

# Penerapan Data Mining Untuk Memprediksi Penerimaan Calon Mahasiswa Baru Fakultas Kedokteran Menggunakan Algoritma K-NN

<sup>1</sup> Amran Yobioktabera, <sup>2</sup>Angga Wahyu Wibowo

<sup>1,2</sup>Program Studi Teknik Informatika Jurusan Teknik Elektro Politeknik Negeri Semarang  
E-mail : <sup>1</sup>amranyobi@polines.ac.id, <sup>2</sup> anggawahyuwibowo@polines.ac.id

## Abstrak

Pada masa penerimaan mahasiswa baru, Fakultas Kedokteran masih menjadi tujuan utama lulusan sekolah menengah atas untuk melanjutkan pendidikannya. Untuk menjadi seorang mahasiswa kedokteran, calon mahasiswa diharuskan melakukan serangkaian tahapan. Selain itu, beberapa hal lain tentunya dapat menjadi poin penilaian dapat diterimanya calon mahasiswa. Penelitian ini berfokus pada prediksi penerimaan calon mahasiswa baru fakultas kedokteran ditinjau dari faktor-faktor yang mempengaruhinya. Untuk dapat memprediksi calon mahasiswa, dilakukan penerapan *data mining* dengan menggunakan algoritma *K-nearest neighbors (KNN)*. Dari percobaan yang dilakukan, diperoleh RMSE terbaik dengan data training 70 data yaitu sebesar 0.218 +/- 0.000. Untuk akurasi terbaik sebesar 76,1%. Dari penelitian ini dapat disimpulkan k-NN dapat digunakan untuk memprediksi penerimaan calon mahasiswa kedokteran, meski hasilnya belum maksimal.

**Kata kunci :** mahasiswa baru, data mining, k-nn

## Abstract

*During the admission period for new students, the Faculty of Medicine is still the main destination for high school graduates to continue their education. To become a medical student, prospective students are required to go through a series of stages. In addition, several other things can of course be an assessment point for the acceptance of prospective students. This study focuses on predicting the admission of prospective medical faculty students in terms of the factors that influence it. To be able to predict prospective students, data mining is implemented using the K-NN algorithm. From the experiments conducted, the best RMSE was obtained with 70 training data, namely 0.218 +/- 0.000. For the best accuracy of 76.1%. From this research, it can be concluded that k-NN can be used to predict the acceptance of prospective medical students, even though the results are not optimal.*

**Keywords :** new students, data mining, k-nn

## I. PENDAHULUAN

Penerimaan mahasiswa baru atau lebih dikenal dengan singkatan PMB merupakan sebuah proses penerimaan mahasiswa yang dilakukan oleh perguruan tinggi berdasarkan data pendaftaran yang dimasukkan oleh calon mahasiswa. Selain Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) yang dilakukan oleh PTN, seleksi juga dilakukan oleh Perguruan Tinggi Swasta (PTS) secara mandiri untuk menjangkau calon mahasiswa yang mendaftar sehingga diperoleh kualitas mahasiswa yang baik.

Salah satu fakultas yang melakukan proses seleksi secara ketat untuk mendapatkan mahasiswa adalah Fakultas Kedokteran. Hal ini menjadi sangat penting dikarenakan minat calon

mahasiswa mendaftar Fakultas Kedokteran juga sangat besar. Pada data PMB tahun 2017/2018 Universitas Muhammadiyah Semarang (Unimus), calon mahasiswa FK yang mendaftar pada tiap gelombang rata-rata mencapai 200 calon mahasiswa dengan Gelombang III menjadi yang terbanyak yaitu sebesar 221 calon mahasiswa.

Dalam proses seleksi yang dilakukan oleh perguruan tinggi, dilakukan beberapa tahapan dan faktor penentu yang digunakan sebagai dasar penyeleksian calon mahasiswa fakultas kedokteran. Akhir dari tahapan penyeleksian mahasiswa adalah penentuan mahasiswa yang diterima sebagai mahasiswa kedokteran.

Proses penentuan mahasiswa hingga pada akhirnya ditentukan pada rapat pleno keputusan diperlukan waktu yang lama, yang sebenarnya hal

ini dapat diprediksi berdasarkan parameter-parameter penentu yang ada. Belum banyak penelitian yang dilakukan berkaitan dengan penerimaan mahasiswa baru khususnya spesifik pada fakultas kedokteran.

