

**Penerapan Algoritma *Modified K-Nearest Neighbour (MKNN)* Pada
Klasifikasi Masa Studi Mahasiswa Teknik Informatika
*Application of the Modified K-Nearest Neighbour (MKNN)
Algorithm In the Classification of the Study Period of Informatics
Engineering Student***

Aldiyan Iman Saputra¹, Hardian Oktavianto², Habibatul Azizah Al Faruq³

¹Mahasiswa Fakultas Teknik, Universitas Muhammadiyah Jember
email: aldisptr87@gmail.com

^{2,3}Dosen Fakultas Teknik, Universitas Muhammadiyah Jember

Abstrak

Mahasiswa dapat diartikan sebagai seseorang yang sedang dalam proses menimba ilmu dan sedang menjalani pendidikan pada salah satu bentuk perguruan tinggi, yang terdiri dari akademik, politeknik, sekolah tinggi, institut dan universitas. Program Sarjana (S1) merupakan jenjang pertama program akademik yang mempunyai beban studi 144 sampai 160 SKS, dijadwalkan sekurang-kurangnya 8 semester dan selama-lamanya 14 semester. Data yang digunakan yaitu data IPS (Indeks Prestasi Semester) dari semester 1 sampai semester 6 yang akan dijadikan dasar perhitungan prediksi dengan harapan dapat menemukan informasi mahasiswa yang terindikasi lulus tepat waktu dan tidak tepat waktu. Model algoritma yang digunakan untuk mendukung prediksi masa studi mahasiswa adalah algoritma Modified K-Nearest Neighbor. Tujuan dari penelitian ini untuk mengetahui dan mengukur tingkat akurasi, presisi dan *recall* pada metode *Modified K-Nearest Neighbor*. Klasifikasi Masa Studi Mahasiswa Teknik Informatika menggunakan metode *Modified K-Nearest Neighbor* dengan data yang digunakan diambil dari UPT Pusat Data dan Informasi Universitas Muhammadiyah Jember, dengan jumlah data 260 data dan 2 kelas *output*. Berdasarkan hasil pengujian yang telah dilakukan didapatkan hasil akurasi tertinggi sebesar 84,62%, hasil presisi tertinggi sebesar 100%, dan hasil *recall* tertinggi sebesar 75% pada $K = 3$.

Keywords: Mahasiswa, Indeks Prestasi Semester, Metode *Modified K-Nearest Neighbor*.

Abstract

A student can be defined as someone who is in the process of gaining knowledge and is undergoing education at one form of higher education, which consists of academics, polytechnics, high schools, institutes and universities. The Undergraduate Program (S1) is the first level of an academic program that has a study load of 144 to 160 credits, is scheduled for at least 8 semesters and a maximum of 14 semesters. The data used is IPS (Semester Achievement Indeks) data form semester 1 to semester 6 which will be used as the basis for calculating predictions in the hope of finding information on students who are indicated to have graduated on time and not on time. The algorithm model used to support the prediction of student study period is the Modified K-Nearest Neighbor algorithm. This research is to determine the study length of students using the Modified K-Nearest Neighbor method. The study purpose was to determine and measure the level of accuracy, precision and recall in the Modified K-Nearest Neighbor method. The study period of classification informatics engineering students uses the Modified K-Nearest Neighbor method with the data used taken from the UPT Data and Information Center, University of Muhammadiyah Jember, with a total of 260 data and 2 output classes. Based on the results of the test which have been done, it was found out that the highest accuracy results were 84,62%, highest precision results were 100% and highest recall results were 75% at $K = 3$.

