



SNESTIK

Seminar Nasional Teknik Elektro, Sistem Informasi,
dan Teknik Informatika

<https://ejurnal.itats.ac.id/snestik> dan <https://snestik.itats.ac.id>



Informasi Pelaksanaan :

SNESTIK III - Surabaya, 11 Maret 2023

Ruang Seminar Gedung A, Kampus Institut Teknologi Adhi Tama Surabaya

Informasi Artikel:

DOI : 10.31284/p.snestik.2023.4388

Prosiding ISSN 2775-5126

Fakultas Teknik Elektro dan Teknologi Informasi-Institut Teknologi Adhi Tama Surabaya
Gedung A-ITATS, Jl. Arief Rachman Hakim 100 Surabaya 60117 Telp. (031) 5945043
Email : snestik@itats.ac.id

Prediksi Kelulusan Mahasiswa dengan Metode K-Nearest Neighbor pada Jurusan Sistem Informasi Institut XYZ

Oriene Sativa Disya Putri, Anggi Yhurinda Perdana Putri

Institut Teknologi Adhi Tama Surabaya

e-mail: oriensativaputri@gmail.com

ABSTRACT

The timely graduation rate of students is very important for a university. The graduation rate is an indicator of the level of success of a tertiary institution in carrying out its teaching and learning process. Student graduation is one of the instruments to determine the accreditation of a university. There are two graduate profile points, namely the percentage of graduates on time and DO (Drop Out). The method used is K-Nearest Neighbor (K-NN) this method is a method used to classify objects based on learning data that is closest to the object. In research that uses (K-NN) to predict student graduation time, the result is that the K-NN approach in this study produces the greatest accuracy value, namely 95.15%. The result is a test with a K value of up to 20 times from K = 5, K = 7, K = 11, K = 15 to K = 51 using 144 training data of ITATS information system students who have graduated from class of 2014 to class of 2017, then data testing is tested with normalized train data to find the euclidian distance and determine K to carry out repetition tests of up to 20 times the predicted result is that there is no change in value even though the K value changes.

Keywords: Data Mining; Classification; K-Nearest Neighbor; Graduation; College.

ABSTRAK

Semakin banyaknya mahasiswa dalam suatu universitas menyelesaikan studinya tepat waktu, maka semakin baik kinerja universitas tersebut. Terdapat dua lulusan menurut Badan Akreditasi Nasional Perguruan Tinggi, yaitu kelulusan tepat waktu dan Drop Out (DO) atau berhenti. Metode yang digunakan adalah K-Nearest Neighbor (K-NN) metode ini merupakan metode yang digunakan untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Tujuan

penelitian mampu memprediksi mengenai kelulusan Mahasiswa Jurusan Sistem Informasi di Institut Teknologi Adhi Tama Surabaya, apakah bernalil lulusan tepat waktu atau lulus tidak tepat waktu. Dalam penelitian menggunakan (K-NN) untuk meramalkan kelulusan mahasiswa mendapatkan hasil bahwa pendekatan K-NN dalam penelitian ini menghasilkan nilai akurasi sebesar yaitu 95,15%. Hasilnya adalah pengujian dengan nilai K hingga 20 kali dari $K = 5$, $K = 11$ hingga $K = 51$ dengan menggunakan 144 data train mahasiswa sistem informasi ITATS yang telah lulus dari angkatan 2014 hingga angkatan 2017, lalu data testing diujikan dengan data train yang telah ternormalisasi hingga pencarian ecludian distance dan penentuan K hingga melakukan pengujian pengulangan hingga 20 kali hasil prediksi nya ialah tidak ada perubahan nilai meski merubah nilai K tersebut.

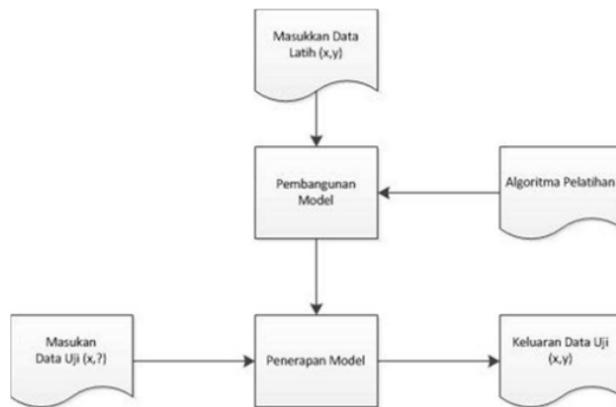
Kata kunci: Data Mining; Klasifikasi; *K-Nearest Neighbor*; Kelulusan; Perguruan Tinggi.

