

# KLASIFIKASI KELULUSAN MAHASISWA PROGRAM STUDI TEKNIK INFORMATIKA MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR

Sofyan Vivi Dwiyanu<sup>1</sup>, Agung Nilogiri<sup>2</sup>

Jurusan Teknik Informatika Fakultas Teknik

Universitas Muhammadiyah Jember

[Sofyanwidianar25@gmail.com](mailto:Sofyanwidianar25@gmail.com), [Agungnilogiri@unmuhjember.ac.id](mailto:Agungnilogiri@unmuhjember.ac.id)

## ABSTRAK

Kelulusan tepat waktu dalam perkuliahan merupakan hal yang sangat diinginkan oleh mahasiswa, dosen maupun segenap civitas akademik. Tak sedikit dari mahasiswa tidak dapat lulus tepat waktu yang disebabkan banyak hal. Dalam hal ini penulis melakukan penelitian yaitu mengklasifikasikan mahasiswa lulus tepat waktu dan tidak tepat waktu pada mahasiswa jurusan Teknik Informatika Universitas Muhammadiyah Jember menggunakan metode K-Nearest Neighbor. Pada data mahasiswa dari angkatan 2011, 2012 sampai 2013 yang dijadikan data latih dan data uji dengan skenario uji menggunakan K-Fold Cross Validation sebanyak 5 kali yang menghasilkan bahwa model dengan nilai kurasi terbaik ada pada skenario ke 4. Dari data validasi berjumlah 50 dari total data 150. Dengan menggunakan jumlah tetangga terdekat = 3 dan pengukuran menggunakan Confusion Matrix didapat kesimpulan bahwa nilai akurasi yang didapat adalah 84% dan nilai presisi yang didapat adalah 62,5%.

Kata Kunci: K-Nearest Neighbor, Klasifikasi, K-Fold Cross Validation

### 1. Pendahuluan

Machine Learning merupakan suatu teknologi baru yang dikembangkan untuk persoalan pembelajaran data dalam jumlah sedikit. Machine Learning atau pembelajaran mesin merupakan pendekatan dalam Artificial Intelligence yang banyak digunakan untuk menggantikan atau menirukan perilaku manusia untuk menyelesaikan masalah atau melakukan otomatisasi. Informasi ini dapat dijadikan alternatif solusi dalam proses pengambilan keputusan. Penelitian yang dilakukan (Zul, 2015) mengajukan penggunaan algoritma K-Nearest Neighbor (K-NN) dan Naïve Bayes untuk prediksi nilai akhir semester yang sedang berjalan. Penggunaan algoritma K-Nearest Neighbor dan Naive Bayes untuk proses mencari rule dari klasifikasi nilai akhir semester yang sedang berjalan, sehingga dapat dimanfaatkan untuk rekomendasi pertimbangan untuk menentukan nilai akhir mahasiswa. Pada penelitian yang dilakukan, menyimpulkan bahwa algoritma terbaik yang didapatkan adalah K-NN dengan  $k = 15$ . Akurasi pengujian dengan menggunakan algoritma ini mencapai 95%. Berdasarkan latar belakang tersebut, penelitian Tugas Akhir ini akan membangun sebuah penelitian machine learning dengan judul “Klasifikasi Mahasiswa Program Studi Teknik Informatika Menggunakan Algoritma K-Nearest Neighbor” untuk membantu Program Studi dalam

memantau dan mengevaluasi perkembangan studi mahasiswa.

### 2. Kajian Pustaka

#### a. K-Nearest Neighbor

Metode K-Nearest Neighbour (KNN) merupakan salah satu metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Tujuannya adalah untuk mengklasifikasikan objek baru berdasarkan atribut dan data training. Klasifikasi dilakukan tanpa menggunakan model tetapi hanya berdasarkan memori. Algoritma K-Nearest Neighbor menggunakan klasifikasi ketetanggaan sebagai prediksi terhadap data baru.

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2}$$

Keterangan :

$d(x_i, x_j)$  : Jarak Euclidean (Euclidean Distance).

$(x_i)$  : record ke-  $i$

$(x_j)$  : record ke-  $j$

$(a_r)$  : data ke- $r$

$i, j : 1, 2, 3, \dots, n$

#### b. 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:

		Predicted Class	
		Class = yes	Class = no
Observes Class	Class = yes	TP	FN
	Class = no	FP	TN

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

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

Keterangan:

TP = jumlah true positive

TN = jumlah true negative

FP = jumlah false positive

FN = jumlah false negative

### 3. Metode Penelitian

#### a. Identifikasi data

Identifikasi adalah satu proses penelitian yang bisa dikatakan langkah penting pada proses identifikasi. Penelitian berupa hasil studi Mahasiswa angkatan 2011, 2012, dan 2013 Program Studi Teknik Informatika Universitas Muhammadiyah Jember.

