



IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOUR UNTUK PREDIKSI KETEPATAN KELULUSAN

Manarul Hidayat^{1)*}, Ahmad Faqih²⁾, Tati Suprapti³⁾

^{1,2,3)} Teknik Informatika, STMIK IKMI Cirebon, Kota Cirebon, Jawa Barat
email: [arul151299@gmail.com^{1\)}](mailto:arul151299@gmail.com), [Ahmadfaqih367@gmail.com^{2\)}](mailto:Ahmadfaqih367@gmail.com),
[tatisuprapti00@gmail.com^{3\)}](mailto:tatisuprapti00@gmail.com)

Abstrak

Dalam sebuah sistem pendidikan, mahasiswa merupakan aset penting dari sebuah perguruan tinggi dan oleh karena itu, penting untuk memperhatikan persentase mahasiswa yang lulus tepat waktu. Namun, ada ketidakseimbangan antara input dan output dari mahasiswa selesai. Mahasiswa yang mendaftar dalam jumlah besar, namun siswa yang lulus tepat waktu dibandingkan dengan yang terlambat sesuai peraturan lebih sedikit. Dalam penelitian ini penulis bertujuan untuk menerapkan metode K-NN menggunakan cross validation untuk prediksi tingkat kelulusan mahasiswa di STMIK IKMI. Hasil penelitian ini berupa model dan evaluasi hasil prediksi kelulusan mahasiswa, apakah lulus tepat waktu atau tidak tepat waktu. Berdasarkan hasil perancangan, implementasi, pengujian menggunakan program RapidMiner untuk prediksi kelulusan mahasiswa menggunakan metode k-NN dengan Cross Validation menghasilkan akurasi sebesar 70,28%, error sebesar 29,78%, dan AUC 0,739.

Kata Kunci : Kelulusan, Mahasiswa, K-NN, Cross Validation.

Abstract

In an education system, students are an important asset of a college and therefore, it is important to pay attention to the percentage of students who graduate on time. However, there is an imbalance between the inputs and outputs of the completed students. Students who enroll in large numbers, but students who graduate on time compared to those who are late according to regulations are fewer. In this study, the author aims to apply the K-NN method using cross validation to predict student graduation rates at STMIK IKMI. The results of this study are in the form of models and evaluations of student graduation predictions, whether they graduate on time or not on time. Based on the results of the design, implementation, testing using the RapidMiner program for predicting student graduation using the k-NN method with Cross Validation resulting in an accuracy of 70.28%, an error of 29.78%, and AUC of 0.739.

Keywords: Graduation, Student, K-NN, Cross Validation.

PENDAHULUAN

Bertambahnya mahasiswa tiap tahunnya akan menyebabkan bertambah pula data mahasiswa di program studi. Prediksi ketepatan kelulusan

mahasiswa mendukung prodi agar mahasiswa lulus tepat waktu. Lulus tepat waktu adalah keinginan semua mahasiswa. Penyelesaian tepat waktu bermanfaat bagi kedua belah pihak. Pertama, perusahaan



cenderung mencari lulusan baru, yang memudahkan mahasiswa untuk mencari pekerjaan. Kedua, untuk perguruan tinggi karena membantu meningkatkan kualitas perguruan tinggi, misalnya melalui peningkatan akreditasi (Suriana et al., 2017)

Salah satu standar yang ditetapkan oleh program studi Teknik Informatika STMIK IKMI Cirebon adalah ditetapkan untuk lulusan teknik informatika menghasilkan lulusan tepat waktu yaitu 4 tahun dengan maksimal delapan semester dengan IPK minimal 3,00. Dengan terus meningkatnya jumlah mahasiswa serta standar IPK lulusan yang terus menjadi besar pula, hingga dibutuhkan sesuatu penelitian untuk prediksi nilai serta waktu lulusan mahasiswa bersumber pada data-data yang telah didapat. Atas dasar permasalahan ini lah, penulis melakukan penelitian dengan memprediksi kelulusan mahasiswa dengan klasifikasi data mining khususnya menggunakan algoritma K-Nearest Neighbors.

Banyak penelitian yang mengkaji dan analisis kelulusan mahasiswa menggunakan metode klasifikasi data mining dengan berbagai algoritma, salah satunya dalam penelitian prediksi kelulusan mahasiswa tepat waktu berdasarkan usia, jenis kelamin dan indeks prestasi menggunakan algoritma Decision Tree menghasilkan nilai akurasi 90,77 % dengan jumlah data 389 orang (Romadhona et al., 2017). Pada penelitian sebelumnya oleh (Fatoni & Noviandha, 2018) penelitian mengenai difteri menggunakan metode KNN dilakukan perhitungan similaritas pada kasus lama dengan kasus baru. Hasil diagnose penyakit difteri berdasarkan gejala yang dialami mendapatkan hasil tingkat akurasi sebesar 95,17%. Penelitian yang dilakukan oleh (Susanto et al., 2018), bertujuan untuk mengetahui uji kelayakan prediksi

kelulusan mahasiswa dengan K-Nearest Neighbor. Hasil pengujian menunjukkan bahwa nilai akurasi mencapai 98.46% dengan precision mencapai 99.53%.

