

PREDIKSI KELULUSAN MAHASISWA MAGISTER TEKNIK INFORMATIKA UNIVERSITAS AMIKOM YOGYAKARTA MENGGUNAKAN METODE K-NEAREST NEIGHBOR

Eri Sasmita Susanto¹, Kusirini², Hanif Al Fatta³

^{1,2,3}Magister Teknik Informatika Universitas Amikom Yogyakarta

Jl. Ring Road Utar, Condong Catur, Depok, Sleman, Yogyakarta 55281

¹92.eriss@gmail.com, ²kusirini@amikom.ac.id, ³hanif.a@amikom.ac.id

INTISARI

Penelitian ini difokuskan untuk mengetahui uji kelayakan prediksi kelulusan mahasiswa Universitas AMIKOM Yogyakarta. Dalam hal ini penulis memilih algoritma *K-Nearest Neighbors* (K-NN) karena *K-Nearest Neighbors* (K-NN) merupakan algoritma yang bisa digunakan untuk mengolah data yang bersifat numerik dan tidak membutuhkan skema estimasi parameter perulangan yang rumit, ini berarti bisa diaplikasikan untuk dataset berukuran besar.

Input dari sistem ini adalah Data sampel berupa data mahasiswa tahun 2014-2015. pengujian pada penelitian ini menggunakan dua pengujian yaitu data testing dan data training. Kriteria yang digunakan dalam penelitian ini adalah, IP Semester 1-4, capaian SKS, Status Kelulusan. Output dari sistem ini berupa hasil prediksi kelulusan mahasiswa yang terbagi menjadi dua yaitu tepat waktu dan kelulusan tidak tepat waktu.

Hasil pengujian menunjukkan bahwa Berdasarkan penerapan $k=14$ dan $k\text{-fold}=5$ menghasilkan performa yang terbaik dalam memprediksi kelulusan mahasiswa dengan metode K-Nearest Neighbor menggunakan indeks prestasi 4 semester dengan nilai akurasi= 98,46%, precision= 99.53% dan recall =97.64%.

Kata kunci: Algoritma *K-Nearest Neighbors*, Prediksi Kelulusan, Data Testing, Data Training

ABSTRACT

This research is focused on knowing the feasibility test of students' graduation prediction of AMIKOM University Yogyakarta. In this case the authors chose the K-Nearest Neighbors (K-NN) algorithm because K-Nearest Neighbors (K-NN) is an algorithm that can be used to process data that is numerical and does not require complicated repetitive parameter estimation scheme, this means it can be applied for large datasets.

The input of this system is the sample data in the form of student data from 2014-2015. test in this research use two test that is data testing and training data. The criteria used in this study are, IP Semester 1-4, achievement of SKS, Graduation Status. The output of this system in the form of predicted results of student graduation which is divided into two that is timely and graduation is not timely.

The result of the test shows that based on the application of $k = 14$ and $k\text{-fold} = 5$, the best performance in predicting the students' graduation using K-Nearest Neighbor method uses 4 semester achievement index with accuracy value = 98,46%, precision = 99.53% and recall = 97.64%.

Keywords: *K-Nearest Neighbors Algorithm, Graduation Prediction, Testing Data, Training Data*

I. PENDAHULUAN

Dalam sistem pendidikan, mahasiswa merupakan aset penting yang dimiliki oleh institusi penyelenggara pendidikan. Lama masa studi mahasiswa dalam menyelesaikan studinya merupakan salah satu elemen penilaian akreditasi. Angka kelulusan yang tinggi merupakan salah satu indikator keberhasilan dalam proses penyelenggaraan pendidikan di perguruan tinggi. Oleh karena itu perlu adanya pemantauan serta evaluasi terhadap ketepatan kelulusan masa studi mahasiswa.

Teknik dalam memanfaatkan data dalam jumlah yang besar untuk memperoleh informasi berharga yang belum diketahui agar dapat dimanfaatkan untuk pengambilan keputusan dikenal dengan istilah Data Mining. Data mining memanfaatkan pengalaman atau bahkan kesalahan di masa lalu untuk meningkatkan kualitas dari model maupun hasil analisisnya, salah satunya dengan kemampuan yang dimiliki teknik data mining yaitu klasifikasi.

