

Penerapan Metode K Nearest Neighbord Dalam Proses Seleksi Penerima Beasiswa

Moenawar Kholil¹⁾, Kusri²⁾, Henderi³⁾

Magister Teknik Informatika Universitas AMIKOM Yogyakarta
Jl. Ringriad Utara, Catur tunggal depok Sleman Yogyakarta
e-mail: moenawarkholil@gmail.com

Abstrak

Beasiswa merupakan hal yang sangat didambakan oleh mahasiswa. Dengan banyaknya peminat beasiswa, maka secara tidak langsung akan mengakibatkan proses seleksi yang lama. Selain subyektifitas, beberapa mahasiswa penerima beasiswa tidak dapat menyelesaikan proses perkuliahan dengan tepat waktu. Untuk mengatasi permasalahan tersebut salah satu alternatif yang diusulkan dalam penelitian ini adalah melakukan seleksi penerima beasiswa dengan menggunakan metode KNN, metode KNN ini di gunakan sebagai klasifikasi dari pendaftar beasiswa menjadi 2 kriteria klasifikasi, yaitu lulus tepat waktu dan lulus tidak tepat waktu. Dalam penelitian ini menggunakan data training sejumlah 500 data dan data testing sejumlah 125 data pendaftar beasiswa . Masing-masing data training akan digunakan sebagai data uji dari ke 125 data pendaftar, sehingga di harapkan prediksi yang dihasilkan akan akurat. Pengujian pada metode yang akan dibangun, sekaligus mengetahui berapa nilai tetangga terdekat (K) dilakukan pengujian KFold Croos Validation dengan cara membagi semua data training menjadi 10 blok data. Pada setiap blok data akan dilakukan pengujian dengan 10 skenario. Kemudian akan dihitung presentase kesamaan hasil prediksi pada masing masing blok uji. Nilai terbaik di dapati pada pada nilai K5 dengan jumlah kesamaan 90%.

Kata kunci: KNN, K Fold Cross Validation, Beasiswa

1. Pendahuluan

Beasiswa merupakan hal yang sangat didambakan oleh mahasiswa. Akan tetapi untuk mendapatkan beasiswa tersebut mahasiswa harus memiliki kriteria-kriteria tertentu yang akan dinilai oleh pihak universitas. Dengan banyaknya peminat beasiswa, maka secara tidak langsung akan mengakibatkan proses seleksi yang lama, hal ini dikarenakan data yang banyak dan prosedur penilaian masih diproses secara manual. Hal tersebut dapat juga berakibat pada kesalahan dalam menentukan penerima beasiswa. Tidak dapat dipungkiri nilai subyektifitas dalam melakukan proses seleksi penerima beasiswa tersebut. Selain subyektifitas, didapati bahwa beberapa mahasiswa penerima beasiswa tidak dapat menyelesaikan proses perkuliahan dengan tepat waktu dengan IPK yang baik. Untuk mengatasi permasalahan tersebut salah satu alternatif yang diusulkan dalam

