

Aplikasi K-Nearest Neighbor Dengan Fungsi Jarak Gower Dalam Klasifikasi Kelulusan Mahasiswa (Studi Kasus : Mahasiswa Program Studi Statistika, Jurusan Matematika, Fakultas Matematika Dan Ilmu Pengetahuan Alam, Universitas Mulawarman)

K-Nearest Neighbor Application With Gower Distance Function In Student Graduation Classification (Case Study of Student of Statistics Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Mulawarman University)

Irfan Fadil, Rito Goejantoro, Surya Prangga

Laboratorium Statistika Komputasi FMIPA Universitas Mulawarman

E-mail: Irfanfadil024@gmail.com

ABSTRACT

The results of the reaccreditation of the Statistics Study Program, Mulawarman University in 2019 remain accredited B. One of the assessment indicators used in reaccreditation is the student's timely graduation status. Therefore, it is necessary to predict the graduation status of Statistics students, Mulawarman University.. The prediction method used in this research is K-Nearest Neighbor (K-NN). K-NN is a classification method based on studying previously classified data. This method is very easy to understand, easy to applied and also non-parametric method, so that no certain assumptions are needed in the process. The independent variables used in this study were student profiles, including gender, regional origin, cumulative Grade Point Average (GPA) and single tuition fee. The dependent variable in this study is the graduation status of students, namely graduating on time and not graduating on time. The data used were students of the Mulawarman University, Statistics Study Program in 2014, 2015, and 2016. The results showed at $k = 7$ and the distribution of training and testing data with the proportion of 80:20 obtained optimal accuracy of 0,909 with a TPrate of 0.500, a TNrate. in the amount of 1,000 and AUC value of 0,75 that means fair classification.

Keywords : Classification, graduate, K-Nearest Neighbor.

Pendahuluan

Program Studi (Prodi) Statistika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA), Universitas Mulawarman merupakan Program Studi yang berdiri pada tahun 2001. Prodi tersebut telah berakreditasi B. Pada tahun 2019, Program Studi tersebut melakukan reakreditasi. Hasil yang didapat adalah tetap berakreditasi B.

Indikator yang digunakan dalam pengakreditasi oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) yang berkaitan dengan lama studi adalah lamanya masa studi tiga tahun terakhir dan persentase kelulusan tepat waktu mahasiswa. Dalam upaya peningkatan akreditasi perlu dilakukan pengoptimalan dalam masa studi mahasiswa.

Salah satu cara yang bisa dilakukan adalah melakukan pengelompokan mahasiswa berdasarkan karakteristik tertentu. Pengelompokan dilakukan untuk mengetahui mahasiswa yang berpotensi tidak lulus tepat waktu sehingga dapat diberikan tindakan pencegahan agar mahasiswa tersebut bisa lulus tepat waktu. Pengelompokan ini dibagi menjadi dua kelompok yaitu lulus tepat waktu dan tidak lulus tepat waktu. Kelompok yang lulus tepat waktu adalah kelompok mahasiswa yang lulus tepat empat tahun. Kelompok yang tidak lulus

tepat waktu adalah kelompok mahasiswa yang lulus lebih dari empat tahun. Pengelompokan ini dapat ditempuh dengan teknik klasifikasi.

Teknik klasifikasi mempunyai banyak metode, diantaranya adalah klasifikasi dengan pohon keputusan, klasifikasi *naive bayes*, regresi logistik, *support vector machine*, *K-Nearest Neighbor* (K-NN) dan masih banyak lagi metode lainnya. Metode yang digunakan pada penelitian ini adalah K-NN.

Variabel bebas yang digunakan adalah jenis kelamin, asal daerah, IPK terakhir, dan golongan UKT. Nilai K yang digunakan adalah $K = 1$, $K = 3$, $K = 5$, $K = 7$, dan $K = 9$.

Tujuan yang ingin dicapai adalah memberikan informasi dan juga masukan kepada Program Studi Statistika berupa model klasifikasi K-NN yang dapat digunakan untuk memprediksi kelulusan mahasiswa sehingga bisa dilakukan langkah pencegahan agar mahasiswa dapat lulus tepat waktu. Memberikan pengetahuan terkait K-NN dan juga Fungsi jarak Gower serta dapat dijadikan rujukan untuk penelitian selanjutnya.

Data Mining

Data Mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data dengan melakukan penggalian

pola-pola data. Secara sederhana, *Data Mining* berarti penggalian atau penambangan pengetahuan dari sejumlah besar data (Han & Kamber, 2006).

