# IMPLEMENTASI K-NEAREST NEIGHBORD PADA RAPIDMINER UNTUK PREDIKSI KELULUSAN MAHASISWA

# Sumarlin, Dewi Anggraini

<sup>1</sup>Program Studi Teknik Informatika STIKOM Uyelindo Kupang Email : *Shumarlin@gmail.com* <sup>2</sup>Program Studi Sistem Informasi STIKOM Uyelindo Kupang Email : *thewi 1205@yahoo.com* 

#### **ABSTRACT**

Data on graduate students is an important part in determining the quality of a private and public university. Graduate data is included in important assessments in the accreditation process. Data from Uvelindo Kupang STIKOM graduates every year will continue to grow and accumulate like neglected data because it is rarely used. To maximize student data into information that can be used by universities, the data must be processed in this case used as training data in a study using data mining to obtain information in the form of predictions of graduation from Kupang Uyelindo STIKOM students. The method used in this study is K-Nearest Neighbor using rapidminer software to measure K-Nearest Neighbor's accuracy against student graduate data. The criteria used were in the form of student names, gender, cumulative achievement index (GPA) from semester 1 to 6. In applying the K-Nearest Neighbor algorithm can be used to produce predictions of student graduation. To measure the performance of the k-nearest neighbor algorithm, the Cross Validation, Confusion Matrix and ROC Curves methods are used, in this study using a 5-fold cross validation to predict student graduation. From 100 student dataset records Uyelindo Kupang STIKOM graduates obtained accuracy rate reached 82% and included a very good classification because it has an AUC value between 0.90-1.00, which is 0.971, so it can be concluded that the accuracy of testing of student graduation models using K-Nearest Neighbor (K-NN) algorithm is influenced by the number of data clusters. Accuracy and the highest AUC value of 5-fold validation is to cluster data k = 4 with the accuracy value of 90%.

Keywords:: K-Nearest Neighbor, Cross Validation, Confusion Matrix, Kurva ROC.

#### 1. PENDAHULUAN

# 1.1 Latar Belakang

STIKOM Üyelindo Kupang merupakan kampus IT Pertama di NTT, memiliki 3 Program Studi yang sudah terakreditasi B oleh BAN PT. Jumlah mahasiswa aktif sebanyak 325, setiap tahun STIKOM Uyelindo Kupang meluluskan Lebih dari 100 mahasiswa yang terdiri dari 3 prodi. Jumlah lulusan tiga tahun terakhir: 2015 (194). 2016 (224), 2017 (230).

Lembaga pendidikan tertinggi adalah perguruan tinggi yang merupakan penyelenggara pendidikan akademik bagi mahasiswa. Mahasiswa sering disebut kelompok masyarakat yang memiliki cirri intelektualitas yang lebih luas dibandingkan dengan kelompok seusia mereka yang bukan mahasiswa ataupun kelompok usia lain yang dibawah mereka. Dengan intelektualitasnya mahasiswa akan mampu menghadapi dan mencari permasalahan secara sistematis yang nantinya diterapkan dalam kehidupan sehari-hari agar bisa bersaing dalam dunia kerja [1].

Presentase naik turunnya kemampuan mahasiswa untuk menyelesaikan studi tepat waktu merupakan salah satu elemen penilaian akreditasi Perguruan Tinggi (Buku VI Matriks Penilaian Instrumen Akreditasi Program Studi Badan Akreditasi Nasional Perguruan Tinggi, 2008).

Untuk itu perlu adanya pemantauan maupun evaluasi terhadapat kecenderungan mahasiswa lulus tepat waktu atau tidak.

Suatu perguruan tinggi menyimpan data akademik, administrasi, biodata mahasiswa dan lain-lain. Data tersebut akan sangat bermakna jika digali dengan tepat sehingga dapat diketahui pola atau pengetahuan untuk mengambil keputusan [10]. Dengan menerapkan data mining dapat memecahkan masalah dengan menganalisis data yang dimiliki perguruan tinggi.