penelitian ini adalah melakukan seleksi penerima beasiswa dengan menggunakan metode KNN, dimana metode KNN ini di gunakan sebagai klasifikasi dari pendaftar beasiswa menjadi 2 kriteria klasifikasi, yaitu lulus tepat waktu dan lulus tidak tepat waktu. Berikut beberapa uraian singkat penelitian terdahulu yang relevan untuk memperkuat alasan dan motivasi perlunya diadakan penelitian seperti yang diajukan. Penelitian menggunakan metode KNN untuk melakukan klasifikasi peserta yang akan diterima dengan cara mengklasifikasikan tinggi badan pada setiap peserta Seleksi Penerimaan Anggota Paskibraka Provinsi Bengkulu dan metode SAW digunakan untuk melakukan perbandingan [1]. Penelitian ini menggunakan metode KNN dan SAW untuk proses seleksi anggota paskibraka. Dalam sistem penyeleksian sebelumnya ditemukan kasus dimana peserta dengan tinggi badan paling maksimal pastikan lulus seleksi, sedangkan yang terjadi dilapangan, peserta yang diambil adalah peserta dengan tinggi badan yang sejajar dengan peserta lainnya sehingga tidak ada kesenjangan dalam barisan. Metode *K-Nearest Neighbor* digunakan untuk mengatasi kasus diatas yaitu dengan cara mengklasifikasikan tinggi badan pada setiap peserta. Dengan membandingkan jarak kedekatan antara data training dan data testing. Data training merupakan data peserta yang telah lulus di tahun sebelumnya sedangkan data testing merupakan data peserta yang akan diseleksi. Metode *K-Nearest Neighbor* pada penelitian ini menggunakan data keseluruhan hasil penilaian peserta seleksi setelah itu sistem akan menentukan data training dan data testing. Pengujian dilakukan dengan data sebanyak 134 data peserta seleksi tahun 2016 dengan kuota yang diterima yaitu sebanyak 56 orang peserta. Berdasarkan uji akurasi sistem, didapat bahwa hasil akurasi sistem memiliki tingkat akurasi sebesar 82,14% [1]. Penelitian selanjutnya menggunakan metode *K-Nearest Neighbor* (KNN) dan *Weighted Product* (WP) untuk penentuan siswa berprestasi [2]. Pada penelitian ini dilakukan perhitungan *K-Nearest Neighbor* dengan cara petugas memasukkan nilai k setelah dilakukan perhitungan *K-Nearest Neighbor* keluar nilai jarak *Euclidean* dilanjut perhitungan *Weighted Product* dengan mengambil data nilai dari setiap kriteria yang telah disimpan dalam tabel tabel yang terdapat pada *database*. Keluaran dari proses perhitungan *K-Nearest Neighbor* dan *Weighted Product* adalah menampilkan hasil perhitungan penentuan siswa

berprestasi. Pengujian tingkat akurasi dilakukan untuk mengetahui kesesuaian sistem dalam memberikan keluaran yang berupa urutan nama siswa berdasarkan nilai rekomendasi tertinggi. Data yang digunakan berjumlah 30 data uji dan 30 data latih. Hasil rekomendasi yang diperoleh dari perhitungan di sistem, dicocokkan dengan hasil dari sekolah. Pengujian yang dilakukan dengan cara membandingkan hasil perbandingan data pakar dengan pihak sekolah SMP Negeri 3 Mejayan dan perbandingan anantara data pakar dengan hasil perhitungan menggunakan metode *K-Nearest Neighbor* dan *Weighted Product*. Hasil tingkat Akurasi yang dihasilkan berdasarkan pengujian akurasi dengan membandingkan hasil perhitungan dari sekolah dengan perhitungan dari pakar memiliki tingkat akurasi sebesar 56.67% dan hasil dari pakar dengan hasil perhitungan metode *K-Nearest Neighbor* memiliki tingkat akurasi sebesar 76.67% [2]. Penelitian berikutnya untuk memprediksi kelulusan mahasiswa diambil dari basis pengetahuan data mahasiswa dengan menggunakan perbandingan algoritma *backpropagation neural network* dan algoritma *K-Nearest Neighbor* [3]. Penelitian ini dilakukan dengan maksud untuk memprediksi kelulusan mahasiswa diambil dari basis pengetahuan data mahasiswa. Basis pengetahuan tersebut adalah pola algoritma yang di hasilkan dari proses perbandingan algoritma *Backpropagation Neural Network* dan algoritma *K-Nearest Neighbor*. Pada penelitian ini data mahasiswa dalam kondisi acak dan tidak semua data mahasiswa akan dicari hubungannya dengan data kelulusan, hanya beberapa atribut yang berguna. Data mahasiswa yang akan dicari hubungannya meliputi jenis kelamin, usia saat mendaftar, kota asal mahasiswa, jurusan asal sekolah, agama, IPK dan fakultas. Penelitian ini mencari hubungan beberapa atribut dari data induk mahasiswa dengan tingkat kelulusan. Karena tidak semua tabel digunakan maka perlu dilakukan pembersihan data agar data yang akan diolah benar-benar relevan dengan yang dibutuhkan. Setelah melakukan proses pembersihan data, didapatkan data lulusan dan data mahasiswa dengan tidak menampilkan data NPM (Nomor Pokok Mahasiswa) dan nama dari mahasiswa dengan alasan merupakan data pribadi. Dari jumlah sampel sebanyak 382 data yang terbagi atas data training dan data testing, dengan pembagian 75% sebagai data training dan 25% data testing. Jumlah data training sebanyak 286 data lulusan dan jumlah data testing sebanyak 96 data mahasiswa. Pengujian metode dilakukan dengan cara *confusion matrix* dan *k-fold cross validation*, untuk menguji metode yang diuji menggunakan tools RapidMiner 5.3. Dari proses pengujian dengan *confusion matrix* diketahui bahwa metode algoritma *K-Nearest Neighbor* menghasilkan nilai akurasi tertinggi yang diperoleh dengan nilai jarak (k-100) serta menggunakan 7 variabel terhadap data baru yaitu sebesar 97,90 %. Penerapan rule dari algoritma *K-Nearest Neighbor* yang digunakan dalam klasifikasi kelulusan mahasiswa kedalam aplikasi *knowledge based system* menggunakan CLIPS, dapat membantu dalam

