



JESCE

(Journal of Electrical and System Control Engineering)

Available online <http://ojs.uma.ac.id/index.php/jesce>

Penerapan K-Optimal Pada Algoritma KNN Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Teknik Informatika

Application of K-Optimal in the KNN Algorithm to Predict Timely Graduation for Students of the Informatics Engineering Study Program

Atri Gadi Ana Amas^{1*)}, Gergorius Kopong Pati²⁾, Felysitas Ema Ose Sanga³⁾
1,2,3) Teknik Informatika, STIMIKOM Stella Maris Sumba, Indonesia

Diterima: Oktober 2023; Direview: November 2023; Dipublikasi: Februari 2024

*Corresponding Email: atrigadia@gmail.com, gregkopong80@gmail.com

Abstrak

Sistem Pengambilan Keputusan (DSS) adalah satu-satunya metode komputernisasi yang paling banyak digunakan saat ini. Kepuasan mahasiswa ditentukan oleh kualitas yang diinginkan mahasiswa, sehingga setiap perguruan tinggi harus memastikan bahwa kualitas ini tersedia. Program Teknik Informatika STIMIKOM Stella Maris Sumba dapat memberikan sebuah informasi baru yang belum diketahui sebelumnya menggunakan teknik data mining yang dapat digunakan untuk memprediksi kelulusan tepat waktu mahasiswa. Penelitian menggunakan metode k-Nearest Neighbor yang merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data training yang jaraknya paling dekat dengan objek tersebut. Pemilihan nilai k pada algoritma kNN menjadi hal yang penting karena akan mempengaruhi kinerja dari algoritma kNN, oleh karena itu perlu diketahui berapa nilai k dan tingkat akurasi. Metode k-Fold Cross Validation dan Uji Akurasi digunakan untuk mengetahui nilai k-Optimal. Hasil yang didapat adalah nilai k=5 dengan tingkat akurasi sebesar 80.00% yang ditetapkan sebagai k-Optimal. Nilai k=5 diterapkan pada algoritma kNN untuk prediksi kelulusan tepat waktu mahasiswa berdasarkan IP sampai dengan semester 4.

Kata kunci: kNN, k-Optimal, Classification, Data mining, k-Fold Cross Validation method.

Abstract

Decision Making Systems (DSS) are the single most widely used computerized method today. Student satisfaction is determined by the quality that students want, so every university must ensure that this quality is available. The STIMIKOM Stella Maris Sumba Informatics Engineering Program can provide new information that was not previously known using data mining techniques that can be used to predict students' timely graduation. The research uses the k-Nearest Neighbor method, which is a method for classifying objects based on training data that is closest to the object. Choosing the k value in the kNN algorithm is important because it will affect the performance of the kNN algorithm, therefore it is necessary to know what the k value is and the level of accuracy. The k-Fold Cross Validation method and Accuracy Test are used to determine the k-Optimal value. The result obtained is the value k=5 with an accuracy level of 80.00% which is designated as k-Optimal. The value k=5 is applied to the kNN algorithm to predict students' on-time graduation based on GPA up to semester 4.

Keywords: kNN, k-Optimal, Classification, Data mining, k-Fold Cross Validation method.



PENDAHULUAN

Metode komputernisasi yang paling banyak digunakan saat ini. Pengambilan keputusan adalah proses memilih antara berbagai pilihan untuk mencapai satu atau lebih tujuan. Kepuasan mahasiswa ditentukan oleh kualitas yang diinginkan mahasiswa, sehingga setiap perguruan tinggi harus memastikan bahwa kualitas ini tersedia, yang saat ini digunakan sebagai salah satu cara perguruan tinggi lebih berkompetisi (Asep Saefulloh, 2013).