PENDAHULUAN

Semakin banyaknya mahasiswa dalam suatu universitas menyelesaikan studinya tepat waktu, maka semakin baik kinerja universitas tersebut, sehingga tingkat ketepatan waktu menjadi salah satu kriteria penilaian akreditasi suatu universitas atau suatu program studi [1]. Akreditasi universitas meningkat seiring dengan jumlah lulusan. Menurut (BAN-PT, 2019) kelulusan mahasiswa merupakan salah satu instrumen untuk menentukan akreditasi suatu universitas. *Data mining* adalah salah satu hal cabang ilmu komputer yang bisa memprediksi, penggunaan ilmu data mining dapat digunakan sebagai pertimbangan dalam mengambil keputusan lebih lanjut tentang faktor yang mempengaruhi kelulusan khususnya faktor dalam data induk mahasiswa [2]. Pada penelitian sebelumnya, bahwa faktor yang paling berpengaruh dalam penentuan klasifikasi kinerja akademik mahasiswa adalah Indeks Prestasi Kumulatif (IPK), Indeks Prestasi Semester (IPS) 1 (satu) hingga 4 (empat) dan Jenis Kelamin [3]. Keunggulan metode K-NN dalam memprediksi yaitu tangguh terhadap training data yang noise dan efektif apabila training data-nya besar. Lalu pada tahun 2021 mahasiswa sistem informasi angkatan 2017 mengalami penurunan kelulusan mahasiswa yaitu sebanyak 33,78% yang dimana mahasiswa yang lulus tepat waktu adalah 25 mahasiswa dari 74 mahasiswa aktif. Data yang digunakan dalam penelitian ini adalah data mahasiswa sistem informasi angkatan 2014 hingga 2017 sebagai data train dan data mahasiswa sistem informasi angkatan 2018 dan 2019 sebagai data testing. Sedangkan atribut yang digunakan adalah IPK, Cuti Semester, Status Bekerja, Mengulang Mata Kuliah, Jumlah Organisasi. Dalam penelitian yang menggunakan K-Nearest Neighbor (K-NN) untuk meramalkan waktu kelulusan mahasiswa, penelitian tersebut mendapatkan hasil bahwa pendekatan K-NN dalam penelitian ini menghasilkan nilai akurasi terbesar yaitu 95,15% [4]. Terdapat dua tujuan dari penelitian ini yaitu dapat memprediksi kelulusan mahasiswa menggunakan algoritma *K-Nearest Neighbor*, dan mampu memprediksi mengenai kelulusan Mahasiswa Jurusan Sistem Informasi di Institut Teknologi Adhi Tama Surabaya, apakah bernalil lulusan tepat waktu atau lulus tidak tepat waktu.

METODE KLASIFIKASI

Klasifikasi didefinisikan secara detail sebagai suatu kegiatan atau pekerjaan yang melakukan pelatihan atau pembelajaran terhadap fungsi target f yang memetakan untuk setiap vektor (set fitur) x ke dalam satu dari sejumlah label kelas y yang tersedia. Pekerjaan pelatihan tersebut akan menghasilkan suatu model yang kemudian disimpan sebagai memori [5]. Metode klasifikasi yang diterapkan dalam penelitian ini dapat dilihat pada Gambar 1 berikut.



Gambar 1. Metode Klasifikasi.

Sumber: (Prasetyo, 2014)

Beberapa metode pengembangan yang digunakan dalam mengembangkan aplikasi prediksi kelulusan mahasiswa yaitu meliputi :

1. Identifikasi Masalah

memprediksi kelulusan mahasiswa berdasarkan parameter yang sudah dijelaskan sebelumnya pada batasan masalah.

2. Observasi

Menyebar kuesioner terhadap seluruh mahasiswa Jurusan Sistem Informasi baik yang sudah lulus ataupun yang belum lulus dan membagi menjadi data *testing* ataupun data *training*.