Tujuannya untuk mengklasifikasi Kelulusan Mahasiswa Program Studi Teknik Informatika Universitas Muhammadiyah Jember.

#### b. Pengambilan data

Total data berjumlah 778 record yang didapat dari Pusat Data Informasi Universitas Jember, data tersebut adalah data mahasiswa yang lulus sampai tahun 2018. Gambaran data tersebut berupa identitas Mahasiswa seperti Tahun angkatan, NIM, Nama, tahun masuk kuliah, jenis kelamin, IP Semester, IP Kumulatif, lama studi dan tahun lulus kuliah. Dengan atribut yang tersedia peneliti akan menggunakan data tersebut untuk mengidentifikasi kelulusan mahasiswa yang akan datang tepat waktu atau tidak.

#### c. Preprocessing

Agar proses identifikasi kelulusan tersebut berjalan dengan baik., maka diperlukan proses preprocessing. Dalam proses ini akan dilakukan pembersihan noise, Atribut – atribut yang digunakan adalah IP semester 1, IP semester 2, IP semester 3, IP semester 4, IP semester 5, IP semester 6, tahun masuk kuliah, dan tahun lulus kuliah.

d. Penerapan Metode K-Nearest Neighbor Selanjutnya adalah tahap penerapan metode, dari hasil preprocessing akan dibagi menjadi 2 class : jumlah mahasiswa lulus studi tepat waktu dan jumlah lulus studi tidak lulus tepat waktu menggunakan metode K-Nearest Neighbor. Langkah – langkah pengimplementasian K-NN sebagai berikut :

- 1) Menentukan parameter k (jumlah tetangga paling dekat).
- 2) Menghitung kuadrat jarak euclidean objek terhadap data training yang diberikan.
- 3) Mengurutkan hasil b secara ascending.
- 4) Mengumpulkan kategori Y dan N (Klasifikasi K-nearest neighbor berdasarkan nilai k).
- 5) Dengan menggunakan kategori K-nearest neighbor yang paling mayoritas maka dapat diprediksikan kategori objek .

#### e. Hasil

Hasil dari penelitian ini berupa nilai tingkat akurasi dan presisi dari metode K-NN serta uji coba sederhana untuk mengetahui jumlah kelulusan mahasiswa dan lulus tepat waktu mahasiswa.

### 4. Implementasi dan Pengujian

#### a. Gambaran data

No	Nama	Keterangan
1	Nama Data	Data Kelulusan Mahasiswa
2	Atribut	- Tahun Angkatan - Nim - Nama - Jenis Kelamin - IP Semester 1 – 14 - Semester - Sks - Tanggal Masuk Kuliah - Tanggal Lulus Kuliah - Keterangan
3	Output	Selisih antara (tahun masuk – tahun lulus) jika kelulusan kurang dari 4 tahun maka dinyatakan lulus tepat waktu
4	Jumlah	778 data
5	Sumber Data	Pusat Data Informasi Universitas Muhammadiyah Jember
6	Tahun Data	2011 – 2013

## b. Implementasi

### 1) K-Fold Cross Validation

Pengujian model dalam penelitian ini menggunakan K-Fold Cross Validation adalah teknik validasi dengan membagi data secara acak kedalam k bagian. Dengan data berjumlah 125 record, data tersebut di bagi menjadi “k” bagian dan K - fold Cross, proses ini menggunakan percobaan sebanyak “k = i”. Dalam penelitian ini nilai k yang digunakan berjumlah 5 bagian karna untuk dipartisi menjadi data training dan data test atau 5-fold Cross Validation. Sebagai contoh pengerjaan dapat dilihat pada tabel 4.5.

Tabel 4.5 Skenario *K-Fold Cross Validation*

Skenario 1	Uji	Latih	Latih	Latih	Latih
Skenario 2	Latih	Uji	Latih	Latih	Latih
Skenario 3	Latih	Latih	Uji	Latih	Latih
Skenario 4	Latih	Latih	Latih	Uji	Latih
Skenario 5	Latih	Latih	Latih	Latih	Uji

Dari skenario 5-Fold Cross Validation diatas terhadap dataset ini menghasilkan nilai akurasi dan presisi. Dari kelima nilai tersebut akan diambil nilai akurasi dan presisi tertinggi akan menjadi model yang digunakan. Berikut hasil yang didapat :

	Akurasi	presisi
Skenario 1	76	0
Skenario 2	76	33,33
Skenario 3	64	0
Skenario 4	88	100
Skenario 5	76	33,33

### 2) Data validasi random

Tahap selanjutnya adalah pengujian data validasi dengan data latih yang telah melalui proses K-Fold Cross Valodation. Data latih tersebut berjumlah 100 record. Data record yang dihasilkan diatas akan diujikan terhadap 50 data validasi. Berikut adalah gambaran tabel data hasil hitung validasi.