proses prediksi kelulusan mahasiswa [3]. Penelitian berikutnya menyelesaikan permasalahan bagaimana menerapkan metode *K-Nearest Neighbor* untuk mengklasifikasikan pemilihan kategori judul skripsi mahasiswa berdasarkan bidang minat mereka dan bidang keahlian serta bagaimana menerapkan metode *K-Nearest Neighbor* dengan tepat [4]. Tujuan dari penelitian ini adalah untuk menerapkan metode *K-Nearest Neighbor* untuk klasifikasi kategori judul skripsi berdasarkan bidang minat dan bidang keahlian masing-masing mahasiswa dalam Program Studi Teknik Informatika Fakultas Teknik Universitas Nusantara PGRI Kediri. Sistem klasifikasi skripsi mahasiswa dalam penelitian ini dibuat dengan menggunakan metode *K-Nearest Neighbor*. [4] Metode *K-Nearest Neighbor* dipilih karena merupakan metode klasifikasi yang tepat yang digunakan untuk mengklasifikasikan kategori judul skripsi berdasarkan nilai mahasiswa pada mata kuliah tertentu dan berdasarkan minat mahasiswa dalam kategori bidang tertentu. [4]

Berdasarkan beberapa penelitian sebelumnya dan latar belakang masalah diatas, maka peneliti melakukan sebuah penelitian dengan menggunakan metode metode KNN untuk mengetahui penerima beasiswa, agar mahasiswa yang menerima beasiswa di harapkan dapat lulus dengan tepat waktu dan memiliki nilai IPK yang baik serta dalam proses seleksi penerima beasiswa terbebas dari unsur subyektifitas.

Metode KNN digunakan untuk memetakan (mengklasifikasikan) data ke dalam satu atau beberapa kelas yang sudah didefinisikan sebelumnya. Metode klasifikasi algoritma *K-Nearest Neighbor* merupakan salah satu metode pengklasifikasian data yang memiliki konsistensi yang kuat, dengan cara mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama berdasarkan pencocokan bobot [5]. Klasifikasi metode KNN berdasarkan data training yang diambil dari data mahasiswa penerima beasiswa PPA sebelumnya (dataset) dengan variabel penentu adalah IPK dan keaktifan tiap semester, sedangkan variabel target adalah mahasiswa tersebut dapat menyelesaikan kuliah tepat waktu atau tidak. Untuk menguji metode metode KNN maka akan dilakukan pengujian dengan menggunakan metode pengujian K-Fold Cross Validation. Pengujian ini dilakukan dengan mengambil data sampel dari mahasiswa yang telah lulus sehingga hasil dari prediksi pendaftar beasiswa akan akurat.

