

PERBANDINGAN AKURASI SOFTWARE RAPIDMINER DAN WEKA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (K-NN)

Devi Damayanti

Universitas Pamulang (UNPAM) Tangerang Selatan Banten, Indonesia
Email: dosen02390@unpam.ac.id

INFO ARTIKEL

Diterima
3 Juni 2021
Direvisi
8 Juni 2021
Disetujui
21 Juni 2021

Keywords:

*prediction; data mining;
k-nn; rapidminer; weka;
accuracy*

ABSTRACT

One of the keys to a person's success in taking education is the period of study. Likewise, universities are required to provide quality education for students because students are assets for an educational institution so that it is necessary to pay attention to the graduation rate of their students so that they are on time. The research was conducted at Sahid University Jakarta, Faculty of Economics, Management Study Program. From the data obtained, it shows that many students do not graduate on time, which is 4 years or 8 semesters. The delay in graduating students can be caused by several factors, one of which is the disruption of the lecture process. There are many aspects that support the disruption of the lecture process, including the lack of student ability in the course, the lack of student attendance in the lecture process and various other aspects that cannot be measured. The purpose of this study is to compare the accuracy results produced by the two software, Rapidminer and WEKA using the K-NN algorithm with the fold cross validation method, processing previous data that will be used as predictions for graduates in the coming years. The attributes used are IPS 1 to IPS 8, IPK semester 2 to IPK semester 8 and compulsory subjects. Making eight data samples to be tested with different value limits based on the Regulation of the Minister of Education and Culture of the Republic of Indonesia Number 49 of 2014 concerning National Standards for Higher Education. The results showed that the WEKA software with the K-NN algorithm produced a higher accuracy value than Rapidminer, which was 78.25%. So can be used as an alternative method to predict the punctuality of graduation at Sahid University, Jakarta.

ABSTRAK

Salah satu kunci keberhasilan seseorang dalam

How to cite:	Damayanti, Devi (2021). Perbandingan Akurasi Software Rapidminer dan Weka Menggunakan Algoritma K-Nearest Neighbor (K-NN). <i>Jurnal Syntax Admiration</i> 2(6). https://doi.org/10.46799/jsa.v2i6.247
E-ISSN:	2722-5356
Published by:	Ridwan Institute

menempuh Pendidikan adalah masa studi. Sama halnya dengan perguruan tinggi dituntut untuk menyelenggarakan Pendidikan yang berkualitas bagi mahasiswa karena mahasiswa adalah aset bagi sebuah institusi pendidikan sehingga perlu diperhatikan tingkat kelulusan mahasiswanya agar tepat pada waktunya. Penelitian dilaksanakan di Universitas Sahid Jakarta Fakultas Ekonomi Prodi Manajemen. Dari data yang didapatkan menunjukkan bahwa banyak mahasiswa yang lulus tidak tepat pada waktu yang telah ditentukan yaitu 4 tahun atau 8 semester. Keterlambatan waktu lulus mahasiswa dapat disebabkan oleh beberapa faktor, salah satunya proses perkuliahan yang terganggu. Banyak aspek yang mendukung terganggunya proses perkuliahan diantaranya kurangnya kemampuan mahasiswa terhadap mata kuliah, kurangnya kehadiran mahasiswa dalam mengikuti proses perkuliahan dan berbagai aspek lainnya yang tidak dapat diukur. Tujuan Penelitian ini adalah membandingkan hasil akurasi yang dihasilkan oleh kedua *software* yaitu Rapidminer dan WEKA menggunakan algoritma K-NN dengan metode *fold cross validation*, mengolah data sebelumnya yang akan dijadikan prediksi untuk lulusan tahun-tahun yang akan datang. Atribut yang digunakan adalah IPS 1 sampai IPS 8, IPK semester 2 sampai IPK semester 8 dan mata kuliah wajib. Membuat delapan *sampling* data untuk diujikan dengan batasan nilai yang berbeda-beda berpacu pada Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 49 Tahun 2014 Tentang Standar Nasional Pendidikan Tinggi. Hasil penelitian menunjukkan bahwa *software* WEKA dengan algoritma K-NN menghasilkan nilai akurasi yang lebih tinggi dibandingkan dengan Rapidminer yaitu sebesar 78,25%. Sehingga dapat digunakan menjadi alternatif metode untuk memprediksi ketepatan waktu kelulusan di Universitas Sahid Jakarta.