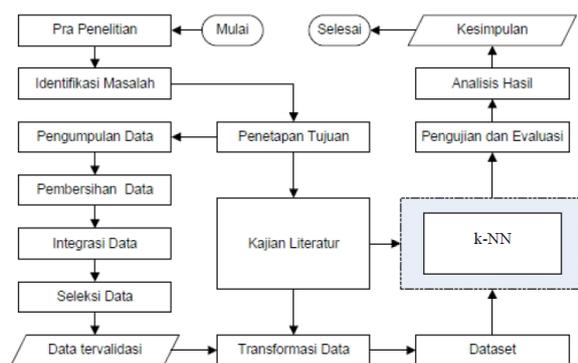
Beberapa penelitian berkaitan dengan student lebih banyak mengarah pada prediksi kelulusan mahasiswa berdasarkan faktor-faktor penentu kelulusan.

Pada penelitian dengan judul “Predicting and preventing student failure – using the k-nearest neighbour method to predict student performance in an online course environment”, Penulis melakukan pengujian menggunakan metode k-NN dengan pengukuran terbaik sebesar 69,6% (Tuomas, 2010).

Sedangkan pada penelitian berjudul “Graduation Prediction of Gunadarma University Students Using Naive Bayes Algorithm and C4.5 Algorithm”, penulis membandingkan antara metode Naive Bayes dan C4.5. Pada akhir penelitian ini, diperoleh hasil akurasi Naive Bayes sebesar 80,85% dan C4.5 sebesar 85,7% (Suhartinah, 2010).

Penelitian dengan judul “Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier” diperoleh nilai precision, recall dan accuracy masing-masing sebesar 83%, 50% dan 70% (Mujib, 2013).

## II. METODE PENELITIAN



Gambar 1. Metode Penelitian

Secara umum, tahapan penelitian dilakukan dengan melakukan identifikasi masalah, pengumpulan data, integrasi dan seleksi data sehingga diperoleh data yang tervalidasi. Selanjutnya data tersebut ditransformasikan menjadi sebuah *Dataset*. Selanjutnya *dataset* tersebut diolah menggunakan algoritma *k-NN*, yang akan menghasilkan prediksi. Hasil tersebut selanjutnya diuji, dievaluasi dan dianalisis.

Selanjutnya hasil analisis tersebut digunakan untuk menarik kesimpulan dari penelitian ini.

## III. HASIL DAN PEMBAHASAN

Data Set pada penelitian ini diperoleh dari Data Penerimaan Mahasiswa Baru (PMB) Universitas Muhammadiyah Semarang Gelombang III tahun 2014/2015 dengan uraian sebagai berikut

- Total data yang ada adalah sebanyak 211 pendaftar, dengan total data valid sebesar 146. Data dinyatakan valid jika calon mahasiswa melakukan semua rangkaian seleksi, sehingga parameter-parameter yang ada terisi dengan benar.
- *Data Training*, diambil beberapa data dengan jumlah yang berbeda dalam 4 tahap, yaitu 20 data, 50 data, 70 data dan 100 data sebagai *data training* dengan proses acak
- *Data Testing*, diambil dari sisa *data training* terbesar yaitu sebanyak 46 data.

Parameter yang digunakan dalam penelitian ini adalah sebagai berikut :

Variable Independent :

- **Akademik** : nilai ujian akademik  
Rentang nilai : 0 - 100
- **Psikiatri** : hasil ujian psikiatri  
(Disarankan, Tidak Valid, Tidak Disarankan, Dipertimbangkan)
- **Kesehatan\_Umum** : hasil tes kesehatan umum  
Lulus (L) dan Tidak Lulus (L)
- **Kesehatan\_Mata** : hasil tes kesehatan mata  
Lulus (L), Tidak Lulus (L) dan Visus OS
- **Kesehatan\_THT** : hasil tes kesehatan THT  
Lulus (L) dan Tidak Lulus (L)
- **Infaq** : besaran infaq yang dapat diberikan  
0 hingga 500.000.000
- **Agama**  
(Islam, Katolik, Protestan, Hindu, Budha, Kong Hu Chu)

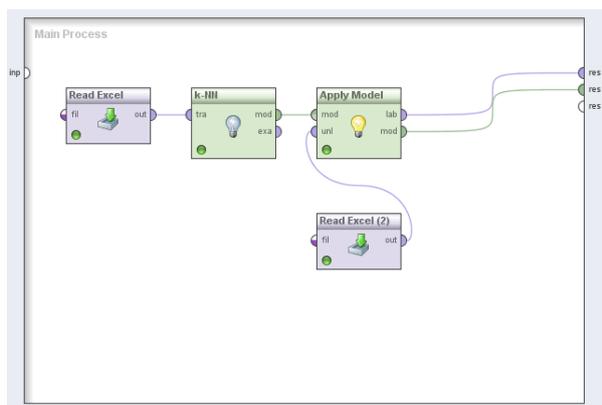
Varibel Dependent :