Data akademik atau informasi yang ada pada suatu perguruan tinggi akan semakin bertambah seiring dengan berlangsungnya proses kegiatan akademik. Hal ini menciptakan kondisi yang membuat adanya suatu tumpukan data (Moerdyanto & Nuryana, 2023). Pada sistem informasi akademik program studi Teknik Informatika Mangkurat terdapat berbagai macam kumpulan data didalamnya. Sampai saat ini, data mahasiswa yang ada belum dimanfaatkan secara maksimal sehingga perlu diolah untuk menemukan sebuah informasi baru. Data yang digunakan adalah data IP mahasiswa mulai dari semester satu sampai dengan semester empat, menggunakan salah satu fungsi *data mining* yaitu fungsi klasifikasi dengan menggunakan algoritma *k-Nearest*

Neighbor dengan harapan dapat memprediksi kelulusan tepat waktu mahasiswa sehingga dapat digunakan oleh pihak program studi untuk mencari solusi atau kebijakan dalam proses evaluasi pembelajaran di program studi ilmu komputer (Subha & Nambi, 2012).

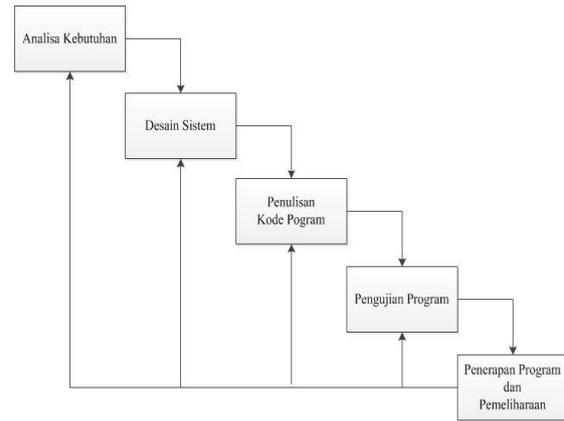
Tujuan K-Nearest Neighbor Algorithm (K-NN) adalah untuk mengklasifikasikan objek berdasarkan atribut dan data latih; tidak memiliki model yang dapat dipelajari dari data uji (Isnain et al., 2021). Oleh karena itu pembelajaran hanya sebatas contoh uji. Deep Belief Network (DBN) mengungguli Nave Bayes dan Support Vector Machine (SVM) dalam analisis tweet berbahasa Indonesia, dan hasilnya lebih unggul (Yunus et al., 2021).

Tujuan K-Nearest Neighbor Algorithm (K-NN) adalah untuk mengklasifikasikan objek berdasarkan atribut dan data latih tidak memiliki model yang dapat dipelajari dari data uji (Hidayati & Hermawan, 2021). Oleh karena itu pembelajaran hanya sebatas contoh uji (Intan Monica Hanmastiana et al., 2022). Deep Belief Network (DBN) mengungguli Nave Bayes dan Support Vector Machine (SVM) dalam analisis tweet berbahasa Indonesia, dan hasilnya lebih unggul (Grassi et al., 2019).

Penelitian sekarang ini mencoba untuk mengetahui nilai k-Optimal dan tingkat akurasi pada algoritma kNN untuk prediksi kelulusan tepat waktu mahasiswa berdasarkan IP sampai dengan semester 4. Data yang digunakan untuk penelitian yaitu data nilai IP mahasiswa program studi Teknik Informatika STIMIKOM Stella Maris Sumba yang telah lulus dari angkatan 2019 hingga 2022. Data mahasiswa tersebut selanjutnya diklasifikasikan menjadi “Ya” dan “Tidak”. Jika mahasiswa tersebut lulus tepat waktu (≤ 5 Tahun) maka masuk dalam klasifikasi “Ya”, sebaliknya jika mahasiswa lulus selama >5 Tahun maka termasuk dalam klasifikasi “Tidak” atau tidak tepat waktu.

METODE PENELITIAN

Metode yang digunakan dalam penelitian ini yaitu metode *waterfall*. Model SDLC air terjun (*waterfall*). Sering juga disebut model sekuensial linear (*sequentially linear*) atau alur hidup klasik (*classic life cycle*). Model air terjun (*waterfall*) menyediakan pendekatan alur hidup perangkat lunak secara sekuensial atau terurut dimulai analisis, desain, pengkodean, pengujian, dan tahap pendukung (*support*). Berikut ini adalah gambar 1 model air terjun (*waterfall*).