Kata Kunci:

prediksi; data mining; K-NN; *rapidminer*; WEKA; akurasi

Pendahuluan

Salah satu hal yang dapat dijadikan sebagai ukuran keberhasilan seseorang dalam menempuh pendidikan adalah masa studi. Masa studi merupakan masa untuk menyelesaikan beban studi dalam mengikuti proses pendidikan pada program studinya. Sama halnya dengan perguruan tinggi dituntut untuk menyelenggarakan pendidikan yang berkualitas bagi mahasiswa sehingga menghasilkan sumber daya manusia yang

berilmu, cakap, kreatif dan bersaing. Dalam sistem pendidikan mahasiswa adalah aset penting bagi sebuah institusi pendidikan dan untuk itu perlu diperhatikan tingkat kelulusan mahasiswa tepat pada waktunya.

Persentase naik turunnya kemampuan mahasiswa untuk menyelesaikan studi tepat waktu merupakan salah satu elemen penilaian akreditasi universitas (Tinggi 2008). Sehingga perlu adanya pemantauan dan evaluasi terhadap kecenderungan mahasiswa lulus tepat waktu atau tidak.

Penelitian ini akan dilakukan pada salah satu program studi yang ada di Fakultas Ekonomi (S1) yaitu Program Studi Manajemen. Berdasarkan data yang diperoleh dari Wakil Direktur Administrasi dan Akademik Universitas Sahid Jakarta mulai dari tahun 2011 sampai dengan tahun 2015. Pada tahun 2011 total jumlah kelulusan mahasiswa adalah 96 mahasiswa dengan waktu kelulusan rata-rata 4,9 tahun, tahun 2012 jumlah kelulusan sebanyak 68 mahasiswa dengan waktu kelulusan rata-rata 4,64 tahun, tahun 2013 jumlah kelulusan mahasiswa sebanyak 96 mahasiswa dengan waktu kelulusan rata-rata 4,62 tahun, tahun 2014 jumlah kelulusan mahasiswa sebanyak 94 mahasiswa dengan waktu kelulusan rata-rata 4,72 tahun dan tahun 2015 jumlah kelulusan mahasiswa sebanyak 119 mahasiswa dengan waktu kelulusan rata-rata 4,67 tahun. Dengan rata-rata kelulusan per tahun dalam lima tahun terakhir adalah 95 mahasiswa dalam waktu 4,71 tahun.

Dari data tersebut menunjukkan bahwa banyak mahasiswa yang lulus tidak tepat pada waktu yang telah ditentukan yaitu 4 tahun atau 8 semester. Keterlambatan waktu lulus mahasiswa dapat disebabkan oleh beberapa faktor, salah satunya proses perkuliahan yang terganggu. Banyak aspek yang mendukung terganggunya proses perkuliahan diantaranya kurangnya kemampuan mahasiswa terhadap mata kuliah, kurangnya kehadiran mahasiswa dalam mengikuti proses perkuliahan dan berbagai aspek lainnya yang tidak dapat diukur.

Data mining merupakan proses penemuan pola dalam data dengan cara menganalisa data yang sudah ada dalam *database* (Ngo 2011) sehingga ketepatan waktu kelulusan mahasiswa dapat diklasifikasikan menggunakan teknik data *mining*. Membandingkan hasil akurasi dari kedua *tools* data *Mining Rapidminer* dan WEKA yang masuk ke dalam *Six Of The Best Open Source Data Mining Tools* dari data yang didapatkan pada (Chertchom 2018) diharapkan informasi yang dihasilkan lebih detail dan akurat sehingga ke depannya dapat menjadi alternatif metode untuk memprediksi ketepatan waktu kelulusan di Universitas Sahid Jakarta.