2. Pembahasan

Nearest Neighbor adalah pendekatan untuk proses penyelesaian kasus dengan menghitung pembobotan pada sejumlah fitur yang telah ditentukan. Pendekatan ini ditentukan dengan menghitung kedekatan antara kasus yang baru dengan kasus yang lama [5]. Prinsip kerja *K-Nearest Neighbor* (KNN) adalah mencari jarak terdekat antara data yang akan dilibatkan dalam penentuan label kelas pada data uji. Dari nilai K tetangga

terdekat kemudian akan dilakukan metode voting mayoritas [6]. Teknik *K-Nearest Neighbor* ini termasuk dalam kelompok klasifikasi nonparametric. Di sini kita tidak memperhatikan distribusi dari data yang ingin kita kelompokkan. Teknik ini sangat sederhana dan mudah diimplementasikan. Mirip dengan teknik klastering, kita mengelompokkan suatu data baru berdasarkan jarak data baru itu ke beberapa data/tetangga (*neighbor*) terdekat. [7]. Tujuan algoritma *K-Nearest Neighbor* adalah mengklasifikasikan obyek baru berdasarkan atribut dan *training sample*. *Classifier* tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik *query*, akan ditemukan sejumlah *k* obyek atau titik training yang paling dekat dengan titik *query*. Klasifikasi menggunakan voting terbanyak diantara klasifikasi dari *k* obyek. Algoritma *K-Nearest Neighbor* menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari *query instance* yang baru. Algoritma metode *K-Nearest Neighbor* sangatlah sederhana, bekerja berdasarkan jarak terpendek dari *query instance* ke *training sample* untuk menentukan *K-Nearest Neighbor*-nya.

Nilai *k* yang terbaik untuk algoritma ini tergantung pada data. Secara umum, nilai *k* yang tinggi akan mengurangi efek noise pada klasifikasi, tetapi membuat batasan antara setiap klasifikasi menjadi semakin kabur. Nilai *k* yang bagus dapat dipilih dengan optimasi parameter, misalnya dengan menggunakan *cross-validation*. Kasus khusus dimana klasifikasi diprediksikan berdasarkan training data yang paling dekat (dengan kata lain, $k=1$) disebut algoritma *Nearest Neighbor*. Untuk melakukan metode *K-Nearest Neighbor* perlu diketahui algoritma terlebih dahulu, algoritmanya adalah sebagai berikut :

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

Rumus *K-Nearest Neighbor*:

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

Keterangan:

x_1 = Sampel Data

x_2 = Data Uji / Testing

i = Variabel Data

d = Jarak

p = Dimensi Data

Ketepatan algoritma *K-Nearest Neighbor* oleh ada atau tidak adanya fitur-fitur yang tidak relevan, atau jika bobot fitur tersebut setara dengan relevansinya terhadap klasifikasi. Algoritma *K-Nearest Neighbor* memiliki

kelebihan yaitu dapat menghasilkan data yang kuat atau jelas dan efektif jika digunakan pada data yang besar. Dari beberapa kelebihan tersebut, *K-Nearest Neighbor* juga memiliki kekurangan yaitu membutuhkan nilai *K* sebagai parameter, jarak dari data percobaan tidak dapat jelas dengan tipe jarak yang digunakan dan dengan atribut yang digunakan untuk memperoleh hasil yang terbaik, maka harus menggunakan semua atribut atau hanya satu atribut yang telah pasti, dan perhitungan harga sangat tinggi karena percobaan ini membutuhkan perhitungan jarak dari beberapa kueri untuk semua data percobaan.

Dalam penelitian ini menggunakan data training sejumlah 500 data seperti pada tabel