Data Mining sering juga disebut *Knowledge Discovery (Mining) in Databases (KDD)*, yaitu ekstraksi pengetahuan dari data dan juga merupakan alat yang penting untuk memanipulasi atau untuk penyajian informasi sesuai kebutuhan pengguna (Risa & Risky Aswi, 2017).

Klasifikasi

Klasifikasi adalah bentuk analisis data yang mengekstraksi model yang menggambarkan suatu kelas data. Klasifikasi bertujuan memprediksi label kelas kategorikal. Secara garis besar klasifikasi terdiri atas dua step, yaitu step pembelajaran dan step klasifikasi. Step pembelajaran adalah step di mana model klasifikasi dibangun. Step klasifikasi adalah langkah pengklasifikasian dari model yang terbentuk (Han & Kamber, 2006).

K- Nearest Neighbor

Klasifikasi K-NN dikenalkan pertama kali pada tahun 1950. K-NN merupakan metode pengklasifikasian berdasarkan pembelajaran analogi, yaitu membandingkan data *testing* terhadap data *training* (Han & Kamber, 2006). Proses pengklasifikasian dengan K-NN berdasarkan data *training* yang jaraknya paling dekat dengan suatu objek. Data *training* diproyeksikan ke dalam ruang berdimensi banyak, di mana masing-masing dimensi mempresentasikan fitur dari data. Ruang tersebut dibagi menjadi bagian-bagian berdasarkan klasifikasi data *training* (Yustanti, 2012).

Jauh dekatnya sebuah objek didasarkan pada jarak. Perhitungan jarak antar objek mempunyai beberapa persamaan, diantaranya yaitu jarak Euclid, jarak manhattan, jarak minkowski, jarak chi-square, jarak mahalnobis dan beberapa persamaan jarak lainnya. Jarak yang umum digunakan pada K-NN adalah jarak Euclid (Yustanti, 2012). Jarak Euclid mempunyai persamaan sebagai berikut :

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}; i \neq j \quad (1)$$

Keterangan :

$d(x_i, x_j)$: Jarak antara observasi ke-i dan j

x_{ik} : Observasi ke-i variabel ke-k

x_{jk} : Observasi ke-j variabel ke-k

p : Banyaknya variabel

(Larose & Larose, 2015).

Jarak Euclid tidak dapat digunakan pada data yang bersifat kategorik, sehingga pada penelitian ini digunakan jarak Gower. Jarak Gower adalah

jarak yang ditransformasi dari koefisien kemiripan Gower. Koefisien kemiripan Gower merupakan koefisien kemiripan yang dikemukakan oleh Gower (1971) dalam penelitiannya yang berjudul *A General Coefficient of Similarity and Some of Its Properties*.

Koefisien kemiripan Gower dapat digunakan untuk melihat jarak antar observasi dengan melakukan perhitungan kemiripan pada setiap variabel acak yang ada sesuai dengan skala pengukuran peubah acak tersebut. Koefisien kemiripan s_k pada setiap variabel dihitung berdasarkan skala pengukuran variabel ke-k. Perhitungan nilai s_k pada skala nominal, ordinal, interval, dan rasio secara berurutan ditunjukkan pada persamaan di bawah ini:

Nominal :

$$s_k(x_{ik}, y_{jk}) = \begin{cases} 1, & x_{ik} = x_{jk} \\ 0, & x_{ik} \neq x_{jk} \end{cases} \quad (2)$$

Ordinal:

$$s_k(x_{ik}, y_{jk}) = 1 - \frac{|r_k(x_{ik}) - r_k(y_{jk})|}{\max\{r_k(x_{\bullet k})\} - \min\{r_k(x_{\bullet k})\}} \quad (3)$$

Interval; Rasio :

$$s_k(x_{ik}, y_{jk}) = 1 - \frac{|(x_{ik}) - (x_{jk})|}{\max\{x_{\bullet k}\} - \min\{x_{\bullet k}\}} \quad (4)$$

Keterangan:

$x_{\bullet k}$ Nilai seluruh observasi variabel k

$r_k(x_{\bullet k})$ Rank seluruh observasi variabel ordinal k

$\max\{x_{\bullet k}\}$ Nilai maksimum dari seluruh variabel k

$\min\{x_{\bullet k}\}$ Nilai minimum dari seluruh variabel k

$\max\{r_k(x_{\bullet k})\}$ Rank maksimum dari seluruh variabel ordinal k

$\min\{r_k(x_{\bullet k})\}$ Rank minimum dari seluruh variabel ordinal k

(Gower, 1971).

