

# Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining

(Studi Kasus : Data Akademik Mahasiswa STMIK Dipanegara Makassar)

Muhammad Syukri Mustafa<sup>1</sup>, I Wayan Simpen<sup>2</sup>

<sup>1,2</sup> Teknik Informatika STMIK Dipanegara

<sup>1</sup>[moh.syukri@gmail.com](mailto:moh.syukri@gmail.com), <sup>2</sup>[simpen.dp@gmail.com](mailto:simpen.dp@gmail.com)

## Abstrak

*Penelitian ini dimaksudkan untuk melakukan prediksi terhadap kemungkinan mahasiswa baru dapat menyelesaikan studi tepat waktu dengan menggunakan analisis data mining untuk menggali tumpukan histori data dengan menggunakan algoritma K-Nearest Neighbor (KNN). Aplikasi yang dihasilkan pada penelitian ini akan menggunakan berbagai atribut yang klasifikasikan dalam suatu data mining antara lain nilai ujian nasional (UN), asal sekolah/ daerah, jenis kelamin, pekerjaan dan penghasilan orang tua, jumlah bersaudara, dan lain-lain sehingga dengan menerapkan analisis KNN dapat dilakukan suatu prediksi berdasarkan kedekatan histori data yang ada dengan data yang baru, apakah mahasiswa tersebut berpeluang untuk menyelesaikan studi tepat waktu atau tidak. Dari hasil pengujian dengan menerapkan algoritma KNN dan menggunakan data sampel alumni tahun wisuda 2004 s.d. 2010 untuk kasus lama dan data alumni tahun wisuda 2011 untuk kasus baru diperoleh tingkat akurasi sebesar 83,36%.*

*Kata kunci: data mining, algoritma KNN*

## Abstract

*This research is intended to predict the possibility of new students time to complete studies using data mining analysis to explore the history stack data using K-Nearest Neighbor algorithm (KNN). Applications generated in this study will use a variety of attributes in a data mining classified among other Ujian Nasional scores (UN), the origin of the school / area, gender, occupation and income of parents, number of siblings, and others that by applying the analysis KNN can do a prediction based on historical proximity of existing data with new data, whether the student is likely to complete the study on time or not. From the test results by applying the KNN algorithm and uses sample data alumnus graduation year 2004 sd 2010 for the case of a long and alumni data graduation year 2011 for new cases obtained accuracy rate of 83.36%.*

*Keywords: data mining, KNN algorithm*

## 1. Pendahuluan

Perkembangan teknologi informasi berdampak pada semakin besarnya kapasitas media penyimpanan data sehingga semakin memudahkan manusia untuk mengumpulkan dan menyimpan data dari berbagai sumber dengan jangkauan yang amat luas. Meskipun teknologi basis data modern telah menghasilkan media penyimpanan yang besar, teknologi untuk membantu menganalisis, memahami, atau bahkan memvisualisasikan data belum banyak tersedia. Hal inilah yang melatarbelakangi dikembangkannya konsep data mining.

Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Dipanegara Makassar adalah salah satu perguruan tinggi swasta di KOPERTIS Wilayah IX Sulawesi, yang menyelenggarakan pendidikan dan pengajaran sejak tahun 1994/1995 berdasarkan Surat Keputusan Menteri Pendidikan dan Kebudayaan Nomor 057/D/O/1994 tanggal 7 Juli 1994. Sejak berdirinya dalam rentang waktu

tahun 1998 hingga tahun 2013, STMIK Dipanegara telah menghasilkan gudang data akademik dengan jumlah lulusan sebanyak 10.187 orang dan jumlah mahasiswa baru setiap tahunnya rata-rata 1000 orang. Dari jumlah mahasiswa baru tersebut ternyata jumlah mahasiswa yang dapat menyelesaikan kuliah tepat waktu dalam kurung tahun 2008-2011 kurang dari 10% untuk setiap periode wisuda. Hal ini tentunya mempengaruhi penilaian akreditasi sehingga perlu upaya dini untuk mengetahui parameter apa saja yang mempengaruhi seorang mahasiswa dapat menyelesaikan studi tepat waktu.

