

Penerapan Algoritma k-Nearest Neighbour (kNN) dalam Memprediksi Kelulusan Mahasiswa pada Konteks Praktikum di Laboratorium

Riza Purwandi
Magister Teknik Informatika Universitas Pamulang
e-mail: purwandir@gmail.com

Abstrak—Pentingnya pemahaman mahasiswa dalam mengikuti praktikum menjadi tolak ukur keberhasilan tenaga pengajar dalam menyampaikan materi praktikum. Prediksi kelulusan mahasiswa dirancang untuk mendukung tenaga pengajar untuk membimbing mahasiswa agar memberikan perhatian khusus terhadap mahasiswa yang di prediksi tidak lulus (tepat waktu) dalam praktikum di Laboratorium sehingga mahasiswa tersebut bisa memperbaiki kemampuannya agar dapat lulus praktikum tepat waktu. Dalam penelitian ini untuk memprediksi kelulusan mahasiswa menggunakan algoritma k-Nearest Neighbor. Algoritma kNN merupakan metode klasifikasi dimana kelas yang paling banyak muncul akan menjadi kelas hasil klasifikasi. Penelitian ini mengambil sample data mahasiswa semester VII salah satu Sekolah Tinggi Swasta di Jakarta Selatan, menggunakan 14 data training dan 1 data testing. Dengan menggunakan $k=9$ yang diterapkan menggunakan metode k-Nearest Neighbor untuk memprediksi kelulusan mahasiswa.

Kata Kunci—Algoritma K-Nearest Neighbor; Prediksi; Klasifikasi; Mahasiswa; Laboratorium.

I. PENDAHULUAN

Perguruan tinggi dituntut untuk menyelenggarakan pendidikan yang berkualitas bagi mahasiswa sehingga menghasilkan sumber daya manusia yang memiliki daya saing. Dalam perguruan tinggi mahasiswa merupakan asset yang sangat penting bagi institusi pendidikan oleh karena itu perlu diperhatikan tingkat kelulusan mahasiswa tepat pada waktunya [1]. Pentingnya pemahaman mahasiswa dalam mengikuti praktikum menjadi tolak ukur keberhasilan tenaga pengajar dalam menyampaikan materi praktikum. Prediksi kelulusan mahasiswa dirancang untuk mendukung tenaga pengajar dalam membimbing mahasiswa dalam praktikum agar dapat memberikan perhatian khusus terhadap mahasiswa yang diprediksi tidak lulus (tepat waktu). Dengan prediksi tersebut diharapkan tenaga pengajar dapat memperbaiki nilainya agar dapat lulus tepat waktu.

Klasifikasi adalah jenis analisis data yang dapat membantu orang memprediksi label kelas sampel harus diklasifikasikan. Berbagai macam teknik klasifikasi telah diusulkan dalam bidang-bidang seperti pembelajaran mesin, sistem pakar dan statistik [2]. Algoritma KNN bekerja berdasarkan jarak terpendek dari objek baru ke objek lama dengan menentukan nilai k . Nilai k merupakan parameter untuk menentukan jarak terdekat antara objek baru terhadap objek lama [3].

Dengan adanya prediksi terhadap kelulusan mahasiswa dengan metode klasifikasi Algoritma k-Nearest Neighbor (kNN) diharapkan agar mahasiswa dapat menjalankan praktikum dan lulus tepat waktu.