Tabel 1. Data Training

NO	Mhs	A1	A2	A3	A4	A5	A6	IPK	Ketulusan
1	MHSX1	3.00	3.00	3.00	3.00	2.00	2.00	2.67	tidak Tepat Waktu
2	MHSX2	3.52	2.57	3.00	3.05	3.38	3.05	3.10	tidak Tepat Waktu
3	MHSX3	3.00	3.56	2.45	3.48	3.15	3.18	3.14	Tepat Waktu
4	MHSX4	3.05	3.47	3.48	3.07	3.05	3.09	3.20	tidak Tepat Waktu
5	MHSX5	3.48	3.15	3.48	3.56	3.22	3.52	3.40	Tepat Waktu
6	MHSX6	3.07	3.05	3.07	3.47	3.38	3.00	3.17	tidak Tepat Waktu
7	MHSX7	3.56	0.00	3.56	3.15	3.15	3.05	2.75	tidak Tepat Waktu
....									
491	MHSX491	3.15	3.28	3.28	2.00	3.20	3.28	3.03	Tepat Waktu
492	MHSX492	3.05	3.30	3.30	3.36	3.40	3.36	3.30	Tepat Waktu
493	MHSX493	3.28	3.36	3.36	3.33	3.28	3.33	3.32	Tepat Waktu
494	MHSX494	3.22	3.51	3.51	3.15	3.36	3.15	3.32	Tepat Waktu
495	MHSX495	3.36	3.20	3.20	3.05	3.33	3.05	3.20	Tepat Waktu
496	MHSX496	3.51	3.40	3.40	3.28	3.15	3.28	3.34	Tepat Waktu
497	MHSX497	3.43	3.28	3.28	3.48	3.05	3.48	3.33	Tepat Waktu
498	MHSX498	3.15	3.33	3.33	3.07	3.28	3.07	3.21	Tepat Waktu
499	MHSX499	3.05	3.15	3.15	3.05	3.48	3.05	3.16	Tepat Waktu
500	MHSX500	3.28	3.05	3.05	3.28	3.07	3.28	3.17	Tepat Waktu

Keterangan:

A1 Indeks Prestasi Semester 1

A2 Indeks Prestasi Semester 2

A3 Indeks Prestasi Semester 3

A4 Indeks Prestasi Semester 4

A5 Indeks Prestasi Semester 5

A6 Indeks Prestasi Semester 6

sedangkan data testing sejumlah 125 data pendaftar beasiswa seperti pada tabel 2 . Masing masing data training akan digunaka sebagai data uji dari ke 125 data pendaftar, sehingga di harapkan prediksi yang dihasilkan akan akurat.

Tabel 2. Data pendaftar beasiswa

NO	MHS	SEM	A1	A2	A3	A4	IPK
1	MHS1	4	3,09	3,23	3,78	3,57	3,42
2	MHS2	4	3,05	3,87	3,57	3,34	3,46
3	MHS3	4	3,34	3,54	3,86	3,67	3,60
4	MHS4	4	3,35	3,06	3,24	3,06	3,18
5	MHS5	2	3,40	3,53	0	0	3,47
6	MHS6	2	3,43	3,52	0	0	3,48

Seminar Nasional Sistem Informasi dan Teknologi Informasi 2018

SENSITEK 2018

STMIK Pontianak, 12 Juli 2018

NO	MHS	SEM	A1	A2	A3	A4	IPK
7	MHS7	2	3,47	3,11	0	0	3,29
8	MHS8	2	3,00	3,45	0	0	3,23
9	MHS9	4	3,05	3,86	3,86	3,86	3,66
10	MHS10	4	3,38	3,24	3,24	3,24	3,28
.....							
116	MHS116	2	3,46	3,47	0	0	3,47
117	MHS117	2	3,87	3,15	0	0	3,51
118	MHS118	2	3,17	3,05	0	0	3,11
119	MHS119	2	3,26	3,18	0	0	3,22
120	MHS120	2	3,11	3,76	0	0	3,44
121	MHS121	2	3,45	3,27	0	0	3,36
122	MHS122	2	3,41	3,49	0	0	3,45
123	MHS123	4	3,06	3,11	3,49	3,17	3,21
124	MHS124	4	3,86	3,67	3,53	3,26	3,58
125	MHS125	4	3,24	3,57	3,06	3,11	3,25

Keterangan:

- A1 Indeks Prestasi Semester 1
- A2 Indeks Prestasi Semester 2
- A3 Indeks Prestasi Semester 3
- A4 Indeks Prestasi Semester 4