- **Keterangan**  
Lulus (1) dan Tidak Lulus (0)

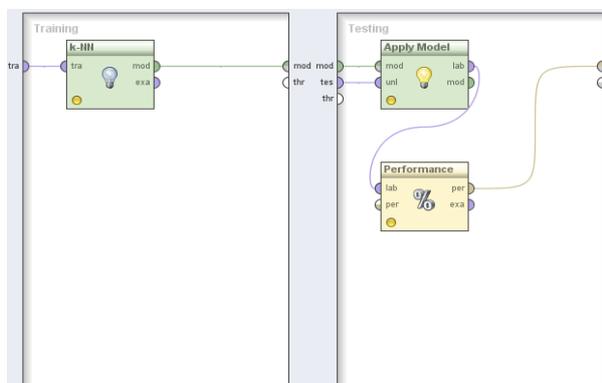
Pengujian dilakukan untuk mengetahui kinerja dari algoritma *k-NN* dalam mengklasifikasikan data yang telah ditentukan sebelumnya. Proses pengujian dilakukan menggunakan *Tool Rapidminer*.

Row No.	KETERANG.	NO.	NAMA PES.	AKADEMIK	PSIKIATRI	KESEHATA.	KESEHATA.	KESEHATA.	INFAQ FK	AGAMA
1	1	1	CM-TR-1	68	Disarankan	L	L	L	150000000	Islam
2	0	2	CM-TR-2	17	Disarankan	TL	TL	TL	0	Islam
3	0	3	CM-TR-3	28	Tidak Valid	L	L	L	100000000	Islam
4	0	4	CM-TR-4	40	Disarankan	L	L	L	100000000	Islam
5	1	5	CM-TR-5	75	Disarankan	L	L	L	800000000	Islam
6	0	6	CM-TR-6	25	Disarankan	L	L	L	750000000	Islam
7	0	7	CM-TR-7	33	Disarankan	L	L	L	360000000	Islam
8	0	8	CM-TR-8	24	Tidak Valid	L	L	L	750000000	Islam
9	1	9	CM-TR-9	75	Disarankan	L	L	L	1000000000	Islam
10	0	10	CM-TR-10	28	Disarankan	L	TL	L	225000000	Islam
11	0	11	CM-TR-11	32	Tidak Disars	L	L	L	1000000000	Islam
12	0	12	CM-TR-12	33	Diperlimban	L	L	L	1200000000	Islam
13	1	13	CM-TR-13	48	Disarankan	L	L	L	750000000	Islam
14	0	14	CM-TR-14	25	Tidak Valid	TL	TL	TL	750000000	Islam
15	0	15	CM-TR-15	28	Tidak Disars	L	L	L	1700000000	Islam
16	0	16	CM-TR-16	21	Tidak Valid	L	L	L	2250000000	Islam
17	0	17	CM-TR-17	36	Disarankan	L	L	L	750000000	Islam
18	0	18	CM-TR-18	44	Disarankan	L	L	L	400000000	Islam
19	0	19	CM-TR-19	25	Disarankan	L	L	L	1250000000	Islam

Gambar 2. Data Training Pada Tool Rapidminer



Gambar 3. Model Prediksi Pada Tool Rapidminer



Gambar 4. Model Performance

Pengujian ini dilakukan dalam 4 tahapan, tahap pertama akan dimasukkan data training sebesar 20 data, 50 data pada tahap dua, 70 data pada tahap ketiga dan 100 data untuk tahapan keempat. Untuk mengevaluasi tiap tahapan, digunakan *RMSE* sebagai tolak ukurnya. Pada tiap tahapan, dilakukan pengujian terhadap *data testing* yang telah ditentukan sebelumnya yaitu sebesar 46 data.

1. Percobaan ke-1

Menggunakan *data training* sebanyak 20 data, berikut hasil pengujian :

- *RMSE* : *root\_mean\_squared\_error*: 0.816 +/- 0.000

- Kesalahan Prediksi sebanyak 14 dari 46 atau 30,4%, sehingga memiliki tingkat akurasi sebesar 69,6%

2. Percobaan ke-2

Menggunakan data training sebanyak 50 data, berikut hasil pengujiannya :

- *RMSE* : *root\_mean\_squared\_error*: 0.447 +/- 0.000

- Kesalahan Prediksi sebanyak 12 dari 46 atau 26%, sehingga memiliki tingkat akurasi sebesar 74%