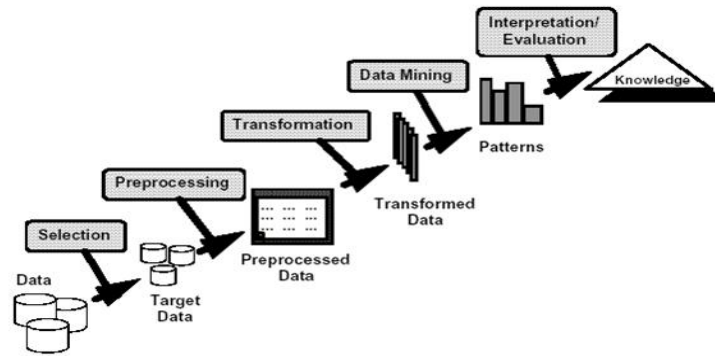
II. METODE PENELITIAN

A. Prediksi

Prediksi (peramalan) adalah upaya untuk menduga atau memperkirakan sesuatu yang akan terjadi di masa mendatang dengan menggunakan metode ilmiah dengan menggunakan berbagai informasi yang relevan dari masa lalu (historis). Tujuan dari prediksi adalah untuk mendapatkan informasi tentang kejadian yang paling mungkin terjadi di masa mendatang. Metode prediksi dapat dilakukan secara kualitatif melalui pendapat para pakar atau secara kuantitatif dengan perhitungan secara matematis. Salah satu metode prediksi kuantitatif adalah menggunakan analisis deret waktu (time series) [4].

B. Data Mining

Data mining merupakan salah satu tahapan yang terdapat dalam proses knowledge discovery in database (KDD) yaitu proses pencarian informasi-informasi baru dan bernilai yang terdapat di dalam suatu kumpulan data atau database. Di dalam penelitian ini, penerapan data mining menggunakan tahapan-tahapan yang terdapat pada KDD. Proses KDD dimulai dari menetapkan tujuan hingga evaluasi [5]. Tahapan KDD (Knowledge discovery in Database) diantaranya adalah *selection*, *preprocessing*, *transformation*, *data mining* dan *interpretation / evaluation* [6]. Berikut ini penjelasan tahapan proses KDD dalam memprediksi kelulusan mahasiswa pada Praktikum Laboratorium di salah satu Sekolah Tinggi Swasta di Jakarta Selatan.



Gambar 1.
Tahapan Proses KDD

1) *Selection*

Data mahasiswa yang didapat dilakukan proses pemilihan terlebih dahulu. Dengan dilakukannya proses seleksi proses pengolahan data akan menjadi baik dan tujuan penelitian dapat dicapai. Pada penelitian ini peneliti menggunakan 15 data mahasiswa.

2) *Preprocessing*

Proses pemeriksaan kembali data mahasiswa untuk membersihkan data-data redundansi, dan bernilai kosong (*null*). Proses ini dilakukan dengan tujuan agar data yang diolah telah bersih sehingga hasil perhitungan algoritma memberikan hasil yang akurat dan sesuai.

3) *Transformation*

Proses pada tahap ini merupakan instruksi untuk merubah input menjadi output yang diinginkan, disesuaikan dengan algoritma klasifikasi yang akan digunakan.

4) *Data Mining*

Dalam memilih algoritma data mining untuk melakukan pencarian proses data mining, ada banyak pilihan. Tujuan dan proses KDD secara keseluruhan menentukan metode atau algoritma yang tepat.

5) *Interpretation / evaluation*

Merupakan proses menginterpretasikan hasil rule yang didapatkan dari teknik *data mining*.

C. *Classification*

Analisis data seperti klasifikasi membantu memprediksi label kelas yang harus diberikan kepada sampel. K-Nearest Neighbor, Bayesian Classifiers, dan Neural adalah beberapa teknik klasifikasi yang umum digunakan dalam bidang seperti pembelajaran mesin, sistem pakar, dan statistik [2].

D. *K-Nearest Neighbor (kNN)*

Salah satu metode klarifikasi data yang efektif adalah algoritma k-nearest neighbor, yang menemukan kasus dengan menggunakan pencocokan bobot untuk menghitung jarak antara kasus baru dan kasus lama[7]. Steps K-Nearest Neighbor[8]:

1. Tentukan parameter K, yang merupakan jumlah tetangga paling dekat. Parameter K pada pengujian ditentukan berdasarkan nilai K ideal pada saat pelatihan.
2. Menghitung kuadrat jarak euclid masing-masing objek terhadap data sampel yang diberikan
3. Mengurutkan objek-objek tersebut ke dalam kelompok dengan jarak euclidian terkecil.
4. Mengumpulkan kategori Y, yang merupakan klasifikasi dekat tetangga terdekat.
5. Dengan menggunakan kategori mayoritas, hasil klasifikasi dapat diperoleh.

