

Algoritma *K-Nearest Neighbor Classification* Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa

Mustakim¹, Giantika Oktaviani F²

^{1,2}Laboratorium Data Mining Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarif Kasim Riau
Jl. HR. Soebrantas No. 155 Simpang Baru, Tampan, Pekanbaru, Riau – Indonesia 28293
Email: ¹mustakim@uin-suska.ac.id, ²giantikaoktaviani@students.uin-suska.ac.id

(Received: 26 Mei 2016; Revised: 20 Juni 2016; Accepted: 20 Juni 2016)

ABSTRAK

Prestasi Mahasiswa merupakan suatu bentuk dari pencapaian hasil selama mengikuti kegiatan Akademik pada sebuah Perguruan Tinggi. Predikat prestasi mahasiswa diperoleh dari hasil sebuah prediksi. Proses prediksi dilakukan dengan menggunakan metode *K-Nearest Neighbor* (KNN). Atribut yang digunakan dalam proses prediksi adalah Jenis Kelamin, Jenis Tinggal, Umur, Jumlah Satuan Kredit Semester (SKS), dan Jumlah Nilai Mutu (NM), sehingga dengan menerapkan algoritma KNN dapat dilakukan sebuah prediksi berdasarkan kedekatan dari histori data lama (*training*) dengan data baru (*testing*). Penentuan atribut ini berdasarkan hasil penelitian terdahulu yang memiliki kesamaan dalam kasus prediksi mahasiswa yang selanjutnya divalidasi oleh bagian Akademik Fakultas Sains dan Teknologi. Proses prediksi dilakukan terhadap Mahasiswa Program Studi Sistem Informasi angkatan 2014/2015 sebagai data *testing* dengan jumlah 50 data, serta berdasarkan dari data angkatan 2012/2013 sebagai data *training* dengan jumlah 165 data yang menghasilkan pengujian akurasi sebesar 82%. Hasil dari perhitungan algoritma KNN diimplementasikan terhadap sebuah *Early Warning System* (EWS). *Output* dari sistem yang dibangun dapat dijadikan sebagai acuan bagi Mahasiswa untuk meningkatkan prestasi dan predikat perkuliahan dimasa yang akan datang.

Kata Kunci : *Early Warning System* (EWS), *K Nearest Neighbor* (KNN), Prediksi Predikat Mahasiswa

ABSTRACT

Students college predicate is a form of achievement during the academic activity at college. This research is intended to make predictions toward predicate students college achievement that will be acquired in the future. The process of predictions by using K-Nearest Neighbor Method (KNN). The attributes that are used in process predictions was gender, kind of stay, age, semester credit unit, and also grade point average. Therefore by applying Al-goritma KNN, the predictions based on the closeness from history of data training to data testing can be done. To determined of this attributes based on the result of previous researches that have similarities of case that validated by academic of Faculty Sains and Technology. The process of predictions toward students information system of 2014/2015 as a sample of data testing. The number of the data was 50. And based on the data of students information system of 2012/2013 as a sample of data training, the number of the data was 165 which produce the accuracy testing was 82%. The result of calculation algoritma KNN is implemented toward Early Morning System (EWS). The output of sytem built to serve as a guide for students to improve the achievement and predicate in the future.

Keywords: *Early Warning System* (EWS), *K Nearest Neighbor* (KNN), *predicated forecasting student collage*

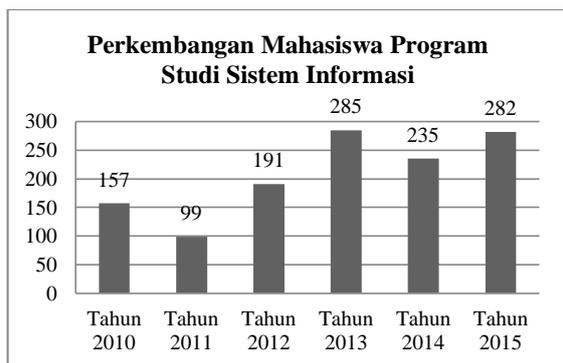
Corresponding Author:

Mustakim,
Laboratorium Data Mining Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarif Kasim Riau,
Email: mustakim@uin-suska.ac.id

Pendahuluan

UIN Sultan Syarif Kasim (Suska) Riau hingga Tahun 2016 memiliki sekitar 28,6 ribu Mahasiswa aktif yang berada pada 8 Fakultas dan 39 Program Studi. Salah satu diantaranya adalah Fakultas Sains dan Teknologi yang telah berdiri sejak tahun 2002. Fakultas ini mempunyai 5 Program Studi yaitu Teknik Informatika, Teknik Industri, Sistem Informasi, Teknik Elektro dan Matematika Terapan [1].