Ada banyak algoritma pada data mining yang dapat digunakan dalam melakukan klasifikasi diantaranya seperti Decision Trees,

Naïve, Neural Networks Nearest Neighbor, Support Vector Machines dan lainnya.

Nearest Neighbor adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pencocokan bobot dari sejumlah fitur yang ada (Kusrini & Luthfi, 2009).

K-Nearest Neighbors (KNN) merupakan algoritma yang bisa digunakan untuk mengolah data yang bersifat numerik dan tidak membutuhkan skema estimasi parameter perulangan yang rumit, ini berarti bisa diaplikasikan untuk dataset berukuran besar (Panoto, Utami, & YS, 2017).

Menurut Pandi algoritma K-Nearest Neighbors (KNN) sebuah data baru diklasifikasikan berdasarkan jarak data baru tersebut dengan tingkat kemiripan data baru terdekat terhadap data pola. Jumlah data tetangga terdekat ditentukan dan dinyatakan dengan k. Penentuan nilai k terbaik dapat ditentukan dengan optimasi parameter, misalnya dengan menggunakan k-Fold Cross Validation yang merupakan salah satu metode yang digunakan untuk mengetahui rata-rata keberhasilan dari suatu sistem dengan cara melakukan perulangan dengan mengacak atribut masukan sehingga sistem tersebut teruji untuk beberapa atribut input yang acak. Nilai k=1 pada algoritme K-Nearest Neighbors (KNN) bukan satu-satunya faktor yang menentukan pemilihan k-Optimal, karena pada k=1 perhitungan jarak dan penentuan hasil prediksi dipengaruhi oleh perhitungan data. Untuk memutuskan satu nilai k terbaik atau k-Optimal maka dilihat dari tingkat akurasi pada hasil pencarian menggunakan metode k-Fold Cross Validation (Banjarsari, Budiman, & Farmadi, 2015).

Berikut ini merupakan uraian singkat beberapa penelitian terdahulu yang relevan untuk memperkuat alasan dan motivasi perlunya diadakan penelitian seperti yang diajukan.

Penelitian (Banjarsari, Budiman, & Farmadi, 2015) dengan judul “Penerapan K-Optimal Pada Algoritma KNN untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer FMIPA UNLAM Berdasarkan IP Sampai Dengan Semester 4” dalam penelitian tersebut disimpulkan nilai k-Optimal pada algoritma KNN untuk prediksi kelulusan tepat waktu mahasiswa berdasarkan indeks prestasi sampai dengan semester 4 adalah k=5. Berdasarkan 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%.

Berdasarkan uraian latar belakang masalah yang telah dipaparkan sebelumnya maka penulis bermaksud untuk melakukan pengujian k-fold cross validation pada algoritma K-Nearest Neighbors dalam memprediksi kelulusan mahasiswa pada mahasiswa Magister Teknik Informatika Universitas AMIKOM Yogyakarta dengan menggunakan atribut indeks prestasi semester 1 hingga 4, guna untuk mengetahui tingkat performa yang didapatkan berdasarkan penerapan antara metode K-Nearest Neighbor.

Tujuan dari penelitian ini adalah :

1. Untuk mengetahui uji kelayakan prediksi kelulusan mahasiswa Universitas AMIKOM Yogyakarta dengan menggunakan Metode *K-Nearest Neighbor* (KNN).
2. Sebagai syarat kelulusan di program Magister Teknik Informatika Universitas AMIKOM Yogyakarta

II. METODOLOGI PENELITIAN

Metode penelitian adalah langkah – langkah yang digunakan agar *hipotesis* penelitian dapat teruji secara *empirik* (Sugiyono, 2003). Metode penelitian yang digunakan oleh penulis adalah Model Penelitian Tindakan (*action research*).

Desain penelitian ini mengikuti siklus penelitian *action research*. Langkah – langkah *action research* menurut Baskerville dalam Lee(2007) yaitu *diagnosing*, *action planning*, *action taking*, *evaluating*, dan *reflection* :

1. Diagnosing

Tahap ini merupakan proses identifikasi masalah-masalah yang menjadi dasar penelitian tentang implementasi sistem pendukung keputusan prediksi kelulusan mahasiswa. Pada tahap ini juga dilakukan proses pengumpulan data, data yang dikumpulkan adalah data hasil observasi dan dokumentasi serta dari literatur.