Tujuan utama penelitian ini yaitu untuk dapat memprediksi ketepatan lulus mahasiswa menggunakan algoritma K-NN dengan program data mining Rapid Miner dan mengimplementasikan Algoritma K-NN sebagai salah satu metode klasifikasi. Hasil dari prediksi diharapkan dapat mengantisipasi mahasiswa-mahasiswa yang tidak sesuai dengan target standar lulusan.

Data Mining

Data mining ialah sebuah proses pengumpulan dan pengolahan data. Tujuan data mining adalah untuk mengekstrak informasi penting tentang data. Data mining digunakan untuk memprediksi kondisi berdasarkan data dan informasi. Data mining adalah salah satu bidang ilmu komputer yang berfokus pada pembelajaran mesin (Muliono et al., 2020). Data mining digunakan untuk memprediksi kondisi berdasarkan data dan informasi.

Knowledge Discovery In Database (KDD)

KDD yakni salah satu tata cara dalam menciptakan, mengganti, serta menyempurnakan informasi serta pola yang bermakna dari database mentah untuk digunakan dalam domain ataupun aplikasi yang berbeda. KDD dalam penambangan informasi merupakan pendekatan terprogram serta analitis memodelkan informasi dari database untuk mengekstrak pengetahuan yang bermanfaat serta berlaku. Informasi mining membentuk tulang punggung KDD serta karenanya sangat berarti untuk segala tata cara. Informasi mining merupakan pusat dari sesuatu proses knowledge discovery in database (KDD)



yang memakai algoritma spesial untuk mengecek informasi, membangun model dan menciptakan pola yang belum ditemui (Zanuardi & Suprayitno, 2018)

K-Nearest Neighbour

K-nearest neighbor adalah salah satu algoritma machine learning dengan pendekatan supervised learning. Algoritma ini menerapkan lazy learning” atau “instant based learning. Algoritma KNN digunakan untuk klasifikasi dan regresi. Algoritma K-Nearest Neighbor adalah klasifikasi metode yang dapat mengklasifikasikan data baru berdasarkan jarak dari data baru ke data terdekat/tetangga dalam data pembelajaran (Atma & Setyanto, 2018).

METODE PENELITIAN

Metode yang digunakan dalam penelitian ini mengikuti tahapan yang ada pada proses knowledge discovery in database (KDD). Berikut ini tahapan proses KDD :

1. Selection

Sumber data yang digunakan dalam penelitian ini adalah data mahasiswa Teknik Informatika pada tahun 2016 hingga tahun 2017. Data nilai berasal dari transkrip nilai mahasiswa yang terdapat pada database bagian akademik perguruan tinggi. Sedangkan data kelulusan berasal dari data wisuda mahasiswa setiap tahunnya.

2. Preprocessing

Merupakan proses untuk mengidentifikasi bagian-bagian data yang tidak lengkap, tidak benar, tidak relevan, tidak akurat atau hilang dan kemudian memperbaiki, mengganti, atau menghapusnya sesuai dengan kebutuhan. Preprocessing penelitian ini

mengganti data di atribut yang kosong .

3. Transformation

Data transformasi merupakan suatu proses untuk mengubah format, struktur, atau nilai data. Pada proses data sudah benar jadi tidak menggunakan data transformation.

4. Data Mining

Data Mining Pada pengolahan data, dilakukan tahap pemodelan untuk proses klasifikasi yaitu menerapkan algoritma kNN. Pengolahan data menggunakan perangkat lunak rapid miner 9.9 Jumlah data yang akan diolah sebanyak 225 data.. Teknik validasi yang digunakan untuk pada proses klasifikasi adalah k-Fold Cross Validation.

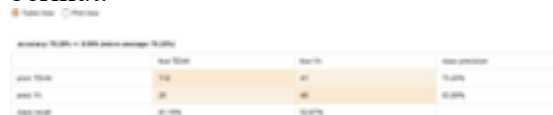
5. Evaluation

Hasil Evaluasi untuk model ini menggunakan Metode evaluasi standard: stratified 10-fold dan menghasilkan nilai akurasi , error , dan Area Under Curve (AUC) ROC (Receiver Operating Characteristic)

HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penelitian. Hasil penelitian dapat dilengkapi dengan tabel, grafik (gambar), dan/atau bagan. Bagian pembahasan memaparkan hasil pengolahan data, menginterpretasikan penemuan secara logis, mengaitkan dengan sumber rujukan yang relevan. [Times New Roman, 12, normal], spasi 1. Format gambar png/jpg.

Hasil Performance dari penerapan algoritma k-NN dapat dilihat sebagai berikut.



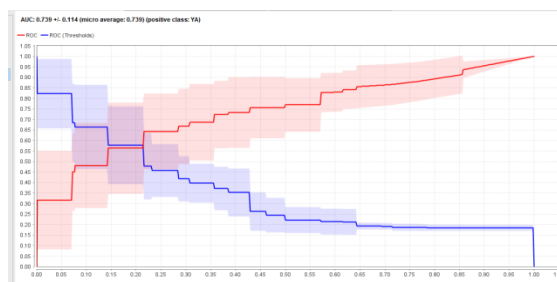
Gambar 1. Confusion Matrix



Hasil diatas menunjukkan nilai rata-rata akurasi dari masing masing percobaan sebanyak 10 percobaan dan pada system menunjukkan nilai akurasi sebesar 70,28%.

