IMPLEMENTASI ORANGE DATA MINING UNTUK KLASIFIKASI KELULUSAN MAHASISWA DENGAN MODEL K-NEAREST NEIGHBOR, DECISION TREE SERTA NAIVE BAYES

ORANGE DATA MINING IMPLEMENTATION FOR STUDENT GRADUATION CLASSIFICATION USING K-NEAREST NEIGHBOR, DECISION TREE AND NAIVE BAYES MODELS

Hozairi¹⁾, Anwari²⁾, Syariful Alim³⁾

¹⁾ Program Studi Teknik Informatika, Fakultas Teknik, Universitas Islam Madura

²⁾ Program Studi Sistem Informasi, Fakultas Teknik, Universitas Islam Madura

³⁾ Program Studi Teknik Informatika, Fakultas Teknik, Universitas Bhayangkara Surabaya
Jl PP.Miftahul Ulum Bettet Pamekasan

Email : dr.hozairi@uim.ac.id¹⁾, anwari.uim@gmail.com²⁾, alim@ubhara.ac.id³⁾

Abstrak

Proses pemantauan dan evaluasi terhadap kelulusan mahasiswa Universitas Islam Madura (UIM) sangat perlu untuk dilakukan karena tingkat kelulusan mahasiswa merupakan salah satu unsur penilaian akreditasi yang sangat penting untuk setiap Program Studi. Data Mining bisa digunakan untuk klasifikasi ketepatan kelulusan mahasiswa, penelitian ini bertujuan untuk menerapkan aplikasi orange data mining dengan menggunakan model K-Nearest Neighbor (K-NN), Decision Tree serta Naive Bayes dan selanjutnya akan dilakukan evaluasi akurasi dari masing-masing model tersebut. Penelitian ini dilakukan di Prodi Teknik Informatika Universitas Islam Madura tahun angkatan 2016, selanjutnya data mahasiswa akan dianalisa menggunakan aplikasi orange data mining dengan menggunakan model K-NN, Decision Tree serta Naive Bayes. Proses pengujian data menerapkan K-Fold Cross Validation (K=5), sedangkan model evaluasi yang digunakan adalah Confusion Matrix dan ROC. Hasil perbandingan ketiga model sebagai berikut, K-NN memiliki tingkat akurasi sebesar 77%, Decision Tree tingkat akurasi sebesar 74%, dan Naive Bayes memiliki tingkat akurasi sebesar 89%. Maka dari itu, untuk klasifikasi tingkat kelulusan mahasiswa Prodi Teknik Informatika Universitas Islam Madura merekomendasikan model Naive Bayes karena memiliki tingkat akurasi lebih baik dibanding K-NN dan Decision Tree.

Kata kunci: Klasifikasi, K-Nearest Neighbor, Naive Bayes, Decision Tree

Abstract

The process of monitoring and evaluating the graduation of students at the Islamic University of Madura (UIM) is very necessary because the student's graduation rate is one of the most important elements of accreditation assessment for each Study Program. Data Mining can be used to classify the accuracy of student graduation, this study aims to apply orange data mining applications using the K-Nearest Neighbor (K-NN), Decision Tree and Naive Bayes models and then evaluate the accuracy of each of these models. This research was conducted at the Informatics Engineering Study Program, Islamic University of Madura in class 2016, then student data will be analyzed using an orange data mining application using the K-NN model, Decision Tree and Naive Bayes. The data testing process applies K-Fold Cross Validation (K=5), while the evaluation model used is Confusion Matrix and ROC. The results of the comparison of the three models are as follows, K-NN has an accuracy rate of 77%, Decision Tree has an accuracy rate of 74%, and Naive Bayes has an accuracy rate of 89%. Therefore, for the classification of the graduation rate of students in the Informatics Engineering Study Program, Islamic University of Madura, it is recommended the Naive Bayes model because it has a better accuracy rate than K-NN and Decision Tree.