Penentuan tetangga terdekat K-NN ditentukan oleh jarak antar observasi, sehingga koefisien kemiripan Gower harus ditransformasi menjadi jarak. Ketika jarak berada pada rentang [0,1], maka kemiripan dapat dihitung dengan dengan persamaan 5.

$$s_k(x_i, y_j) = 1 - d_k(x_i, y_j) \quad (5)$$

Keterangan:

$d_k(x_i, y_j)$: Jarak observasi ke-i dan j pada variabel k.

Berdasarkan persamaan (5), secara umum menghitung jarak berdasarkan kemiripan dapat menggunakan persamaan berikut ini :

$$d_k(x_i, x_j) = 1 - s_k(x_i, x_j) \quad (6)$$

Berdasarkan persamaan (6), perhitungan jarak Gower untuk setiap skala data dapat dihitung dengan persamaan berikut:

$$\text{Nominal : } d_k(x_{ik}, y_{jk}) = \begin{cases} 0, & x_{ik} = x_{jk} \\ 1, & x_{ik} \neq x_{jk} \end{cases} \quad (7)$$

$$\text{Ordinal : } d_k(x_{ik}, y_{jk}) = \frac{|r_k(x_{ik}) - r_k(y_{jk})|}{\max\{r_k(x_{\bullet k})\} - \min\{r_k(x_{\bullet k})\}} \quad (8)$$

$$\text{Interval; Rasio: } d_k(x_{ik}, y_{jk}) = \frac{|(x_{ik}) - (y_{jk})|}{\max\{x_{\bullet k}\} - \min\{x_{\bullet k}\}} \quad (9)$$

Untuk menghitung jarak antara dua objek digunakan persamaan berikut:

$$d(x_i, y_j) = \frac{\sum_{k=1}^p \delta_k(x_{ik}, y_{jk}) d_k(x_{ik}, y_{jk})}{\sum_{k=1}^p \delta_k(x_{ik}, y_{jk})} \quad (10)$$

dengan $\delta_k(x_{ik}, y_{jk}) = \begin{cases} 1, & x_{ik} \text{ atau } x_{jk} \text{ tidak hilang} \\ 0, & \text{lainnya} \end{cases}$.

(Prasath, dkk., 2017).

Validasi Hasil Klasifikasi

Evaluasi terhadap hasil klasifikasi pada suatu penelitian perlu dilakukan untuk mengetahui tingkat akurasi dari suatu prediksi. Untuk mengetahui tingkat akurasi prediksi dari suatu klasifikasi dapat dilakukan dengan membentuk matriks konfusi. Ilustrasi Matriks konfusi dapat dilihat pada Tabel 1.

Tabel 1. Matriks Konfusi

Kelas Aktual	Kelas Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

(Gorenescu, 2011).

True Positive (TP) dan True Negatif (TN) adalah objek yang diklasifikasikan dengan hasil yang sesuai dengan kelasnya. False Negative (FN) dan False Positive (FP) adalah objek yang diklasifikasikan tidak sesuai dengan kelasnya. Akurasi dari prediksi pengklasifikasian dapat dilihat dengan menghitung akurasi, TPRate dan TNRate. Akurasi merupakan tingkat ketepatan objek yang diprediksi benar. TPRate merupakan tingkat ketepatan objek dengan kelas Positif yang diprediksi benar. TNrate merupakan tingkat ketepatan objek dengan kelas negatif yang diprediksi benar.

Akurasi, TPRate, dan TNRate dapat dihitung berdasarkan TP, TN, FP, dan FN.

Persamaan untuk menghitung akurasi, TPRate, dan TNRate dapat dilihat pada persamaan di bawah ini:

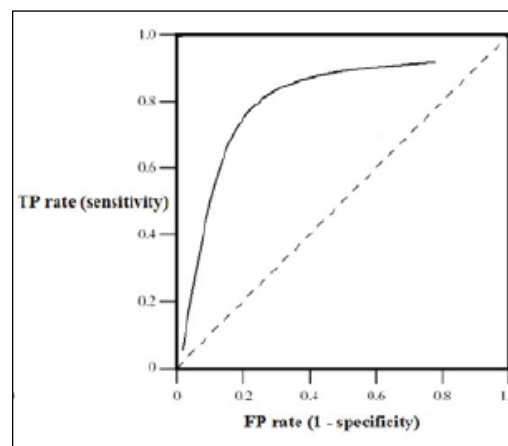
$$\text{Akurasi} = \frac{TP + TN}{TP + FP + TN + FN} \quad (11)$$