Program Studi Sistem Informasi saat ini mengalami peningkatan yang signifikan dibandingkan dengan lima tahun pertama. Jumlah Mahasiswa baru yang semakin banyak tidak menutup kemungkinan Mahasiswa lama dapat menyelesaikan perkuliahan dengan tepat waktu sehingga, mengakibatkan jumlah Mahasiswa Program Studi Sistem Informasi akan semakin banyak. Banyak kemungkinan faktor yang menjadi alasan Mahasiswa tidak dapat menyelesaikan perkuliahan dengan tepat waktu, salah satunya adalah personalan Indeks Prestasi Kumulatif (IPK) Mahasiswa.



Gambar 1. Perkembangan Mahasiswa Program Studi Sistem Informasi 6 Tahun Terakhir

Semakin bertambahnya Mahasiswa maka akan semakin menimbulkan pola yang berfariatif pada atribut data Mahasiswa serta berpotensi untuk dilakukan penggalian data (data mining) yang akan memberikan sebuah pengetahuan baru. Banyak teknik yang dapat diterapkan untuk mengetahui pengetahuan tersebut, diantaranya adalah dengan melakukan klasifikasi. Adapun teknik yang sering digunakan pada klasifikasi adalah algoritma K-Nearest Neighbor (K-NN). Algoritma K-NN adalah suatu metode yang menggunakan algoritma *supervised* [2]. Data yang digunakan adalah data Mahasiswa angkatan 2012/2013 sebagai data *training* dan data Mahasiswa angkatan 2014/2015 sebagai data *testing*. Sedangkan atribut yang akan digunakan adalah Jenis Kelamin, Umur,

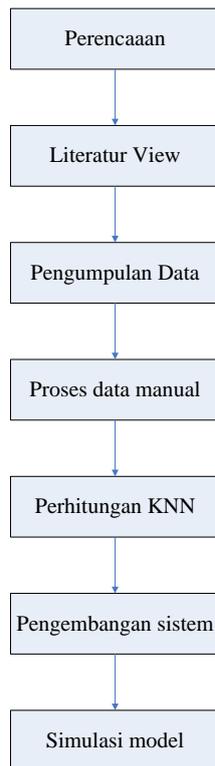
Jenis Tinggal, Jumlah Nilai Mutu, dan Jumlah Satuan Kredit SKS. Sebagai kelas penentu hasil prediksi adalah nilai Indeks Prestasi Kumulatif (IPK) Mahasiswa yang dibagi kedalam empat predikat yaitu Pujian, Sangat memuaskan, Memuaskan, dan Cukup. Hal ini memiliki kedekatan kesamaan atribut yang digunakan pada penelitian sebelumnya.

Pada penelitian sebelumnya algoritma K-NN pernah diterapkan untuk memprediksi kelulusan tepat waktu bagi Mahasiswa baru pada STMIK Diponegara. Atribut prediktor yang digunakan adalah nilai NEM, Jenis Kelamin, Agama, Jurusan dan Provinsi. Data yang digunakan adalah data alumni dan Mahasiswa tahun 2004 sampai dengan 2010 sebagai data lama dan data Mahasiswa Tahun 2011 sebagai data baru. Dari hasil perhitungan pengujian akurasi algoritma K-NN mampu menghasilkan akurasi sebesar 83% [5].

Terdapat dua tujuan dari penelitian ini yaitu memodelkan Algoritma K-NN sebagai salah satu metode prediksi pada kasus penentuan predikat prestasi Mahasiswa. Tujuan yang kedua yaitu membangun *Early Warning System* (EWS) sebagai *forecasting* predikat prestasi Mahasiswa dengan menerapkan algoritma K-NN. Kedua tujuan tersebut nantinya dapat mengatasi beberapa problem yang selama ini dihadapi oleh Penasehat Akademis yang tidak dapat mengetahui secara periodik bagaimana perkembangan kedepan Mahasiswa yang dibimbingnya. Disamping itu, pembuktian sebuah Algoritma K-NN *Classification* yang mampu memodelkan sebuah kasus prediksi dengan akurasi yang tinggi.

Metode Penelitian

Metode penelitian yang diterapkan dalam penelitian ini dapat dilihat pada Gambar 2 berikut:



Gambar 2. Metodologi Penelitian

Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam *database*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *mechine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [6].

Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang tersimpan di dalam *database* besar [7].