Data tentang mahasiswa yang lulus dapat memberikan informasi yang berguna bagi perguruan tinggi jika dimanfaatkan dengan maksimal. Salah satu cara untuk memanfaatkan data tentang mahasiswa yang lulus ini adalah dengan mengolahnya menggunakan data mining. Dengan proses data mining ini dapat ditemukan pola atau aturan yang dapat digunakan untuk menghasilkan suatu informasi seperti prediksi kelulusan mahasiswa.

Jumlah mahasiswa yang lulus dan mahasiswa baru yang masuk setiap tahunnya tidak sebanding, untuk itu perlu adanya analisis untuk memprediksi kelulusan mahasiswa. prediksi kelulusan mahasiswa ini memerlukan informasi yang ada untuk mengetahui apakah seorang mahasiswa dapat lulus tepat waktu atau tidak.

Apabila kelulusan mahasiswa dapat diketahui sejak dini maka pihak akademik dapat menerapkan suatu kebijakan untuk meminimalisir jumlah mahasiswa yang tidak lulus tepat waktu sesuai dengan masa studinya.

Dalam penelitian yang berjudul Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques, Jadhav dan Channe membandingkan performa metode K-NN, Naïve Bayes, dan pohon keputusan dalam berbagai aspek dengan menggunakan berbagai dataset. Dari hasil penelitian tersebut dapat diketahui bahwa pohon keputusan merupakan metode yang paling cepat performanya dibandingkan dengan metode yang lain. Selain itu pohon keputusan lebih akurat dan memiliki error rate yang rendah [13].

Dalam penelitian lain yang berjudul Comparative Analysis of Decision Tree Algorithms for The Prediction of Eligibility of A Man for Availing Bank Loan, Mohankumar dkk membandingkan berbagai algoritma untuk membangun pohon keputusan dan algoritma C4.5 merupakan algoritma dengan performa tercepat dan memiliki akurasi yang paling tinggi [9].

#### 1.2 Perumusan Masalah

Berdasarkan uraian latar belakang, masalah yang ditemukan adalah bagaimana tingkat akurasi algoritma K-Nearest Neighbor dalam memprediksi kelulusan mahasiswa STIKOM Uyelindo Kupang?.

## 1.3 Tujuan Penelitian

Adapun yang menjadi tujuan dalam peelitian ini adalah untuk mengetahui tingkat akurasi algoritma K-Nearest Neigbor dalam memprediksi ketepatan kelulusan mahasiswa STIKOM Uyelindo Kupang.

#### 1.4 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah:

- Memberikan Peringatan dini bagi mahasiswa yang berpotensi tidak lulus tepat waktu
- Membuat kebijakan dalam meningkatkan kelulusan mahasiswa
- 3. Memberikan kontribusi keilmuan pada bidang klasifikasi data mining untuk prediksi kelulusan mahasiswa.

#### 1.5 Ruang Lingkup Penelitian

Agar permasalahan yang dibahas tidak menyimpang dari pokok pembahasan, maka ruang lingkup penelitian adalah sebagai berikut:

- Penelitian dilakukan pada Prodi Teknik Informatika S1, Teknik Informatika D3 dan Sistem Informasi S1.
- Software yang digunakan adalah Rapidminer dan Algortima yang digunakan adalah K-Nearest Neighbor
- 3. Data Yang digunakan adalah data Mahasiswa lulusan STIKOM Uyelindo.

#### 2. METODE PENELITIAN

## 2.1 Algoritma K-Nearest Neighbor

Algoritma K-Nearest Neighbor merupakan metode klasifikasi yang mengelompokan data baru berdasarkan jarak data baru itu kebeberapa data/tetangga (neighbord) terdekat [12].

Teknik K-Nearest Neighbor dengan melakukan langkah-langkah yaitu [12], mulai input: Data training, label data traning, k, data testing

- a. Untuk semua data testing, hitung jaraknya ke setiap data training
- b. Tentukan k data training yang jaraknya paling dekat dengan data
- c. Testing
- d. Periksa label dari k data ini
- e. Tentukan label yang frekuensinya paling banyak
- f. Masukan data testing ke kelas dengan frekuensi paling banyak
- g. Berhenti

Rumus KNN:

$$d_i = \sqrt{\sum_{i=1}^{p} (x_{2i} - x_{1i})^2}$$

Keterangan:

 $x_1 =$ Sampel Data

 $x_2$  = Data Uji / Testing

i = Variabel Data

d = Jarak

p = Dimensi Data

Ada banyak cara untuk mengukur jarak kedekatan antara data baru dengan data lama (data training), diantaranya *euclidean distance* dan manhattan distance (*city block distance*), yang paling sering digunakan adalah euclidean distance [3], yaitu:

Untuk melakukan pembobotan variabel dilakukan dengan metode perbandingan pasangan, bobot variabel ditentukan dengan cara normalisasi vector eigen, yang diasosiasikan dengan nilai eigen maksimum pada suatu matriks rasio sebelum membandingkan harus ditentukan skala nilai pengaruh atau penting antara variabel, seperti yang terlihat pada Tabel 1 [11].

Tabel 1. Skala untuk perbandingan pasangan Nilai Defenisi

Nilai	Defenisi
1	Sama Penting
2	Sama hingga cukup penting
3	Cukup penting
4	Cukup penting hingga tinggi
	kepentingannya
5	Tinggi kepentingannya
6	Tinggi kepentingannya hingga sangat
	tinggi
7	Sangat tinggi kepentingannya
8	Kepentingannya sangat tinggi hingga
	amat sangat tinggi
9	Kepentingannya amat sangat tinggi

#### 2.2 Metode Klasifikasi

Klasifikasi merupakan proses penemuan (fungsi) yang menggambarkan membedakan kelas data atau konsep yang bertujuan agar bias digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [6]. Klasifikasi merupakan bagian dari data mining, dimana Data mining merupakan suatu istilah yang menguraikan digunakan untuk penemuan pengetahuan didalam database. Data mining juga proses vang menggunakan teknik statistik. matematika, kecerdasan buatan, dan machine learning untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [5].

Proses klasifikasi didasarkan pada empat komponen [5].

- 1. Kelas, Variabel dependen yang berupa kategorikal yang merepresentasikan 'label' yang terdapat pada objek. Contohnya: resiko penyakit jantung, resiko kredit, customer loyalty, jenis gempa.
- 2. Predictor, Variabel independen yang direpresentasikan oleh karakteristik (atribut) data. Contohnya: merokok, minum alkohol, tekanan darah, tabungan, aset, gaji.
- 3. Training dataset, Satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan predictor.
- 4. Testing dataset, Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluas.

# 2.3 Metode Evaluasi dan Validasi Algoritma Data Mining

# a. Cross Validation

Cross validation merupakan pengujian standar yang dilakukan untuk memprediksi error. Data training dibagi secara random ke dalam beberapa bagian dengan perbandingan yang sama kemudian error dihitung bagian demi bagian, selanjutnya hitung rata-rata seluruh error untuk mendapatkan error rate secara keseluruhan [7].

Evaluasi digunakan untuk mengukur kinerja metode klasifikasi, dalam penelitian ini digunakan untuk mengukur keakuratan metode klasifikasi yang diukur dengan akurasi, precision dan recall. Recall didefinisikan sebagai persentase antara data kelas data buruk yang dikelaskan dengan benar dan data kelas data buruk yang salah diprediksi ke kelas data baik. Precision adalah persentase dari kelas data buruk yang dikelaskan dengan benar dan kelas yang seharusnya termasuk kelas data baik tetapi dikelaskan sebagai kelas data buruk.

# b. Confusion Matrix

Metode ini menggunakan tabel matriks, jika data set hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang

lainnya negative [3].

Tabel 2. Model Confusion Matrix

Klasifikasi yang	Diklasifikasi sebagai			
benar	+	-		
+	True	False negative		
-	positive False positive	True negative		

#### Keterangan tabel:

- 1. True positives merupakan jumlah record positif yang diklasifikasikan sebagai positif,
- 2. False positives merupakam jumlah record negative yang diklasifikasikan sebagai positif,
- 3. False negatives merupakan jumlah record positif yang diklasifikasikan sebagai negatif,
- 4. True negatives merupakan jumlah record negative yang diklasifikasikan sebagai negative.