2. Action Planning

Tahap ini merupakan penyusunan rencana tindakan yang tepat guna menyelesaikan permasalahan pada penelitian ini, tindakan yang direncanakan meliputi penerapan kaidah prediksi kelulusan mahasiswa dengan menggunakan metode *K-Nearest Neighbor*.

3. Action Taking

Tahap ini merupakan tahap implementasi pengujian menggunakan *Rapidminer*.

4. Evaluation

Tahap ini merupakan fase lanjutan terhadap tujuan data mining. Evaluasi dilakukan secara mendalam dengan tujuan agar hasil pada tahap pemodelan sesuai dengan sasaran yang diinginkan.

Metode Pengumpulan Data yang digunakan dalam penelitian adalah wawancara, *observasi*, dan dokumen. Pengumpulan data dilakukan untuk memperoleh informasi yang dibutuhkan dalam rangka mencapai tujuan penelitian. Metode pengumpulan data dalam penelitian ini adalah sebagai berikut :

1. Observasi

Merupakan metode pengumpulan data dengan cara melakukan pengamatan secara langsung pada objek yang diteliti yaitu pada kelulusan mahasiswa yang meliputi data kualitatif dan data kuantitatif dilakukan observasi untuk dianalisis dengan *algoritme K-Nearest Neighbor*. Pengamatan data peninjauan ke Universitas AMIKOM Yogyakarta untuk melihat dan mengetahui kondisi pembelajaran secara langsung sehingga mengetahui permasalahan yang terjadi.

2. Dokumen

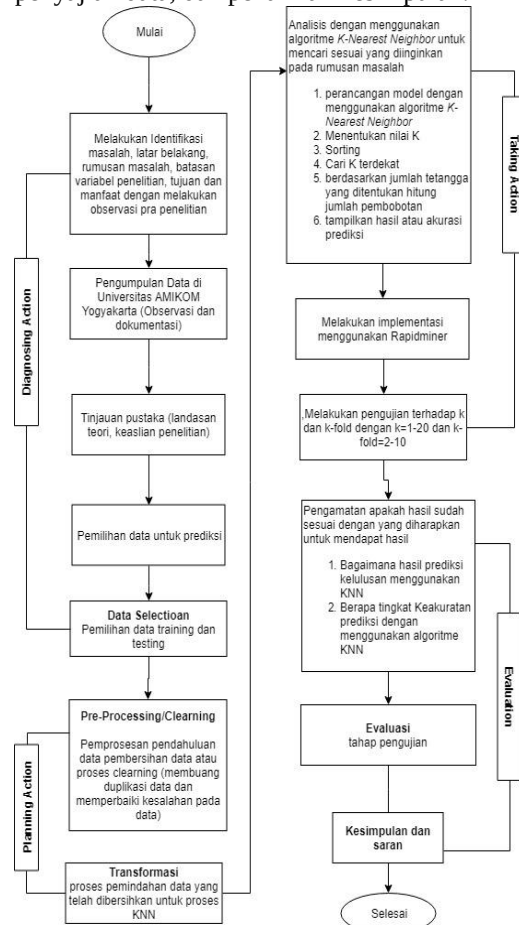
Pengambilan data melalui dokumen tertulis maupun elektronik dari lembaga atau institusi sesuai dengan yang diangkat dalam tesis ini yaitu di Universitas AMIKOM Yogyakarta. Dokumen diperlukan untuk mendukung keperluan data yang lain.

3. Studi pustaka

Mengetahui informasi secara teori mengenai pokok permasalahan dan teori pendukung yang digunakan sebagai dasar pemikiran dalam membahas permasalahan yang ada, membaca referensi yang berhubungan dengan metode *K-Nearest Neighbor*.

Metode analisis data yang digunakan penulis dalam penelitian ini adalah menggunakan Algoritma *K-Nearest Neighbor*. Dalam algoritma *K-Nearest Neighbor* memiliki prinsip kerja mencari jarak terpendek antara data yang akan dievaluasi dengan K terdekat dalam data pelatihannya. Algoritma *K-Nearest Neighbor* memiliki tujuan mengklasifikasi objek baru berdasarkan atribut, algoritma *K-Nearest Neighbor* juga menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari *queri instance* yang baru.