Keywords : Classification, K-Nearest Neighbor, Naive Bayes, Decision Tree

1. PENDAHULUAN

Salah satu evaluasi akreditasi Program Studi di Perguruan Tinggi yaitu ketepatan waktu mahasiswa dalam menuntaskan studinya. Oleh sebab itu, dibutuhkan pemantauan terhadap masa studi mahasiswa [1]. Rata-rata masa studi mahasiswa belajar di Teknik Informatika Universitas Islam Madura masih diatas 4 tahun sehingga butuh dicoba evaluasi dengan memakai cara klasifikasi mahasiswa dengan menggunakan aplikasi *orange* dengan tiga model yaitu *K-Nearest Neighbor* (KNN), *Decision Tree* serta *Naive Bayes*. Lama studi ialah jangka waktu yang dibutuhkan mahasiswa dalam menuntaskan pendidikannya. Lama studi mahasiswa sudah diatur dalam ketetapan Kementerian Pendidikan dan Kebudayaan tentang sistem pendidikan program sarjana (S1) memiliki beban kredit semester yang harus ditempuh antara 144 sampai 160 SKS dengan lama pendidikan di kampus antara 8 sampai 10 semester atau setara antara 4 sampai 5 tahun [2].

Informasi nilai mahasiswa tiap semester serta informasi kelulusan mahasiswa bisa diolah buat menciptakan data yang berarti buat menganalisa ketepatan masa studi mahasiswa [3]. Bersumber pada informasi yang diperoleh dari Program Studi Teknik Informatika Universitas Islam Madura, pada tahun angkatan 2015 hingga dengan 2020 dengan jumlah rata-rata kelulusan 150 orang diperoleh data rata-rata masa studi mahasiswa masih diatas 4 tahun [4]. Oleh sebab itu, butuh dicoba analisis faktor-faktor yang menunjang ketepatan waktu serta keterlambatan masa studi mahasiswa tersebut. Riset terdahulu yang berkaitan dengan prediksi kelulusan mahasiswa serta klasifikasi banyak memakai mode *K-NN*, *SVM*, *Neural Network* serta *Naive Bayes* [5], [6], [7]. Sebaliknya banyak riset terdahulu yang mangulas hasil analisa perbandingan sebagian model data mining yang digunakan buat klasifikasi warga miskin [8], [9], klasifikasi penghasilan UMKM [10], klasifikasi gizi [11], klasifikasi kelulusan mahasiswa [12], serta klasifikasi buat industri pertanian serta kesehatan warga [12], [13].

Penelitian ini bertujuan untuk klasifikasi ketepatan waktu masa studi mahasiswa Prodi Teknik Informatika Universitas Islam Madura dengan menerapkan tiga metode yaitu *KNN*, *Naive Bayes* dan *Decision Tree*. Selanjutnya akan dilakukan analisa perbandingan tiga model dengan menerapkan analisa *Confusion Matrix* dan *ROC* untuk memastikan tingkat akurasi dari tiga metode tersebut. Kontribusi riset ini sangat membantu manajemen Program Studi Teknik Informatika Universitas Islam Madura untuk menyusun strategi meminimalisasi mahasiswa yang tidak tepat waktu dalam menyelesaikan masa studinya dan memberikan kontribusi untuk menentukan kinerja akurasi beberapa metode data mining antara lain *KNN*, *Naive Bayes* dan *Decision Tree*.

2. DASAR TEORI

Data mining ataupun penambangan informasi merupakan fitur lunak yang digunakan buat menciptakan pola tersembunyi, tren, ataupun aturan-aturan yang ada dalam basis berdimensi besar serta menciptakan aturan-aturan yang digunakan buat memperkirakan sikap di masa mendatang [14]. Klasifikasi dalam informasi *mining* ialah tata cara pendidikan informasi buat memprediksi nilai dari sekelompok atribut. Algoritma klasifikasi hendak menciptakan sekumpulan ketentuan yang disebut rule yang hendak digunakan selaku penanda buat bisa memprediksi kelas dari informasi yang mau diprediksi.