Hal inilah yang mendasari peneliti untuk merancang suatu sistem yang dapat melakukan prediksi terhadap mahasiswa baru yang mendaftar apakah berpeluang menyelesaikan kuliah tepat waktu atau tidak dengan menggunakan salah satu model data mining yaitu algoritma K-Nearest Neighbor yaitu melakukan analisis klasifikasi pada obyek yang baru berdasarkan  $K$  pada tetangga terdekatnya.

## 2. Tinjauan Pustaka

### 2.1. Penelitian yang relevan

Arief Junanto melakukan penelitian untuk memprediksi kinerja akademik mahasiswa baru dengan membandingkan antara algoritma nearest neighbor dan SLIQ. Dari sistem yang dihasilkan, disimpulkan bahwa algoritma SLIQ dengan teknik pohon keputusan mempunyai tingkat akurasi prediksi yang lebih rendah dibandingkan dengan tingkat akurasi dari penggunaan algoritma nearest neighbor [1].

Lin Ernawati dalam penelitiannya memprediksi status keaktifan studi mahasiswa dengan Algoritme C5.0 dan *K-Nearest Neighbor* (KNN). Penelitian ini menggunakan perangkat lunak Weka sebagai alat bantu dalam proses klasifikasi memberikan hasil bahwa atribut Indeks Prestasi Kumulatif (IPK) adalah atribut yang menentukan status studi mahasiswa. Hasil percobaan memberikan informasi bahwa Algoritme C5.0 lebih baik dibandingkan algoritme KNN [4].

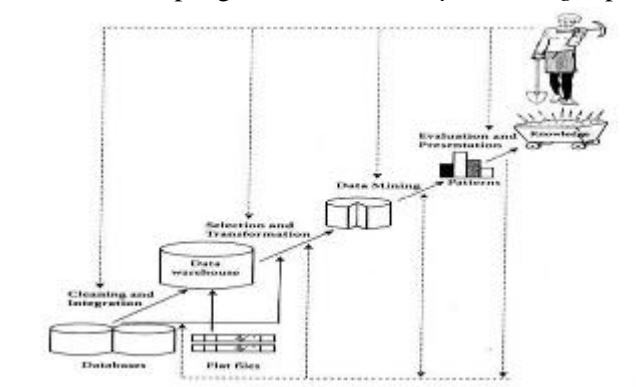
### 2.2. Landasan Teori

#### 2.2.1. Pengertian data Mining

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini

Sedangkan menurut Han dan Kamber, “Data mining adalah proses menambang (*mining*) pengetahuan dari sekumpulan data yang sangat besar”. Data mining merupakan suatu langkah dalam *knowledge discovery in database* (KDD) [3].

Langkah-langkah dalam menemukan pengetahuan (*discovery knowledge*) pada Gambar 1 meliputi :



Gambar 1. Langkah-langkah dalam menemukan pengetahuan  
(Sumber : Han dan Kamber, 2006)

1. *Data cleaning*

Menghilangkan noise dan data yang tidak konsisten.

2. *Data integration*

Menggabungkan berbagai sumber data

3. *Data selection*

Menerima data yang berhubungan dengan analisa dari database.

#### 4. Data transformation

Mengubah data ke bentuk yang sesuai untuk mining dengan melakukan agregasi atau summary.

#### 5. Data mining

Melakukan proses mining untuk mengekstrak data.

#### 6. Pattern evaluation

Mengidentifikasi pola yang menggambarkan pengetahuan (*knowledge*).

#### 7. Knowledge presentation

Menampilkan mined knowledge kepada pengguna.

### 2.2.2. Teknik data Mining

Kegunaan data mining adalah untuk mengklasifikasikan pola yang harus ditemukan dalam data mining. Secara umum, data mining dapat diklasifikasikan dalam dua kategori yaitu deskriptif dan prediktif. [7].

Adapun operasi-operasi dan teknik-teknik yang berhubungan :

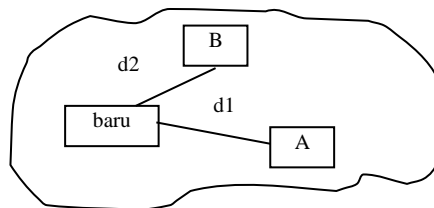
1. Operasi Predictive modeling : (*classification, value prediction*)
2. Database segmentation : (*demographic clustering, neural clustering*)
3. Link Analysis : (*association discovery, sequential pattern discovery, similar timesequence discovery*)
4. Deviation detection : (*statistics, visualization*) [7].