Penelitian terdahulu terkait dengan penelitian mengenai prediksi ketepatan waktu kelulusan, algoritma serta *software data mining* yang sama diantaranya yaitu oleh (Mustafa and Simpen 2014) berjudul Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik *Data Mining* (Studi Kasus: Data Akademik Mahasiswa Stmik Dipanegara Makassar) hasil penelitian akurasi algoritma *K-Nearest Neighbor* adalah 83,36%. (Huda 2020) berjudul Perbandingan Akurasi *K-Nearest Neighbor* Dan *Naïve Bayes* Untuk Algoritma Sistem Prediksi Nilai Akhir Mahasiswa hasil penelitiannya Algoritma terbaik adalah K-NN dengan K=15 dan nilai

akurasinya adalah 95%. (Kamagi and Hansun 2014) berjudul Implementasi *Data Mining* Dengan Algoritma C4.5 Untuk Memprediksi Tingkat Kelulusan Mahasiswa dengan hasil penelitian Tingkat akurasi aplikasi dari hasil prediksi adalah 87,5%. (Nursalim and Himawan 2014) berjudul Klasifikasi Bidang Kerja Lulusan Menggunakan Algoritma *K-Nearest Neighbor* (KNN) dengan hasil penelitian Algoritma KNN memiliki kinerja terbaik untuk klasifikasi bidang kelulusan dengan nilai akurasi 83,33% AUC 0,900. (Rohman 2015) berjudul Model Algoritma *K-Nearest Neighbor* (K-NN) Untuk Prediksi Kelulusan Mahasiswa dengan hasil penelitian Tingkat akurasi tertinggi pada kluster $k=5$ yaitu 85,15%. (Banjarsari, Budiman, and Farmadi 2016) berjudul Penerapan K-Optimal Pada Algoritma Knn Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan Ip Sampai Dengan Semester 4 dengan hasil penelitian Tingkat akurasi tertinggi pada kluster $K=5$ yaitu 80,00%. (Mustakim and Oktaviani 2016) berjudul Algoritma *K-Nearest Neighbor Classification* Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa dengan hasil penelitian Algoritma K-NN menghasilkan akurasi sebesar 82%. (Srianto and Mulyanto 2016) berjudul Perbandingan *K-Nearest Neighbor* Dan *Naïve Bayes* Untuk Klasifikasi Tanah Layak Pohon Jati dengan hasil penelitian K-NN lebih baik dibandingkan dengan *Naïve Bayes* dengan akurasi 96,66% dan 82,63%. algoritma *K-Nearest Neighbor Classification* sebagai sistem prediksi predikat prestasi mahasiswa oleh (Mustakim and Oktaviani 2016) menghasilkan akurasi K-NN sebesar 82%. Dengan hasil akurasi yang di dapat dari penelitian-penelitian sebelumnya peneliti tertarik untuk menggunakan algoritma K-NN yang akan dihitung ke dalam dua *tools data mining* yaitu *Rapidminer* dan *WEKA* dengan tujuan untuk mengetahui diantara kedua *tools* tersebut yang lebih tinggi tingkat akurasinya

(Muin 2016) berjudul Metode *Naïve Bayes* Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi) dengan hasil penelitian Tingkat akurasi *Naïve Bayes* mencapai 94%. Dari beberapa penelitian terdahulu akhirnya peneliti memutuskan melakukan penelitian membandingkan *software* Rapidminer dan *WEKA* dengan menggunakan satu algoritma klasifikasi yaitu algoritma K-NN yang menurut peneliti adalah suatu kebaruan dari penelitian karena peneliti belum menemukan dengan metode yang sama. Selain kebaruan penelitian, tujuan dilakukannya penelitian ini adalah agar mendapatkan nilai akurasi yang paling baik diantara kedua *software data mining* tersebut sehingga hasilnya nanti akan dapat menjadi alternatif metode bagi Universitas Sahid Jakarta untuk memprediksi kelulusan mahasiswanya di tahun – tahun berikutnya.

