

Penerapan Metode *K-Nearest Neighbor* untuk Memprediksi Tingkat Kelulusan Mahasiswa Berbasis Web pada Fakultas Ilmu Komputer UMI

Januaril Aditya Samudra^{a,1,*}, Siska Anraeni^{a,2}, dan Herman^{a,3}

^a Universitas Muslim Indonesia, Jl. Urip Sumoharjo KM.05, Makassar, 90231, Indonesia
¹ januaril760@gmail.com; ² siska.anraeni@umi.ac.id; ³ herman@umi.ac.id
*corresponding author

INFORMASI ARTIKEL	ABSTRAK
Diterima : 05 – 10 – 2020 Direvisi : 20 – 11 – 2020 Diterbitkan : 30 – 11 – 2020	Program studi Teknik Informatika pada Universitas Muslim Indonesia merupakan prodi yang banyak diminati setiap tahun ajaran baru, akan tetapi semakin bertambahnya jumlah mahasiswa tidak berbanding lurus dengan jumlah mahasiswa yang lulus sehingga menyebabkan terjadinya penumpukan data mahasiswa. Berdasarkan data yang ditemukan pada prodi tersebut untuk angkatan 2015 terdapat 186 mahasiswa sedangkan jumlah mahasiswa yang lulus tepat waktu sebanyak 53 mahasiswa. Hal tersebut menjadi kendala kemajuan dan menurunnya tingkat kualitas pada sebuah perguruan tinggi yang mempengaruhi penilaian akreditasi kampus. Pada penelitian ini menggunakan 141 dataset mahasiswa yang diperoleh dari Fakultas Ilmu Komputer pada mahasiswa angkatan 2015 dan 2016. Atribut yang digunakan untuk melakukan proses inputan prediksi kelulusan mahasiswa adalah data akademik mahasiswa seperti: IPS 1-4, Jenis Kelamin, SKS lulus serta beberapa data tambahan yang digunakan yaitu penghasilan orang tua perbulan, dan keaktifan organisasi. Tahap selanjutnya adalah membagi dataset menjadi dua yaitu data training yang merupakan data mahasiswa angkatan 2015 sebanyak 120 dan data testing merupakan data mahasiswa angkatan 2016 sebanyak 12 untuk menentukan mahasiswa yang lulus tepat waktu dan tidak tepat waktu dengan menggunakan algoritma K-Nearest Neighbor. Algoritma ini berperan untuk mendapatkan nilai kedekatan kasus baru terhadap kasus lama, populasi terbanyak pada area K dengan nilai terdekat mahasiswa tersebut di prediksi apakah lulus tepat waktu atau tidak. Hasil penelitian ini menunjukkan hasil prediksi dengan menguji nilai K = 5 hingga K = 11 banyaknya data testing, diperoleh hasil pengujian terbaik pada nilai K = 11 mempunyai accuracy 76%, precision 70%, recall 77% untuk klasifikasi tepat waktu dan tidak tepat waktu.
Kata Kunci: Mahasiswa <i>K-Nearest Neighbor</i> Kelulusan Mahasiswa Klasifikasi	

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



I. Pendahuluan

Universitas Muslim Indonesia merupakan salah satu perguruan tinggi swasta di Indonesia timur yang mempunyai jumlah mahasiswa yang sangat banyak. Pada Fakultas Ilmu Komputer, terdapat 2 (dua) Program Studi yaitu: Teknik Informatika dan Sistem Informasi dimana prodi Teknik Informatika merupakan prodi yang paling banyak diminati pada setiap tahun ajaran baru. Akan tetapi semakin bertambahnya jumlah mahasiswa tidak berbanding lurus dengan jumlah mahasiswa yang lulus sehingga menyebabkan terjadinya penumpukan data mahasiswa. Berdasarkan data peserta wisuda program Sarjana (S1) di Universitas Muslim Indonesia banyak alumni yang menempuh masa studi lebih dari 8 semester. Hal ini menunjukkan bahwa masih banyak mahasiswa Program Sarjana (S1) yang menempuh lama studi lebih dari yang dijadwalkan. Berdasarkan jumlah data mahasiswa aktif pada Fakultas Ilmu Komputer Program Studi Teknik Informatika pada angkatan 2015 yaitu sebanyak 186 mahasiswa sedangkan jumlah data alumni mahasiswa yang lulus tepat waktu pada angkatan 2015 sebanyak 53 mahasiswa.