2.1. *K-Nearest Neighbor* (K-NN)

K-Nearest Neighbor (K-NN) merupakan suatu tata cara *supervised* yang berarti memerlukan informasi *training* buat mengklasifikasi obyek yang jaraknya sangat dekat. Prinsip kerja *K-NN* merupakan mencari jarak terdekat antara informasi yang hendak di evaluasi dengan k orang sebelah (*neighbor*) pada data pelatihan [6], [15]. *Dataset* dikelompokkan secara manual cocok dengan jenis data mahasiswa Prodi Teknik Informatika - Universitas Islam Madura. *Dataset* yang dijadikan acuan merupakan 35 data mahasiswa yang berikutnya hendak dicoba model klasifikasinya. Berikutnya rumus menghitung kemiripan vektor *dataset* dengan setiap

dataset latih yang sudah diklasifikasikan. Teorema K-NN buat menghitung jarak secara *universal* sebagai berikut:

$$d_i = \sqrt{\sum_{j=1}^n (x_{ij} - p_j)^2} \quad (1)$$

Keterangan

d_i = Jarak sampel

x_{ij} = Data sampel pengetahuan

p_j = Data input var ke-j

n = Jumlah sampel

Tahapan proses penerapan metode K-NN sebagai berikut [6]:

1. Menentukan parameter k (jumlah tetangga paling dekat).
2. Menghitung kuadrat jarak *eucliden* objek terhadap data training yang diberikan.
3. Mengurutkan hasil no 2 secara *ascending* (berurutan dari nilai tinggi ke rendah)
4. Mengumpulkan kategori Y (Klasifikasi *nearest neighbor* berdasarkan nilai k)
5. Dengan menggunakan kategori *nearest neighbor* yang paling mayoritas maka dapat diprediksikan kategori objek.

2.2. Naive Bayes

Naive Bayes ialah tata cara pengklasifikasian probabilistik simpel. Model ini hendak menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi serta campuran nilai dari *dataset* yang diberikan. Tata cara *Naive Bayes* berpendapat seluruh atribut pada tiap jenis tidak mempunyai ketergantungan satu sama lain [9], [16]. *Naive Bayes* teruji mempunyai akurasi serta kecepatan yang besar dikala diaplikasikan ke dalam database dengan informasi yang besar. Keuntungan pemakaian *Naive Bayes* ialah cuma membutuhkan beberapa kecil informasi latih buat memastikan parameter *mean* serta *varians* dari variabel yang dibutuhkan buat klasifikasi. *Teorema Bayes* mempunyai wujud universal selaku berikut [17]:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (2)$$

Keterangan

X : Data dengan *class* yang belum diketahui

H : Hipotesis data merupakan suatu *class* spesifik

$P(H|X)$: Probabilitas *hiotesis* H berdasar kondisi X (*Posteriori Probabilitas*)

$P(H)$: Probabilitas Hipotesis H (*Prior Probabilitas*)

$P(X|H)$: Probabilitas X berdasarkan kondisi hipotesis H

$P(X)$: Probabilitas X

2.3. Decision Tree

Decision Tree berdasarkan algoritma C4.5 adalah teknik klasifikasi yang umum digunakan untuk mengekstrak hubungan yang relevan dalam data. Algoritma C4.5 ialah program yang membuat pohon keputusan berdasarkan pada set data input berlabel. Kelebihannya adalah modelnya dapat dengan mudah ditafsirkan dan diimplementasikan dengan nilai kontinyu dan nilai diskrit. Algoritma C4.5 membagi data *training* dengan bantuan perolehan informasi. Atribut yang memiliki frekuensi tinggi dipertimbangkan untuk memisahkan data berdasarkan informasi yang tersedia dalam *dataset* [5], [18]. Pada saat menghitung nilai *gain* maka perlu diketahui *entropy value* yaitu dengan rumus berikut:

$$Entropy(S) = - \sum_{i=1}^n p_i \cdot \log_2 p_i \quad (3)$$

Keterangan:

S Himpunan kasus

- A Atribut
 n Jumlah patisi S
 Pi Proporsi dari S_i terhadap S

Persamaan yang digunakan untuk menghitung *Information Gain*:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (4)$$

Keterangan:

- S Himpunan kasus
 A Atribut
 n Jumlah patisi S
 Pi Proporsi dari S_i terhadap S
 $A|S_i|$ Jumlah kasus pada partisi ke-i
 $|S|$ Jumlah kasus dalam S