Dalam melakukan penyelesaian dengan metode KNN, dilakukan prediksi mahasiswa yang mendaftar beasiswa sebagai data testing di lakukan perhitungan dengan menggunakan data training berupa data mahasiswa yang sudah lulus. Dalam proses perhitungan diambil sample data testing sebanyak 10 data mahasiswa. Nilai K yng digunakan pada penelitian ini adalah sebanyak 10. Setelah itu di lakukan perhitungan dengan memasukan nilai data testing di bandingkan dengan data training. Prediksi diakukan dengan nilai IP persemester pada data testing ke dalam persamaan 1. Perhitungan nilai IP ini memperhatikan jumlah semester dari data testing, apabila semester data testing sebanyak 2 semester, maka akan dibandingkan juga sebanyak 2 semester dari data training, demikian juga apabila data sebanyak 4 semester, maka akan di bandingkan juga sebanyak 4 semester dari data testing. Contoh proses perhitungan pada MHS1 sebagai berikut

1. $\sqrt{(3,05 - 3,09)^2 + (3,00 - 3,23)^2 + (3,00 - 3,78)^2 + (3,00 - 3,57)^2} = 0,99$
2. $\sqrt{(3,52 - 3,09)^2 + (2,57 - 3,23)^2 + (3,00 - 3,78)^2 + (3,05 - 3,57)^2} = 1,22$
3. $\sqrt{(3,00 - 3,09)^2 + (3,56 - 3,23)^2 + (2,45 - 3,78)^2 + (3,05 - 3,57)^2} = 1,38$
4. $\sqrt{(3,05 - 3,09)^2 + (3,47 - 3,23)^2 + (3,48 - 3,78)^2 + (3,07 - 3,57)^2} = 0,89$
-
500. $\sqrt{(3,33 - 3,09)^2 + (3,48 - 3,23)^2 + (3,15 - 3,78)^2 + (3,05 - 3,57)^2} = 0,89$

Dari hasil perhitungan ditersebut di atas, di dapati jarak dari data training ke data testing, jarak kedekatan dari 10 mahasiswa pendaftar beasiswa dengan data training terdapat pada tabel 3 berikut :

Tabel 3 . Jarak data training dengan data testing

NO	Data Training	MHS1	MHS2	MHS3	MHS4	MHS5	MHS6	MHS7	MHS8	MHS9	MHS10
1	MHSX1	0,99	1,09	0,99	0,99	0,23	0,23	0,23	0,23	0,99	0,99

NO	Data Training	MHS1	MHS2	MHS3	MHS4	MHS5	MHS6	MHS7	MHS8	MHS9	MHS10
2	MHSX2	1,22	1,5	1,22	1,22	0,79	0,79	0,79	0,79	1,22	1,22
3	MHSX3	1,38	1,16	1,38	1,38	0,34	0,34	0,34	0,34	1,38	1,38
4	MHSX4	0,63	0,41	0,63	0,63	0,24	0,24	0,24	0,24	0,63	0,63
5	MHSX5	0,8	0,97	0,8	0,8	0,4	0,4	0,4	0,4	0,8	0,8
.....											
496	MHSX496	3,41	3,98	3,41	3,41	3,33	3,33	3,33	3,33	3,41	3,41
497	MHSX497	0,91	0,92	0,91	0,91	0,22	0,22	0,22	0,22	0,91	0,91
498	MHSX498	0,79	0,92	0,79	0,79	0,32	0,32	0,32	0,32	0,79	0,79
499	MHSX499	0,83	0,85	0,83	0,83	0,27	0,27	0,27	0,27	0,83	0,83
500	MHSX500	0,89	0,64	0,89	0,89	0,35	0,35	0,35	0,35	0,89	0,89

Dari hasil perhitungan tersebut maka didapati rangking dari masing masing data training ke data testing, Dari hasil perangkangan tersebut, dengan penentuan nilai K=5 diambil 5 jarak terdekat dengan data training. Nilai K=5 ini di dapatkan dari hasil pengujian metode KNN pada data training. Dengan data sampel 10 mahasiswa di dapatkan hasil sebagai berikut pada tabel 4 berikut :