3. Analisis

Tahapan analisis bertujuan untuk mencari kebutuhan pengguna dan organisasi serta menganalisa kondisi yang ada (sebelum diterapkan sistem yang baru).

4. Desain

Tahapan desain bertujuan untuk menentukan spesifikasi detail dari komponen-komponen program (manusia, perangkat keras, perangkat lunak, jaringan dan data) dan produk-produk informasi yang sesuai dengan hasil tahapan analisis.

5. Implementasi

Tahapan ini merupakan tahapan untuk mendapatkan atau mengembangkan perangkat keras dan perangkat lunak (pengkodean program), melakukan pengujian, pelatihan dan perpindahan ke sistem yang baru.

6. Perawatan

Tahapan ini dilakukan ketika sistem sudah dioperasikan. Pada tahapan perawatan ini dilakukan proses monitoring, evaluasi dan perubahan bila diperlukan.

Data Mining

Nama data Mining sebenarnya mulai dikenal sejak tahun 1990, ketika pekerjaan pemanfaatan data menjadi sesuatu yang penting dalam berbagai bidang, mulai dari bidang akademik, bisnis, hingga medis [6]. Data mining bisa diterapkan pada berbagai bidang yang mempunyai sejumlah data, tetapi karena wilayah penelitian dengan sejarah yang belum lama, dan belum melewati masa ‘remaja’, maka data mining masih menjadi perdebatan pada posisi bidang pengetahuan yang memilikinya. Maka, Daryl Pregibon menyatakan bahwa “data mining adalah campuran dari statistik, kecerdasan buatan, dan riset basis data” yang masih berkembang [7].

Konsep K-Nearest Neighbor (K-NN)

Algoritma K-NN menjadi salah satu metode yang paling tua dan populer. Nilai K yang digunakan disini menyatakan jumlah tetangga terdekat yang bisa dilibatkan dalam penentuan

prediksi pada label kelas data uji. Sebanyak K tetangga terdekat yang terpilih kemudian, akan dilakukan pemungutan suara dari K tetangga terdekat tersebut. Kelas dengan jumlah suara tetangga terbanyak, maka akan dianggap sebagai label kelas hasil prediksi pada data uji tersebut [8]. K-NN termasuk algoritma supervised learning, yang dimana hasil dari query instance baru, diklasifikasikan berdasarkan mayoritas dari kategori pada K-NN. Kelas yang paling banyak muncul, yang akan menjadi kelas hasil klasifikasi maka rumus perhitungan jarak dengan *Euclidean* seperti di bawah ini [9]:

$$d_{xy} = \sqrt{\sum_{i=1}^K (X_i - Y_i)^2} \quad (1)$$

Dimana:

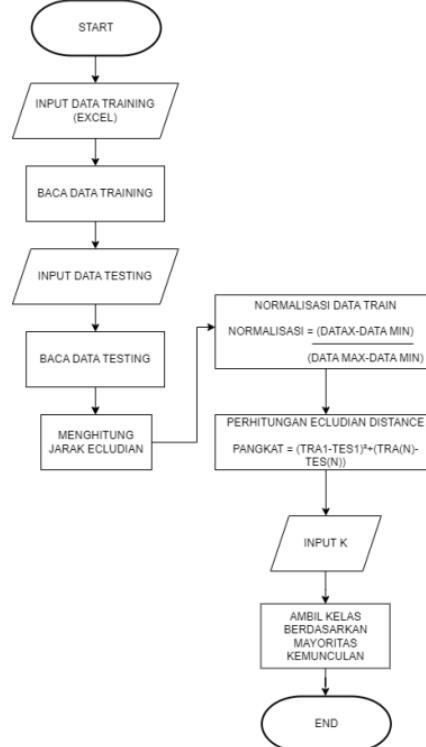
X_i adalah data *training* atau data yang sudah memiliki nilai

Y_i adalah data uji baru yang belum memiliki nilai.