Gambar 1. Model *Waterfall*

Dari Gambar 1 dapat dijelaskan uraiannya sebagai berikut:

1) Analisis (*Analysis*)

Langkah ini merupakan analisa terhadap kebutuhan sistem. Pengumpulan data dalam tahap ini bisa melakukan sebuah penelitian, wawancara atau studi literatur. Sistem analisis akan menggali informasi sebanyak-banyaknya dari *user* sehingga akan tercipta sebuah sistem komputer yang bisa melakukan tugas-tugas yang diinginkan oleh *user* tersebut. Tahapan ini akan menghasilkan dokumen *user requirement* atau dapat dikatakan sebagai data yang berhubungan dengan keinginan *user* dalam pembuatan sistem. Dokumen ini lah yang akan menjadi acuan sistem analisis untuk menerjemahkan ke dalam bahasa pemrogram.

2) Desain (*Design*)

Tahapan dimana dilakukan penuangan pikiran dan perancangan sistem terhadap solusi dari permasalahan

yang ada dengan menggunakan perangkat pemodelan sistem seperti diagram alir data (*data flow diagram*), diagram hubungan entitas (*entity relationship diagram*) serta struktur dan bahasan data.

3) Kode (*code*)

Penulisan kode program atau *coding* merupakan penerjemahan *design* dalam bahasa yang dapat dikenali oleh komputer. Dilakukan oleh *programmer* yang akan menterjemahkan transaksi yang diminta oleh *user*. Tahapan ini lah yang merupakan tahapan secara nyata dalam mengerjakan suatu sistem. Dalam artian penggunaan komputer akan dimaksimalkan dalam tahapan ini. Setelah pengkodean selesai maka akan dilakukan *testing* terhadap sistem yang telah dibuat tadi. Tujuan *testing* adalah menemukan kesalahan-kesalahan terhadap sistem tersebut dan kemudian dapat diperbaiki.

4) Pengujian (*test*)

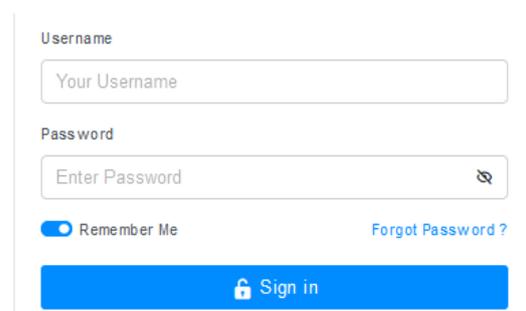
Tahapan akhir dimana sistem yang baru diuji kemampuan dan keefektifannya sehingga didapatkan kekurangan dan kelemahan sistem yang kemudian dilakukan pengkajian ulang dan perbaikan terhadap aplikasi menjadi lebih baik dan sempurna.

5) Penerapan

Perangkat lunak yang sudah disampaikan kepada pelanggan pasti akan mengalami perubahan. Perubahan tersebut bisa karena mengalami kesalahan karena perangkat lunak harus menyesuaikan dengan lingkungan (peripheral atau sistem operasi baru) baru, atau karena pelanggan membutuhkan perkembangan fungsional.

HASIL DAN PEMBAHASAN

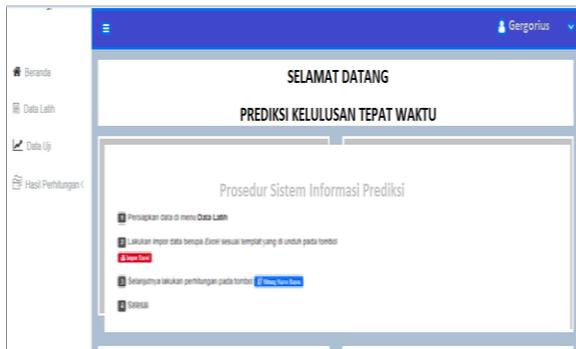
Proses Manupulasi sistem dibuatkan menu secara interaktif dalam mempermudah user dalam melakukan manipulasi data melalui interface yang ada. Pada tahap antar muka login sebagai awal untuk masuk sebagai administrator dilihat pada gambar 2 dibawah.