# c. Kurva ROC

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan confusion matrix. ROC adalah grafik dua dimensi dengan false positives sebagai garis horisontal dan true positives sebagai garis vertikal. The area under curve (AUC) dihitung untuk mengukur perbedaan performansi metode yang digunakan. AUC dihitung menggunakan rumus [8]:

$$\theta^n = \frac{1}{mn} \sum_{j=1}^n \sum_{i=1}^m \psi(xi^r, xj^r)$$

Dimana,

$$K = \begin{cases} 1 & Y < X \\ 1 & Y = X \\ 2 & Y > X \end{cases}$$

# Keterangan:

K = Jumlah algoritma yang dikomparasi

X = Output Positif

Y = Output Negatif

#### 2.4 Kelulusan Mahasiswa

sering Mahasiswa disebut kelompok masyarakat yang memiliki cirri intelektualitas yang lebih luas dibandingkan dengan kelompok seusia mereka yang bukan mahasiswa ataupun kelompok usia lain yang dibawah mereka. Dengan intelektualitasnya mahasiswa akan mampu menghadapi dan mencari permasalahan secara sistematis yang nantinya diterapkan dalam kehidupan sehari-hari agar bisa bersaing dalam dunia kerja [1].

Kelulusan mahasiwa adalah hal yang penting diperhatikan, karena persentase jumlah kelulusan mempengaruhi penilaian pemerintah serta mempengaruhi status akreditasi program studi [15]. Faktor-faktor yang dapat mempengaruhi kelulusan mahasiswa antara lain adalah nilai akhir SMA, Indeks Prestasi Semester (IPS), gaji orang tua dan pekerjaan orang tua [4]. Indeks prestasi sering digunakan sebagai indikator penilaian akademik,

banyak perguruan tinggi member standar minimum yang sulit di peroleh mahasiswa [11]. Adapun variabel yang dapat digunakan dalam prediksi kelulusan mahasiswa seperti umur, status pernikahan, jumlah saudara [16].

Pada penenlitian ini parameter yang digunakan adalah nomor induk, nama mahasiswa, jenis kelamin, indeks prestasi semester satu sampai dengan indeks prestasi semester enam dan status kelulusan mahasiswa.

#### 3. PROSEDUR PENELITIAN

# 3.1 Observasi

Observasi dilakukan dengan mengadakan pengamatan secara langsung terhadap suatu objek yang akan diteliti yaitu pada STIKOM Uyelindo Kupang.

#### 3.2 Wawancara

Melakukan *interview* dengan Wakil Ketua Bidang I dan Ketua Program Studi terkait untuk mendapatkan informasi secara lisan dengan tujuan untuk memperoleh data yang dapat menjelaskan ataupun menjawab sebuah permasalahan dalam penelitian ini.

# 3.3 Kepustakaan

Dilakukan dengan cara mencari literaturliteratur pendukung yang mampu memberikan informasi yang memadai dalam menyelesaikan penelitian ini serta dapat membantu mempertegas teori-teoriyang sudah ada.

### 3.4 BahanPenelitian

Bahan yang digunakan dalam penelitian ini adalah data-data tentang mahasiswa per angkatan dan per program studi. Buku-buku dijadikan sebagai panduan.

### 3.5 Analisis Data

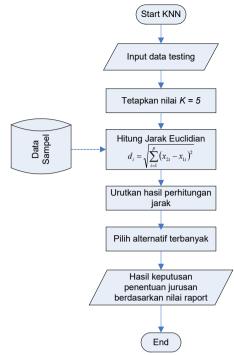
Analisa Data yaitu menganalisa terhadap permasalahan untuk mengetahui dan menentukan batasan-batasan sehigga dapat menentukan cara yang efektif dalam menyelesaikan permasalah tersebut.