III. HASIL DAN PEMBAHASAN

A. *Analisa Masalah*

Penelitian ini menggunakan algoritma k Nearest Neighbor untuk mengumpulkan data dari mahasiswa Semester VII salah satu perguruan tinggi swasta di Jakarta Selatan. Berikut adalah data sampel yang dipakai.

Tabel 1.
Data Mahasiswa Semester VII

NPM	Nama	Nilai Praktikum	Nilai Presensi	Nilai Rata Rata
10413041	Rino Dwi Hutomo	75.75	100	87.875
10413049	Putra Toba	72.75	100	86.375
10413065	Alizar Akbar	74.25	100	87.125
10413081	Dinar Gutama	60.63	75	67.81
10413082	Rita Agustinus	75.375	100	87.6875
10413089	Dony Al Fiqih	53.75	60	56.875
10413090	Risma Oktaviani	70.625	100	85.3125
10413092	Rizky Fauzi	72.5	100	86.25

NPM	Nama	Nilai Praktikum	Nilai Presensi	Nilai Rata Rata
10413098	Ahmad Fikri Mubarak	72.75	100	86.375
10413105	Bagus Arianto	44.5	62.5	53.5
10413151	Zaenal Abidin	45	62.5	53.75
10413172	Ruwanda Yosandi	54.375	50	52.1875
10413181	Marharrisatama Rizal	60.625	75	67.8125
10413190	Rinto Prabowo	79.5	100	89.75
10413201	Dimas Nugraha	85	100	92.5

Setelah sampel data didapatkan, peneliti membagi menjadi 2 bagian yaitu data training dan data testing seperti di bawah ini :

1) *Data Training*

Data training terdiri dari nilai praktikum dan nilai kehadiran mahasiswa selama semester VII dan Keterangan yang berisi kategori kelulusan seperti pada Tabel 2.

Tabel 2.
Data Training

NPM	Nama	Nilai Praktikum	Nilai Presensi	Nilai Rata Rata	Keterangan
10413041	Rino Dwi Hutomo	75.75	100	87.875	LULUS
10413049	Putra Toba	72.75	100	86.375	LULUS
10413065	Alizar Akbar	74.25	100	87.125	LULUS
10413081	Dinar Gutama	60.63	75	67.81	TIDAK LULUS
10413082	Rita Agustinus	75.375	100	87.6875	LULUS
10413089	Dony Al Fiqih	53.75	60	56.875	TIDAK LULUS
10413090	Risma Oktaviani	70.625	100	85.3125	LULUS
10413092	Rizky Fauzi	72.5	100	86.25	LULUS
10413098	Ahmad Fikri Mubarak	72.75	100	86.375	LULUS
10413105	Bagus Arianto	44.5	62.5	53.5	TIDAK LULUS
10413151	Zaenal Abidin	45	62.5	53.75	TIDAK LULUS
10413172	Ruwanda Yosandi	54.375	50	52.1875	TIDAK LULUS
10413181	Marharrisatama Rizal	60.625	75	67.8125	LULUS
10413190	Rinto Prabowo	79.5	100	89.75	LULUS
10413201	Dimas Nugraha	85	100	92.5	LULUS

2) *Data Testing*

Data training terdiri dari nilai praktikum dan nilai kehadiran mahasiswa selama semester VII seperti pada Tabel 3.

Tabel 3.
Data Testing

NPM	Nama	Nilai Praktikum	Nilai Presensi	Nilai Rata Rata	Keterangan
10413181	Marharrisatama Rizal	66.25	87.5	76.86	?

B. *Penerapan Algoritma kNN*

Setelah data training dan data testing disiapkan langkah yang selanjutnya dilakukan dalam algoritma kNN adalah :

1) Menentukan Parameter K

Nilai Parameter K dalam penelitian ini adalah K = 9.