Kriteria	Jumlah
TP	5
TN	37
FP	3
FN	5

Dari hasil penghitungan confusion matrix diatas, menunjukkan bahwa akurasi yang diperoleh sebesar 84% dan presisi diperoleh sebesar 62,5%.

## 5. Penutup

### a. Kesimpulan

Berdasarkan perhitungan menggunakan metode K-Nearest Neighbor maka dapat disimpulkan bahwa telah diperoleh klasifikasi kelulusan mahasiswa Teknik Informatika sebagai berikut :

- 1) Dari proses K-Fold Cross Validation didapatkan tingkat akurasi pada algoritma K-Nearest Neighbor untuk Klasifikasi Kelulusan Mahasiswa didapatkan sebuah hasil bahwa nilai akurasi sebesar 84%.
- 2) Dari proses K-Fold Cross Validation untuk presisi didapatkan sebuah hasil terhadap klasifikasi kelulusan Mahasiswa sebesar 62,5%.

### b. Saran

Bebearapa saran untuk pengembangan penelitian dimasa akan datang adalah sebagai berikut :

- 1) Penelitian ini membahas penggunaan K-Nearest Neighbor dalam penelitian klasifikasi kelulusan mahasiswa, diharapkan dalam penelitian selanjutnya dapat dibandingkan dengan memanfaatkan metode klasifikasi lainnya, sehingga dengan menggunakan banyak metode dapat mengetahui kelebihan masing – masing metode dan metode mana yang menghasilkan nilai akurasi dan presisi terbaik.
- 2) Pengklasifikasian terhadap data Mahasiswa Universitas Muhammadiyah Jember sebaiknya dilakukan secara rutin guna untuk menghindari penurunan kelulusan mahasiswa setiap tahunnya.

### Daftar Pustaka

- Adeniyi, D., Wei, Z., & Yongquan, Y. (2014). Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method. jurnal artikel, 91-106.
- Bablani, A., Edla, D. R., & Dodia, S. (2018). Classification of EEG Data using k-Nearest Neighbor approach for Concealed Information Test. Jurnal, 245.
- Bablania, A., Edla, D. R., & Dodia, S. (2018). Classification of EEG Data using k-Nearest Neighbor approach for Concealed Information Test. Jurnal, 243-248.
- Budyks. (2017, September 05). Pengertian dan Konsep Dasar Machine Learning. Diambil kembali dari Teknosains:

- <http://teknosains.com/others/pengertian-konsep-dasar-machine-learning>
- Defiyanti, S. (2014). Perbandingan: Prediksi Prestasi Belajar mahasiswa Menggunakan Teknik Data Mining (Study Kasus Fasilkom Unsika). *Jurnal*, 1-2.
- Hand, D. J. (2008). *Data Mining: Methods and Models* by D. T. Larose. *Jurnal*, 316.
- Hania, A. A. (2017). Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning. *Jurnal Artikel*, 1-5.
- Herdianto. (2013). *Prediksi Kerusakan Motor Induksi Menggunakan Metode Jaringan Saraf Tiruan Backpropagation*. Tesis.
- Maharani, Hasibuan, N. A., & Silalahi, N. (2017). Implementasi Data Mining Untuk Pengaturan Layout Minimarket Dengan Menerapkan Association Rule. *Jurnal*, 1-2.
- Nugroho, Y. S. (2014). *Data Mining Menggunakan Algoritma Naive Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro*. *Jurnal*, 2.
- Nugroho, Y. S. (2014). Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Mahasiswa Fakultas Komunikasi Dan informatika Universitas Muhammadiyah Surakarta. *Jurnal*, 1.
- Putra, J. W. (2018). *Pengenalan Pembelajaran Mesin dan Deep Learning*. Tokyo: Tokyo Institute of Technology.
- Ridwan, M., Suyono, h., & M.Sarosa. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Jurnal*, 59-.
- Safri, Y. F., Arifudin, R., & Muslim, M. A. (2018). K-Nearest Neighbor and Naive Bayes Classifier Algorithm in Determining The Classification of Healthy Card Indonesia Giving to The Poor. *Jurnal*, 12-16.
- Salmu, S., & Solichin, A. (2017). *Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naive Bayes: Studi Kasus UIN Syarif Hidayatullah Jakarta*. *Jurnal*, 701-703.
- Sumarlin. (2015). Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM. *Jurnal*, 52-62.
- Sumarno, H. (2017). *Komparasi Algoritma Klasifikasi Machine Learning Pada Analisis Sentimen Film Berbahasa Indonesia*. *Jurnal*, 189-196.
- Zul, M. I. (2015). *Prediksi Hasil Penilaian Akhir Mahasiswa pada Matakuliah Tertentu dengan Menggunakan Algoritma k-NN dan Naive Bayes*. *jurnal*, 3-5.