Menurut Miles dan Huberman (2000) dalam Sugiyono (2008) tahap – tahap analisis data yang terjadi dilapangan secara objectif, kemudian hasil pencatatan tersebut dikelompokkan atau dikategorikan secara rinci sesuai dengan kata kunci yang muncul. Setelah itu, peneliti akan merangkum hasil dari pencatatan materi tersebut untuk dipilih dan kemudian difokuskan pada hal – hal yang penting. Selanjutnya, pada setiap kategorinya diberi pernyataan yang menunjukkan hubungan antar kategori, sehingga dapat ditarik sebuah kesimpulan. Secara operasional analisis data kualitatif dilakukan dengan tiga langkah sistematis secara jalin menjalin (Miles, 1992:19), yaitu Reduksi data, penyajian data, dan penarikan kesimpulan.



Gambar 1. Alur Penelitian

III. HASIL DAN PEMBAHASAN

1. Pengumpulan Data

Pengumpulan data dilakukan dengan pengambilan pada Universitas AMIKOM Yogyakarta. Data yang didapat merupakan data mahasiswa program Magister Teknik Informatika tahun 2014 dan 2015. Pengambilan data disesuaikan dengan

kebutuhan pada penelitian. Data yang diperoleh 8015 data.

2. Data Selection

Tahap pemilihan atau penyeleksian terhadap data dilakukan untuk mendapatkan data yang sesuai dengan kebutuhan dalam prediksi kelulusan mahasiswa. Dalam prediksi kelulusan mahasiswa dibutuhkan data mahasiswa yang telah lulus. Data yang dapat digunakan adalah data mahasiswa tahun kelulusan 2014 dan 2015 yang didapat dari pengumpulan data sebelumnya. Pada tahap pemilihan data maka diperoleh sebanyak 388 data. 388 data yang telah diperoleh akan dijadikan sebagai data *training*.

Data yang digunakan merupakan data awal yang terdiri dari NIM, Kode, Matakuliah, Nilai, Tahun Ajaran dan Semester. Selanjutnya data akan mengalami perubahan bentuk yang terdiri dari *field* NIM, Indeks Prestasi Semester 1 (IPS 1), Indeks Prestasi Semester 2 (IPS 2), Indeks Prestasi Semester 3 (IPS 3) dan Indeks Prestasi Semester 4 (IPS 4), Jumlah SKS dan Klasifikasi.

Indeks Prestasi Semester merupakan transformasi nilai huruf menjadi angka yang di akumulasi dengan jumlah SKS setiap mata kuliahnya. Klasifikasi berdasarkan jumlah SKS yang telah ditempuh. Jika jumlah SKS dalam 4 semester lebih sama dengan 42 maka "TEPAT" yang artinya lulus dengan tepat waktu sedangkan jika jumlah SKS yang telah ditempuh dalam 4 semester kurang dari 42 maka "TIDAK TEPAT" yang artinya lulus dengan tidak tepat waktu.

3. Data Pre-processing/ Clearing

Tahap ini merupakan proses pembersihan terhadap data yang dilakukan untuk memastikan data yang diperoleh sebelumnya dapat digunakan serta bebas dari duplikasi, kesalahan dan *validation rules* sudah sesuai.

4. Transformasi Data

Tahap ini merupakan proses untuk melakukan perubahan bentuk tabel terhadap data yang telah dipilih, sehingga data yang dipilih sesuai dengan proses yang dilakukan. Pada Tabel 4.1 merupakan tabel yang akan digunakan untuk Prediksi Kelulusan Mahasiswa dengan variable NIM, Indeks Prestasi Semester 1 (IPS 1), Indeks Prestasi Semester 2 (IPS 2), Indeks Prestasi Semester 3 (IPS 3) dan Indeks Prestasi Semester 4 (IPS 4) Jumlah SKS dan Klasifikasi.