2.4. Elemen Pengukuran

Confusion Matrix merupakan pengukuran performa buat permasalahan klasifikasi *machine learning* dimana luaran bisa berbentuk 2 kelas ataupun lebih. *Confusion Matrix* merupakan tabel dengan 4 campuran berbeda dari nilai prediksi serta nilai aktual. Berdasarkan Tabel.2 dapat dijelaskan bahwa terdapat 4 sebutan representasi hasil proses klasifikasi pada *confusion matrix* ialah *True Positif* (TP), *True Negatif* (TN), *False Positif* (FP), serta *False Negatif* (FN) [13], [19]. *Confusion matrix* pula kerap disebut *error matrix*. Pada dasarnya *confusion matrix* membagikan data perbandingan hasil klasifikasi yang dicoba oleh sistem dengan hasil klasifikasi sesungguhnya.

Tabel 2. Pengujian *Confusion Matrix*

Classification	Predicted Class	
	True	False
Actual: True	True Positif (TP)	False Negatif (FN)
Actual: False	False Positif (FP)	True Negatif (TN)

Accuracy didefinisikan selaku jenjang korelasi antara nilai prediksi dengan nilai aktual. *Precision* merupakan tingkatan ketepatan antara data yang diharapkan oleh pengguna dengan jawaban yang diberikan oleh sistem. Sebaliknya *recall* merupakan tingkatan keberhasilan sistem dalam menciptakan kembali suatu data.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

3. METODOLOGI PENELITIAN

Penelitian ini bertujuan untuk melakukan analisa perbandingan metode *KNN*, *Naive Bayes* dan *Decision Tree* yang digunakan untuk mengklasifikasi kelulusan mahasiswa Prodi Teknik Informatika Universitas Islam Madura, aplikasi yang digunakan untuk simulasi adalah *Orange Data Mining* yaitu aplikasi *data mining open source* yang terbukti mampu membantu peneliti menganalisa datanya. Tahapan proses pada riset ini bisa dilihat pada Gambar 1.

Gambar 1. Tahapan penelitian

Sesuai Gambar 1 tersebut, langkah pertama adalah identifikasi masalah, perumusan dan kajian pustaka hal ini dilakukan pertama kali untuk menyusun tujuan riset dan kontribusi riset. Kedua adalah proses *collecting data* yaitu menyusun data latih dan data uji sebagai sumber klasifikasi data. Ketiga adalah proses perancangan *widget orange data mining* untuk proses klasifikasi kelulusan mahasiswa dan perbandingan metode. Keempat adalah proses klasifikasi kelulusan mahasiswa Prodi Teknik Informatika - Universitas Islam Madura menggunakan model *KNN*, *Decision Tree*, serta *Naive Bayes*. Kelima adalah proses evaluasi kinerja metode klasifikasi dan menganalisa hasil perbandingan metode tersebut.

3.1. Atribut Penelitian

Kumpulan data mahasiswa Teknik Informatika Universitas Islam Madura berupa data primer diperoleh melalui wawancara dengan bagian Akademik Universitas Islam Madura. Atribut awal dari data ini adalah 1 atribut tujuan dan 35 data *instance*. Deskripsi data set dijelaskan pada Tabel 1.

Tabel 1. Attribute data mahasiswa

No	Attribute	Type	Description
1	NIM	Teks	Nomer Induk Mahasiswa
2	Nama	Teks	Nama Mahasiswa
3	JKL	Teks	Jenis Kelamin
4	Usia	Numeric	Usia Mahasiswa
5	IPS1	Numeric	IP Sementara semester ke-1
6	IPS2	Numeric	IP Sementara semester ke-2
7	IPS3	Numeric	IP Sementara semester ke-3
8	IPS4	Numeric	IP Sementara semester ke- 4
9	IPS5	Numeric	IP Sementara semester ke-5
10	IPS6	Numeric	IP Sementara semester ke- 6
11	IPS7	Numeric	IP Sementara semester ke-7
12	IPS5	Numeric	IP Sementara semester ke-8
13	Status	Categorical	Status Kelulusan

3.2. Data Selection Process / Preprocessing

Pada saat proses *preprocessing dataset* mahasiswa Teknik Informatika Universitas Islam Madura ini tidak ada yang mengalami *missing value* pada datanya sehingga hanya dilakukan pemilihan data saja. Nilai yang hilang dalam data *instance* akan mengganggu proses klasifikasi. Beberapa model pada klasifikasi tidak dapat diproses karena ada data dan nilai yang hilang.