Hal tersebut menjadi kendala kemajuan dan menurunnya tingkat kualitas pada sebuah perguruan tinggi [1]. Presentasi naik turunnya kemampuan mahasiswa untuk menyelesaikan studi tepat waktu merupakan salah satu elemen penilaian akreditasi kampus [2]. Perguruan tinggi saat ini dituntut untuk memiliki keunggulan bersaing dengan memanfaatkan semua sumber daya yang dimiliki serta perlu memperhatikan terhadap tingkat kelulusan mahasiswa tepat pada waktunya [3][4]. Selain perguruan tinggi, program studi juga harus bertindak kritis terhadap tingkat kelulusan mahasiswanya untuk mencegah terjadinya kemunduran waktu studi dari mahasiswa

tersebut. Adapun dua faktor utama yang mempengaruhi kelulusan mahasiswa yaitu faktor internal dan eksternal. Faktor internal adalah faktor dari dalam diri yang meliputi kecerdasan mahasiswa, penangkapan materi saat dikelas serta kemampuan mahasiswa itu sendiri, sedangkan faktor eksternal adalah faktor dari luar diri yang meliputi kondisi ekonomi, kondisi sosial, organisasi, peminatan, serta pergaulan. Selain mempengaruhi kelulusan, faktor tersebut juga akan mempengaruhi ekonomi mahasiswa seperti harus membayar uang semester lebih untuk menanggung matakuliah yang tidak lulus atau ketertinggalan mata kuliah [5].

Berdasarkan masalah yang telah diuraikan di atas, maka peneliti membuat sebuah aplikasi yang dapat digunakan oleh prodi Teknik Informatika di Universitas Muslim Indonesia untuk melakukan prediksi tingkat kelulusan mahasiswa Teknik Informatika dan melakukan tindakan ataupun penanganan yang diperlukan agar mahasiswa dapat lulus tepat waktu dengan menggunakan metode *K-Nearest Neighbour* (KNN). Metode ini dipilih karena termasuk dalam top 10 metode data mining yang paling populer dan berdasarkan pada penelitian yang pernah dilakukan dengan membandingkan metode KNN, *Random Forest*, dan *Support Vector Machine* (SVM) diperoleh hasil bahwa metode KNN sangat direkomendasikan untuk digunakan pada sistem pengklasifikasian untuk studi kasus pada penelitian tersebut [6].

II. Metode

K-Nearest Neighbor (KNN) adalah algoritma yang digunakan untuk melakukan klasifikasi suatu objek, berdasarkan k buah data latih yang jaraknya paling dekat dengan objek tersebut sehingga dapat digunakan dalam penentuan prediksi pada label kelas data uji [7]. 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. Perhitungan jarak ketetanggan menggunakan menggunakan *euclidean distance* ditunjukkan pada berikut [8]:

$$d_{(x,y)} = \sqrt{\sum_{j=1}^n (x_i - y_i)^2} \quad (1)$$

Keterangan:

$d_{(x,y)}$ = Jarak Euclidean (*Euclidean Distance*)

