tetangga biasanya dihitung berdasarkan *Euclidean Distance*

ISSN 2502-1613/E-ISSN 2541-3740



Gambar1. Cara Kerja Algoritma KNN

1. Mentukan parameter K, K merupakan Tetangga terdekat
2. Setelah mendapatkan K, kemudian hitung jarak antara data baru dengan semua data training
3. Urutkan jarak tersebut dan tetapkan tetangga terdekat berdasarkan jarak minimum ke-K
4. Periksa kelas dari tetangga terdekat
5. Gunakan mayoritas sederhana dari kelas tetangga terdekat sebagai nilai prediksi data baru.

Berikut merupakan penjelasan dari parameter yang digunakan dalam persamaan K-NN [5]

Tabel 1 Parameter- Parameter K-NN

Parameter	Keterangan
Sample	Matriks dimana baris merupakan data, kolom merupakan fiture. Sample merupakan data uji yang akan diklasifikasikan ke dalam kelas. Matriks Sample harus mempunya jumlah kolom (fitur) yang sama dengan matriks *traning*
Traning	Matriks yang digunakan untuk mengelompokan baris di dalam matriks sample. Matriks traning harus mempunya jumlah kolom yang sama dengan matriks sample. Setiap baris dalam matriks traning mempunyai relasi kelas pada baris yang sama di matriks group.
Group	Vector (matriks 1 kolom) yang setiap barisnya menyatakan kelas dari baris yang sama matriks traning
K	Jumlah tetangga terdekat yang digunakan untuk klasifikasi. Nilai defaulnya dalah 1
Distance	String yang menyatakan matriks jarak yang digunakan untuk mencari tetangga terdekat. Pilihannya „euclidean", jarak Euclidean (default)"cityblok" jarak Manhattan atau jumlah absolute perbedaan nilai antar fitur „cosine" jarak 1 – cos (sudut antara dua titik) „correlation", jarak 1 – korelasi di antara titik (nilai sekuen) „hamming", jarak presentase bit yang berbeda (cocok untuk data biner).

C. Particle Swarm Optimization (PSO)

Algoritma *Particle Swarm Optimization.* (PSO) telah terinspirasi dari tingkah laku hewan koloni seperti, lebah, burung dan rayap. Ada beberapa parameter pada PSO yaitu berupa kecepatan maksimum, posisi, berat inersia kecepatan dan konstanta percepatan. Khoirul Mu"arif [3] mengatakan PSO adalah model optimasi heuristic global yang diperkenalkan oleh Kennedy Eberthart pada tahun 1995. Perilaku terhadap kawasan burung dan ikan. Partikel pada PSO juga sering dikaitkan dengan kecepatan partikel terbang melalui ruang pencarian dengan kecepatan dinamis yang disesuaikan untuk

perilaku histori mereka. Setiap partikel dalam PSO Juga dikaitkan dengan kecepatan partikel terbang melalui ruang pencarian dengan kecepatan yang dinamis disesuaikan untuk perilaku histori mereka. Oleh karena itu, partikel memiliki kecendurungan untuk terbang menujudaerah pencarian yang lebih baik dan lebih baik selama proses pencarian. Secara singkat PSO dimulai dari inisialisasi populasi hingga penghentiankomputasi, seperti algoritma berikut [5].

a. Inisialisai populasi (populasi dan kecepatan acak) dalam *hyperspace*
b. Evaluasi *fitness* partikel individu
c. Memodifikasi kecepatan berdasarkan terbaik sebelumnya (previous best: pbest) dan terbaik global atau local (global or neighborhood best: gbest or lbest)
d. Hentikan berdasarkan beberapa kondisi
e. Kembali ke langkah (b)

Pencarian solusi optimal dalm PSO akan dilakukan sampai semua memiliki skemasolusi yang sama atau ketika iterasi tertinggi telah tercapai.