TABEL I.
TRANSFORMASI DATA TRAINING

No	NIM	IPS 1	IPS 2	IPS 3	IPS 4	JML SKS	Klasifikasi
1	14.51.0532	3.8	2.95	1.82	4	42	TEPAT
2	14.51.0534	3.6	3.88	1.82	0	36	TIDAK TEPAT
3	14.51.0535	3.6	4	1.82	0	36	TIDAK TEPAT
4	14.51.0536	3.6	3.33	1.18	0	38	TIDAK TEPAT
5	14.51.0537	3.6	4	1.55	4	42	TEPAT
6	14.51.0539	3.8	3.69	1.82	0	36	TIDAK TEPAT
7	14.51.0545	4	4	1.82	4	42	TEPAT
8	14.51.0549	3.6	3	1.82	0	36	TIDAK TEPAT
9	14.51.0553	3.6	3.19	4	0	33	TIDAK TEPAT
10	14.51.0555	3.6	3	1.82	4	42	TEPAT
11	14.51.0556	3.8	3.81	1.82	0	36	TIDAK TEPAT
12	14.51.0560	3.8	2.95	1.21	0	39	TIDAK TEPAT
13	14.51.0561	3.6	3.17	1.82	0	36	TIDAK TEPAT
14	14.51.0563	3.6	3.5	4	0	38	TIDAK TEPAT
15	14.51.0566	3.8	3.5	1.55	0	36	TIDAK TEPAT
16	14.51.0568	3.6	3.19	1.82	0	36	TIDAK TEPAT
17	14.51.0572	3.6	3.13	1.55	0	36	TIDAK TEPAT
18	14.51.0573	3.8	3.63	1.82	4	42	TEPAT
19	14.51.0576	3.2	3.37	1	0	36	TIDAK TEPAT
20	14.51.0580	3.8	2.68	1.82	4	42	TEPAT
21	14.51.0585	3.4	3.56	1.82	0	36	TIDAK TEPAT
22	14.51.0586	3.4	3.25	1.82	0	36	TIDAK TEPAT
23	14.51.0588	3.4	3.69	1.82	3	42	TEPAT
24	14.51.0589	3.4	3.81	4		42	TEPAT
25	14.51.0593	3.8	4	1.82	0	36	TIDAK TEPAT
26	14.51.0600	3.4	3.25	1.82	0	36	TIDAK TEPAT
27	14.51.0602	3.8	3.19	1.82	0	36	TIDAK TEPAT
28	14.51.0609	3.6	4	4		42	TEPAT
29	14.51.0611	3.6	3.31	1.55	0	36	TIDAK TEPAT
30	14.51.0612	3.6	3.5	1.82	0	36	TIDAK TEPAT
31	14.51.0613	3.8	3.56	1.55	0	36	TIDAK TEPAT
32	14.51.0614	3.8	3.69	1.82	0	36	TIDAK TEPAT
33	14.51.0616	3.6	3.56	4		42	TEPAT

5. Pemodelan

Pemodelan data merupakan tahapan yang secara langsung melibatkan teknik data mining. Teknik pemilihan data mining pada Prediksi Kelulusan Mahasiswa Magister Teknik Informatika yaitu adalah algoritme *K-Nearest Neighbor* atau yang dikenal dengan singkatan KNN

6. Perhitungan Manual K-Nearest Neighbor

Tahapan perhitungan dalam Prediksi Kelulusan Mahasiswa Magister Teknik Informatika pada Universitas AMIKOM Yogyakarta menggunakan metode *K-Nearest Neighbor* adalah sebagai berikut ini:

1. Menetapkan data training

Data training yang digunakan merupakan data berdasarkan pada Tabel 4.1. *Data training* merupakan data yang sudah memiliki Klasifikasi. Klasifikasi ditetapkan berdasarkan masa studi, dimana jika masa studi 4 semester maka "TEPAT" dan jika masa studi lebih dari 4 semester maka "TIDAK TEPAT".

2. Menetapkan data testing

Data Testing merupakan data IP mahasiswa yang akan dilakukan klasifikasi, apakah lulus dengan masa studi ≥ 4 semester atau masa studi ≤ 4 semester.

TABEL II.
DATA TESTING

NIM	IPS 1	IPS 2	IPS 3	IPS 4	SKS	Klasifikasi
16.51.0796	3.8	3.5	1.82	0	36	?