Metode Penelitian

Data yang digunakan untuk penelitian adalah data mahasiswa Fakultas Ekonomi program studi Manajemen (S1) dalam waktu kelulusan mulai tahun 2011. Dari jumlah mahasiswa yang lulus tepat waktu dan tidak tepat waktu dalam 5 tahun sampai dengan tahun 2015 berjumlah 149 data mahasiswa. Dari data awal yang dikumpulkan sebanyak 473 data mahasiswa yang kemudian melalui 6 tahapan teknik analisis yaitu *domain*

understanding, selection and addition, data cleaning, transformation, evaluation and interpretation, dan discover knowledge sehingga menghasilkan data lengkap hanya berjumlah 149 data mahasiswa.

Tabel 1
Pemilihan Atribut

No.	Atribut	Keterangan	Detail
1	IPS	Indeks Prestasi Sementara	Indeks Prestasi Sementara Semester 1 sampai dengan Semester 8
2	IPK	Indeks Prestasi Kumulatif	Indeks Prestasi Kumulatif Semester 2 sampai dengan Semester 8
3	MK	Mata Kuliah	Mata Kuliah Wajib (Pengantar Manajemen, Manajemen Operasi, Manajemen SDM)
4	KT	Status Keterangan Lulus	Tepat Waktu dan Tidak Tepat Waktu

Pemilihan atribut yang terdiri dari kumpulan nilai disebut dengan Indeks Prestasi per semester dan Indeks Prestasi Kumulatif serta nilai mata kuliah wajib ini diharapkan dapat meningkatkan nilai akurasi yang diperoleh dari perhitungan algoritma K-NN serta bertujuan untuk mengumpulkan nilai-nilai agar dapat dijadikan acuan prediksi kelulusan tepat waktu atau tidak tepat waktu.

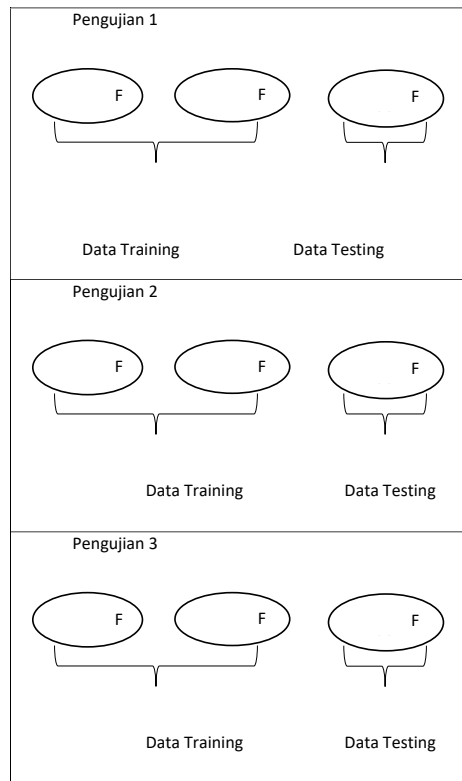
Mencari tingkat akurasi yang paling baik, peneliti menggunakan 8 data *sampling* dengan masing-masing batasan nilai yang berbeda. Batasan-batasan nilai yang digunakan adalah sebagai berikut:

Tabel 2
Batasan Data Set untuk Pembuatan Sampling Pengujian Algoritma

<i>Sampling</i>	Batasan Nilai Atribut IPS / IPK	Predikat	Batasan Nilai Atribut Mata Kuliah Wajib	Predikat
1	2,00	Cukup	2 / C	Cukup
2	2,76	Memuaskan	2 / C	Cukup
3	3,01	Sangat memuaskan	2 / C	Cukup
4	3,50	Pujian	2 / C	Cukup
5	2,00	Cukup	3 / B	Baik
6	2,76	Memuaskan	3 / B	Baik
7	3,01	Sangat memuaskan	3 / B	Baik
8	3,50	Pujian	3 / B	Baik

Menguji pola klasifikasi pada penelitian ini adalah metode *K-Fold Cross Validation*. Dalam *K-Fold Cross Validation*, data dibagi menjadi k bagian, D1, D2, ... Dk, dan masing-masing D memiliki jumlah data yang sama. Kemudian dilakukan proses pengulangan sebanyak K, dimana dalam setiap pengulangan ke-I, D akan dijadikan data *testing*, dan sisanya akan dijadikan data *training*. Sebagai contoh misalkan akan dilakukan metode *cross validation* dengan menggunakan *3-fold*. Pertama pilih salah satu *fold* menjadi data *testing*, kemudian *fold* sisanya menjadi data *training*.