3. Percobaan ke-3

Menggunakan data training sebanyak 70 data, berikut hasil pengujiannya :

- *RMSE* : *root\_mean\_squared\_error*: 0.218 +/- 0.000

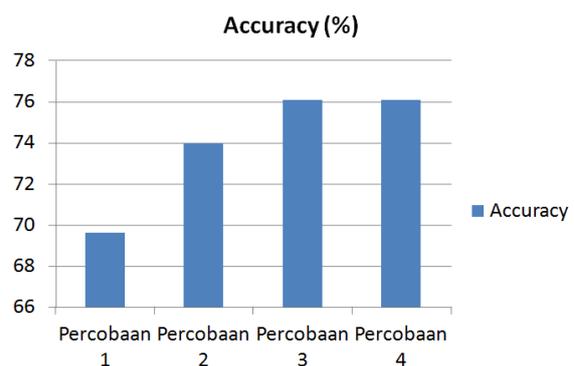
- Kesalahan Prediksi sebanyak 11 dari 46 atau 23,9%, sehingga memiliki tingkat akurasi sebesar 76,1%

4. Percobaan ke-4

Menggunakan data training sebanyak 100 data, berikut hasil pengujiannya :

- *RMSE* : *root\_mean\_squared\_error*: 0.316 +/- 0.000

- Kesalahan Prediksi sebanyak 11 dari 46 atau 23,9%, sehingga memiliki tingkat akurasi sebesar 76,1%



Gambar 5. Grafik Tingkat Akurasi

Dari 4 percobaan yang dilakukan diatas, dapat diperoleh *RMSE* terbaik pada percobaan ke-3 dengan *data training* sebanyak 70 data yaitu sebesar 0.218 +/- 0.000. Tingkat akurasi terbaik dapat terlihat pada percobaan ke-3 dan ke-4 dengan akurasi 76,1%.

Dari data *rule* dengan tingkat akurasi terbaik nantinya akan digunakan sebagai model dalam penentuan prediksi penerimaan calon mahasiswa.

#### IV. KESIMPULAN

Dari pembahasan diatas maka dapat diambil kesimpulan bahwa, *k-NN* dapat digunakan untuk memprediksi penerimaan calon mahasiswa kedokteran, meski hasilnya belum maksimal. Hasil terbaik penerapan *k-NN* untuk memprediksi calon mahasiswa adalah sebesar 76,1% yaitu pada percobaan ke-3 dan ke-4. Dari hasil percobaan, dapat terlihat bahwa grafik akurasi cenderung naik jika ditambahkan *data training*.

Secara umum, hasil dari pengujian menunjukkan nilai yang sedang dan belum maksimal. Hal ini dikarenakan faktor penentu diterimanya calon mahasiswa memiliki nilai yang kurang konsisten.

Penentuan *data training* dapat mempengaruhi hasil pengujian. Pola dari *data training* tersebut digunakan sebagai *rule* dalam proses *testing*.

Using Naive Bayes Algorithm and C4.5 Algorithm.

- [9] Mujib Ridwan, Hadi Suyono dan M.Sarosa. 2013. Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier

#### DAFTAR PUSTAKA

- [1] Erdogan, S.Z. dan M. Timor. 2005. A Data Mining Application In A Student Database. *Journal Of Aeronautics And Space Technologies*. Volume 2 Number 2: 53-57.
- [1] Tuomas Tanner, Hannu Toivonen. 2010. Predicting and preventing student failure – using the k-nearest neighbour method to predict student performance in an online course environment.
- [2] Kabakchieva, D. 2012. Student Performance Prediction by Using Data Mining Classification Algorithms. *IJCSMR*. Vol 1 Issue 4: 686-690.
- [3] Larose, D.T. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*. John Willey & Sons, Inc.
- [4] McLeod, Jr.R. dan G.P. Schell. 2007. *Management Information System*. 10<sup>th</sup> ed. Pearson Education, Inc. Ali Akbar Yulianto dan Afia R. Fitriati (penterjemah). 2008. *Sistem Informasi Manajemen*. Edisi 10. Nina Setyaningsih (editor). Salemba Empat. Jakarta.
- [5] Ogor, E.N. 2007. Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques. *IEEE Computer Society*.
- [6] Santosa, B. 2007. *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Graha Ilmu. Yogyakarta.
- [7] Shovon, Md.H.I. dan M. Haque. 2012. Prediction of Student Academic Performance by an Application of K-Means Clustering Algorithm. *IJARCSSE*. Volume 2 Issue 7: 353-355.
- [8] Suhatinah, M.S. dan Ernastuti. 2010. Graduation Prediction of Gunadarma University Students