#### 3.6 Penerapan Metode K-NN

Cara Kerja Algortima KNN:

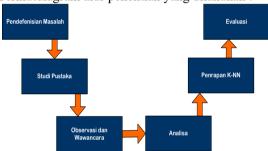
- 1. Tentukan parameter *K*
- 2. Hitung jarak antara data yang akan dievaluasi dengan semua pelatihan
- 3. Urutkan jarak yang terbentuk (urut naik)
- 4. Tentukan jarak terdekat sampai urutan K
- 5. Pasangkan kelas yang bersesuaian
- 6. Cari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi

Di bawah ini merupakan flowchart dari metode KNN:



Gambar 1. Flowchart Algoritma KNN

Berikutdiagram alur penelitian yang dilakukan:



Gambar 2. Diagram Alur Penelitian

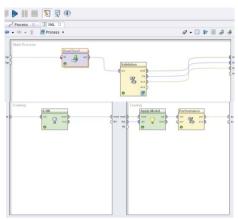
# 4. METODOLOGI PENELITIAN

## 4.1 Implementasi Rapid Miner

Rapid Miner merupakan software/perangkat lunak untuk pengolahan data. Dengan menggunakan prinsip dan algoritma data mining, RapidMiner mengekstrak pola-pola dari data set yang besar dengan mengkombinasikan metode statistika, kecerdasan buatan dan database.

Rapid Miner memudahkan penggunanya dalam melakukan perhitungan data yang sangat banyak dengan menggunakan operator-operator. Operator ini berfungsi untuk memodifikasi data. Data dihubungkan dengan node-node pada operator kemudian menghubungkan ke node hasil untuk melihat hasilnya. Hasil yang diperlihatkan RapidMiner dapat ditampilkan secara visual dengan grafik. RapidMiner merupakan salah satu software pilihan untuk melakukan prediksi data dengan metode-metode data mining.

Ada gambar berikut merupakan contoh kasus prediksi kelulusan mahasiswa menggunakan konfigurasi data dan operator.



Gambar 3. Konfigurasi RapidMiner

#### 4.2 Pengolahan Data Awal

Data yang digunakan dalam penelitian ini diperoleh dari observasi peneliti, dalam penelitian ini menggunakan data mahasiswa lulusan STIKOM Uyelindo Kupang dengan jumlah sampel yang digunakan sebanyak 100 mahasiswa yang terdiri dari 3 Program Studi. Variable yang digunakan terdiri dari nomor induk mahasiswa, nama mahasiswa, jenis kelamin, IPK semester 1 – 6 dan status kelulusan mahasiswa.

Tabel 3. Data Penelitian

Nim	Nama	JK	S1	52	S3	S4	55	56	Keterangan
12120026	CHARLES FEBYANTO BOUK	Laki-Laki	3.48	3.61	3.46	3.39	3.79	3.57	Tepat
12120028	DESSY LEONARTI POLLO	Perempuan	4	3.91	3.75	3.7	3.79	3.85	Tepat
12120033	ELISABETH ROS BULEN	Perempuan	3.57	3.07	2.9	3	2.76	3	Tepat
12120034	ERVINA ROSVITA MEO DASO	Perempuan	3.13	2.74	2.7	2.8	2.62	3	Tepat
12120035	EVANTRI ABAOK	Laki-Laki	2.91	2.93	2.87	2.84	2.6	2.79	Tepat
12120037	FAHRISAL ATU	Laki-Laki	2.87	2.93	2.97	3.03	2.8	2.9	Tidak Tepat
12120038	FAIZAL MAIMAN	Laki-Laki	3	3.04	2.96	2.97	2.71	2.7	Tidak Tepat
12120039	FIRMANSYAH JWIYONO	Laki-Laki	3.35	3.37	3.31	3.28	2.95	3.25	Tepat
12120041	FRANSISKUS ZAVERIUS GANGGUR	Laki-Laki	3.26	3.04	3.03	2.91	2.17	3	Tidak Tepat
12120042	FREDIRIKUS MANEK	Laki-Laki	3.48	3.52	3.39	3.31	3.12	3.28	Tepat

Untuk mendapatkan data yang berkualitas, beberapa teknik yang dilakukan adalah sebagai berikut [14].

- 1. Data validation, untuk mengidentifikasi dan menghapus data yang ganjil (outlier/noise), data yang tidak konsisten, dan data yang tidak lengkap (missing value). Missing data terlihat
- Data integration and Transformation, untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini bernilai kategorikal. Data ditransformasikan ke dalam software RapidMiner.
- 3. Data size reduction and dicrtization, untuk memperoleh data set dengan jumlah atribut dan record yang lebih sedikit tetapi bersifat informatif. Dalam penelitian ini atribut yang tidak relevan seperti nim, nama, jurusan, indeks prestasi semester lima, enam, tujuh dan delapan dihapuskan.