Gambar 2. Proses pemilihan data menggunakan widget

Berdasarkan Gambar 2 tersebut ialah proses pemilihan data menggunakan *widget*, pertama pilih kolom dimana *atribute* yang akan digunakan adalah Usia, Nilai IP Sementara semester 1 sampai semester 8 dengan *atribute* target adalah status kelulusan mahasiswa.

3.3. Data Mining Process

Dalam menganalisa performa beberapa model klasifikasi pada *orange tool*, dilakukan perbandingan beberapa metode *data mining* untuk memilih metode yang terbaik dengan akurasi yang tinggi, dalam mengklasifikasi *dataset* status kelulusan mahasiswa Teknik Informatika Universitas Islam Madura seperti terlihat pada Gambar 3.

Gambar 3. Design widget model klasifikasi status kelulusan mahasiswa

Pada Gambar 3 merupakan perancangan *widget* memakai model klasifikasi pada *software Orange data mining* berupa K-NN, Decision Tree serta Naive Bayes yang diinputkan *dataset* yang telah diolah sebelumnya. Kemudian *dataset* tersebut diproses kedalam mode klasifikasi.

3.4. Proses Pengujian Model Klasifikasi

Dalam proses pengujian model klasifikasi yang telah dibuat sebelumnya, dibutuhkan kumpulan data uji untuk mengetahui hasil klasifikasi seperti terlihat di Gambar 4.

Gambar 4. *Design widget model klasifikasi dataset status kelulusan mahasiswa*

Pada Gambar 4 adalah desain *widget* yang telah ditambahkan proses uji coba klasifikasi untuk model klasifikasi. Pada gambar kotak merah adalah sekumpulan data uji coba yang dimasukkan ke dalam proses klasifikasi untuk mengetahui hasil klasifikasi kelulusan mahasiswa Teknik Informatika Universitas Islam Madura.

3.5. *Proses Evaluasi Hasil Perbandingan Model Klasifikasi*

Proses selanjutnya adalah melakukan proses perbandingan model klasifikasi dengan menggunakan *Test and Score* diperlukan untuk menghitung tingkat keberhasilan antara masing-masing model klasifikasi di data mining *Orange* seperti terlihat di Gambar 5.

Gambar 5. *Design widget untuk menghitung keberhasilan model klasifikasi*

Pada Gambar 5 merupakan desain *widget* yang telah ditambahkan proses perhitungan tingkat keberhasilan model klasifikasi dengan menggunakan *widget Test and Score* yang selanjutnya akan dilakukan evaluasi akurasi menggunakan *Confusion Matrix* dan *ROC Analysis*.

4. PENGUJIAN DAN PEMBAHASAN

4.1. Hasil simulasi 3 model klasifikasi

Hasil simulasi model klasifikasi dilakukan dengan menggunakan kumpulan data uji dengan 1 atribut sebagai target, 9 *atribute numeric* yaitu Umur, Indeks Prestasi Sementara (IPS) mahasiswa mulai semester 1 sampai semester 8, sehingga diperoleh hasil *test score* seperti terlihat pada Gambar 6.

Gambar 6. Hasil widget test and score

Berdasarkan 35 data mahasiswa yang telah diuji, diperoleh hasil perhitungan *Precision*, *recall*, *accuracy* dari masing-masing model seperti terlihat pada Gambar 6. Hasil klasifikasi model *K-NN*, *Decision Tree* serta *Naive Bayes* menunjukkan bahwa nilai akurasi *Naive Bayes* paling tinggi yaitu 86%.

Berdasarkan Gambar 6 juga memperlihatkan perbandingan 3 model AUC, diketahui bahwa nilai AUC yang paling tinggi adalah metode *Naive Bayes* yaitu 0.987. AUC dipakai buat mengukur kinerja diskriminatif dengan memperkirakan probabilitas output dari ilustrasi yang diseleksi secara acak dari populasi positif ataupun *negatif*, semakin besar AUC, semakin baik hasil klasifikasi yang digunakan.