Keywords: Students, Semester Achievement Index, Modified K-Nearest Neighbor Method

1. PENDAHULUAN

Mahasiswa dapat diartikan sebagai seseorang yang sedang dalam proses menimba ilmu dan sedang menjalani pendidikan pada salah satu bentuk perguruan tinggi, yang terdiri dari akademik, politeknik, sekolah tinggi, institut dan universitas (Hartaji & Sedjo, 2012). Sesuai dengan Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia (Permendikbud) No. 49 tahun 2014 tentang Standar Nasional Pendidikan Tinggi dalam pasal 17 ayat 2 berbunyi “144 sks untuk diploma empat dan program sarjana” dan ayat 3 berbunyi “4 sampai 5 tahun untuk program diploma empat dan program sarjana”. Namun, tidak semua mahasiswa dapat menyelesaikan masa studinya dengan tepat waktu. Untuk memprediksi masa studi mahasiswa, maka dibutuhkan sebuah teknik data mining klasifikasi menggunakan algoritma *Modified K-Nearest Neighbor*. Data yang digunakan yaitu data IPS (Indeks Prestasi Semester) dari semester 1 sampai semester 6, yang akan dijadikan sebagai dasar perhitungan prediksi yang dilakukan dengan harapan dapat menemukan informasi mahasiswa yang terindikasi lulus tepat waktu dan mahasiswa yang lulus tidak tepat waktu. Terdapat beberapa penelitian sebelumnya yang dilakukan oleh (Muhammad, 2015) dengan judul *Modified K-Nearest Neighbor* untuk Prediksi Curah Hujan. Dari hasil penelitian dapat disimpulkan bahwa akurasi terbaik dihasilkan oleh BPNN-1m sebesar 82,46%. Berdasarkan paparan di atas, penulis tertarik untuk mengklasifikasikan lamastudi mahasiswa melalui penelitian dengan judul “Penerapan Algoritma *Modified K-Nearest Neighbor* (MKNN) Pada Klasifikasi Masa Studi Mahasiswa Teknik Informatika”. Penggunaan metode *Modified K-Nearest Neighbor* diharapkan dapat menghasilkan nilai akurasi, presisi dan *recall* yang tinggi.

2. METODOLOGI PENELITIAN

a. Data Penelitian

Pada penelitian ini data yang digunakan merupakan data sekunder yang diperoleh dari UPT. Pusat Data dan Informasi

Universitas Muhammadiyah Jember berupa data alumni mahasiswa Teknik Informatika tahun 2010-2016 berjumlah 260 data, sebagai kasus yang dibahas dalam tugas akhir ini dan berhubungan dengan algoritma yang diterapkan.

b. *Modified K-Nearest Neighbor*

Suatu algoritma yang dikembangkan dari algoritma *K-Nearest Neighbor* dengan menambahkan proses baru yaitu proses validitas data pada setiap dataset yang telah dilakukan perhitungan jarak *euclidean* antar pada data training, dan juga penambahan proses *weight voting* yang didapatkan dari hasil nilai validitas data dikalikan dengan hasil *euclidean distance* data training dan data testing (Parvin dkk, 2010).

c. *Confusion Matrix*

Confusion matrix memeberikan keputusan yang diperoleh dalam training dan testing, *confusion matrix* memberikan penilaian performance klasifikasi berdasarkan objek dengan benar atau salah (Gorunescu, 2011) *Confusion matrix* berisi informasi aktual (*actual*) dan prediksi (*predicted*) pada sistem klasifikasi.

Tabel 1. *Confusion Matrix*

		Kelas Prediksi	
		1	2
Kelas Aktual	1	TP	FP
	2	FN	TN

Sumber : Hasil Pengamatan

Untuk perhiutngan akurasi, presisi, recall digunakan persamaan :

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Presisi = \frac{TP}{TP + FP} \times 100\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

3. HASIL DAN PEMBAHSAN

A. Gambaran Data

Data yang digunakan diambil dari UPT Pusat Data dan Informasi Universitas Muhammadiyah Jember berupa data alumni

mahasiswa Teknik Informatika tahun 2010-2016. Pengumpulan data didapat sebanyak 260 data, untuk melakukan oengujian terhadap hasil presiksi yang terdiri dari 2 klasifikasi yaitu lulus tepat waktu, dan tidak lulus tepat waktu.

B. Praproses Data

Pada tahap ini data yang dipersiapkan dapat memepermudah proses penambangan (*mining*) diantaranya memilih atribut pada dataset tersebut.