D. Confusion Matrix

Dalam melakukan evaluasi terhadap suatu model klasifikasi berdasarkan perhitungan objek testing mana yang diprediksi benar dan mana yang diprediksi tidak benar. Perhitungan tersebut digambarkan dalam tabel yang disebut *confusion matrix*. *Confusion matrix* merupakan data set yang hanya memiliki dua kelas, kelas yang satu sebagai positif dan kelas yang lain sebagai negative. Terlihat pada gambar di bawah tabel *confusion matrix* [4]*:*

Tabel 2 *Confusion Matrix* untuk dua kelas

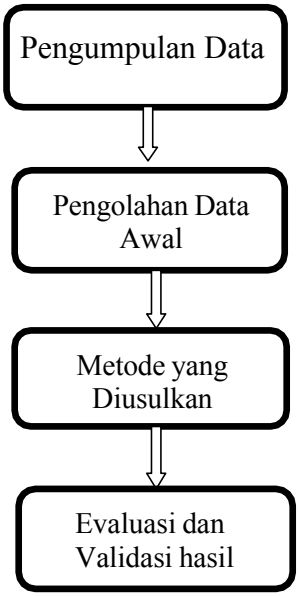
CLASSIFICATION		PREDICTED CLASS	
		Class = Yes	Class = No
OBSERVED CLASS	**Class = Yes**	a (*true positive* – TP)	b (*false negative* – FN)
	Class = No	c (*false positive* – FP)	d (*true negatife* – TP)

Berdasarkan tabel di atas *true positive* adalah jumlah *record positive* yang diklasifikasikan sebagai *positive*, *false positive* adalah jumlah *record negative* yang diklasifikasikan sebagai *positive*, *false negatives* adalah jumlah *record positive* yang diklasifikasikan sebagai *negative*, *true negatives* adalah jumlah *record negatif* yang diklasifikasikan sebagai *negative*, setelah itu masukkan dataset. Setelah dataset dimasukkan ke dalam *confusion matrix*, akan menghasilkan nilai seperti jumlah *sensitivity* (*recall*), *specificity*, *precision* dan *accuracy*. *Sensitivity* digunakan untuk membandingkan jumlah TP terhadap jumlah *record* yang positif sedangkan *specificity* adalah perbandingan jumlah TN terhadap jumlah *record* yang negatif.

ISSN 2502-1613/E-ISSN 2541-3740

III. Metode Penelitian

Pada penelitian ini digunakan metode ekperimen, dimana dapat dilihat pada gambar di bawah ini:



Gambar 2. Langkah Metode Penelitian

A. *Pengumpulan Data*

Data yang digunakan pada penelitian ini berasal dari penelitian Wandira Irene [3] untuk mendapatkan data mahasiswa. Dari data yang telah dikumpulkan terdapat jumlah variable data, untuk Fakultas Ilmu Pendidikan 4242 *record* dan Fakultas Teknik 862 *record* yang masing-masing fakultas memiliki atribut bernilai nominal dan pastinya setiap tahun mengalami perubahan pada jumlah *record*. Dan dataset dari kedua fakultas ini telah digabungkan menjadi satu yang memiliki jumlah 5.104 *record.*

B. *Pengolahan Data Awal*

Setelah melakukan pengumpulan data yang menghasilkan 4242 *record* untuk Fakultas Ilmu Pendidikan dan 862 *record* untuk Fakultas Teknik yang seluruhannya berjumlah 5.104 *record*. Dalam pengolahan data (*Preparation Data)* awal ada beberapa proses yang harus di lakukan yaitu:

1. Pembersihan Data Pembersihan Data
   Dalam proses ini dilakukan pembersian data dengan cara menghapus data yang kosong atau tidak terisi pada beberapa atribut, dan data yang sulit diolah saat melakukan pengujian, pembersian data ini dilakukan untuk mendapatkan dataset yang berkualitas.

ISSN 2502-1613/E-ISSN 2541-3740

2. Menentukan Atribut

Setelah melakukan penghapusan data, langkah selanjutnya yaitu menentukan atribut yang digunakan saat pengujian nanti. Sebelumnya pada dataset mahasiswa ini terdapat 16 atribut yang diantranya Nama, Nim, Alamat, dan seterusnya. 16 atribut tersebut tidak semuanya digunakan karena dalam memprediksi nanti atribut seperti Nama, Nim dan Alamat tidak mempengaruhi hasil prediksi. Atribut- atribut yang digunakan terlihat pada tabel di bawah ini:

Tabel 3. Kedudukan Atribut yang akan digunakan

Atribut	**Type**	
Nama Mhs	×	Nominal
Strata	√	Nominal
Sex	√	Nominal
Jurusan	√	Nominal
Kelas	√	Nominal
Seleksi	√	Nominal
Pekerjaan Ayah	×	Nominal
Pekerjaan Ibu	×	Nominal
Pendidikan Ayah	×	Nominal
Pendidikan Ibu	×	Nominal
Penghasilan Ayah	√	Nominal
Penghasilan Ibu	√	Nominal
Asal Sekolah	×	Nominal
Asal Daerah	×	Nominal
IPK	√	Nominal
Lama Studi	√	Nominal

C. .Evaluasi dan Validasi Hasil

Pada tahap evaluasi akurasi diukur dengan *confusion matrix* untuk pengukuran model, Hasil dari *confusion matrix* dapat menggambarkan berupa hasil akurasi yaitu, tepat waktu dan tidak tepat waktu. Analisis terhadap hasil dari hasil evaluasi *confusion matrix* berupa nilai *Accuracy* yang dapat dihitung dengan persamaan **(2.10)**, dan tabel *Confusion Matrix* terlihat pada gambar di bawah ini:

Tabel 4. Contoh *confusion matrix* K-NN

Confusion Matrix	True TEPAT WAKTU	True TIDAK TEPAT WAKTU
Pred. TEPAT WAKTU	TP	FN
Pred. TIDAK TEPAT WAKTU	FP	TP

IV. Hasil Penelitian

A. Sebelum menentukan bobot perlu dilakukan proses penentuan *traning* dan *testing* mengunakan Algoritma *K-Nearest Neighbor*. Dalam pengujian ini peneliti menggunakan *Cross Validation*, dimana teknik pengujiannya dengan cara membagi dua dataset secara acak yang dalam pencariannya menghasilkan nilai *accuracy* pada algoritma *K-Nearest Neighbor* dengan menggunakan *number of validation.* Hasil dari pengujian berdasarkan perolehan nilai *accuracy* terlihat dari nilai *accuracy* K tertinggi terdapat pada K ke 14 yaitu berjumlah 88.58% yang tingkat akurasinya dipengaruhi oleh dataset yang digunakan, dan hasil *accuracy* tertinggi dapat digambarkan dalam tabel *confusion* matrix sebagai berikut:

Tabel 5. Tingkat *accuracy* Algoritma *K-Nearest Neighbor* dengan Nilai K = 14

accuracy: 88.58% +/- 1.97% (mikro: 88.58%)			
Klasifikasi Nilai K=14	true TIDAK TEPAT WAKTU	true TEPAT WAKTU	class precision
pred. TIDAK TEPAT WAKTU	272	70	79.53%
pred. TEPAT WAKTU	193	1768	90.16%
class recall	58.49%	96.19%	

B. Algoritma *Particle Swarm Optimization* digunakan untuk pembobotan atribut dengan melakukan proses pencarian nilai K terbaik berdasarkan algoritma *K-Nearest Neighbor.* Setelah mendapatkan nilai *accuracy* terbaik dengan mengubah parameter pada *K-Nearest Neighbor* dilanjutkan dengan menentukan inisialisasi partikel pada data (x), inisialisasi kecepatan partikel (v) secara random. Untuk mendapatkan parameter terbaik yang terbentuk perlu di lakukan Evaluasi *fitness* dari masing-masing partikel berdasarkan posisinya, kemudian tentukan partikel dengan *fitness* terbaik, dan tetapkan sebagai *Gbest*. Untuk setiap partikel, *Pbest* awal akan sama dengan posisi awal. Dalam Proses pembobotan atribut yang berjumlah 2312 *record* dengan menggunakan status kelulusan mahasiswa yang menjadi lebel dari 9 atribut yang terpilih dalam menentukan hasil klasifikasi dalam memprediksi kelulusan mahasiswa. Atribut-atribut tersebut memiliki bobot disetiap atribut yaitu seperti di bawah ini:

Tabel 6. Hasil pembobotan atribut mengunakan K terbaik pada Algoritma *Particle Swarm Optimization*

Atribute	Weight
Strata	1.0
Sex	0.689
Jurusan	0.952
Kelas	1.0
Seleksi	0.084
Penghasilan Ayah	0.696
Penghasilan Ibu	0.566
IPK	0.944