$$\text{TPRate} = \frac{TP}{TP + FN} \quad (12)$$

$$\text{TNRate} = \frac{TN}{FP + TN} \quad (13)$$

Akurasi adalah nilai menyatakan tingkat keakuratan dari suatu pengklasifikasian. True Positive Rate (TPRate), yang biasa juga disebut Sensitivitas atau Recall adalah nilai akurasi objek yang mempunyai kelas positif. True Negatif Rate (TNRate) atau yang juga disebut spesifitas adalah akurasi objek dengan kelas negatif (Gorenescu, 2011).

Untuk menggambarkan kebaikan hasil klasifikasi dapat dilihat menggunakan Kurva Receiver Operating Characteristic (ROC). Kurva ROC adalah kurva yang dua dimensi yang berasal dari tarik ulur antara TPrate dan TNrate. X-axis pada kurva ROC adalah Fprate (1-TNrate) dan Y-axis adalah TPrate. Contoh kurva ROC dapat dilihat pada Gambar 1.



Sumber : (Gorenescu, 2011)

Gambar 1. Kurva ROC

Kurva ROC menghasilkan Luasan di bawah kurva yang disebut Area Under the Curve (AUC) yang mempunyai besaran antara 0,50 hingga 1,00. Interpretasi nilai AUC sebagai berikut :

- 0,90 – 1,00 = Sangat baik
- 0,80 – 0,90 = Baik
- 0,70 – 0,80 = Sedang
- 0,60 – 0,70 = Lemah
- 0,50 – 0,60 = Sangat lemah

(Gorenescu, 2011).

Kelulusan Mahasiswa

Berdasarkan Peraturan Rektor Universitas Mulawarman tahun 2018, masa Studi Sarjana 1 adalah 8 hingga 14 semester dengan beban SKS sebesar 144 SKS atau lebih. Mahasiswa dinyatakan lulus jika memenuhi syarat sebagai berikut :

1. Telah menyelesaikan semua mata kuliah yang ditetapkan pada kurikulum Program Studi bersangkutan dengan IPK $\geq 2,00$ tanpa nilai huruf E;
 2. Lulus *Mulawarman University English Proficiency Test (MU-EPT)* atau *TOEFL Prediction* yang diakui oleh UPT Bahasa Unmul dengan skor minimal 400;
 3. Membuat minimal 1 (satu) artikel ilmiah yang siap untuk dipublikasikan dan telah disetujui oleh Pembimbing;
 4. Telah lulus ujian tugas akhir/skripsi; dan
 5. Menyelesaikan persyaratan lain yang ditetapkan oleh Fakultas masing-masing.
- (Universitas Mulawarman, 2018).

Hasil dan Pembahasan

Pada penelitian ini, populasinya adalah mahasiswa maupun alumni Prodi Statistika, FMIPA, yang telah melalui delapan semester dan yang menjadi sampel adalah mahasiswa maupun alumni Prodi Statistika, FMIPA, Universitas Mulawarman angkatan 2014, 2015, dan 2016.

Variabel yang digunakan pada penelitian ada lima, yang terbagi atas satu variabel terikat dan empat variabel bebas. Variabel terikat merupakan status kelulusan mahasiswa dan variabel bebasnya merupakan profil mahasiswa. Variabel yang digunakan pada penelitian ini tersaji pada Tabel 2.

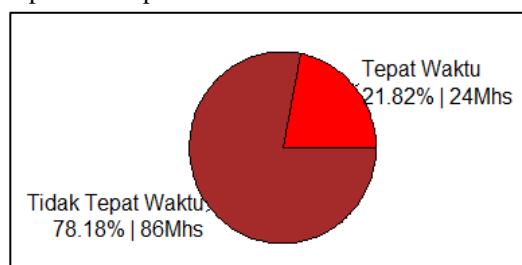
Tabel 2. Variabel Penelitian

Variabel Penelitian	Keterangan	Skala Data
Status Kelulusan	Tepat Waktu : Lulus ≤ 8 Semester (1) Tidak Tepat Waktu : Lulus > 8 Semester (2)	Nominal
Jenis Kelamin	Laki-Laki dan Perempuan	Nominal
IPK	Indeks Prestasi Kumulatif terakhir	Interval
Asal Daerah	Samarinda dan Luar Samarinda	Nominal
Golongan UKT	Golongan 0, Golongan 1, Golongan 2, Golongan 3, Golongan 4 dan Golongan 5	Ordinal

Analisis pada penelitian ini dimulai dengan mendeskripsikan data lalu membuat model K-NN. Pada metode K-NN data akan dibagi menjadi dua yaitu data *training* dan data *testing*. Proporsi pembagiannya yaitu 70:30 dan 80:20. Sebelum dilakukan pembagian data, data terlebih dahulu diacak menggunakan fungsi *sample* pada *software R*.