Tabel 2. Atribut Dataset

Nama Atribut	Keterangan
NIM	Nomor Induk Mahasiswa
IPS-1	Indek Prestasi Semester Satu
IPS-2	Indek Prestasi Semester Dua
IPS-3	Indek Prestasi Semester Tiga
IPS4	Indek Prestasi Semester Empat
IPS5	Indek Prestasi Semester Lima
IPS-6	Indek Prestasi Semester Enam
Masa Studi	Masa Studi Terjadwal
Class	Tepat waktu (1), Tidak tepat waktu (2)

Sumber : Hasil Penelitian

Setelah dilakukan pengumpulan data, selanjutnya melakukan tahap *Pre-pocessing* data sebelum data tersebut digunakan pada proses pengujian . Dalam proses ini akan dilakukan pembersihan nilang kosong (*missing values* atau *noise*) Tidak semua atribut dalam tabel data digunakan, hanya atribut yang dianggap berpengaruh saja yang akan digunakan dalam penelitian. Setelah dilakukan pengumpulan data, selanjutnya melakukan tahap *Pre-pocessing* data sebelum data tersebut digunakan pada proses pengujian . Dalam proses ini akan dilakukan pembersihan nilang kosong (*missing values* atau *noise*) Tidak semua atribut dalam tabel data digunakan, hanya atribut yang dianggap berpengaruh saja yang akan digunakan dalam penelitian.

Tabel 3. Precosing pengujian klasifikasi Modified K-Nearest Neighbor

No	IPS1	IPS2	IPS3	IPS4	IPS5	IPS6
1	1,72	1,93	2,50	2,67	2,21	2,58
2	3,00	3,04	2,74	3,13	2,17	2,59
3	3,17	2,58	3,19	3,04	2,50	3,00
4	2,50	2,52	2,10	1,48	2,33	2,54
5	2,67	2,71	2,71	1,65	2,14	2,50
...
260	3,69	3,66	3,65	3,07	3,61	3,47

Sumber : Hasil Penelitian

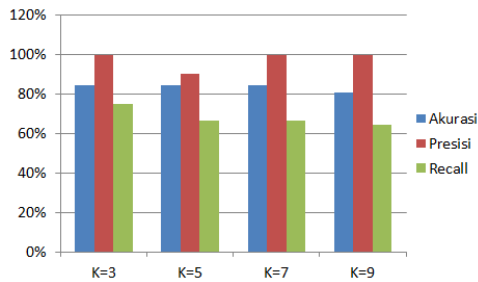
C. Hasil Pengujian nilai Akurasi, Presisi, Recall pada k-fold

Pengujian k-fold dilakukan dengan menggunakan *2-fold*, *4-fold*, *5 fold*, *10-fold* dengan jumlah data 260 data. Setelah dilakukan pengujian didapatkan hasil akurasi, presisi dan *recall*. Dalam pengujian nilai akurasi didapatkan persentase tertinggi pada percobaan k-fold 10 dari proses percobaan sebanyak 10 kali dengan data yang diacak, untuk K=3 menghasilkan akurasi 84,62% pada skenario 8, K=5 menghasilkan akurasi 84,62 pada skenario 2, K=7 menghasilkan akurasi 84,62 pada skenario 8, K=9 menghasilkan akurasi terendah yaitu 80,77% pada skenario 9, nilai presisi tertinggi untuk K=3 menghasilkan presisi 100% pada skenario 10, K=5 menghasilkan nilai presisi paling rendah yaitu 90% pada skenario 10, K=7 menghasilkan nilai presisi 100% pada skenario 10, dan K=9 menghasilkan nilai presisi 100% pada skenario 10, nilai recall tertinggi untuk K=3 menghasilkan *recall* 75% pada skenario 8, K=5 menghasilkan nilai *recall* 66,67% skenario 7, K=7 menghasilkan nilai *recall* 66,67% pada skenario 3, dan K=9 menghasilkan nilai *recall* terendah yaitu 64,29% pada skenario 10.