### 2.2.3. Algoritma K-Nearest Neighbor (KNN)

#### A. Pengertian Algoritma KNN

Algoritma *K-Nearest Neighbor* (KNN) adalah merupakan sebuah metode untuk melakukan klasifikasi terhadap obyek baru berdasarkan (K) tetangga terdekatnya. (Gorunescu, 2011). KNN termasuk algoritma *supervised learning*, dimana hasil dari query instance yang baru, diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Kelas yang paling banyak muncul yang akan menjadi kelas hasil klasifikasi [2].

Nearest Neighbor adalah suatu pendekatan untuk menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada. Ilustrasi kedekatan kasus pada Gambar 2. memberikan gambaran tentang proses mencari solusi terhadap seorang pasien baru dengan menggunakan mengacu pada solusi dari pasien terdahulu. Untuk mencari kasus pasien mana yang akan digunakan, maka dihitung kedekatan anatara kasus pasien baru dengan semua kasus pasien lama. Kasus pasien lama dengan kedekatan terbesar-lah yang akan diambil solusinya untuk digunakan pada kasus pasien baru [6].



**Gambar 2. Ilustrasi kedekatan kasus**

(Sumber : Kursini & Emha Taufiq Luthfi, 2009)

Rumus untuk melakukan penghitungan kedekatan antara kedua dua kasus adalah sebagai berikut : [6]

$$\text{Similarity}(T,S) : \frac{\sum_{i=1}^n f(T_i, S_i) * w_i}{w_i}$$

Dimana :

T : kasus baru

S : kasus dalam penyimpanan

n : jumlah atribut tiap kasus

$i$  : atribut individu 1 s/d  $n$

$f$  : fungsi kesamaan atribut  $i$  antara kasus  $T$  dan  $S$

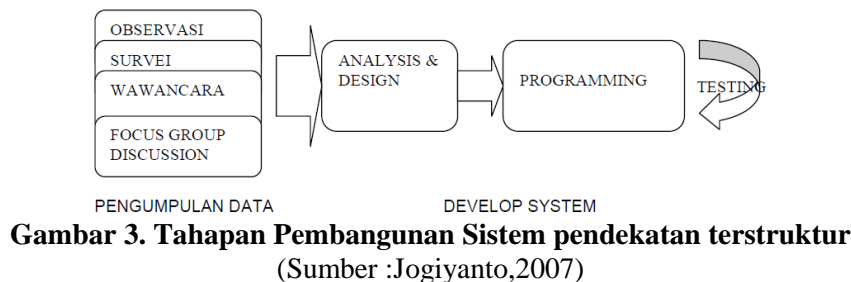
$w$  : bobot pada atribut yang ke  $i$

#### B. Urutan Algoritma KNN [6]

1. Menentukan parameter  $K$  (jumlah tetangga paling dekat)
2. Menghitung kuadrat jarak euclidian (*euclidean distance*) masing-masing obyek terhadap data sampel yang diberikan
$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$
3. Mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak euclid terkecil
4. Mengumpulkan kategori  $Y$  (klasifikasi *nearest neighbor*)
5. Dengan menggunakan kategori mayoritas, maka dapat diprediksikan nilai query instance yang telah dihitung

### 2.3. Metode Perancangan

Metodologi yang digunakan dalam penelitian adalah menggunakan pendekatan terstruktur yang dengan tahapan terdiri dari pengumpulan data, kemudian develop sistem informasi seperti terlihat pada Gambar 3.



Kegiatan dalam Pengumpulan Data adalah :

1. Observasi, yakni dengan melakukan pengamatan terhadap proses-proses yang sedang berlangsung/berjalan. Kegiatan ini didahului dengan identifikasi dan inventarisasi kondisi organisasi, teknologi informasi dan sumber daya manusia yang ada.
2. Survei, bertujuan untuk memperoleh gambaran kondisi kesiapan fungsional dan non fungsional akan kebutuhan sistem untuk tujuan migrasi ke sistem baru.
3. Analisa Dokumen, dilakukan dengan tujuan mempelajari arus dokumen yang menggambarkan sistem yang sedang berjalan.
4. Focus Group Discussion, yaitu diskusi ahli yang memiliki kompetensi dibidangnya guna mendapatkan masukan dan respon bagi penyempurnaan pada tahap pengumpulan data [5].