Gambar 2. Antar Muka Login Admin

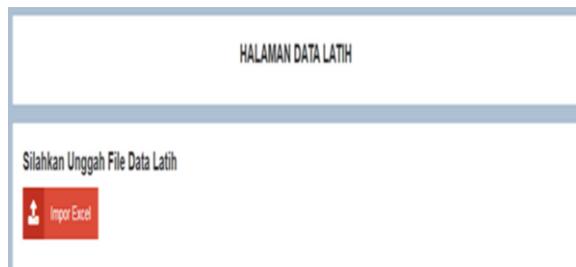
Kemudian ditambahkan pada tampilan antar muka menu utama sebagai tampilan awal Web. Pada menu utama dari aplikasi ini, terdapat empat menu diantaranya yaitu: beranda, Data latih, data uji, hasil perhitungan. Tampilan antar

muka menu utama dapat dilihat pada gambar 3 berikut.



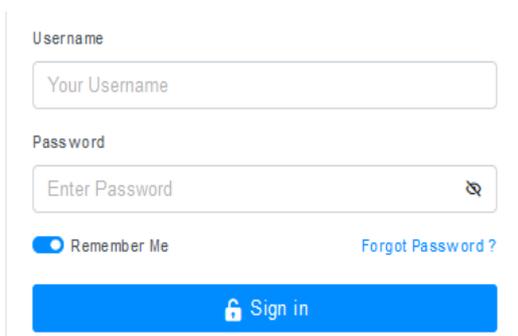
Gambar 3. Tampilan antar muka menu utama

Kemudian pada data latih dalam form ini terdapat satu field yakni import excel. Form ini digunakan untuk memasukan data excel. Tampilan form dapat dilihat pada gambar 4. berikut.



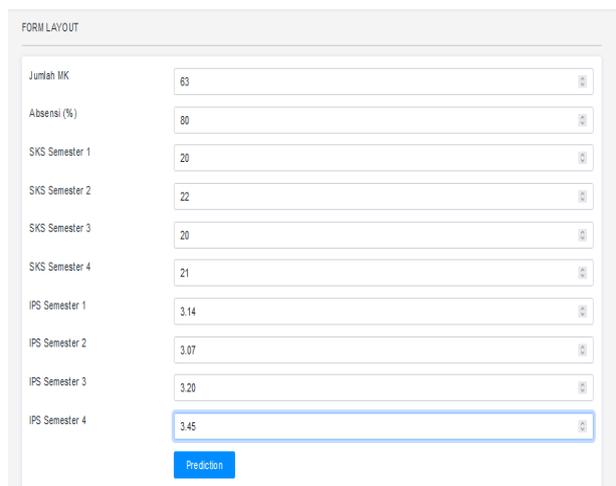
Gambar 4. Data Latih

Kemudian data login user Dalam antar muka login user sebagai awal untuk masuk sebagai user dapat dilihat pada gambar 5 berikut.