Tabel 4. hasil prediksi

	MHS1	MHS2	MHS3	MHS4	MHS5	MHS6	MHS7	MHS8	MHS9	MHS10
1	0	0	1	0	1	1	0	1	0	1
2	1	1	1	0	1	1	1	0	1	1
3	0	1	1	1	1	1	1	0	0	1
4	1	1	1	1	1	1	1	1	1	1
5	0	1	1	1	1	1	1	1	0	1
AVG	0	1	1	1	1	1	1	1	0	1

Keterangan:

- 0 Lulus tidak tepat waktu
- 1 Lulus tepat waktu

Dari hasil pada tabel dapat di ketahui hasil perdiskinya pada tabel 5 di bawah ini :

Tabel 5. Hasil prediksi

	AVG	Prediksi
MHS1	0	TidakTepatWaktu
MHS2	1	Tepat Waktu
MHS3	1	Tepat Waktu
MHS4	1	Tepat Waktu
MHS5	1	Tepat Waktu
MHS6	1	Tepat Waktu
MHS7	1	Tepat Waktu
MHS8	1	Tepat Waktu
MHS9	0	TidakTepatWaktu
MHS10	1	Tepat Waktu

Dalam melakukan pengujian apakah metoda yang akan di bangun benar, sekaligus mengetahui berapa nilai tetangga terdekat (K) pada metode KNN di lakukan pengujian Kfold cross validation. Pengujian ini dilakukan dengan membagi semua data taring menjadi 10 blok data, yang masing-masing blok data terdapat 50 data. Kemudian dari setiap blok data akan dilakukan pengujian dengan membuat 10 skenario. Skenario yang akan dilakukan tersebut, terdapat pada tabel 6.

Tabel 6. Skenario pengujian

Skenario	Data Training dan data Testing									
1	1	2	3	4	5	6	7	8	9	10
2	1	2	3	4	5	6	7	8	9	10
3	1	2	3	4	5	6	7	8	9	10
4	1	2	3	4	5	6	7	8	9	10
5	1	2	3	4	5	6	7	8	9	10
6	1	2	3	4	5	6	7	8	9	10
7	1	2	3	4	5	6	7	8	9	10
8	1	2	3	4	5	6	7	8	9	10
9	1	2	3	4	5	6	7	8	9	10
10	1	2	3	4	5	6	7	8	9	10

Keterangan:

	Data Testing
	Data Training

Hasil pengujian pada masing masing skenario diukur dengan menghitung tingkat akurasi dari data uji. Akurasi dihitung dengan cara jumlah data uji yang benar dibagi dengan jumlah data uji kemudian dikalikan dengan 100%. Jumlah tetangga terdekat (K) yang digunakan adalah K1 sampai K10, atau tetangga terdekat ke 1 sampai ke 10. Hasil pengujian di atas maka akan didapatkan nilai akurasi dari masing masing skenario, nilai akurasi dari masing masing skenario, dapat di lihat pada 7.

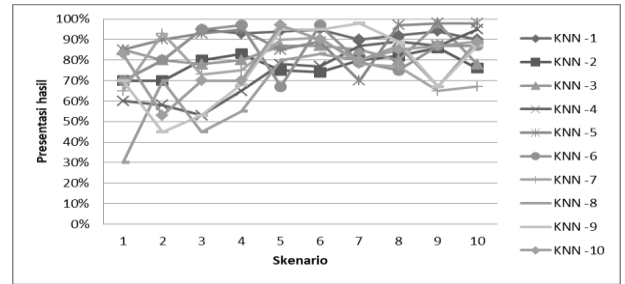
Tabel 7. Hasil pengujian K-Fold Cross Validation

	K 1	K 2	K 3	K 4	K 5	K 6	K 7	K 8	K 9	K 10
1	70%	70%	85%	60%	85%	70%	65%	30%	70%	83%
2	80%	70%	80%	58%	90%	80%	93%	70%	45%	53%
3	95%	80%	78%	53%	93%	95%	73%	45%	53%	70%
4	93%	83%	80%	65%	95%	97%	75%	55%	68%	70%
5	94%	75%	87%	78%	85%	67%	90%	80%	95%	97%
6	95%	74%	87%	77%	89%	97%	91%	83%	95%	90%
7	90%	80%	85%	87%	70%	79%	78%	80%	98%	80%
8	92%	82%	80%	89%	97%	75%	77%	87%	89%	85%
9	94%	86%	98%	87%	98%	87%	65%	67%	67%	87%
10	90%	76%	78%	95%	98%	87%	67%	89%	87%	89%
Hasil	89%	78%	84%	75%	90%	83%	77%	69%	77%	80%