Kegiatan dalam Tahap Development Sistem adalah sebagai berikut :

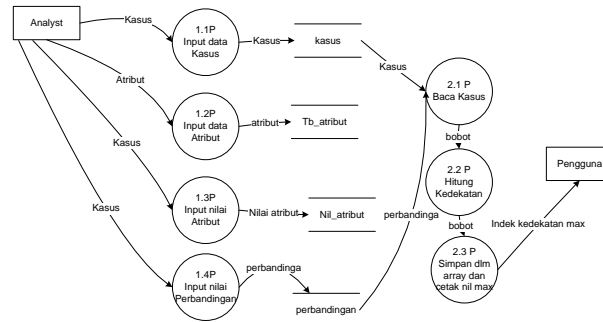
1. Analisis Sistem, dilakukan berdasarkan dari hasil pada tahap pengumpulan data. Hal ini dilakukan dengan memperhatikan permasalahan yang terjadi, tujuan dibangunnya sistem informasi, identifikasi input dan output yang dibutuhkan pada laporan/tampilan informasi.
2. Desain Sistem, pada tahap ini akan dibuat rancangan sistem informasi yang detail berdasarkan kebutuhan yang diinginkan pada tahap analisis. Pada tahap ini juga dilakukan penetapan standar baku yang diutuhkan dalam perencanaan basis data.
3. Programming, yaitu proses implementasi dari tahap desain dengan menggunakan bahasa pemrograman komputer.
4. Testing, dilakukan untuk pengujian modul dan sistem keseluruhan (unit testing and sistem testing) [5].

### 3. Hasil dan Pembahasan

#### 3.1. Diagram Arus Data Level 2 Proses Mining

Diagram arus data level 2 proses mining memperlihatkan adanya 2 kelompok proses 1, yaitu kelompok proses input yang terdiri dari input data kasus, input data atribut, input nilai atribut dan input data perbandingan dan kelompok proses 2 yang terdiri dari proses baca kasus baru, hitung kedekatan dan simpan dalam array, kemudian ambil nilai array tertinggi untuk menentukan kasus lama mana yang memiliki nilai kedekatan paling tinggi untuk menentukan hasil prediksi kelulusan tepat waktu dari kasus baru.

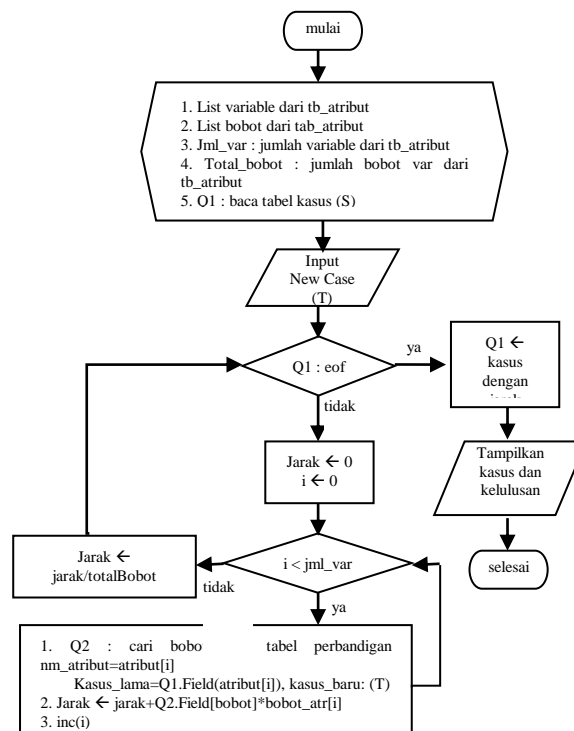
Diagram arus data level 2 proses mining dapat dilihat pada Gambar 4



Gambar 4. Bagan Alir Dokumen Level 2 proses data mining

#### 3.2. Bagan Alir Proses Data Mining K-Nearest Neighbor

Algoritma yang digunakan untuk menghitung besarnya nilai kedekatan antara kasus baru dengan kasus-kasus yang sudah ada dalam database kasus, dapat dilihat pada Gambar 5. berikut :