Deskripsi Data Penelitian

Total data yang digunakan pada penelitian ini sebanyak 110. Deskripsi data status kelulusan dapat dilihat pada Gambar 2.



Gambar 2. Status Kelulusan

Gambar 2 menunjukkan bahwa dari keseluruhan 110 data mahasiswa hanya 21,82% atau 24 Mahasiswa yang lulus tepat waktu dan 78,18% atau 86 Mahasiswa tidak lulus tepat waktu.

Menghitung Jarak Gower

-Perhitungan jarak pada variabel pertama

Karena skala data pada variabel pertama adalah nominal maka persamaan yang digunakan adalah persamaan (7).

$$d_k(x_{ik}, y_{jk}) = \begin{cases} 0, & x_{ik} = y_{jk} \\ 1, & x_{ik} \neq y_{jk} \end{cases}$$

Karena

$$x_{11} = P \neq y_{11} = L$$

maka

$$d_1(x_{11}, y_{11}) = 1$$

-Perhitungan Jarak pada variabel kedua

Karena skala data pada variabel kedua adalah interval maka persamaan yang digunakan adalah persamaan (9).

$$d_2(x_{12}, y_{12}) = \frac{|(x_{12}) - (y_{12})|}{\max\{x_{.2}\} - \min\{x_{.2}\}}$$

$$d_2(x_{12}, y_{12}) = \frac{|(3,48) - (3,15)|}{3,8 - 2,8}$$

$$d_2(x_{12}, y_{12}) = 0,33$$

-Perhitungan jarak pada variabel ketiga

karena skala data pada variabel ketiga adalah nominal maka persamaan yang digunakan adalah persamaan (7).

$$d_3(x_{13}, y_{13}) = \begin{cases} 0, & x_{13} = y_{13} \\ 1, & x_{13} \neq y_{13} \end{cases}$$

Karena

$$x_{13} = \text{LUAR SAMARINDA} \neq y_{13} \\ = \text{SAMARINDA}$$

maka

$$d_3(x_{13}, y_{13}) = 1$$

-Perhitungan jarak pada variabel keempat

Karena skala data pada variabel keempat adalah ordinal maka persamaan yang digunakan adalah persamaan (8).

$$d_4(x_{14}, y_{14}) = \frac{|r_4(x_{14}) - r_4(y_{14})|}{\max\{r_4(x_{.4})\} - \min\{r_4(x_{.4})\}}$$

$$d_4(x_{14}, y_{14}) = \frac{|3 - 3|}{5 - 0}$$

$$d_4(x_{14}, y_{14}) = 0$$

Setelah selesai menghitung jarak antara objek pervariabel, dilanjutkan dengan menghitung jarak antara dua objek menggunakan persamaan (10).

$$d(x_1, y_1) = \frac{\sum_{k=1}^4 d_k(x_{1k}, y_{1k})}{4}$$

$$d(x_1, y_1) = \frac{1 + 0,33 + 1 + 0}{4}$$

$$d(x_1, y_1) = 0,5825$$

Klasifikasi dengan K-NN

Hasil klasifikasi pada proporsi 70:30 untuk setiap nilai K dapat dilihat pada Tabel 3.

Tabel 3. Hasil Klasifikasi Proporsi 70:30

Data Testing	Prediksi K-NN					Klasifikasi Pada Data Asli
	1	3	5	7	9	
16	2	2	2	2	2	2
83	1	1	2	1	1	1
14	2	2	2	2	2	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
68	2	2	2	2	2	2
Prediksi Salah	8	6	6	6	5	

Hasil klasifikasi status kelulusan Mahasiswa pada proporsi 70:30 berdasarkan tabel 4.6 didapatkan hasil bahwa pada nilai K = 1 terdapat 8 objek yang diprediksi berbeda dari kelas aslinya, pada K = 3, K = 5, dan K = 7, terdapat 6 objek yang diprediksi berbeda dari kelas aslinya, dan pada K = 9 terdapat 5 objek yang diprediksi berbeda dari data aslinya.