4.2. Hasil Evaluasi dengan *Confusion Matrix*

Confusion Matrix merupakan pengukuran performa buat permasalahan klasifikasi *machine learning* dimana keluaran bisa berbentuk 2 kelas ataupun lebih. *Confusion Matrix* merupakan tabel dengan 4 campuran berbeda dari nilai prediksi serta nilai aktual. Hasil evaluasi untuk masing-masing model klasifikasi dapat dilihat pada Gambar 7 untuk model *K-NN*, sedangkan hasil *Confusion Matrix* model *Decision Tree* dapat dilihat pada Gambar 8 serta nilai *Confusion Matrix* model *Naive Bayes* dapat dilihat pada Gambar 9.

		TEPAT	TERLAMBAT	Σ
Actual	TEPAT	26	0	26
	TERLAMBAT	8	1	9
Σ		34	1	35

Gambar 7. Nilai *Confusion Matrix* metode *K-NN*

Pada Gambar 7 menunjukkan bahwa nilai dari *True Positif* (TP) adalah 26, *True Negatif* (TN) adalah 1, *False Positif* (FP) adalah 0, dan *False Negatif* (FN) adalah 8. Maka nilai *Accuracy*, *Precision* dan *Recall* dari metode *K-NN* adalah sebagai berikut:

$$Accuracy = \frac{(26 + 1)}{(26 + 8 + 1 + 0)} \times 100\% \quad \text{Maka nilai Accuracy} = 77\%$$

$$Precision = \frac{(26)}{(26 + 8)} \times 100\% \quad \text{Maka nilai Precision} = 76\%$$

$$Recall = \frac{(26)}{(26 + 0)} \times 100\% \quad \text{Maka nilai Recall} = 100\%$$

		TEPAT	TERLAMBAT	Σ
Actual	TEPAT	22	4	26
	TERLAMBAT	4	5	9
	Σ	26	9	35

Gambar 8. Nilai Confusion Matrix metode Decision Tree

Pada Gambar 8 menunjukkan bahwa nilai dari *True Positif* (TP) adalah 22, *True Negatif* (TN) adalah 4, *False Positif* (FP) adalah 4, dan *False Negatif* (FN) adalah 5. Maka nilai *Accuracy*, *Precision* dan *Recall* dari metode *Decision Tree* adalah sebagai berikut:

$$Accuracy = \frac{(22 + 4)}{(22 + 4 + 4 + 5)} \times 100\% \quad \text{Maka nilai Accuracy} = 74\%$$

$$Precision = \frac{(22)}{(22 + 4)} \times 100\% \quad \text{Maka nilai Precision} = 85\%$$

$$Recall = \frac{(22)}{(22 + 4)} \times 100\% \quad \text{Maka nilai Recall} = 85\%$$

		TEPAT	TERLAMBAT	Σ
Actual	TEPAT	23	3	26
	TERLAMBAT	1	8	9
	Σ	24	11	35

Gambar 9. Nilai Confusion Matrix metode Naive Bayes

Pada Gambar 9 menunjukkan bahwa nilai dari *True Positif* (TP) adalah 23, *True Negatif* (TN) adalah 8, *False Positif* (FP) adalah 3, dan *False Negatif* (FN) adalah 1. Maka nilai *Accuracy*, *Precision* dan *Recall* dari metode *Naive Bayes* adalah sebagai berikut:

$$Accuracy = \frac{(23 + 8)}{(23 + 8 + 3 + 1)} \times 100\% \quad \text{Maka nilai Accuracy} = 89\%$$

$$Precision = \frac{(23)}{(23 + 3)} \times 100\% \quad \text{Maka nilai Precision} = 88\%$$

$$Recall = \frac{(23)}{(23 + 1)} \times 100\% \quad \text{Maka nilai Recall} = 96\%$$

Berdasarkan hasil evaluasi dan validasi dengan menggunakan *Confusion Matrix* diperoleh nilai perbandingan *Accuracy*, *Precision* dan *Recall* dari 3 metode *K-NN*, *Naive Bayes*, dan *Decision Tree* seperti terlihat pada Tabel 2.