Hal ini dilakukan berulang untuk semua kombinasi data *training-testing*. Sebagai ilustrasi, perhatikan gambar dibawah ini.

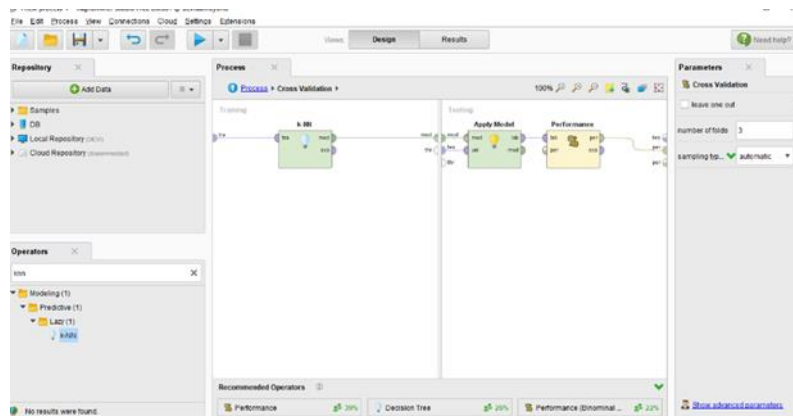


Gambar 1
Ilustrasi 3-Fold Cross Validation

Sistem dalam klasifikasi diharapkan mampu melakukan semua set data dengan benar, namun tidak dapat dipungkiri bahwa kesalahan pasti akan terjadi dalam proses pengklasifikasian. Untuk itu perlunya dilakukan pengukuran kinerja klasifikasi dengan matriks konfusi (*confusion matrix*). Matrik konfusi adalah tabel pencatat hasil kerja klasifikasi yang terdapat pada *Rapidminer* dan *WEKA*.

Setelah mendapatkan 8 data sampling, pengujian keakuratan dari suatu klasifikasi, dengan ketentuan memakai 3, 5, 10, 15 dan 20 *fold cross validation* yang diharapkan akan menemukan tingkat akurasi yang terbaik.

Pengujian algoritma K-NN menggunakan *Rapidminer* contoh menggunakan *sampling* data ke-1 dan metode pengujian 3 *fold cross validation*.



Gambar 2
Subproses *Training* dan *Testing* K-NN

ExampleSet yang sudah dilatih sebelumnya akan dibangun dengan metode K-NN. *Apply* model berfungsi mempelajari *ExampleSet* yang sudah dilatih sedangkan *performance* ini digunakan untuk evaluasi statistik dari kinerja klasifikasi dan memberikan daftar nilai kriteria dari kinerja klasifikasi tersebut.

Algoritma K-NN pada subproses *training* sedangkan *apply* model dan *performance* pada subproses *testing*, kemudian masing-masing dihubungkan ke *port*. Setelah itu mulai lah proses pengujian dengan menggunakan algoritma K-NN.

Hasil akurasi pengujian data sampling ke-1 dengan algoritma K-NN menggunakan 3 fold cross validation adalah sebagai berikut:

Table View Plot View

accuracy: 42.20% +/- 21.24% (mikro: 42.28%)

	true Tidak Tepat Waktu	true Tepat Waktu	class precision
pred. Tidak Tepat Waktu	17	73	18.89%
pred. Tepat Waktu	13	46	77.97%
class recall	56.67%	38.66%	

Gambar 3
Hasil Akurasi *Sampling* ke-1 Algoritma K-NN menggunakan 3 Fold Cross Validation