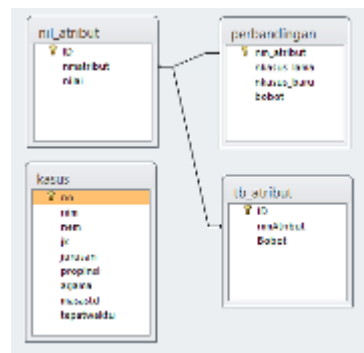
Gambar 5. Bagan Alir Proses Dataming K-Neares Neighbor

Program diawali dengan mengaktifkan list variable dan jum\_var dari tabel tb\_atribut, list bobot variable pada tabel nilai\_atribut, menghitung Total\_bobot dari tabel tb-atribut dan menyiapkan tabel kasus yang berisi yang berisi data akademik alumni yang sudah menyelesaikan studi. Dalam penelitian ini, penulis menggunakan data sampel sebanyak 541 data alumni dengan sebaran angkatan dari 2004 hingga 2010.

Proses berikutnya adalah menyimpan semua data kasus dan kedekatan kedalam variable array. Selanjutnya pengguna menginput data kasus baru untuk dilakukan perhitungan jarak antar atribut data testing terhadap data kasus. Record data kasus ke i akan dibandingkan terhadap seluruh data kasus. Hasil perhitungan jarak disimpan dalam array dengan menggunakan metode sort maximum yang kemudian menjadi dasar dalam menentukan kasus mana yang memiliki nilai kedekatan tertinggi sehingga menjadi acuan dalam menentukan hasil prediksi apakah mahasiswa baru tersebut dapat menyelesaikan studi “tepat waktu “ atau “tidak” dengan mengacu pada kolom atribut kelulusan dari tabel kasus tersebut.

### 3.3. Desain Database

Database yang digunakan dalam produk aplikasi datamining ini terdiri dari 4 tabel yaitu tabel tb\_atribut, tabel nil\_atribut, tabel perbandingan dan tabel kasus. Desain database ini dapat dilihat pada gambar 6



**Gambar 6. Desain database**

### 3.4. Implementasi Sistem

Perangkat lunak yang dihasilkan dari penelitian ini dibangun dengan menggunakan bahasa pemrograman PHP 5.0 dan database engine MySQL. Pada bagian ini akan dijelaskan tahapan implementasi dari penggunaan perangkat lunak aplikasi untuk memprediksi kelulusan tepat waktu mahasiswa baru meliputi pembentukan data kasus yang diambil dari data akademik mahasiswa yang telah menyelesaikan studi, pembentukan data perbandingan, pembentukan atribut list dan proses sorting, perhitungan nilai index setiap atribut, penyusunan aturan serta implementasi beberapa rancangan form.

#### A. Input data Atribut, Bobot dan Kedekatan

Pada tampilan awal aplikasi ini terdapat menu untuk penginputan data atribut, bobot atribut, dan nilai kedekatan kasus lama dengan kasus baru dapat dilihat pada gambar 7, gambar 8 dan gambar 9. Data tentang bobot adalah data angka yang menunjukkan tentang nilai bobot dari sebuah atribut dalam penentuan label kelas prediksi, dimana dalam hal ini ditentukan berdasarkan urutan nilai kepentingan terhadap penentuan nilai suatu label kelas yang berhubungan dengan kelulusan tepat waktu dari seorang mahasiswa baru. Kedekatan nilai antar atribut pada aplikasi ini didasarkan pada asumsi dari peneliti, sehingga nilai kedekatan antar nilai atribut hanya diberikan sejumlah nilai yang sekiranya mendekati sesuai kondisi sebenarnya. Namun demikian nilai kedekatan antar nilai atribut dapat diperbaiki (edit) melalui sebuah jendela/form yang telah disiapkan.

**Aplikasi Prediksi Kelulusan Tepat Waktu**

Variabel

Ganti nilai bobot untuk atribut yang akan digunakan

No.	Nama Variabel Bobot Atribut	Bobot	
<input type="checkbox"/> 1	NEM	0.5	✖
<input type="checkbox"/> 2	JK	0.5	✖
<input type="checkbox"/> 3	Agama	0.4	✖
<input type="checkbox"/> 4	Jurusan	0.7	✖
<input type="checkbox"/> 5	Provinsi	0.6	✖

© Copyright 2020. ITN KIC Computer. Powered by M. Firdaus Mubandari dan N. Nurul Huda. [Logout](#)