K-Nearest Neighbor (KNN)

Algoritma *K-Nearest Neighbor*(K-NN) adalah suatu metode yang menggunakan algoritma *supervised*[2]. K-NN termasuk kelompok *instance-based learning*. Algoritma ini juga merupakan salah satu teknik *lazy learning*. K-NN dilakukan dengan mencari kelompok k objek dalam data *training* yang paling dekat (mirip) dengan objek pada data baru atau data *testing*[4].

Secara umum untuk mendefinisikan jarak antara dua objek *x* dan *y*, digunakan rumus jarak *Euclidean* pada persamaan 2.1 [5].

$$d_{xy} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Algoritma K-NN dapat secara umum dapat digambarkan dengan *Flowchart* pada Gambar 3.

Pengujian Akurasi

Confusion matrix adalah tool yang digunakan untuk evaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah. Sebuah matrix dari prediksi yang akan dibandingkan dengan kelas yang asli dari inputan atau dengan kata lain berisi informasi nilai aktual dan prediksi pada klasifikasi.

Adapun perhitungan tingkat akurasi pada *confusion matrix*:

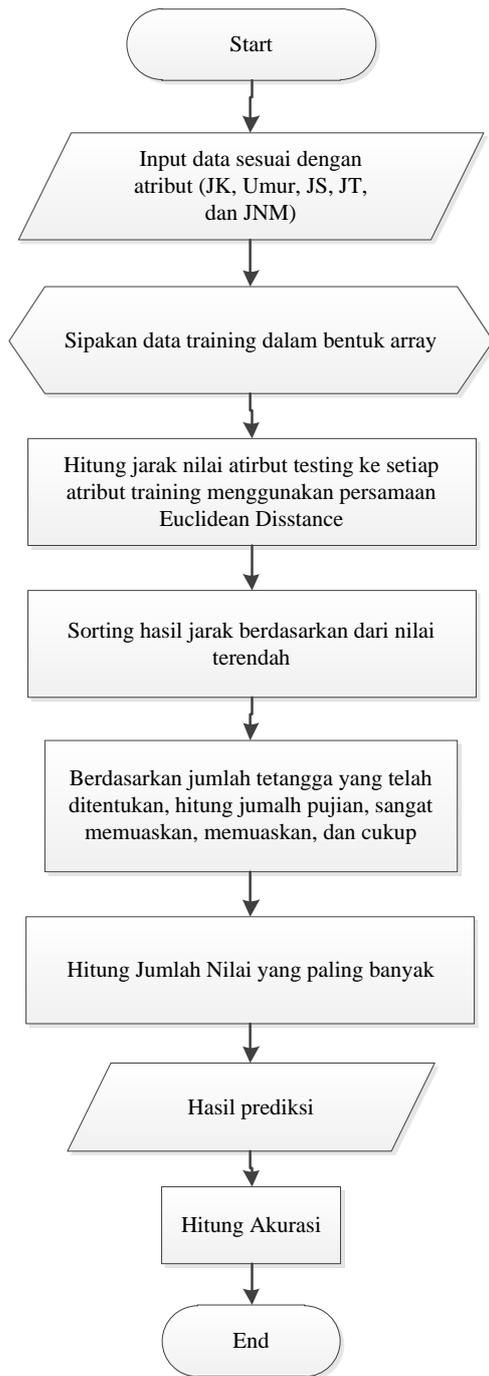
$$Accuracy = \frac{\text{Jumlah nilai benar}}{\text{Jumlah data keseluruhan}} \times 100\% \quad (2)$$

Hasil dan Pembahasan

Sesuai dengan metodologi penelitian pada bab sebelumnya, beberapa hal penting yang akan dilakukan dalam menyelesaikan penelitian ini terdiri atas perencanaan, pengumpulan dan pengolahan data, hasil dan analisis, dan dokumentasi.

Pengumpulan Data

Dalam pengumpulan data terdapat sumber data, sumber data yang dihimpun langsung oleh peneliti disebut dengan sumber primer, sedangkan apabila melalui tangan kedua disebut sumber sekunder [7].



Gambar 3. Flowchat K-NN

Sebagai metode dalam pengumpulan data Mahasiswa yang menjadi landasan utama dalam proses prediksi ini adalah menggunakan wawancara dan penyebaran kuesioner *online*. Wawancara dilakukan dengan pihak Admin Program Studi Sistem Informasi dan diperoleh data Mahasiswa aktif Angkatan 2012 yang berjumlah 170 data. Sedangkan kuesioner *online* dilakukan

menggunakan *google form* yang disebarakan kepada 3 kelas Angkatan 2014.