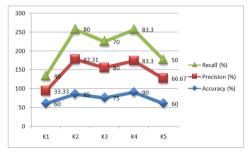
Dalam penelitian ini akan dilakukan eksperimen dengan menggunakan metode klasifikasi data mining K-Nearest Neighbor terhadap data mahasiswa yang terkait dengan kelulusan mahasiswa. Data akan diolah dengan menggunakan algoritma K-Nearest Neighbor dan menghasilkan model, maka terhadap model yang dihasilkan tersebut dilakukan pengujian menggunakankan validation, k-fold cross

kemudian dilakukan evaluasi dan validasi hasil dengan *confusion matrix* dan kurva ROC.

## 4.3 Pengujian Algoritma

#### a. Cross Validation

Dalam penelitian ini digunakan 5 Fold Validation dimana 100 record data training dibagi secara random kedalam 5 bagian perbandingan yang sama kemudian error rate dihitung bagian demi bagian, selanjutnya hitung rata-rata seluruh error rate untuk mendapatkan error rate secara keseluruhan. Data diuii tingkat eror-nva menggunakan teknik cross validation dengan 5 fold dan terdapat 20 record fold ke-1 sampai ke-5. Hasil pengujian menunjukkan persentase precision data terendah berada pada K1 yaitu 33,33% dan akurasi masing-masing mencapai 60% dan recall 40%.



Gambar 4. Hasil proses 5 Folf Cross Validation

### b. Confusion Matriks

Perhitungan kedekatan kasus lama pada data training dengan kasus baru pada data testing, diketahui dari 100 record data, 46 diklasifikasikan Tepat, 11 data diklasifikasi tetap tetapi ternyata tidak, 36 data class tidak diklasifikasi sesuai, dan 7 data diklasifikasi tidak ternyata tepat. Tingkat akurasi penerapan algoritma k-nearest neighbor ini sebesar 82%, precision 77,14% dan recall 83,50%.

Tabel 4. Confusion Matriks Untuk K-NN

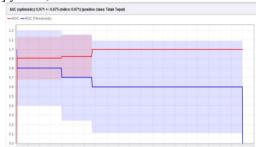
accuracy: 82.00% +1. 11.66% (mikro: 82.00%)							
	true Tepat	true Tidak Tepat	class precision				
pred. Tepat	46	7	86.79%				
pred. Tidak Tepat	11	36	76.60%				
class recall	80.70%	83.72%					

# c. Kurva ROC

Kurva ROC digunakan untuk menunjukan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*. ROC merupakan grafik dua dimensi dengan *false positive* sebagai garis horizontal dan true positive sebagai garis vertical. Hasil perhitungan divisualisasikan dengan kurva ROC.

Terdapat dua titik garis pada kurva ROC yaitu garis true negative rate specificity dan garistrue negative rate sensitivity. Specificity didefenisikan sebagai kemampuan alat tes untuk mendeteksi data negatif sedangkan sensitivity didefenisikan sebagai kemampuan alat tes untuk mendeteksi data positif. Garis specificity juga

disebut sebagai batas atas normal dan garis sensitivity disebut sebagai garis batas bawah normal, tingkat kehandalan diperoleh dengan menghitung luas area dibawah garis batas normal atas kurva. Berdasarkan pengelompokkan di atas maka dapat disimpukan bahwa metode k-nearest neighbor termasuk klasifikasi baik untuk prediksi kelulusan mahasiswa STIKOM Uyelindo Kupang karena memiliki nilai AUC (Area Under Curva) antara 0.90-1.00 kategori *Excellent classification* [5] yaitu 0,971.