Hasil akurasi dengan perhitungan nilai *confusion matrix* terhadap algoritma K-NN menggunakan *Rapidminer* dengan 3 fold cross validation yang menghasilkan tingkat akurasi sebesar 42,28% adalah sebagai berikut:

Pred Tidak Tepat Waktu-true Tidak Tepat Waktu = jumlah data yang diprediksi Tidak Tepat Waktu dan kenyataannya Tidak Tepat Waktu (TP)

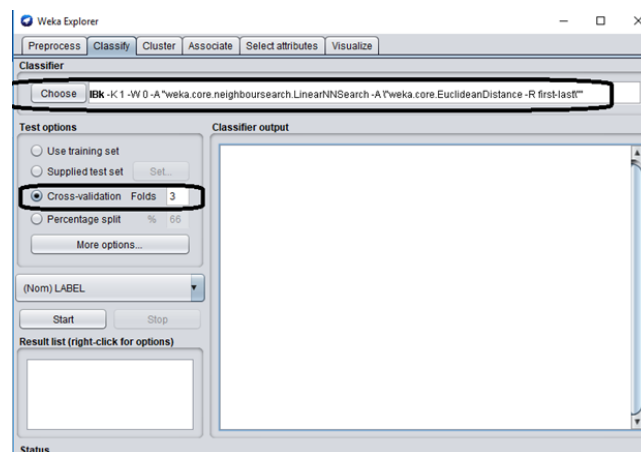
1. Pred Tepat Waktu – true Tepat Waktu = jumlah data yang diprediksi Tepat Waktu dan kenyataannya Tepat Waktu (TN)
2. Pred Tidak Tepat Waktu – true Tepat Waktu = jumlah data yang diprediksi Tidak Tepat Waktu tapi kenyataannya Tepat Waktu (FP)

3. Pred Tepat Waktu – *true* Tidak Tepat Waktu = jumlah data yang diprediksi Tepat Waktu tapi kenyataanya Tidak Tepat Waktu (FN)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
$$Accuracy = \frac{17 + 46}{17 + 46 + 73 + 13}$$
$$Accuracy = \frac{63}{149} = 42,28 \%$$

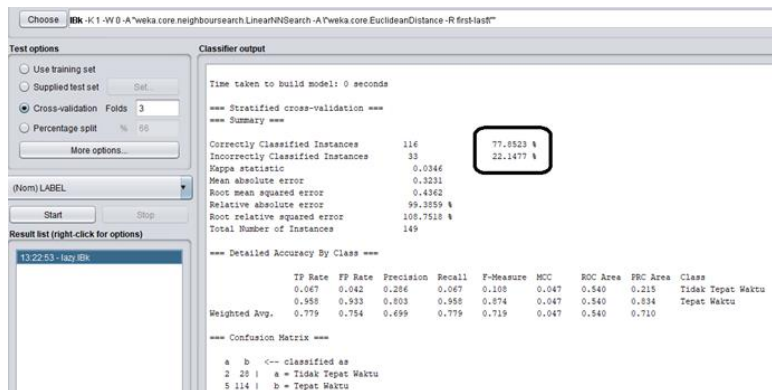
Hasil perhitungan dengan rumus akurasi, terlihat bahwa hasil yang didapat dengan hasil pada *Rapidminer* adalah sama.

Setelah dilakukan pengujian dengan menggunakan tool *Rapidminer* selanjutnya ke delapan sampling data tersebut akan kembali diuji dengan menggunakan tool WEKA. Pada pengujian algoritma K-NN menggunakan tool WEKA juga dilakukan dengan menggunakan *cross-validation* dengan 3, 5, 10, 15 dan 20 fold. Sebagai contoh akan menggunakan 3 fold, seperti pada gambar di bawah ini.



Gambar 4 Pemilihan Metode Pengujian

Setelah semua selesai langkah terakhir adalah menguji data sampling ke-1 dengan meng klik *Start* dan berikut adalah hasil akurasi data *sampling* ke-1:



Gambar 5
Hasil Akurasi *Sampling* ke-1 Algoritma K-NN menggunakan 3 *Fold Cross Validation*

Berikut perhitungan nilai akurasi *confusion matrix* yang menghasilkan tingkat akurasi sebesar 77,85% dengan klasifikasi a adalah Tidak Tepat Waktu dan b adalah Tepat Waktu.