Tabel 2. Perbandingan Kinerja

Metode	Accuracy	Precision	Recall
<i>K-NN</i>	77%	76%	100%
<i>Decision Tree</i>	74%	85%	85%
<i>Naive Bayes</i>	89%	88%	96%

Berdasarkan Tabel 2 dapat diketahui bahwa kinerja dari model *Naive Bayes* lebih baik dari model *K-NN* dan *Decision Tree*. Akurasi klasifikasi tidak bisa mencapai hasil yang sempurna karena pasti ada nilai *error*. Hal tersebut dipengaruhi oleh banyaknya data uji dan data latih yang digunakan dalam proses simulasi yang dilakukan.

4.3. Hasil Evaluasi dengan ROC Curve

Nilai akurasi secara manual bisa dilakukan dengan melihat perbandingan curve ROC yang divisualisasi dari *Confusion Matrix*. Model melihat kurva ROC adalah cara yang paling mudah terlihat untuk membandingkan nilai akurasi masing-masing model klasifikasi secara grafis. Hasil grafis ROC dapat dilihat pada Gambar 10 dan 11.

Pada Gambar 10 menunjukkan bahwa hasil analisis ROC kelulusan mahasiswa Teknik Informatika-Universitas Islam Madura yang TEPAT pada masing-masing model sebagai berikut: (1) *K-NN* adalah 0.454, (2) *Naive Bayes* adalah 0.698, dan (3) *Decision Tree* adalah 0.955. Oleh karena itu, untuk studi kasus ini model yang memiliki nilai akurasi paling baik adalah *Naive Bayes* dan *K-NN* karena kurvanya mendekati titik 0.1.

Gambar 10. Analisa ROC dengan target kelulusan mahasiswa TEPAT

Pada Gambar 11 menunjukkan bahwa hasil analisis ROC kelulusan mahasiswa Teknik Informatika Universitas Islam Madura yang TERLAMBAT pada masing-masing model klasifikasi sebagai berikut: (1) *K-NN* adalah 0.546, (2) *Naive Bayes* adalah 0.562, dan (3) *Decision Tree* adalah 0.045. Maka dari itu, riset klasifikasi dengan menggunakan 3 model dengan studi kasus Prodi Teknik Informatika – Universitas Islam Madura sangat direkomendasikan menggunakan model *Naive Bayes* dan *K-NN* karena kurvanya mendekati titik 0.1.

Gambar 11. Analisa ROC dengan target kelulusan mahasiswa TERLAMBAT

5. KESIMPULAN

Hasil penelitian ini menunjukkan bahwa setelah menggunakan model *K-Nearest Neighbor*, *Decision Tree* serta *Naive Bayes* untuk mengklasifikasi status kelulusan mahasiswa Teknik Informatika Universitas Islam Madura diperoleh hasil bahwa kinerja *Naive Bayes* lebih unggul dari *K-Nearest Neighbor* serta *Decision Tree*. Terbukti bahwa dari 35 data uji yang digunakan *Naive Bayes* memiliki nilai akurasi 89%, presisi 88% sedangkan *K-Nearest Neighbor* memiliki nilai akurasi 77% , presisi 76% dan *Decision Tree* memiliki nilai akurasi 74% dan presisi 84%. Kontribusi riset ini bisa digunakan oleh manajemen Prodi Teknik Informatika Universitas Islam Madura untuk mendeteksi sejak awal kondisi mahasiswa supaya tidak kelulusannya tidak terlambat dan mempengaruhi nilai akreditasi Program Studi Teknik Informatika Universitas Islam Madura.