3. Menetapkan nilai k

Dalam perhitungan data testing dengan NIM 16.51.0796 menetapkan k=1 atau dapat menetapkan dengan nilai k lainnya

4. Perhitungan jarak dengan rumus

Perhitungan dengan rumus

$$D = \sqrt{\sum_{i=1}^p (X_{2i} - X_{1i})^2}$$

TABEL III.

PERHITUNGAN NILAI JARAK ANTAR VARIABEL

No.		D	Klasifikasi
1	$= \sqrt{(3.8 - 3.8)^2 + (2.95 - 3.5)^2 + (1.82 - 1.82)^2 + (4 - 0)^2}$	4.037635	TEPAT
2	$= \sqrt{(3.6 - 3.8)^2 + (3.88 - 3.5)^2 + (1.82 - 1.82)^2 + (4 - 0)^2}$	0.429418	TIDAK TEPAT
3	$= \sqrt{(3.6 - 3.8)^2 + (4 - 3.5)^2 + (1.82 - 1.82)^2 + (0 - 0)^2}$	0.538516	TIDAK TEPAT
4	$= \sqrt{(3.6 - 3.8)^2 + (3.33 - 3.5)^2 + (1.82 - 1.82)^2 + (-2)^2}$	0.691737	TIDAK TEPAT
....			
...			
....			
138	$= \sqrt{(3.6 - 3.8)^2 + (4 - 3.5)^2 + (4 - 1.82)^2 + (0 - 2)^2}$	2.245529	TEPAT
...			
...			
388	$= \sqrt{(3.6 - 3.8)^2 + (3.5 - 3.5)^2 + (2.06 - 1.82)^2 + (0 - 0)^2}$	0.31241	TEPAT

Dari keseluruhan hasil perhitungan jarak dengan data training yang berjumlah 388 data, hasil perhitungan nilai jarak (D) diurutkan mulai terkecil hingga terbesar. Setelah diurutkan, dilihat mayoritas klasifikasi yang muncul dari perhitungan jarak yang terkecil atau yang pertama sampai dengan yang terakhir. Dapat dilihat pada Tabel 4.4 mayoritas klasifikasi pada data training dengan NIM 16.51.0796 yang muncul adalah “TIDAK TEPAT”, sehingga data testing yang ingin diklasifikasikan termasuk kedalam klasifikasi “TIDAK TEPAT” atau tidak tepat waktu.

7. Implementasi Dengan Rapidminer

Pada algoritma *K-Nearest Neighbor* dilakukan pengujian terhadap jumlah klasifikasi dengan nilai k=1 hingga k=20 dan *k-fold cross validation* 2-10. Pengujian klasifikasi dilakukan untuk mendapatkan tingkat akurasi, precision dan recall algoritma *K-Nearest Neighbor* yang terbaik

TABEL IV.

HASIL PERFORMA PENGUJIAN K-NEAREST
NEIGHBOR

K-fold	K	Akurasi	Precision	Recall
2	1	96.13%	95.80%	97.16%
2	2	96.39%	99.00%	94.30%
2	3	97.94%	98.12%	98.10%
2	4	98.20%	99.52%	97.16%
2	5	97.94%	98.57%	97.63%
2	6	98.20%	99.05%	97.63%
2	7	98.20%	99.05%	97.63%
2	8	98.20%	99.05%	97.63%
2	9	97.94%	98.58%	97.63%
2	10	98.45%	99.52%	97.63%
2	11	98.20%	99.05%	97.63%
2	12	98.20%	99.05%	97.63%
2	13	97.94%	98.58%	97.63%
2	14	97.94%	98.58%	97.63%
2	15	97.68%	98.10%	97.63%
2	16	97.94%	98.58%	97.63%
2	17	97.94%	98.58%	97.63%
2	18	98.20%	99.05%	97.63%
2	19	97.94%	98.58%	97.63%
2	20	98.20%	99.05%	97.63%
3	1	96.91%	96.71%	97.63%
3	2	96.65%	99.49%	94.31%
3	3	97.68%	97.66%	98.10%
3	4	98.19%	99.03%	97.63%
3	5	97.94%	98.12%	98.10%
3	6	97.68%	98.11%	97.63%
3	7	97.16%	97.20%	97.63%
3	8	97.94%	98.58%	97.63%
3	9	97.68%	98.11%	97.63%
3	10	97.94%	98.58%	97.63%
3	11	97.68%	98.11%	97.63%
3	12	97.94%	98.58%	97.63%
3	13	97.68%	98.11%	97.63%
3	14	97.94%	98.58%	97.63%
3	15	97.68%	98.11%	97.63%
3	16	97.68%	98.11%	97.63%
3	17	97.68%	98.11%	97.63%
3	18	97.94%	98.58%	97.63%