C. Dengan menggunakan *populasi size 5*, *maximum number of generations 30* (default), *inertia weight 1.0* (default), *local bestweight 1.0* (default), *global best weight 1.0* (default) dan *min weight 0.0* (default) mendapatkan hasil untuk *weight* 1 merupakan atribut yang paling berpengaruh pada proses klasifikasi dalam memprediksi kelulusan mahasiswa seperti atribut Strata dan kelas, untuk atribut Jurusan dan IPK juga berpengaruh yang terlihat pada hasil *weight* 0.9 dan untuk atribut lainnya kurang berpengaruh karena memiliki *weight* di bawah 0.9. Hasil dari pembobotan tersebut selanjutnya digunakan untuk pengujian nilai terbaik pada *Particle Swarm Optimization*. Dan untuk hasil akurasi pada algoritma *Particle Swarm Optimization* menggunakan nilai K terbaik dengan mengubah *Populasi Size* yang dapat di lihat pada tabel C*onfusion Matrix* di bawah ini:

Tabel 6. *Confusion Matrix* (Accuracy) data Mahasiswa dengan *Population Size* 10

accuracy: 89.14% +/- 1.03% (mikro: 89.14%)			
Population Size Ke 10	true. TIDAK TEPAT WAKTU	true. TEPAT WAKTU	class precision
pred. TIDAK TEPAT WAKTU	266	51	83.91%
pred. TEPAT WAKTU	199	1787	89.98%
Class recall	57.20%	97.23%	

Untuk memperoleh persentase hasil akurasi Algoritma *Particle Swarm Optimization* dengan Nilai K = 14 dan mengubah *populasi size* 10 dapat dicari dengan persamaan seperti pada pencarian akurasi Algoritma *K-Nearest Neighbor* dengan Nilai K= 14 di atas.

ISSN 2502-1613/E-ISSN 2541-3740

V. Kesimpulan

Berdasarkan hasil tahapan pengujian yang dilakukakan untuk mencari nilai optimasi terbaik dari dataset yang diteliti, dapat disimpulkan nilai terbaik saat menggunakan Algoritma *K-Nearest Neighbord* terdapat pada K=14 yaitu sejumlah 88.58%. Namun saat menggunakan Algoritma *Particle Swarm Optimization* dengan patameter *defult* dapa lebih meningkatkan nilai yaitu sejumlah 88.97%. Namun, saat menggunakan nilai K terbaik dan merubah parameter *populasi size* pada algoritma *Particle Swarm Optimization* untuk mencari nilai terbaik yang terdapat pada *populasi size* ke 10 mendapatkan nilai *accuracy* lebih tinggi sebesar 89.14%, jadi saat menggunakan Algoritma *K-Nearest Neighbord* dan Algoritma *K-Nearest Neighbord* berbasis *Particle Swarm Optimization* dengan mencari nilai terbaik meningkat sebesar 0.56%, terlihat dari hasil *accuracy* 88.58% dan meningkat menjadi 89.14%. Tinggkat keakurasian biasanya dapat berubah – ubah sesuai dataset yang kita gunakan, semakin berkualitas sebuh dataset, semakin tinggi nilai *accuracy* yang didapatkan.

ISSN 2502-1613/E-ISSN 2541-3740

VI. REFERENSI

[1] UNIVERSITAS NEGERI GORONTALO, *Pedoman Akademik*. UNGPress, 2013.
[2] M. A. Banjarsari, H. I. Budiman, and A. Farmadi, "Penerapan K-OptimalPada Algoritma Knn untuk Prediksi Kelulusan Tepat Waktu MahasiswaProgram Studi Ilmu Komputer Fmipa Unlam Berdasarkan IPSampai Dengan Semester 4," vol. 2, no. 2, pp. 50–64, 2015.
[3] L. H. Wandira Irene, Mukhlisulfatih Latief, "Penerapan Algoritma C5.0 Dalam Pengklasifikasian Data Mahasiswa Universitas Negeri Gorontalo," 2014.
[4] K. Mu‟arif, "Pemodelan data menggunakan c4.5 dan c4.5 berbasis particle swarm optimization untuk memprediksi kelulusanmahasiswa."
[5] A. Nirfah, "Klasifikasi Resiko Kredit Sepeda Motor Menggunakan Algoritma *K-Nearest Neighbor* Berbasis *Particle SwarmOptimization*," Universitas Dian Nuswantoro, 2016.
[6] https://ilmudatapy.com/algoritma-k-nearest-neighbor-knn-untuk-klasifikasi/.