**Gambar 7 Form input Atribut**

**Aplikasi Prediksi Kelulusan Tepat Waktu**

Variabel

Nama Atribut/Variabel :

Provinsi

Bobot :

0.5

Submit Batal

**Gambar 8. Form input/edit bobot atribut**

**Aplikasi Prediksi Kelulusan Tepat Waktu**

Nilai Atribut

Karakteristik digunakan untuk menginput variabel yang akan digunakan

No.	Nama Atribut	Nilai	Aktif
<input type="checkbox"/> 1	NEM	1	✖
<input type="checkbox"/> 2	NEM	2	✖
<input type="checkbox"/> 3	NEM	3	✖
<input type="checkbox"/> 4	NEM	4	✖
<input type="checkbox"/> 5	NEM	5	✖
<input type="checkbox"/> 6	NEM	6	✖
<input type="checkbox"/> 7	JK	Prta	✖
<input type="checkbox"/> 8	JK	Wangsa	✖
<input type="checkbox"/> 9	Jurusan	IPA	✖
<input type="checkbox"/> 10	Jurusan	IPS	✖
<input type="checkbox"/> 11	Jurusan	Ekstensi	✖

**Gambar 9. Form input nilai atribut**

## B. Menghitung Prediksi Kelulusan tepat waktu mahasiswa baru

Menu Prediksi pada gambar 10 digunakan untuk melakukan prediksi dengan terlebih dahulu mengisi data identitas mahasiswa baru berupa nama, nim dan variable yang diperlukan untuk proses perhitungan yang terdiri dari atribut NEM, Jenis kelamin, Agama, Jurusan SMA, dan Provinsi. Selanjutnya Tekan tombol Submit untuk melakukan perhitungan jarak kedekatan antara kasus baru dengan kasus yang lama. Aplikasi akan menampilkan hasil perhitungan berupa Kasus dengan jarak terdekat, bobot kedekatan dan prediksi kelulusan tepat waktu Ya atau Tidak.

**DATA MINING**

**Aplikasi Prediksi Kelulusan Tepat Waktu**

Form Input Data:

- Nama: Mr. X
- NIM: 123
- Jenis Kelamin: Laki-Laki
- Agama: Islam
- Usia: 40
- Jurusan: IPA
- Provinsi: Sulsel
- Jarak: 409

Prediksi: Ya

**Gambar 10. Prediksi Kelulusan tepat waktu**

Dengan menggunakan aplikasi ini, dapat ditampilkan tabel hasil perhitungan kedekatan dengan contoh kasus berikut :

Jenis kelamin : laki-laki  
 Agama : Islam  
 Nem : 40  
 Jurusan : IPA  
 Provinsi SMA : Sulsel

Dengan menekan tombol Submit, Maka diperoleh hasil sbb :

Jarak terdekat pada kasus ke : 409

Dengan Jarak : 0.576923076923

Prediksi Lulus Tepat Waktu : Ya

Tabel Hasil Perhitungan kedekatan per kasus sebanyak 541 kasus sebelum diranking dapat dilihat pada gambar 11. Untuk hasil perhitungan Kedekatan setelah diranking dapat dilihat pada gambar 12.

No.	No. Kasus	Kedekatan
1	1	0.44
2	2	0.366666666667
3	3	0.385714285714
4	4	0.36
5	5	0.378787878788
6	6	0.35
7	7	0.385714285714
8	8	0.446428571429
9	9	0.446428571429
10	10	0.378787878788
11	11	0.446428571429
12	12	0.446428571429
13	13	0.385714285714

**Gambar 11. Hasil perhitungan Kedekatan per kasus**



	538	538	0.384010384010
	539	539	0.300523809524
	540	540	0.483870867742
	541	541	0.364516129032
SETELAH DIRANGKING			
	No.	No. Kasus	Kedekatan
	1	409	0.579923079923
	2	227	0.547819047819
	3	189	0.547819047819
	4	351	0.538095238095
	5	488	0.485714285714
	6	409	0.484848484848
	7	115	0.484848484848
	8	113	0.484848484848
	9	82	0.484848484848
	10	48	0.484848484848
	11	94	0.484848484848
	12	54	0.484848484848
	13	50	0.484375

**Gambar 12. Hasil perhitungan Kedekatan Per Kasus Setelah Diranking**

#### 4. Pengujian

Dalam melakukan pengujian khususnya terhadap tingkat keakurasian hasil implementasi algoritma NKK, digunakan data alumni STMIK Dipanegara tahun wisuda 2004 s.d. 2010 untuk kasus lama dan untuk kasus baru digunakan data alumni tahun wisuda 2011 dengan mengambil masing-masing jumlah sampel 541. Kedua kasus ini masing-masing sudah diketahui masa studinya *tepat waktu* atau *tidak tepat waktu*.