Berdasarkan pengujian terhadap akurasi, precision dan recall metode KNN dengan pengujian menggunakan k sebanyak 1-20 dan pengujian dengan k-fold sebanyak 2-10 maka didapatkan hasil yang paling maksimal (tertinggi) presentasi akurasi, precision dan recall adalah pada pengujian k-fold=5 dengan k=14 dan k=18, maka selanjutnya dilakukan pengamatan antara k=14 dan k=18 yang sering memiliki nilai terbesar antara kedua nilai k tersebut adalah pada k=14. Jadi diketahui bahwa k-optimal untuk metode KKN untuk penerapan prediksi kelulusan mahasiswa menggunakan indeks prestasi 4 semester adalah dengan k=14 dan k-fold cross

validation 5 dengan hasil akurasi= 98,46%, precision= 99.53% dan recall =97.64%.

IV. KESIMPULAN

Dari hasil pengujian prediksi kelulusan mahasiswa dengan metode *K-Nearest Neighbor* berdasarkan nilai 4 semester menggunakan *Rapidminer* didapatkan kesimpulan sebagai berikut ini:

1. K-optimal dan *k-fold cross validation* dalam prediksi kelulusan mahasiswa dengan metode *K-Nearest Neighbor* berdasarkan nilai 4 semester adalah $k=14$ dan $k\text{-fold}=5$.
2. Berdasarkan penerapan $k=14$ dan $k\text{-fold}=5$ menghasilkan performa yang terbaik dalam memprediksi kelulusan mahasiswa dengan metode *K-Nearest Neighbor* menggunakan indeks prestasi 4 semester dengan nilai akurasi= 98,46%, precision= 99.53% dan recall =97.64%.

REFERENSI

- [1] Hermawati, F. A. (2013). Data Mining, Edisi 1. Yogyakarta: ANDI.
- [2] Kusriani, & Luthfi, E. T. (2009). Algoritma Data Mining. Yogyakarta: ANDI.
- [3] Sugiono, (2015). Metode Penelitian Kuantitatif, Kualitatif dan R&D. Bandung: ALFABETA
- [4] Turban, E., & dkk. (2005). Decision Support System and Intelligent Systems. Yogyakarta: Andi Offset.
- [5] Widodo, P. P., Handayanto, R. T., & Herlawati. (2013). Penerapan Data Mining Dengan Matlab. Bandung: Rekayasa Sains.

PUSTAKA MAJALAH, JURNAL ILMIAH ATAU PROSIDING

- [1] Banjarsari, M. A., Budiman, H. I., & Farmadi, A. (2015). Penerapan K-Optimal Pada Algoritma Knn untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam. Kumpulan jurnal Ilmu Komputer (KLIK) , ISSN: 2406-7857, Vol. 02, No.02 September 2015, Hal. 50-64.
- [2] Foeh, R. S., Pandie, E. S., & Mola, S. A. (2017). Aplikasi Prediksi Status Kelulusan Mahasiswa Menggunakan K-Nearest Neighbor. Jurnal Informatika dan Komputer - ICON , ISSN 2337-7631, Vol. 3 No. 2, Oktober 2015, Hal. 66 ~ 77.
- [3] Ginting, S. L., Zarman, W., & Darmawan, A. (2014). Teknik data Mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritme K-Nearest Neighbor. Jurnal Teknik Komputer Unikom – Komputika , Volume 3, No.2 - 2014, Hal-29-34.
- [4] Masse, F. A., & Sudarmawan, I. W. (2017). Analisis Tingkat Kelulusan Mahasiswa Sekolah Tinggi Manajemen Informatika dan.

Jurnal Elektronik Sistem Informasi dan Komputer (JESIK) , ISSN: 2477-5290
Print, ISSN: 2502-2148 Online, Vol.3 No.2
Juli-Desember 2017, Hal. 27-37.