1. Pred Tidak Tepat Waktu-*true* Tidak Tepat Waktu = jumlah data yang diprediksi Tidak Tepat Waktu dan kenyataannya Tidak Tepat Waktu (TP)
2. Pred Tepat Waktu-*true* Tepat Waktu = jumlah data yang diprediksi Tepat Waktu dan kenyataannya Tepat Waktu (TN)
3. Pred Tidak Tepat Waktu-*true* Tepat Waktu = jumlah data yang diprediksi Tidak Tepat Waktu tapi kenyataannya Tepat Waktu (FP)
4. Pred Tepat Waktu-*true* Tidak Tepat Waktu = jumlah data yang diprediksi Tepat Waktu tapi kenyataannya Tidak Tepat Waktu (FN)

$$Accuracy = (TP+TN)/(TP+TN+FP+FN)$$

$$Accuracy = (2+114)/(2+114+28+5)$$

$$Accuracy = 116/(149) = 77,85 \%$$

Dari hasil perhitungan dengan rumus dan hasil WEKA menunjukkan bahwa hasil akurasi yang didapatkan adalah sama.

Hasil dan Pembahasan

Delapan data *sampling* sudah diujikan dengan langkah-langkah seperti yang sudah dijelaskan sebelumnya. Delapan data *sampling* diuji dengan 3, 5, 10, 15 dan 20 *fold cross validation*. Ada alasan yang menjadi dasar pemilihan 3, 5, 10, 15 dan 20 *fold cross validation*, karena menurut (Witten et al. 2005) 10 *fold* memberikan hasil prediksi terbaik dalam percobaan dengan beragam data dan beragam algoritma sehingga 10 *fold* menjadi standar dalam melakukan pengujian tetapi pendapat ini masih sering didebatkan, sehingga peneliti ingin menambahkan 3, 5, 15 dan 20 *fold* dengan harapan mendapatkan *fold* yang menghasilkan akurasi terbaik dalam pengujian. Berikut adalah

hasil akurasi algoritma K-NN menggunakan *tool* Rapidminer dan WEKA dengan delapan data sampling menggunakan 3, 5, 10, 15 dan 20 *fold cross validation*.

Tabel 3
Hasil Pengujian Algoritma K-NN Menggunakan *Rapidminer* dan WEKA

Hasil Akurasi Algoritma K-NN			
<i>Sampling</i>	Metode Pengujian <i>Cross Validation</i>	<i>Rapidminer</i>	WEKA
1	3 Fold	42.20%	77.85%
2		70.50%	71.14%
3		63.06%	69.80%
4		53.62%	77.85%
5		66.44%	75.84%
6		72.54%	69.80%
7		66.45%	65.10%
8		71.84%	75.84%
Rata – Rata		63.33%	72.90%
1	5 Fold	39.68%	76.51%
2		65.79%	73.82%
3		64.34%	76.51%
4		50.23%	75.84%
5		63.77%	71.81%
6		64.46%	72.48%
7		68.46%	63.76%
8		69.82%	77.18%
Rata - Rata		60.82%	73.49%
1	10 Fold	32.29%	78.52%
2		67.00%	72.49%
3		63.19%	76.51%
4		43.71%	77.18%
5		62.42%	71.81%
6		67.11%	72.48%
7		67.24%	64.43%
8		70.48%	77.18%
Rata – Rata		59.18%	73.83%
1	15 Fold	29.48%	77.18%
2		68.59%	73.15%
3		65.11%	75.84%
4		42.81%	75.84%
5		62.52%	71.81%
6		67.33%	69.13%
7		68.52%	63.76%
8		72.52%	77.18%
Rata – Rata		59.61%	72.99%
1	20 Fold	29.48%	78.52%
2		68.59%	72.48%
3		65.11%	77.18%
4		42.81%	77.18%
5		62.52%	73.15%
6		67.33%	72.48%
7		68.52%	65.10%
8		72.52%	77.85%
Rata – Rata		59.61%	74.24%