Daftar Pustaka

- [1] D. Kartini, R. A. Nugroho, and M. R. Faisal, "Klasifikasi Kelulusan Mahasiswa Menggunakan Algoritma Learning Vector Quantization," *POSITIF J. Sist. dan Teknol. Inf.*, vol. 3, no. 2, p. 93, 2017.
- [2] Fakultas Teknik Universitas Islam Madura, "Rencana Strategis," 2018.
- [3] A. Trimanto, F. Faqih, I. M. Irfani, and S. Timur, "Penerapan Data Mining Untuk Evaluasi Status Kelulusan Mahasiswa Fakultas Teknologi Pertanian Tahun 2015 Menggunakan Algoritma Naive Bayes Classifier," in *Seminar Nasional Informatika, Amikom University*, 2015.
- [4] Fakultas Teknik Universitas Islam Madura, "Rencana Operasional Fakultas Teknik Universitas Islam Madura," 2020.
- [5] E. P. Rohmawan, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Desicion Tree Dan Artificial Neural Network," *J. Ilm. MATRIK*, vol. 20, no. 1, pp. 21–30, 2018.
- [6] I. A. A. Amra and A. Y. A. Maghari, "Students performance prediction using KNN and Naive Bayes ian," in *ICIT 2017 - 8th International Conference on Information Technology, Proceedings*, 2017, no. December 2018, pp. 909–913.
- [7] S. P. Nabila, N. Ulinuha, and A. Yusuf, "Model Prediksi Kelulusan Tepat Waktu Dengan Metode Fuzzy C-Means Dan K-Nearest Neighbors Menggunakan Data Registrasi Mahasiswa," *Netw. Eng. Res. Oper.*, vol. 6, no. 1, p. 39, 2021.
- [8] D. Iskandar and Y. K. Suprpto, "Perbandingan Akurasi Klasifikasi Tingkat," *Netw. Eng. Res. Oper.*, vol. 2, no. 1, pp. 37–43, 2015.
- [9] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018.

- [10] L. Yuwono, M. E. Fadillah, M. Indrayani, W. Maesarah, A. Ramadhan, and S. F. Panjaitan, "Klasifikasi Pendapatan Pedagang Kaki Lima Dan Pelaku Usaha Online Akibat Dampak Covid-19 Menggunakan Metode Naive Bayes," *Bull. Appl. Ind. Engineering Theory*, vol. 2, no. 1, pp. 1–6, 2021.
- [11] H. Hafizan and A. N. Putri, "Penerapan Metode Klasifikasi Decision Tree Pada Status Gizi Balita Di Kabupaten Simalungun," *KESATRIA J. Penerapan Sist. Inf. (Komputer Manajemen)*, vol. 1, no. 2, pp. 68–72, 2020.
- [12] S. Dewi, "Komparasi 5 Metode Algoritma Klasifikasi Data Mining Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan," *Techno Nusa Mandiri*, vol. XIII, no. 1, pp. 60–66, 2016.
- [13] I. B. P. Jayawiguna, "Comparison of Model Prediction for Tile Production in Tabanan Regency with Orange Data Mining Tool," in *International Journal of Engineering and Emerging Technology*, 2020, vol. 5, no. 2, pp. 72–76.
- [14] I. A. Nikmatun and I. Waspada, "Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *J. SIMETRIS*, vol. 10, no. 2, pp. 421–432, 2019.
- [15] S. H. Rukmawan, F. R. Aszhari, Z. Rustam, and J. Pandelaki, "Cerebral Infarction Classification Using the K-Nearest Neighbor and Naive Bayes Classifier," *J. Phys. Conf. Ser.*, vol. 1752, no. 1, 2021.
- [16] B. Krithiga, P. Sabari, I. Jayasri, and I. Anjali, "Early detection of coronary heart disease by using naive bayes algorithm," *J. Phys. Conf. Ser.*, vol. 1717, no. 1, 2021.
- [17] Bustami, "Penerapan Algoritma Naive Bayes Untuk Nasabah Asuransi," *J. Inform.*, vol. 8, no. 1, pp. 884–898, 2014.
- [18] T. Hardiani, "Comparison of Naive Bayes Method, K-NN (K-Nearest Neighbor) and Decision Tree for Predicting the Graduation of 'Aisyiyah University Students of Yogyakarta," *Int. J. Heal. Sci. Technol.*, vol. 2, no. 1, 2021.
- [19] R. Puspita and A. Widodo, "Perbandingan Metode KNN, Decision Tree, dan Naive Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS," *J. Inform. Univ. Pamulang*, vol. 5, no. 4, p. 646, 2021.