x = Sampel data

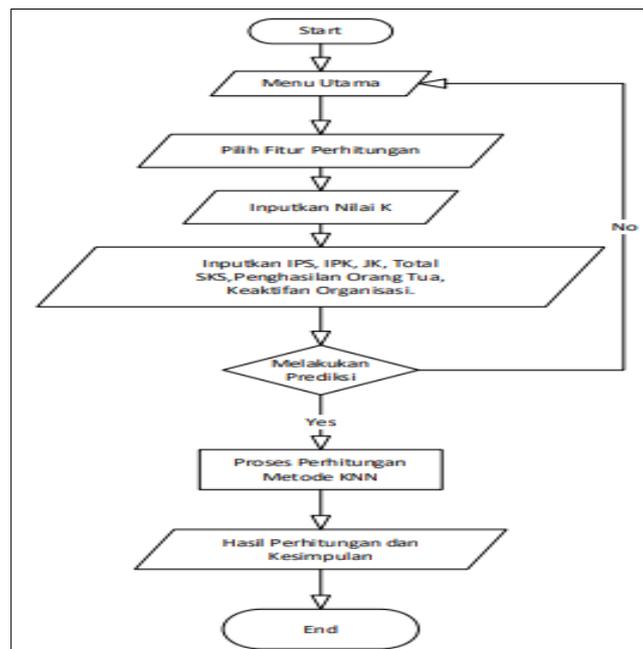
y = data uji/testing

i = variabel data

n = dimensi data

A. Perancangan Sistem

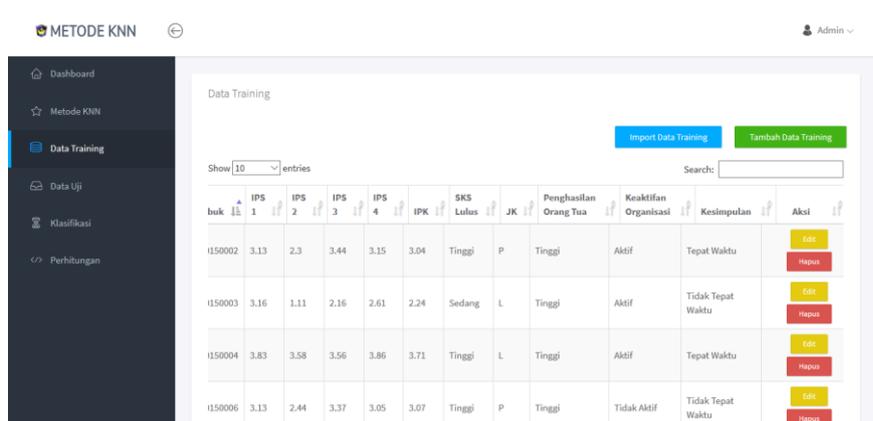
Adapun alur perancangan sistem yang akan dilakukan pada penelitian ini ditunjukkan pada Gambar 1 sebagai berikut:



Gambar 1. Flowchart Sistem

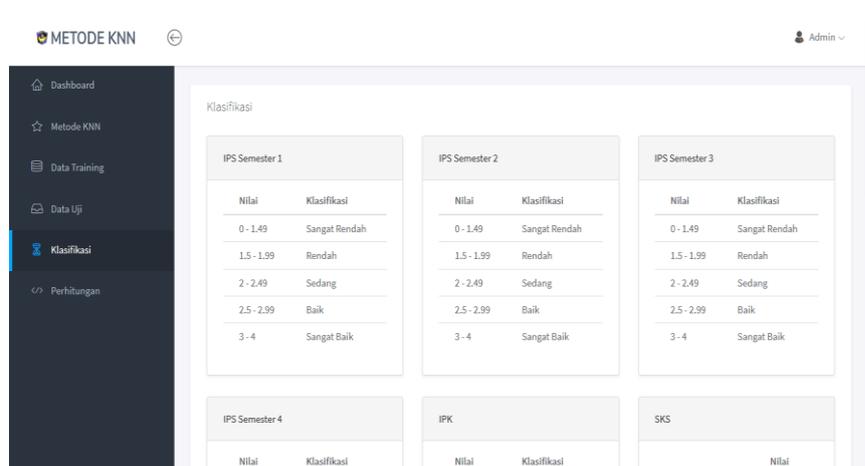
III. Hasil dan Pembahasan

A. Implementasi Interface Sistem



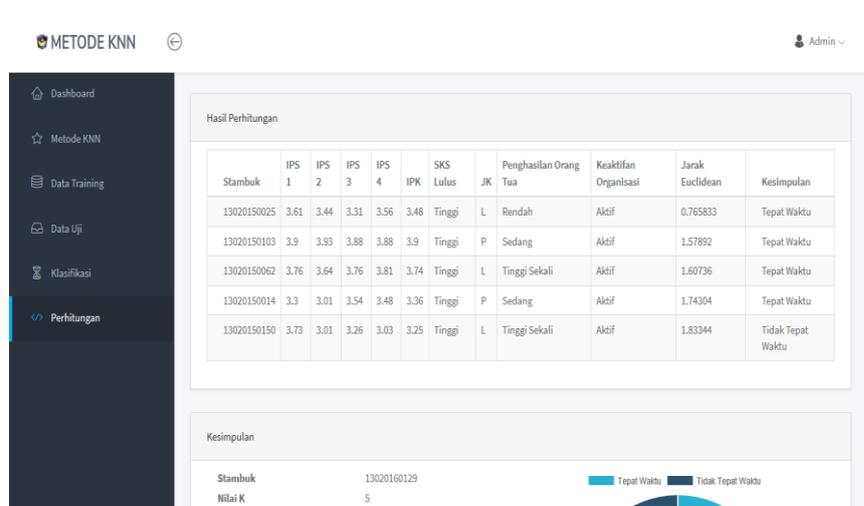
Gambar 2. Interface Data Training

Interface pada Gambar 2 berfungsi untuk melakukan penginputan, menghapus dan mengubah data training pada aplikasi kemudian menampilkannya dalam bentuk tabel.

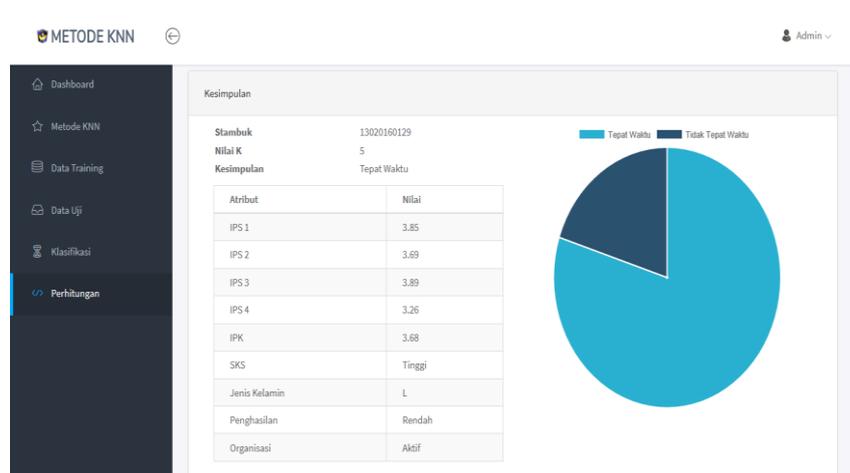


Gambar 3. Interface Klasifikasi

Interface pada Gambar 3 berfungsi untuk menampilkan nilai baru pada data kualitatif seperti Jenis Kelamin, Keaktifan Organisasi, Asal Daerah, dan Penghasilan Orang Tua.



Gambar 4. Interface Perhitungan Metode



Gambar 5. Interface Kesimpulan Dari Perhitungan Metode

Interface pada Gambar 4 dan gambar 5 menampilkan hasil perhitungan jarak Euclidean distance dan mahattan distance berdasarkan jarak yang dipilih serta menampilkan diagram hasil dari nilai dominan tepat waktu dan tidak tepat waktu berdasarkan jumlah nilai K yang diinputkan.

B. Implementasi Metode

Implementasi metode *K-Nearest Neighbor* (KNN) merupakan contoh proses perhitungan manual mulai dari penetapan data training dan data testing. Pada Tabel 1 menunjukkan 10 sample data training dan pada Tabel 2 menunjukkan 2 sample data testing. 12 sample tersebut yang akan diterapkan untuk perhitungan manual. Pada Tabel 3 sampai dengan tabel 6 menunjukkan perhitungan data dan hasil perhitungan data.

Tabel 1. Sampel Data Training

Stambuk	IPS 1	IPS 2	IPS 3	IPS 4	IPK	JK	SKS	Penghasilan	Organisasi	Keputusan
13020150002	3,13	2,3	3,44	3,15	3,04	2	3	3,5	1	Tepat Waktu
13020150004	3,83	3,58	3,56	3,86	3,71	1	3	3,5	1	Tepat Waktu
13020150009	3,58	1,94	3,3	3,65	3,1	1	2	2,5	3	Tepat Waktu
13020150014	3,3	3,01	3,54	3,48	3,36	2	3	2,5	1	Tepat Waktu
13020150023	3,25	3,23	3,04	3,43	3,24	1	3	4	1	Tepat Waktu
13020150003	3,16	1,11	2,16	2,61	2,24	1	2	3,5	1	Tidak Tepat Waktu
13020150006	3,13	2,44	3,37	3,05	3,07	2	3	3,5	3	Tidak Tepat Waktu
13020150010	1,98	1,39	2,13	2,49	1,97	1	1	3,5	1	Tidak Tepat Waktu
13020150013	2,99	0,18	2,24	2	1,81	1	1	3,5	1	Tidak Tepat Waktu
13020150015	2,45	1,61	1,24	1	1,59	2	1	4	3	Tidak Tepat Waktu