2) Menghitung kuadrat jarak (*euclidean distance*) terkecil pada masing-masing object terhadap data sampel yang diberikan dengan rumus:

- Rumus 1 Bentuk Umum *euclidean distance*

$$d(x, y) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + \dots + (y_n - x_n)^2}$$

- Rumus 2 Rumus *euclidean distance*

$$d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2}$$

Keterangan :

- d_{ij} = jarak euclidean obyek data ke-i dan obyek data ke-j
- m = banyaknya parameter yang digunakan
- x_{ik} = Objek data ke-i pada parameter ke-k
- x_{jk} = Objek data ke-j pada parameter ke-k

Tabel 4.
Data Training dan Testing

NPM	Nama	Nilai Praktikum	Nilai Presensi	Nilai Rata Rata	Keterangan
10413041	Rino Dwi Hutomo	75.75	100	87.86	LULUS
10413049	Putra Toba	72.75	100	86.38	LULUS
10413065	Alizar Akbar	74.25	100	87.13	LULUS
10413081	Dinar Gutama	60.625	75	67.81	TIDAK LULUS
10413082	Rita Agustinus	75.38	100	87.69	LULUS
10413089	Dony Al Fiqih	53.75	60	56.88	TIDAK LULUS
10413090	Risma Oktaviani	70.63	100	85.31	LULUS
10413092	Rizky Fauzi	72.5	100	86.25	LULUS
10413098	Ahmad Fikri Mubarak	72.75	100	86.38	LULUS

NPM	Nama	Nilai Praktikum	Nilai Presensi	Nilai Rata Rata	Keterangan
10413105	Bagus Arianto	44.5	62.5	53.5	TIDAK LULUS
10413151	Zaenal Abidin	45	62.5	53.75	TIDAK LULUS
10413172	Ruwanda Yosandi	54.38	50	52.19	TIDAK LULUS
10413190	Rinto Prabowo	79.5	100	89.75	LULUS
10413201	Dimas Nugraha	85	100	92.5	LULUS
10413181	Marharrisatama Rizal	66.25	87.5	76.86	?

Contoh penerapan rumus 1 pada data mahasiswa ke 1 terhadap data testing (data mahasiswa ke 15).

$$\begin{aligned}
 d(x, y) &= \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + (y_3 - x_3)^2} \\
 d(1,15) &= \sqrt{(66.25 - 75.75)^2 + (87.5 - 100)^2 + (76.86 - 87.86)^2} \\
 &= \sqrt{(-9.5)^2 + (-12.5)^2 + (-11)^2} \\
 &= \sqrt{90.25 + 156.25 + 121} \\
 &= \sqrt{367.5} \\
 &= 19.17
 \end{aligned}$$

Berikut adalah hasil penerapan rumus 1 pada data training terhadap data testing.

Tabel 5.
 Hasil Penerapan Rumus 1 Pada Data Testing

NPM	Nama	Nilai Praktikum	Nilai Presensi	Nilai Rata Rata	Keterangan	Euclidean Distance Data Testing
10413041	Rino Dwi Hutomo	75.75	100	87.86	LULUS	19.17
10413049	Putra Toba	72.75	100	86.38	LULUS	17.00
10413065	Alizar Akbar	74.25	100	87.13	LULUS	18.05
10413081	Dinar Gutama	60.625	75	67.81	TIDAK LULUS	16.43
10413082	Rita Agustinus	75.38	100	87.69	LULUS	18.89
10413089	Dony Al Fiqih	53.75	60	56.88	TIDAK LULUS	36.23
10413090	Risma Oktaviani	70.63	100	85.31	LULUS	15.71
10413092	Rizky Fauzi	72.5	100	86.25	LULUS	16.84
10413098	Ahmad Fikri Mubarok	72.75	100	86.38	LULUS	17.00
10413105	Bagus Arianto	44.5	62.5	53.5	TIDAK LULUS	40.54
10413151	Zaenal Abidin	45	62.5	53.75	TIDAK LULUS	40.13
10413172	Ruwanda Yosandi	54.38	50	52.19	TIDAK LULUS	46.43
10413190	Rinto Prabowo	79.5	100	89.75	LULUS	22.32
10413201	Dimas Nugraha	85	100	92.5	LULUS	27.43

- 3) Mengurutkan objek-objek pada tabel 5 ke dalam kelompok yang memiliki jarak *euclidian* terkecil ke terbesar, hasil pengurutan tersebut seperti yang dapat di lihat pada Tabel 6.