Tahap pertama adalah harus menyiapkan data latih terlebih dahulu, data latih adalah data atau informasi yang diambil dari waktu sebelumnya yang sudah diketahui kelas atau labelnya. Tahap selanjutnya adalah buat tabel untuk perhitungan normalisasinya. Rumus dari normalisasinya sendiri adalah

$$\text{Normalisasi} = \frac{\text{DataX} - \text{DataMin}}{\text{DataMax} - \text{DataMin}} \quad (2)$$

Algoritma K-NN dapat secara umum dapat digambarkan dengan Flowchart pada Gambar 2.



Gambar 2. Flowchart K-NN.

Sebagai metode dalam pengumpulan suatu data mahasiswa yang menjadi landasan utama dalam proses prediksi kelulusan mahasiswa ini adalah menggunakan wawancara dan penyebaran kuesioner secara online. Wawancara dilakukan dengan pihak sekretaris jurusan Program Studi Sistem Informasi ITATS dan diperoleh data mahasiswa 144. Sedangkan kuesioner secara online dilakukan menggunakan google form yang disebarluaskan kepada seluruh angkatan 2018 dan 2019 mahasiswa Sistem Informasi ITATS.

PERHITUNGAN K-NN

Sebelum melakukan proses perhitungan K-NN perlu disiapkan data mahasiswa yang telah lulus(training), dan data mahasiswa yang belum lulus(testing). Tabel data train bisa dilihat pada tabel 1.

Tabel 1. Data Training

No	NPM	IPK	Status Bekerja	Cuti Semester	Jumlah MK Mengulang	Jumlah Organisasi	Keterangan Lulus
1.	13.2014.1.00408	3,51	0	0	1	3	1
2.	14.2016.1.00315	3,34	0	0	1	2	1
3.	13.2016.1.00502	3,55	0	0	0	1	1
4.	13.2017.1.00626	3,26	1	0	2	0	2
5.	13.2016.1.00494	3,72	0	0	0	2	1
6.	13.2016.1.00506	3,5	0	0	0	1	1
7.	13.2016.1.00514	3,55	0	0	0	1	1
8.	13.2017.1.00650	3,3	1	0	2	2	2
9.	13.2017.1.00651	3,31	1	0	1	1	2
10.	13.2017.1.00657	3,19	1	1	1	1	2
		MA X	MAX	MAX	MAX	MAX	MAX
		3,72	1	1	2	3	2
		MIN	MIN	MIN	MIN	MIN	MIN
		3,19	0	0	0	0	1

10 dari 144 data mahasiswa yang telah lulus digunakan sebagai data training pada simulasi pada metode K-NN. Data tersebut akan dinormalisasi menggunakan rumus persamaan (1). Untuk tabel data testing bisa dilihat pada tabel 2.

Tabel 2. Data Testing

no	NPM	IPK	Status Bekerja	Cuti semester	jumlah MK yang mengulang	jumlah Organisasi	Keterangan Lulus

1.	13.2019.1.00774	3,75	0	0	0	2	
2.	13.2019.1.00832	3,29	0	0	0	1	
3.	13.2019.1.00837	3,02	0	0	1	0	
4.	13.2019.1.00781	3,61	0	0	0	0	
5.	13.2019.1.00870	3,5	1	2	0	1	

Tabel 2 menunjukkan 5 dari 100 data testing sebagai pengujian yang akan dihasilkan oleh K-NN. Dari data ini akan dihasilkan kesimpulan sebuah data akan masuk kedalam kelas lulus tepat waktu(1), atau lulus tidak tepat waktu(2). Selanjutnya proses penormalisasi data fungsi dari penormalisasi data ini bertujuan untuk mengurangi redundansi data serta menstandarisasi data agar dapat diolah dengan baik, rumus bisa dilihat pada persamaan (1), tabel data yang telah di normalisasi bisa dilihat pada tabel 3.

Tabel 3. Normalisasi Data.