Tabel 2. Sampel Data Testing

Stambuk	IPS 1	IPS 2	IPS 3	IPS 4	IPK	JK	SKS	Penghasilan	Organisasi	Keputusan
13020160173	2,93	3,56	3,79	3,61	3,48	2	3	4	1	Tepat Waktu
13020160140	2,8	1,02	1,76	0,69	1,52	1	1	3,5	1	Tidak Tepat Waktu

Tabel 3. Perhitungan Data Testing I (Euclidean Distance)

Stambuk	IPS 1	IPS 2	IPS 3	IPS 4	IPK	JK	SKS	Penghasilan	Organisasi	Keputusan
13020160140	2,8	1,02	1,76	0,69	1,52	1	1	3,5	1	Tidak Tepat Waktu

Tabel 4. Hasil Perhitungan dan Urutan Jarak

Hasil Data Testing	K=5	Urutan Jarak Terdekat	Keputusan
1,550902963	3	1,365210606	Tepat Waktu
1,492883117	2	1,492883117	Tepat Waktu
3,418040374	6	1,550902963	Tepat Waktu

Tabel 5. Hasil Perhitungan dan Urutan Jarak

Hasil Data Testing	K=5	Urutan Jarak Terdekat	Keputusan
1,66829254	4	1,66829254	Tepat Waktu
1,365210606	1	2,490481881	Tidak tepat Waktu
3,674221006	7	3,418040374	Tepat Waktu
2,490481881	5	3,674221006	Tidak tepat Waktu
4,141436949	8	4,141436949	Tidak tepat Waktu
4,945856852	9	4,945856852	Tidak tepat Waktu
5,377694673	10	5,377694673	Tidak tepat Waktu

Kesimpulan pada Tabel 4.4 yaitu data testing Stambuk=13020160173 yang dihitung menggunakan euclidean distance dengan nilai K=5 mempunyai hasil prediksi (klasifikasi) yang tepat, dimana nilai yang paling dominan yaitu Tepat Waktu.

C. Pengujian Sistem Black Box

Pengujian *black box* berfokus pada persyaratan fungsional perangkat lunak dengan menguji *interface* apakah berjalan dengan baik dan efisien sehingga mudah digunakan oleh pengguna. Pengujian ini menggunakan dua kasus uji yaitu apabila sistem berjalan sesuai dengan harapan dan apabila terjadi kesalahan input. Berikut tabel hasil pengujian *black box*:

Tabel 6. Hasil Pengujian Black Box

Deskripsi	Langkah Uji	Hasil Harapan	Hasil Yang Diperoleh	Status
Login user	Masukkan username dan password	Masuk ke halaman dashboard	Berhasil masuk kehalaman dashboard	OK
Tambah Data Training	Masukkan data IPS 1- 4, jenis kelamin, penghasilan orang tua, keaktifan organisasi, asal daerah	Data training berhasil ditambahkan	Data training berhasil ditambahkan	Ok
Import data training	Pilih file xls	File berhasil di tambahkan	File berhasil ditambahkan	OK
Perhitungan Metode	Pilih jarak, jarak terdekat, masukkan data IPS 1- 4, jenis kelamin, penghasilan orang tua,	Data berhasil dihitung dan tampil hasil dan kesimpulan	Data berhasil dihitung dan tampil hasil dan kesimpulan	OK
Login User	Salah menginputkan username dan password	Tampilkan pesan kesalahan	Tampilkan pesan kesalahan	OK
Tambah data training	Tidak mengisi data, data tidak lengkap, atau data sama	Tampil pesan kesalahan	Tampil pesan kesalahan	OK
Import data training	Import file selain dari xls	Tampil pesan file tidak ditemukan	Tampil pesan file tidak ditemukan	OK
Perhitungan Metode	Tidak mengisi data, data tidak lengkap, atau data sama	Tampil pesan kesalahan	Tampil pesan kesalahan	OK