Gambar 5. Grafik ROC Curve Algoritma KNearest Neighbor

#### 5. KESIMPULAN

Penelitian ini dilakukan dengan mengimplementasikan model algoritma k-nearest neighbor untuk prediksi kelulusan mahasiswa STIKOM Uyelindo Kupang. Untuk mencari data yang berkualitas, dilakukan preprosesing sebelum diterapkan ke dalam algoritma. Kedekatan antara kasus baru dengan kasus lama dilakukan untuk menentukan pada kelas mana kasus baru akan diklasifikasikan. Dengan membangun sistem pendukung keputusan untuk memprediksi kelulusan mahsiswa yang ditentukan atau dianalisis dengan menggunakan algoritma k-nearest neighbor. Ada lima variable yang digunakan yaitu nim, nama, jenis kelamin, IPK semester 1 sampai – 6 dan ketepatan lulusan mahasiswa.

Untuk mengukur kinerja algoritma knearest neighbor digunakan metode *Cross Validation, Confusion Matrix* dan *Kurva ROC*, dalam penelitian ini menggunakan 5-fold cross validation untuk prediksi kelulusan mahasiswa. Dari 100 *record dataset* mahsiswa lulusan STIKOM Uyelindo Kupang diperoleh tingkat accuracy mencapai 82% dan termasuk klasifikasi sangat baik karena memiliki nilai AUC berada diantara 0,90-1,00, yaitu 0,971,

Maka dengan demikian dapat disimpulkan tingkat akurasi pengujian model kelulusan mahasiswa dengan menggunakan algoritma K-Nearest Neighbor (K-NN) dipengaruhi oleh jumlah klastering data. Akurasi dan nilai AUC paling tinggi dari 5-fold validation adalah dengan mengklaster data k=4 dengan nilai accuracy sebesar 90%.

#### DAFTAR PUSTAKA

- [1] Azwar. (2004). Penyusunan Skala Psikologi. Yogyakarta: Pustaka Pelajar.
- [2] Buku VI Matriks Penilaian Instrumen Akreditasi Program Studi Badan Akreditasi Nasional Perguruan Tinggi. (2008).
- [3] Bramer, M., 2007. Principles of Data Mining, Springer, London.
- [4] Ernastuti, S. &. (2010). Graduation Prediction of Gunadarma University Students Using Algorithm and Naïve Bayes C4.5 Algoritmh
- [5] Gorunescu, F., 2011. Data Mining: Concepts, Models, and Techniques. Verlag Berlin Heidelberg, Springer.
- [6] Han, J., Kamber, M., 2006. Data Mining Concept and Tehniques, San Fransisco, Morgan Kauffman.
- [7] Kohavi, R., 1995. A study of cross validation and bootstrap for accuracy estimation and model selection. Proceedings of the International JointConference on Articial Intelligence, vol. 2, 1137-1143.
- [8] Liao., 2007. Recent Advances in Data Mining of Enterprise Data Algoritms and Application. World Scientific Publishing, Singapore.
- [9] M. Mohankumar, S. Amuthakkani and G. Jeyamala, "Comparative Analysis of Decision Tree Algorithms for The Prediction of Eligibility of A Man for Availing Bank Loan," International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST), vol. 2, no. 15, pp. 360-366, 2016.
- [10] Ogor. (2007). Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques.
- [11] Oyelade, O. &. (2010). Application of kmeans
  Clustering algorithm for predicting of
  Students Academic Performace.
  International Journal of Computer Science
  and Information Security, 292-295.
- [11] Saaty, T.L., 1980. The Analytic Hierarchy Process, McGraw-Hill, New York
- [12] Santoso, B. (2007). Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis (1 ed.). Yogyakarta: Graha Ilmu.
- [13] S. D. Jadhav and H. Channe, "Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques," International Journal of Science and Research (IJSR), vol. 5, no. 1, pp. 1842-1845, January 2016.
- [14] Vercellis. (2009). Business Intelligence: Data Mining and Optimization for Decision Making Decision Making. John Willey & Sons Inc: Southern Gate.
- [15] Vrettos, K. &. (2009). Sentivity Analysis of Neural Network for Identifying the Factors

for Collage Students Success. World Congress on Computer Science and Information Engineering. (978-0-7695-3507-4).

[16] Yingkuachat, P. K. (2007). An Application Probabilitic Model to the Prediction of Student Graduation Using Bayesian Belief Network. ECTI Transaction on Computer and Technology, 63-71.