Pembersihan data dilakukan untuk mengurangi efek *noise* pada saat proses perhitungan dan menghilangkan atribut yang tidak digunakan. Selanjutnya proses normalisasi data. Hasil normalisasi dapat dilihat pada Tabel 1.

Table 1. Normalisasi Data

NIM	JK	Umur	JS	JT	JNM	Predikat
11253100111	0	0	3	0	1	2
11253100147	0	0	3	0	1	3
11253100228	0	1	2	1	1	2
11253100230	0	0	2	1	1	2
11253100296	0	1	2	1	2	1
11253100301	0	1	2	0	1	2
11253100309	0	1	3	0	1	2
11253100311	0	1	2	0	1	3
11253100499	0	1	2	0	2	2

Perhitungan KNN

Sebelum melakukan proses perhitungan KNN perlu disiapkan data kelas/ label, data atribut, data lama (*training*), dan data baru (*testing*).

Tabel 2. Atribut Penentu

Nilai IPK	Predikat
3.50 – 4.00	Pujian
3.00 – 3.49	Sangat memuaskan
2.50 – 2.99	Memuaskan
2.00 – 2.49	Cukup

Tabel 2 merupakan jangkauan dan atribut penentu yang digunakan untuk menentukan kelas pada klasifikasi.

Tabel 3. Atribut Prediktor

JK	Umur	JS	JT	JNM
LK	<21	<81	Wali	<200
PR	=21	=81	Kost	=200
	>21	>81	Orang Tua	>200

Tabel 3 diatas memiliki 2 nilai atribut pada Jenis Kelamin dan 3 nilai atribut pada Umur, Jumlah Satuan Kredit SKS, Jenis Tinggal dan Jumlah Nilai Mutu.

Tabel 4. Data *Training*

NIM	JK	JT	UMUR	JS	JNM	Nilai
11253100014	0	0	1	1	0	2
11253100111	0	2	0	0	0	2
11253100147	0	0	2	0	0	3
11253100228	0	1	1	1	0	2
11253100230	0	1	2	1	0	2
11253100296	0	1	1	1	2	1
11253100717	0	0	2	1	0	3
11253100763	0	0	1	1	0	3
....
11253205000	1	0	2	0	0	3

10 dari 250 Data awal digunakan sebagai data *training* pada simulasi pada metode K-NN pada kasus prestasi mahasiswa. Data tersebut telah dinormalisasi menggunakan *min-max normalization*.

Tabel 5. Data *Testing*

NIM	JK	JT	UMUR	JS	JNM	Predikat
11453101916	0	0	0	2	2	?
11453205235	1	2	0	2	2	?
11453105423	0	2	0	2	2	?
11453101884	0	0	0	2	2	?
11453201892	1	0	0	2	2	?
11453204106	1	0	0	2	2	?
11453205368	1	2	0	2	2	?
11453204106	1	0	0	2	2	?
11453205559	1	0	0	2	2	?
..
11453201785	1	0	0	2	0	?

Tabel 5 menunjukkan 10 dari 50 data testing sebagai pengujian hasil model terbaik yang dihasilkan K-NN. Dari data ini akan dihasilkan kesimpulan sebuah data akan masuk kedalam kelas pertama, kedua, ketiga ataupun yang keempat berdasarkan nilai ketetapan yang telah proses.

Selanjutnya proses perhitungan K-NN dilakukan dengan menggunakan persamaan

Eulidean Distance. Sebagai contoh perhitungan data baru yang akan dihitung adalah data pertama dengan NIM 11453101916. Proses perhitungan dilakukan kesetiap data lama sehingga nantinya akan menghasilkan nilai jarak sesuai dengan jumlah data lama. Dari perhitungan tersebut diperoleh hasil jarak seperti pada Tabel 6.

Tabel 6. Hasil Perhitungan KNN

NIM	JK	JT	UMUR	JS	JNM	Nilai	Jarak
11253100014	0	0	1	1	0	2	2,45
11253100111	0	2	0	0	0	2	3,46
11253100147	0	0	2	0	0	3	3,46
11253100228	0	1	1	1	0	2	2,65
11253100230	0	1	2	1	0	2	3,16
..
11253205000	1	0	2	0	0	3	3,60

Setelah mendapatkan hasil jarak, selanjutnya diurutkan dari nilai terkecil hingga nilai terbesar seperti pada Tabel 7.