IPK	Status Kerja	Cuti	Jumlah Matkul	Jumlah Organisasi	Ket Lulus
0,603773585	0	0	0,5	1	0
0,283018868	0	0	0,5	0,666666667	0
0,679245283	0	0	0	0,333333333	0
0,132075472	1	0	1	0	1
1	0	0	0	0,666666667	0
0,58490566	0	0	0	0,333333333	0
0,679245283	0	0	0	0,333333333	0
0,20754717	1	0	1	0,666666667	1
0,226415094	1	0	0,5	0,333333333	1
0	1	1	0,5	0,333333333	1

Setelah proses penormalisasi data dilanjut dengan menggunakan persamaan Euclidean Distance. Sebagai contoh data pertama pada data testing yang akan dihitung dengan NPM 13.2019.1.00774. Proses perhitungan setiap atribut akan diujikan dengan seluruh data train yang telah ternormalisasi sehingga nantinya akan menghasilkan nilai jarak sesuai dengan jumlah data lama, perhitungan bisa dilihat pada persamaan (1). Dari perhitungan tersebut diperolehlah hasil jarak bisa dilihat Tabel 4. Berdasarkan nilai K yang telah ditetapkan adalah 5, maka jarak yang diambil adalah 5 data terkecil atau 5 data yang terdekat dengan data training seperti Tabel 4.

Tabel 4. Euclidean Distance dan Nilai K

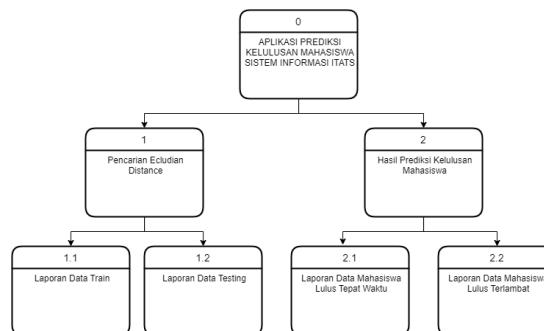
Distance Data 1	K= 5
3,3389729940563	1
3,7480309427678	
3,4938964365933	1
4,3691392621990	
3,0561868034820	1

3,5770937863703	1
3,4938964365933	1
4,0406373051648	
4,0550497364614	
4,3692422429728	

Berdasarkan hasil diatas diketahui dari 10 Euclidean Distance pada data 1 atau NPM 13.2019.1.00774 diketahui jarak 5 terkecil lebih banyak muncul nilai 1 yaitu lulus tepat waktu.

HASIL DAN PEMBAHASAN

Sebelum memasuki hasil dan pembahasan akan dilakukan perancangan dan implementasi sistem yang dimana untuk menerapkan dan mengimplementasikan algoritma yang telah dibahas sebelumnya agar dapat digunakan dan diimplementasikan kedalam bahasa pemrograman seperti perancangan diagram jenjang bisa dilihat pada Gambar 3.



Gambar 3. Diagram Jenjang

Berdasarkan gambar 2 diatas diagram konteks yang dibangun hanya untuk 1 user yang dimana user ini adalah Kepala Program Studi Jurusan Sistem Informasi ITATS yang bisa menggunakan. User memiliki hak akses untuk update data mahasiswa dan seluruh konten yang ada pada sistem. Sistem yang dikelola dengan hak akses user ini dapat ditunjukkan pada gambar 3 untuk CRUD Data testing dan gambar 4 untuk hasil prediksi kelulusan mahasiswa yang belum lulus.

No.	NPM	IPK	Status Belajar	Gaji Semester	Jumlah MA Menganggur	Jumlah Organisasi
1	13.2019.1.0023	3.7	1	0	0	1
2	13.2019.1.0012	3.89	1	0	1	3
3	13.2019.1.0074	3.8	0	0	0	1
4	13.2019.1.0046	3.8	0	0	0	1
5	13.2019.1.0074	3.75	0	0	0	2
6	13.2019.1.0032	3.29	0	0	0	1
7	13.2019.1.0087	3.82	0	0	1	0

Jumlah Lulus Tepat Waktu = 73	
Persentase Lulus Tepat Waktu	60,83333333333333%
Persentase Lulus Terlambat	39,16666666666667%

(a)

(b)