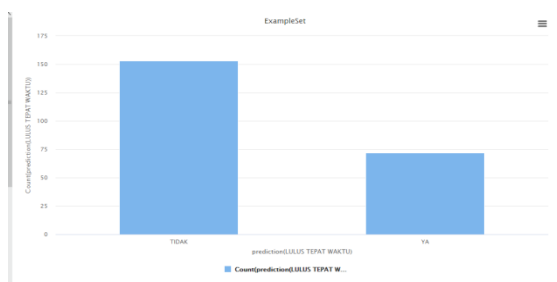
$$Error = \frac{FP + FN}{TP + TN + FN + FP} = \frac{41 + 26}{112 + 46 + 41 + 26} = 29,78\%$$

Hasil Perhitungan menunjukkan nilai error sebesar 29,78%.



Gambar 2. Grafik ROC K-NN

AUC (the area under curve) dihitung untuk mengukur perbedaan performansi. Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual dengan false positive rate (specificity) sebagai garis horizontal dan true positive rate (sensitivity) sebagai garis vertical. Berdasarkan gambar diperoleh nilai AUC sebesar 0,739.



Gambar 2. Grafik Prediction KNN

Berdasarkan Gambar 3 menghasilkan prediksi lulus tepat waktu (*label*) sebesar

153 untuk TIDAK (*negative*), 72 untuk YA (*positive*)

KESIMPULAN

1. Pengujian dengan metode kNN didapatkan akurasi pada model dengan akurasi 70,28% ,error 29,78% dan auc sebesar 0,739
2. Berdasarkan model KNN untuk pengujian bahwa menunjukkan hasil prediksi kecenderungan kelulusan mahasiswa S1 Teknik Informatika angkatan 206-2017 sebesar 72 orang diprediksi lulus tepat waktu dan 153 orang diprediksi tidak lulus tepat waktu.
3. Semakin tidak teratur data maka semakin rendah tingkat akurasinya, sebaliknya semakin data teratur dan rapih semakin tinggi akurasinya. Hal ini menunjukkan bahwa algoritma KNN diimplementasikan pada kasus dimana prediksi kelulusan mahasiswa sangat dipengaruhi oleh data.

UCAPAN TERIMA KASIH

Pertama-tama peneliti ucapkan terima kasih kepada Orang tua yang selalu mendukung dan mendoakan sehingga dapat hasil yang maksimal dalam penelitian ini, tidak lupa juga rasa terima kasih pada pembimbing yang sangat membantu dalam mengarahkan penelitian ini sehingga mencapai hasil yang optimal, dan tak luput juga pihak pihak lain yang bersangkutan.

DAFTAR PUSTAKA

- [1] Atma, Y. D., & Setyanto, A. (2018). Perbandingan algoritma c4.5 dan k-nn dalam identifikasi mahasiswa berpotensi drop out. *Metik Jurnal*, 2(2), 31–37.
- [2] Fatoni, C. S., & Noviandha, F. D.



-
- (2018). Case Based Reasoning Diagnosis Penyakit Difteri dengan Algoritma K-Nearest Neighbor. *Creative Information Technology Journal*, 4(3), 220. <https://doi.org/10.24076/citec.2017v4i3.112>
- [3] Muliono, R., Lubis, J. H., & Khairina, N. (2020). Analysis K-Nearest Neighbor Algorithm for Improving Prediction Student Graduation Time. *Sinkron*, 4(2), 42. <https://doi.org/10.33395/sinkron.v4i2.10480>
- [4] Romadhona, A., Suprapedi, & Himawan, H. (2017). Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Usia, Jenis Kelamin, Dan Indeks Prestasi Menggunakan Algoritma Decision Tree. *Jurnal Teknologi Informasi*, 13, 69–83.
- [5] Suriana, S., Bintari, E. D., & Kurniawan, D. (2017). Desain Aplikasi Klasifikasi Kelulusan Mahasiswa Menggunakan Metode Naive Bayes dan Algoritma C4.5. *Journal of Big Data Analytic and Artificial Intelligence*, 3(1), 31–41. <https://www.neliti.com/publications/308842/>
- [6] Susanto, E. S., Kusrini, & Fatta, H. Al. (2018). Prediksi Kelulusan Mahasiswa Magister Teknik Informatika Universitas Amikom Yogyakarta Menggunakan Metode K-Nearest Neighbor. *Jurnal Teknologi Informasi*, XIII(2), 67–72.
- [7] Zauardi, A., & Suprayitno, H. (2018). Analisa Karakteristik Kecelakaan Lalu Lintas di Jalan Ahmad Yani Surabaya melalui Pendekatan Knowledge Discovery in Database. *Jurnal Manajemen Aset Infrastruktur & Fasilitas*, 2(1), 45–55. <https://doi.org/10.12962/j26151847.v2i1.3767>.