Akurasi Prediksi

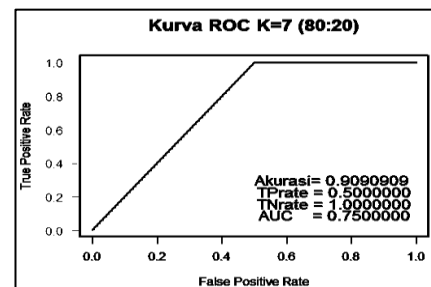
Akurasi yang dihasilkan metode K-NN jarak Gower pada klasifikasi status kelulusan mahasiswa untuk setiap nilai K pada proporsi 70:30 maupun 80:20 dapat dilihat pada Tabel 4.

Tabel 4. Akurasi Prediksi

Proporsi	K-NN	Akurasi	TPrate	TNrate
70:30	1-NN	0,757	0,286	0,885
	3-NN	0,818	0,429	0,923
	5-NN	0,818	0,286	0,962
	7-NN	0,818	0,429	0,923
	9-NN	0,848	0,429	0,962
80:20	1-NN	0,773	0,250	0,888
	3-NN	0,864	0,500	0,944
	5-NN	0,864	0,500	0,944
	7-NN	0,909*	0,500*	1,000*
	9-NN	0,909	0,500	1,000

Berdasarkan Tabel 4 didapatkan akurasi, TPrate, dan TNrate terbaik berada pada proporsi 80:20 dengan nilai K = 7 yaitu masing-masing sebesar 0,909, 0,500 dan 1,000. Akurasi sebesar 0,909 mengindikasikan bahwa peluang suatu objek diprediksi sesuai dengan data aslinya sebesar 0,909. TPrate sebesar 0,500 mengindikasikan bahwa suatu objek dengan kelas tidak lulus tepat waktu diprediksi lulus tepat waktu dengan peluang sebesar 0,500. TNrate sebesar 1,000 mengindikasikan bahwa objek yang diprediksi lulus tidak tepat waktu benar-benar lulus tidak tepat waktu.

Kurva ROC untuk nilai K=7 pada proporsi 80:20 dapat dilihat pada gambar 3.



Gambar 3. ROC K=7 Proporsi 80:20

Nilai AUC pada K = 7 proporsi 80:20 sebesar 0,75 yang artinya klasifikasi yang dihasilkan berada pada tingkat sedang.

Kesimpulan

Klasifikasi dengan metode K-NN Jarak Gower pada Status Kelulusan Mahasiswa Prodi Statistika Universitas Mulawarman angkatan 2014, 2015 dan 2016 pada K optimal = 7 dengan proporsi data training dan testing 80:20 didapatkan 20 hasil prediksi benar dan 2 prediksi salah. Akurasi yang dihasilkan adalah 0,909. Klasifikasi yang dihasilkan berada pada tingkat sedang.

Daftar Pustaka

- BAN-PT. (2019). *Akreditasi Program Studi: Matriks Penilaian Evaluasi Diri dan Laporan Kinerja Program Studi*. Jakarta: BAN-PT.
- BAN-PT. (2019). *Kriteria dan Prosedur-Instrumen Akreditasi Program Studi* versi 4.0. Jakarta: BAN-PT.
- Gorenescu, F. (2011). *Data Mining: Concept, Model, And Technique*. Verlag Berlin Heidelberg: Springer.
- Gower, J. C. (1971). *A General Coefficient of Similarity And Some of Its Properties*. *Biometrics*. 27(4), 857-871.
- Han, J., & Kamber, M. (2006). *Data Mining: Concepts And Techniques*. San Fransisco: Morgan Kaufmann.
- Larose, D. T., & Larose, C. D. (2015). *Data Mining And Predictive Analytcs*. New Jersey: John Wiley And Sons.
- Prasath, V. B., Alfeilat, H. A., Lasassmeh, O., & Hassanat, A. B. (2017). *Distance and Similarity Measures Effect on the Performances of K- Nearest Neighbor Classifier - A Review*. 1-50.
- Risa, H., & Risky Aswi, R. S. (2017). *Data Mining K- Nearest Neighbor*. Kediri: Universitas Nusantara PGRI Kediri.
- Universitas Mulawarman. (2018). *Peraturan Rektor Universitas Mulawarman*. Samarinda: Universitas Mulawarman.
- Yustanti, W. (2012). *Algoritma K- Nearest Neighbor Untuk Memprediksi Harga Jual Tanah*. *Jurnal Matematika, Statistika, & Komputasi*, 57-68.