Tabel 6.
 Hasil Penerapan Rumus 1 Pada Data Testing

NPM	Nama	Keterangan	Euclidean Distance Data Testing	Peringkat Jarak Terkecil
10413041	Rino Dwi Hutomo	LULUS	19.17	8
10413049	Putra Toba	LULUS	17.00	4
10413065	Alizar Akbar	LULUS	18.05	6
10413081	Dinar Gutama	TIDAK LULUS	16.43	2
10413082	Rita Agustinus	LULUS	18.89	7
10413089	Dony Al Fiqih	TIDAK LULUS	36.23	11
10413090	Risma Oktaviani	LULUS	15.71	1
10413092	Rizky Fauzi	LULUS	16.84	3
10413098	Ahmad Fikri Mubarok	LULUS	17.00	5
10413105	Bagus Arianto	TIDAK LULUS	40.54	13
10413151	Zaenal Abidin	TIDAK LULUS	40.13	12
10413172	Ruwanda Yosandi	TIDAK LULUS	46.43	14
10413190	Rinto Prabowo	LULUS	22.32	9
10413201	Dimas Nugraha	LULUS	27.43	10

- 4) Mengumpulkan kategori Y (klasifikasi *nearest neighbor*)
 Langkah berikutnya yang dilakukan setelah mendapatkan peringkat pada jarak adalah mengumpulkan kategori Y (klasifikasi *nearest neighbor*). Nilai Parameter yang ditentukan dalam penelitian ini sesuai dengan pada langka 1 adalah 5. Hasil klasifikasi kNN seperti terlihat pada Tabel 7.

Tabel 7.
 Hasil Penerapan Rumus 1 Pada Data Testing

NPM	Nama	Keterangan	Euclidean Distance Data Testing	Peringkat Jarak	Klasifikasi
10413041	Rino Dwi Hutomo	LULUS	19.17	8	Tidak
10413049	Putra Toba	LULUS	17.00	4	Ya
10413065	Alizar Akbar	LULUS	18.05	6	Tidak
10413081	Dinar Gutama	TIDAK LULUS	16.43	2	Ya
10413082	Rita Agustinus	LULUS	18.89	7	Tidak
10413089	Dony Al Fiqih	TIDAK LULUS	36.23	11	Tidak
10413090	Risma Oktaviani	LULUS	15.71	1	Ya
10413092	Rizky Fauzi	LULUS	16.84	3	Ya
10413098	Ahmad Fikri Mubarak	LULUS	17.00	5	Ya
10413105	Bagus Arianto	TIDAK LULUS	40.54	13	Tidak
10413151	Zaenal Abidin	TIDAK LULUS	40.13	12	Tidak
10413172	Ruwanda Yosandi	TIDAK LULUS	46.43	14	Tidak
10413190	Rinto Prabowo	LULUS	22.32	9	Tidak
10413201	Dimas Nugraha	LULUS	27.43	10	Tidak

- 5) Menggunakan data mayoritas untuk mendapatkan hasil klasifikasi Langkah selanjutnya setelah mendapatkan kategori Y dengan $K = 9$ adalah mengumpulkan kategori mayoritas. Hasil dari pengumpulan kategori mayoritas dapat dilihat pada Tabel 8.

Tabel 8.
 Hasil Kategori Mayoritas

NPM	Nama	Keterangan	Euclidean Distance Data Testing	Peringkat Jarak	Klasifikasi
10413041	Rino Dwi Hutomo	LULUS	19.17	8	Ya
10413049	Putra Toba	LULUS	17.00	4	Ya
10413065	Alizar Akbar	LULUS	18.05	6	Ya
10413081	Dinar Gutama	TIDAK LULUS	16.43	2	Ya
10413082	Rita Agustinus	LULUS	18.89	7	Ya
10413090	Risma Oktaviani	LULUS	15.71	1	Ya
10413092	Rizky Fauzi	LULUS	16.84	3	Ya
10413098	Ahmad Fikri Mubarak	LULUS	17.00	5	Ya
10413190	Rinto Prabowo	LULUS	22.32	9	Ya

Dari hasil Tabel 8 di atas data testing termasuk dalam kategori Lulus dikarenakan jumlah klasifikasi bernilai Lulus berjumlah 8 data dan 1 data bernilai Tidak Lulus. Hasil dari data testing tersebut dapat dilihat pada Tabel 9.