Keterangan

1,2,3,4,5,6,7,8,9,10 = Skenario pengujian
 K1, K2, K3, K4, K5, K6, K7, K8, K9, K10 = KNN

Pada tabel 7 diatas, hasil akurasi untuk setiap skenario pada setiap nilai K pada nilai KNN, hasil terbesar pada K5, sehingga akurasi terbaik untuk perhitungan metode KNN pada data serupa dengan data pada penelitian ini adalah K5 atau 5 data dengan jarak terdekat pada data training. Nilai ini juga yang akan digunakan sebagai nilai KNN untuk perhitungan pada data yang sebenarnya. Secara grafik, hasil uji akurasi pada metode KNN dapat di lihat pada gambar 1.



Gambar 1. Grafik hasil pengujian

3. Kesimpulan

Berdasarkan hasil penelitian yang telah dijelaskan pada bab sebelumnya, maka kesimpulan pada penelitian ini adalah sebagai berikut:

- Ketika di lakukan pengujian K Fold Cross validation sejumlah 10 Fold untuk dengan metode KNN pada data uji sejumlah 500 data, maka didapati nilai K terbaik adalah pada metode KNN adalah 5. Pengujian ini dilakukan dengan membagi 500 data menjadi 10 blok uji dengan masing-masing blok akan di uji dengan data yang lain. Kemudian akan dihitung presentase kesamaan hasil prediksi pada masing masing blok uji. Nilai terbaik di dapati pada pada nilai K5 dengan jumlah kesamaan 90%.
- Hasil dari pengujian ini akan digunakan sebagai nilai KNN dalam pengujian data pendaftaran, sehingga hasil prediksi yang didapatkan akan optimal.
- Hasil preediksi dengan KNN dapat dilakukan pengolahan lebih lanjut dengan menggunakan Sistem pendukung keputusan untuk merangking hasil dari prediksi yang telah dilakukan.

Daftar Pustaka

- [1] A. Johar.T, D. Yanosma dan K. Anggriani, "Implementasi Metode K-Nearest Neighbor (KNN) Dan Simple Additive Weighting (SAW) Dalam Pengambilan Keputusan Seleksi Penerimaan Anggota Paskibraka (Studi Kasus : Dinas Pemuda dan Olahraga Provinsi Bengkulu)," *Jurnal Pseudocode, Volume III Nomor 2*, pp. 98 - 112, 2016.
- [2] J. I. Kartika, E. Santoso dan Sutrisno, "Penentuan Siswa Berprestasi Menggunakan Metode K-Nearest Neighbor dan Weighted Product (Studi Kasus : SMP Negeri 3 Mejayan)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, Vol. 1, No. 5*, pp. 352-360, 2017.
- [3] T. F. Prasetyo, D. Susandi dan I. S. Widianingrum, "Prediksi Kelulusan Mahasiswa Pada Perguruan Tinggi Kabupaten Majalengka Berbasis Knowledge Based System," *Seminar Nasional Telekomunikasi dan Informatika, ISSN : 2503-2844*, pp. 32-37, 2016.

Seminar Nasional Sistem Informasi dan Teknologi Informasi 2018

SENSITEK 2018

STMIK Pontianak, 12 Juli 2018

- [4] R. K. Niswatin dan A. Sanjaya, "Classification Of Category Selection Title Undergraduate Thesis Using K-Nearest Neighbor Method," *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 7, No. 3, pp. 846 - 854, 2017.
- [5] Kusriani dan E. T. Luthfi, *Algoritma Data Mining*, Yogyakarta: Penerbit Andi, 2009.
- [6] E. Prasetyo, *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*, Yogyakarta: Andi, 2014.
- [7] B. Santosa, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Yogyakarta: Graha Ilmu, 2007.