D. Pengujian Performa Metode KNN

Pengujian performa yang dilakukan pada penelitian ini menggunakan metode *confusion Matrix* dengan menghitung nilai *accuracy*, *precision*, dan *recall* untuk memvisualisasikan kinerja dari algoritma KNN yang digunakan. Setiap pengujian nilai *K* memiliki *Confusion matrix* masing-masing sebagai berikut:

1) Pengujian confusion matrix untuk nilai K=3

Tabel 7. Confusion Matrix Nilai K=3

	Predicted Tepat Waktu	Predicted Tidak Tepat Waktu
Actual Tepat Waktu	TP = 7	FP = 7
Actual Tidak Tepat Waktu	FN = 2	TN = 5

Setelah nilai TP, FP, FN, dan TN diketahui, maka dapat dihitung nilai performa sebagai berikut:

$$Akurasi = \frac{(7+5)}{(7+5+7+2)} = 0,57 = 57\%$$

$$\text{Presisi} = \frac{7}{(7+7)} = 0,50 = 50\%$$

$$\text{Recall} = \frac{7}{(7+2)} = 0,77 = 77\%$$

2) Pengujian confusion matrix untuk nilai $K=5$

Tabel 8. Confussion Matrix Nilai $K=5$

	<i>Predicted</i> Tepat Waktu	<i>Predicted</i> Tidak Tepat Waktu
<i>Actual</i> Tepat Waktu	TP = 7	FP = 5
<i>Actual</i> Tidak Tepat Waktu	FN = 2	TN = 7

Setelah nilai TP, FP, FN, dan TN diketahui, maka dapat dihitung nilai performa sebagai berikut:

$$\text{Akurasi} = \frac{(7+7)}{(7+7+5+2)} = 0,66 = 66\%$$

$$\text{Presisi} = \frac{7}{(7+5)} = 0,58 = 58\%$$

$$\text{Recall} = \frac{7}{(7+2)} = 0,77 = 77\%$$

3) Pengujian confusion matrix untuk nilai $K=7$

Tabel 9. Confussion Matrix Nilai $K=7$

	<i>Predicted</i> Tepat Waktu	<i>Predicted</i> Tidak Tepat Waktu
<i>Actual</i> Tepat Waktu	TP = 7	FP = 5
<i>Actual</i> Tidak Tepat Waktu	FN = 2	TN = 7

Setelah nilai TP, FP, FN, dan TN diketahui, maka dapat dihitung nilai performa sebagai berikut:

$$\text{Akurasi} = \frac{(7+7)}{(7+7+5+2)} = 0,66 = 66\%$$

$$\text{Presisi} = \frac{7}{(7+5)} = 0,58 = 58\%$$

$$\text{Recall} = \frac{7}{(7+2)} = 0,77 = 77\%$$

4) Pengujian confusion matrix untuk nilai $K=9$

Tabel 10. Confussion Matrix Nilai $K=9$

	<i>Predicted</i> Tepat Waktu	<i>Predicted</i> Tidak Tepat Waktu
<i>Actual</i> Tepat Waktu	TP = 7	FP = 4
<i>Actual</i> Tidak Tepat Waktu	FN = 2	TN = 8

Setelah nilai TP, FP, FN, dan TN diketahui, maka dapat dihitung nilai performa sebagai berikut:

$$\text{Akurasi} = \frac{(7+8)}{(7+8+4+2)} = 0,71 = 71\%$$

$$\text{Presisi} = \frac{7}{(7+4)} = 0,63 = 63\%$$

$$\text{Recall} = \frac{7}{(7+2)} = 0,77 = 77\%$$

5) Pengujian confusion matrix untuk nilai $K=11$ Tabel 11. Confussion Matrix Nilai $K=11$

	<i>Predicted</i> Tepat Waktu	<i>Predicted</i> Tidak Tepat Waktu
<i>Actual</i> Tepat Waktu	TP = 7	FP = 3
<i>Actual</i> Tidak Tepat Waktu	FN = 2	TN = 9