Gambar 4. a) Data Testing, b) Hasil Prediksi

Hasil pengujian dilakukan dengan tujuan untuk mengetahui kualitas dari suatu perangkat lunak. pengujian dilakukan terhadap beberapa aspek berdasarkan metode k-nearest neighbor yang memiliki variabel data yang sudah ditentukan. Variabel yang akan digunakan dalam pengujian

adalah IPK, Status bekerja, Status cuti, Jumlah mata kuliah mengulang, Jumlah organisasi dan Keterangan lulus variabel data mahasiswa ini akan dibedakan menjadi data testing (data mahasiswa yang belum lulus) dan data training (data mahasiswa yang telah lulus), setelah dibedakan maka akan melewati sebuah proses penormalisasi data, pencarian euclidean distance, dan penentuan nilai k. Hasil pengujian menggunakan $K=5$ atau 5 tetangga terdekat untuk mengambil nilai lulus tepat waktu atau tidak di setiap masing-masing data mahasiswa ini. Diketahui dari 106 mahasiswa angkatan 2018 dan 2019 hanya 33 mahasiswa yang lulus tepat waktu, dan mahasiswa yang tidak tepat waktu sebanyak 73 mahasiswa atau sebesar 68.867924528302% dan setiap perulangan tidak menunjukkan hasil yang berbeda namun hanya ada perubahan terhadap jumlah nilai K.

KESIMPULAN

Pada penelitian ini berhasil dibangun sebuah sistem berbasis website untuk memprediksi kelulusan mahasiswa sistem informasi ITATS menggunakan metode algoritma K-Nearest Neighbor untuk membantu Institut Teknologi AdhiTama Surabaya dalam memprediksi kelulusan mahasiswanya, dalam penelitian tersebut mendapatkan hasil bahwa pendekatan K-NN dalam penelitian ini menghasilkan nilai akurasi terbesar yaitu 95,15%. Tujuan dalam melakukan penelitian ini adalah Mampu memprediksi mengenai kelulusan Mahasiswa Jurusan Sistem Informasi di Institut Teknologi Adhi Tama Surabaya, apakah bermilai lulusan tepat waktu atau lulus tidak tepat waktu. Pengujian K hingga 20 kali menggunakan sistem prediksi ini hingga $K=51$. Setelah melewati serangkaian proses yang telah terlewati akan dilanjut mengujikan seluruh data mahasiswa yang belum lulus menggunakan data 106 mahasiswa untuk diujikan dan hasil kesimpulan nya mahasiswa belum lulus hanya 33 mahasiswa yang lulus tepat waktu, dan mahasiswa yang tidak tepat waktu sebanyak 73 mahasiswa atau sebesar 68.867924528302% selama perulangan hingga 20kali tidak ada perubahan hasil sama sekali.

DAFTAR PUSTAKA

- [1] M. Broto Legowo and B. Indiarto, "Model Sistem Penjaminan Mutu Berbasis Integrasi Standar Akreditasi BAN-PT dan ISO 9001:2008," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 2, pp. 90–98, 2017, doi: 10.29207/resti.v1i2.51.
- [2] B. Zahedi, B. Nahid-Mobarakeh, S. Pierfederici, and L. E. Norum, "A robust active stabilization technique for dc microgrids with tightly controlled loads," *Proc. - 2016 IEEE Int. Power Electron. Motion Control Conf. PEMC 2016*, vol. VI, no. 1, pp. 254–260, 2016, doi: 10.1109/EPEPEMC.2016.7752007.
- [3] M. Ridwan, H. Suyono, and M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Eeccis*, vol. 7, no. 1, pp. 59–64, 2013, doi: 10.1038/hdy.2009.180.
- [4] A. Rohman, "Model Algoritma K-nearest Neighbor untuk memprediksi kelulusan mahasiswa," *Foreign Aff.*, vol. 91, no. 5, pp. 1–9, 2012.
- [5] E. Prasetyo, *Konsep dan Aplikasi Menggunakan Matlab*. CV Andi Offset, 2012.
- [6] F. Gorunescu, *Data Mining Concepts, Models and Techniques*, vol. 21, no. 1. 2010. [Online]. Available: <http://journal.um-surabaya.ac.id/index.php/JKM/article/view/2203>
- [7] P. Smyth and D. Pregibon, "Data-Driven Evolution of Data Mining Algorithms," vol. 45, no. 8, pp. 33–37, 2002.
- [8] S. T.P and K. M, "Data Mining : Data Lecture Notes for Chapter 2 Introduction to Data Mining by What is Data ?," *Ratio*, 2004.
- [9] S. Sayad, "K Nearest Neighbors regression," pp. 1–18, 2010.