Tabel 7. Pengurutan nilai jarak

NIM	JK	JT	UMUR	JS	JNM	Nilai	Jarak
11253100499	0	0	1	1	2	2	1,41
11253103223	0	0	1	1	2	2	1,41
11253104887	0	0	1	1	2	2	1,41
11253201210	1	0	0	1	2	2	1,41
11253202195	1	0	0	1	2	2	1,41
11253100296	0	1	1	1	2	1	1,73
11253100691	0	1	1	1	2	1	1,73
11253102164	0	1	1	1	2	2	1,73
11253104985	0	1	1	1	2	2	1,73
..
11253201977	1	2	2	0	0	2	4,12

Berdasarkan nilai k yang telah ditetapkan yaitu 5, maka nilai jarak yang diambil adalah 5 terkecil seperti pada Tabel 8.

Tabel 8. Nilai jarak 5 terkecil

NIM	JK	JT	UMUR	JS	JNM	Predikat	Jarak
11253100499	0	0	1	1	2	2	1,41
11253103223	0	0	1	1	2	2	1,41
11253104887	0	0	1	1	2	2	1,41
11253201210	1	0	0	1	2	2	1,41
11253202195	1	0	0	1	2	2	1,41

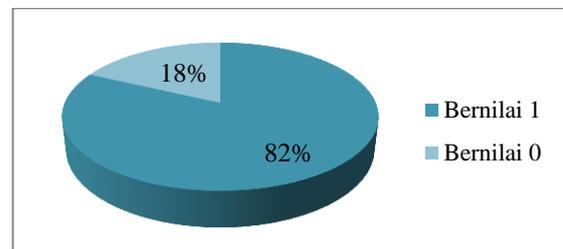
Dari 5 nilai jarak diatas, data menunjukkan nilai predikat bernilai sama, sehingga sebagai nilai prediksi untuk data baru yang pertama adalah predikat nomor 2 yaitu Sangat Memuaskan. Dari hasil prediksi tersebut dilakukan pengujian menggunakan *confusionmatriks* membandingkan nilai predikat sebelumnya dengan nilai predikat hasil prediksi yang menghasilkan akurasi.

Tabel 9. Perbandingan Nilai Predikat

NIM	Predikat Sem 3	Predikat Prediksi	Point
11453101916	Sangat Memuaskan	Sangat Memuaskan	1
11453205235	Pujian	Sangat Memuaskan	0
11453105423	Pujian	Sangat Memuaskan	0
11453101884	Sangat Memuaskan	Sangat Memuaskan	1
11453201892	Sangat Memuaskan	Sangat Memuaskan	1
11453204106	Pujian	Sangat Memuaskan	0
..
11453201785	Sangat Memuaskan	Sangat Memuaskan	1

Dari Tabel diatas hitung jumlah point selanjutnya dibagi dengan jumlah data dan dikali 100% sehingga didapat hasil sebesar 82% yang

dipresentasikan kedalam sebuah diagram seperti pada Gambar 4 berikut:



Gambar 4. Hasil akurasi perhitungan K-NN

Perancangan dan Implementasi Sistem

Perancangan dan implementasi sistem dimaksudkan untuk menerapkan dan mengimplementasikan algoritma yang telah dibahas sebelumnya untuk dapat digunakan dan diimplementasikan kedalam bahasa pemrograman.

Referensi

- [1] Hidayat, Amir Syarif. Panduan dan Informasi Akademik 2012/ 2013 UIN Suska Riau. Pekanbaru.
- [2] Han J and Kamber M. *Data Mining: Concept and Techniques*. New York: Morgan Kaufmann Publisher ;2006.
- [3] Jayanti, Ririn Dwi. Aplikasi Metode K-Nearest Neighbor Dan Analisa Diskriminan Untuk Analisa Resiko Kredit Pada Koperasi Simpan Pinjam Di Kopinkra Sumber Rejeki. Prosiding Seminar Nasional Aplikasi Sains dan Teknologi (SNAST). Yogyakarta. 2014
- [4] Leidiyana. Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor. Jurnal Penelitian Ilmu Komputer, System Embedded & Logic, Vol : 1. STMIK Nusa Mandiri. 2010
- [5] Mustafa. Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining (Studi Kasus : Data Akademik Mahasiswa STMIL Dipanegara Makassar). Citec Jurnal Vol : 1. STMIK Dipanegara. 2014
- [6] Mustakim. Pemetaan Digital dan Pengelompokan Lahan Hijau di Wilayah Provinsi Riau Berdasarkan Knowledge Discovery in Database (KDD) dengan Teknik K-Means Mining. Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 4, Pekanbaru, 3 Oktober 2012
- [7] Turban, E dkk . *Decicion support systems and intelligent system*. Yogyakarta: andi Offset. 2005
- [8] Riduwan. (2008). *Metode dan Teknik Menyusun Tesis*. Bandung: Alfabeta.