Tabel 9.
 Hasil Data Testing

NPM	Nama	Nilai Praktikum	Nilai Presensi	Nilai Rata Rata	Keterangan
10413041	Rino Dwi Hutomo	75.75	100	87.86	LULUS
10413049	Putra Toba	72.75	100	86.38	LULUS
10413065	Alizar Akbar	74.25	100	87.13	LULUS
10413081	Dinar Gutama	60.625	75	67.81	TIDAK LULUS
10413082	Rita Agustinus	75.38	100	87.69	LULUS
10413089	Dony Al Fiqih	53.75	60	56.88	TIDAK LULUS
10413090	Risma Oktaviani	70.63	100	85.31	LULUS
10413092	Rizky Fauzi	72.5	100	86.25	LULUS
10413098	Ahmad Fikri Mubarak	72.75	100	86.38	LULUS
10413105	Bagus Arianto	44.5	62.5	53.5	TIDAK LULUS
10413151	Zaenal Abidin	45	62.5	53.75	TIDAK LULUS
10413172	Ruwanda Yosandi	54.38	50	52.19	TIDAK LULUS
10413190	Rinto Prabowo	79.5	100	89.75	LULUS
10413201	Dimas Nugraha	85	100	92.5	LULUS
10413181	Marharrisatama Rizal	66.25	87.5	76.86	LULUS

IV. KESIMPULAN

Berdasarkan teknik klasifikasi menggunakan metode algoritma kNN maka peneliti mengambil kesimpulan sebagai berikut:

1. Teknik klasifikasi menggunakan metode algoritma kNN dapat memprediksi Kelulusan Mahasiswa dalam praktikum di laboratorium.
2. Berdasarkan teknik klasifikasi kNN dengan $K=9$ maka mahasiswa bernama Marharrisatama Rizal mendapatkan keterangan Lulus.
3. Banyaknya data training yang digunakan dapat mempengaruhi hasil dari metode kNN.

DAFTAR PUSTAKA

- [1] A. Rohman, 2015 "Model Algoritma K-Nearest Neighbor (K-Nn) Untuk Prediksi Kelulusan Mahasiswa, Fakultas Teknik Universitas Pandanaran Semarang."
- [2] Yu, L., Chen, G., Koronios, A., Zhu, S., & Guo, X. (n.d.). Application and Comparison of Classification Techniques in Controlling Credit Risk. World, 2007–2007.
- [3] Samsugi, S., Neneng, N., & Suprpto, G. N. F. (2021). Otomatisasi Pakan Kucing Berbasis Mikrokontroler Intel Galileo Dengan Interface Android. J-SAKTI (Jurnal Sains Komputer Dan Informatika), 5(1), 143–152.
- [4] Nurmahaludin, "Analisis Perbandingan Metode Jaringan Syaraf Tiruan Dan Regresi Linear Berganda Pada Prakiraan Cuaca," Jurnal INTEKNA, no. 2, 2014.
- [5] Tomar, D., dan Agarwal, S., 2013, A Survey On Data Mining Approaches For Healthcare, Int. J. Bio-Science Bio-Technology, No. 5, Vol. 5, Hal 241–266.
- [6] Fayyad, U., Shapiro-Piatetsky, G. & Smyth, P., 1996. From Data Mining to Knowledge Discovery in Database. AI Magazine, pp. 37-54.
- [7] R. I. Ndaumanu and M. R. Arief, Kusri, 2014 "Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa dengan Metode K-Nearest Neighbor," Jatisi, vol. 1, no. 1, pp. 1–15.
- [8] M. S. Mustafa and I. W. Simpen, 2014 "Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining (Studi Kasus : Data Akademik Mahasiswa STMIK Dipanegara Makassar)," Citec J., vol. Vol. 1, No, pp. 270–281.