Hasil dari kedekatan setelah dikelompokkan berdasarkan kedekatan pada penerapan algoritma NKK dari aplikasi ini dapat dilihat pada Tabel 1 berikut:

**Tabel 1. Hasil Pengujian Kedekatan**

Kedekatan	Tepat Waktu	Jumlah
0.49	Tidak	1
0.49	Ya	3
0.48	Ya	9
0.47	Ya	42
0.46	Tidak	54
0.46	Ya	220
0.45	Tidak	10
0.45	Ya	170
0.44	Ya	19
0.43	Ya	3
0.42	Ya	4
0.41	Ya	3
0.40	Tidak	3
Total		541

Hasil pengujian dengan menerapkan algoritma NKK dari kasus lama dan kasus baru terhadap alumni, diperoleh tingkat akurasi tepat waktu dan tidak tepat waktu studi mahasiswa diperoleh 83% seperti terlihat pada Tabel 2.

**Tabel 2. Persentase Akurasi Hasil Pengujian Tepat Waktu Studi Mahasiswa**

Uji Kasus	kasus lama	kasus baru	Jumlah	Persentase
Tepat Waktu	Tidak	Ya	90	16.64%
	Ya	Tidak		
	Tidak	Tidak	451	83.36%

	Ya	Ya		
Total			541	100.00%

## 5. Kesimpulan

Dengan menggunakan data mining, khususnya penerapan algoritma NKK, kita dapat mengetahui hubungan kedekatan antara kasus yang baru dengan kasus yang telah ada dalam suatu gudang data (*data warehouse*) sehingga dapat menjadi acuan untuk memprediksi kelulusan seorang mahasiswa baru apakah dapat menyelesaikan kuliahnya dengan tepat waktu atau tidak berdasarkan kedekatan data yang sudah ada. Dari hasil pengujian dengan menerapkan algoritma KNN dan menggunakan data sampel alumni tahun wisuda 2004 s.d. 2010 untuk kasus lama dan data alumni tahun wisuda 2011 untuk kasus baru diperoleh tingkat akurasi sebesar 83,36%.

## 6. Saran

Saran yang diajukan dalam penelitian ini diantaranya adalah penelitian ini dapat diimplementasikan oleh pengelola kampus sehingga dapat diambil tindakan preventif terhadap mahasiswa baru yang diprediksi kelulusannya tidak tepat waktu. Selain itu saran peneliti juga kiranya penelitian ini dapat dikembangkan oleh peneliti berikutnya kaitannya dengan prediksi keberhasilan mahasiswa melewati batas drop out tiga semester pertama.

## Daftar Pustaka

- [1] Arief Jananto., 2010., *Perbandingan Performansi Algoritma Nearest Neighbor dan SLIQ untuk Prediksi Kinerja Akademik Mahasiswa Baru.*, Jurnal Teknologi Informasi DINAMIK Volume XV, No.2, Juli 2010 : 157-169 ISSN : 0854-9524
- [2] Florin Gorunescu, 2011, *Data Mining: Concepts, Models and Techniques*, Springer.
- [3] Han, J., dan Kamber, M., 2006, *Data Mining: Concepts and Techniques*, SunFransisco : Morgan Kaufmann Publishers
- [4] Iin Ernawati, 2008, online di <http://repository.ipb.ac.id/bitstream/handle/9558/ier.pdf> diunduh tanggal 30 April 2012
- [5] Jogiyanto, HM, 2007, *Analisis dan Desain Sistem Informasi, Pendekatan Terstruktur*, Andi, Yogyakarta
- [6] Kursini & Emha Taufiq Luthfi, 2009, *Algoritma data Mining*, ANDI, Yogyakarta
- [7] Ian H. Witten, Frank Eibe, Mark A. Hall, *Data mining: Practical Machine Learning Tools and Techniques 3rd Edition*, Elsevier, 2011.