Tabel 4. Sumber Hasil Perhitungan Nilai akurasi, presisi, *recall* dengan menggunakan confusion matrix pada jupyter notebook

Nilai K	Akurasi	Presisi	Recall
3	84,62%	100%	75%
5	84,62%	90%	66,67%
7	84,62%	100%	66,67%
9	80,77%	100%	64,29%

Sumber : Hasil Perhitungan



Gambar 1. Hasil Pengamatan Grafik pada setiap nilai k

Sumber : Hasil Pengamatan

Berdasarkan tabel diatas, hasil akurasi, presisi, *recall* untuk dataset Masa Studi Mahasiswa Teknik Informatika dilakukan pencarian dengan menggunakan nilai k beragam yang mampu menghasilkan nilai akurasi, presisi, *recall* tertinggi dan didapatkan nilai tertinggi pada percobaan $k=3$ dengan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya pada *confusion matrix*. Nilai persentase akurasi sebesar 84,62%, presisi sebesar 100%, dan *recall* sebesar 75%.

4. KESIMPULAN DAN SARAN

a. Kesimpulan

Berdasarkan penelitian untuk melakukan klasifikasi masa studi mahasiswa Teknik Informatika menggunakan metode Modified K-Nearest Neighbor maka dapat disimpulkan bahwa :

1. Hasil pengujian akurasi tertinggi yang diperoleh dari metode *Modified K-Nearest Neighbor* pada klasifikasi masa studi mahasiswa Teknik Informatika berdasarkan Index Prestasi Semester 1 sampai 6 memperoleh nilai persentase akurasi tertinggi yaitu sebesar 84,62% pada percobaan dengan nilai $k=3$.
2. Hasil presisi tertinggi yang diperoleh dari metode Modified K-Nearest Neighbor pada klasifikasi masa studi mahasiswa Teknik Informatika berdasarkan Index Prestasi Semester 1 sampai 6 yaitu sebesar 100%.
3. Hasil *recall* tertinggi yang diperoleh dari metode Modified K-Nearest

Neighbor pada klasifikasi masa studi mahasiswa Teknik Informatika berdasarkan Index Prestasi Semester 1 sampai 6 yaitu sebesar 75%.

b. Saran

Berdasarkan hasil penelitian yang telah dilakukan, adapun saran yang diberikan untuk penelitian selanjutnya adalah sebagai berikut :

1. Penelitian ini dapat dijadikan bahan referensi untuk penelitian selanjutnya.
2. Untuk penelitian selanjutnya jumlah data yang digunakan sebaiknya ditingkatkan
3. Dapat menggunakan nilai k dan k-fold yang lebih beragam.

5. DAFTAR PUSTAKA

- Gorunescu, F. (2011). Data mining: Concepts, models and techniques. *Intelligent Systems Reference Library*. <https://doi.org/10.1007/978-3-642-19721-5>
- Hartaji, R. D. A., & Sedjo, P. (2012). Berkuliah Dengan Jurusan Pilihan Orang Tua. *Skripsi. Fakultas Psikologi Universitas Gunadarma*.
- Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*, 2(2), 213–219.
- Muhammad, B. L. (2015). Modified Nearest Neighbor Untuk Prediksi Curah Hujan. *Konferensi Nasional Sistem Dan Informatika 2015*, 1, 272–277.
- Parvin, H., Alizadeh, H., & Minati, B. (2010). A Modification on K-Nearest Neighbor Classifier. *Global Journal of Computer Science and Technology*.
- Ramadina.,(2015). Penerapan Fungsi Data Mining Klasifikasi Untuk Prediksi Masa Studi Mahasiswa Tepat Waktu Pada Sistem Informasi Akademik Perguruan Tinggi.
- Wafiyah Fakihatin., Hidayat Nurul., & Perdana Rizal Setya. (2017). Implementasi Algoritma Modified K-

Nearest Neighbor (MKNN) untuk klasifikasi Penyakit demam.

- Amelia, M. winny, Lumenta, A. S. ., & Jacobus, A. (2017). Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naïve Bayes. *Jurnal Teknik Informatika*, 11(1).
- Fitri, A. D., Eka, R. D., & Wahyu, W. A. (2017). Deteksi Penyakit Kucing dengan Menggunakan Modified K-Nearest Neighbor Teroptimasi (Studi Kasus : Puskesmas Klinik Hewan dan Satwa Sehat Kota Kediri). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 1(11), 1295–1301.