Setelah nilai TP, FP, FN, dan TN diketahui, maka dapat dihitung nilai performa sebagai berikut:

$$\text{Akurasi} = \frac{(7+9)}{(7+0+3+2)} = 0,76 = 76\%$$

$$\text{Presisi} = \frac{7}{(7+3)} = 0,70 = 70\%$$

$$\text{Recall} = \frac{7}{(7+2)} = 0,77 = 77\%$$

Berikut tabel perhitungan nilai performa (accuracy, precision, dan recall) untuk setiap nilai K yang digunakan yaitu:

Tabel 12. Hasil Uji Confusion Matrix Metode KNN

Euclidean Distance			
Nilai Performa			
Nilai K	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
3	57%	50%	77%
5	66%	58%	77%
7	66%	58%	77%
9	71%	63%	77%
11	76%	70%	77%

Berdasarkan Tabel 9 maka dapat disimpulkan bahwa dengan menguji nilai $K = 5$ hingga $K = 11$ banyaknya data training, diperoleh hasil performa terbaik pada nilai $K = 11$ dengan nilai accuracy sebesar 76%, precision 70% dan recall 77% untuk klasifikasi tepat waktu dan tidak tepat waktu.

IV. Kesimpulan dan saran

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan maka dapat ditarik kesimpulan sebagai berikut:

- 1) Perhitungan performa metode KNN menggunakan *Confussion Matrix* dengan menguji nilai $K=5$ hingga $K=11$ diperoleh hasil performa terbaik pada nilai $K=11$ dengan nilai *accuracy* sebesar 76%, *precision* 70% dan *recall* 77% untuk klasifikasi tepat waktu dan tidak tepat waktu.
- 2) Dengan menggunakan pengujian *black box* diketahui bahwa sistem yang dibangun telah berjalan sesuai dengan fungsinya.

Daftar Pustaka

- [1] A. Budiyantara, I. Irwansyah, E. Prengki, P. A. Pratama, and N. Wiliani, "Komparasi Algoritma Decision Tree, Naive Bayes dan K-Nearest Neighbor untuk Memprediksi Mahasiswa Lulus Tepat Waktu," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 5, no. 2, pp. 265–270, 2020.
- [2] R. Muliono, J. H. Lubis, and N. Khairina, "Analisis Algoritma K-Nearest Neighbors dalam Prediksi Waktu Kelulusan Mahasiswa," *Semant. (Seminar Nas. Tek. Inform.)*, vol. 2, no. 1, pp. 12–16, 2019.
- [3] A. Y. Saputra and Y. Primadasa, "Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *Techno.Com*, vol. 17, no. 4, pp. 395–403, 2018.
- [4] M. A. Maricar and Dian Pramana, "Perbandingan Akurasi Naive Bayes dan K-Nearest Neighbor pada Klasifikasi untuk Meramalkan Status Pekerjaan Alumni ITB STIKOM Bali," *J. Sist. dan Inform.*, vol. 14, no. 1, pp. 16–22, 2019.
- [5] P. S. C. Moonallika, K. Q. Fredlina, and I. B. K. Sudiarmika, "Penerapan Data Mining Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Algoritma Naive Bayes Classifier (Studi Kasus

-
- STMIK Primakara),” *J. Ilm. Komput.*, vol. 6, no. 1, pp. 47–56, 2020.
- [6] H. Sujaini, “Klasifikasi Citra Alat Musik Tradisional dengan Metode k-Nearest Neighbor, Random Forest, dan Support Vector Machine,” *J. Sist. Inf. Bisnis*, vol. 9, no. 2, p. 185, 2019.
- [7] M. Rivki and A. M. Bachtiar, “Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia,” *J. Sist. Inf.*, vol. 13, no. 1, p. 31, 2017.
- [8] K. Fitra, K. Jhoni, F. Ichsan, M. Sitti, “Klasifikasi Keluarga Miskin Menggunakan Metode K-Nearest Neighbor berbasis Euclidean Distance,” *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI)*, vol. 11, pp. 230-239, 2019.
- [9] R. Muliono, J. H. Lubis, and N. Khairina, “Analisis Algoritma K-Nearest Neighbors dalam Prediksi Waktu Kelulusan Mahasiswa,” *Semant. (Seminar Nas. Tek. Inform.)*, vol. 2, no. 1, pp. 12–16, 2019.
- [10] A. Y. Saputra and Y. Primadasa, “Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbor,” *Techno.Com*, vol. 17, no. 4, pp. 395–403, 2018.