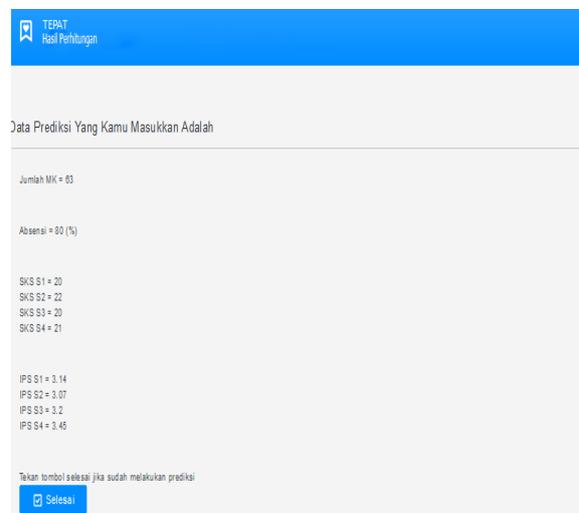
Gambar 5. Antar Muka Login User

Pada Form Prediksi yaitu terdapat sepuluh field yakni jumlah mata kuliah, absensi, sks semester 1 sampai semester 4, IPK semester 1 sampai semester 4. Form ini digunakan untuk mengisi data prediksi. Tampilan form dapat dilihat pada gambar 6 berikut.



Gambar 6. Form Prediksi

Tahap terakhir merupakan hasil prediksi yaitu dalam form ini digunakan untuk melihat hasil prediksi ketepatan perkuliahan. Tampilan form dapat dilihat pada gambar 7 berikut.



Gambar 7. Data Hasil Prediksi

KESIMPULAN

Kesimpulan yang diperoleh dari penelitian ini adalah: a. Nilai k-Optimal pada algoritma kNN untuk prediksi kelulusan tepat waktu mahasiswa berdasarkan IP sampai dengan semester 4 adalah k=5. b. Dari proses k-Fold Cross Validation didapatkan tingkat akurasi untuk k=5 pada algoritma kNN untuk prediksi kelulusan tepat waktu mahasiswa berdasarkan IP sampai dengan semester 4 adalah sebesar 80,00%.

(*Indonesian Journal of Computing and Cybernetics Systems*), 15(2).
<https://doi.org/10.22146/ijccs.65176>

- Moerdyanto, O. P., & Nuryana, I. K. D. (2023). Prediksi Kelulusan Tepat Waktu Menggunakan Pendekatan Pohon Keputusan Algoritma Decision Tree. *Journal of Informatics and Computer Science*, 05(1).
- Subha, M. V., & Nambi, S. T. (2012). Classification of stock index movement using k-nearest neighbours (k-NN) algorithm. *WSEAS Transactions on Information Science and Applications*, 9(9).
- Yunus, R., Ulfa, U., & Safitri, M. D. (2021). Application of the K-Nearest Neighbors (K-NN) Algorithm for Classification of Heart Failure. *Journal of Applied Intelligent System*, 6(1).<https://doi.org/10.33633/jais.v6i1.4513>

DAFTAR PUSTAKA

- Asep Saefulloh, M. (2013). Penerapan Metode Klasifikasi Data Mining untuk Prediksi Kelulusan Tepat Waktu. *InfoSys Journal*, 2.
- Grassi, S., Benedetti, S., Opizzio, M., Di Nardo, E., & Buratti, S. (2019). Meat and fish freshness assessment by a portable and simplified electronic nose system (Mastersense). *Sensors (Switzerland)*, 19(14).
<https://doi.org/10.3390/s19143225>
- Hidayati, N., & Hermawan, A. (2021). K-Nearest Neighbor (K-NN) algorithm with Euclidean and Manhattan in classification of student graduation. *Journal of Engineering and Applied Technology*, 2(2).
<https://doi.org/10.21831/jeatech.v2i2.42777>
- Intan Monica Hanmastiana, Budi Warsito, Rita Rahmawati, Hasbi Yasin, & Puspita Kartikasari. (2022). Classification of Public Opinion on Social Media Twitter concerning the Education in Indonesia Using the K-Nearest Neighbors (K-NN) Algorithm and K-Fold Cross Validation. *STATISTIKA Journal of Theoretical Statistics and Its Applications*, 21(2).
<https://doi.org/10.29313/statistika.v21i2.297>
- Isnain, A. R., Supriyanto, J., & Kharisma, M. P. (2021). Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning. *IJCCS*