Tabel 4
Hasil Akurasi Tertinggi

HASIL AKURASI TERTINGGI ALGORITMA K-NN					
<i>RAPIDMINER</i>	72,54%	<i>SAMPLING</i>	6	<i>FOLD</i>	3
<i>WEKA</i>	78,52%	<i>SAMPLING</i>	1	<i>FOLD</i>	10, 20

Dari hasil pengujian algoritma K-NN dengan *Rapidminer* dan *WEKA* menggunakan 3, 5, 10, 15, dan 20 *fold cross validation* seperti yang telah dijabarkan pada tabel di atas, di dapatkan hasil akurasi tertinggi dengan *Rapidminer* yaitu sebesar 72,54% pada data *sampling* ke-6 dengan menggunakan 3 *fold cross validation*. Sedangkan pada *WEKA*, akurasi tertinggi sebesar 78,52% pada data *sampling* ke-1 dan ada pada *fold* ke 10 dan 20.

Kesimpulan

Kesimpulan yang diperoleh dari penelitian ini adalah hasil akurasi *software* *WEKA* menggunakan algoritma K-NN lebih tinggi yaitu sebesar 78,25% dibandingkan dengan *Rapidminer* menggunakan algoritma K-NN yaitu sebesar 72,54% sehingga *software* *WEKA* menggunakan algoritma K-NN dapat menjadi alternatif metode atau metode yang bisa digunakan khususnya pada Universitas Sahid Jakarta dalam memprediksi ketepatan waktu kelulusan mahasiswanya untuk tahun- tahun berikutnya.

BIBLIOGRAFI

- Banjarsari, Mutiara Ayu, Irwan Budiman, and Andi Farmadi. 2016. "Penerapan K-Optimal Pada Algoritma Knn Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan Ip Sampai Dengan Semester 4." *Klik-Kumpulan Jurnal Ilmu Komputer* 2(2): 159–73.[Google Scholar](#)
- Chertchom, Prajak. 2018. "A Comparison Study between Data Mining Tools over Regression Methods: Recommendation for SMEs." In *2018 5th International Conference on Business and Industrial Research (ICBIR)*, IEEE, 46–50.[Google Scholar](#)
- Huda, Nuzulul. 2020. "Klasifikasi Berita Menggunakan Metode Learning Vector Quantization." [Google Scholar](#)
- Kamagi, David Hartanto, and Seng Hansun. 2014. "Implementasi Data Mining Dengan Algoritma C4. 5 Untuk Memprediksi Tingkat Kelulusan Mahasiswa." *Ultimatics: Jurnal Teknik Informatika* 6(1): 15–20.[Google Scholar](#)
- Muin, Asrul Ashari. 2016. "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)." *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar* 2(1): 22–26.[Google scholar](#)
- Mustafa, Muhammad Syukri, and I Wayan Simpen. 2014. "Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining (Studi Kasus: Data Akademik Mahasiswa STMIK Dipanegara Makassar)." *Creative Information Technology Journal* 1(4): 270–81.[Google Scholar](#)
- Mustakim, Mustakim, and Giantika Oktaviani. 2016. "Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa." *Jurnal Sains, Teknologi dan Industri* 13(2): 195–202.[Google Scholar](#)
- Ngo, Terry. 2011. "Data Mining: Practical Machine Learning Tools and Technique, by Ian h. Witten, Eibe Frank, Mark a. Hell." *ACM Sigsoft Software Engineering Notes* 36(5): 51–52.[Google Scholar](#)
- Nursalim, Suprapedi, and H Himawan. 2014. "Klasifikasi Bidang Kerja Lulusan Menggunakan Algoritma K-Nearest Neighbor." *Jurnal Teknologi Informasi* 10(1): 31–43.[Google Scholar](#)
- Rohman, Abdul. 2015. "Model Algoritma K-Nearest Neighbor (K-Nn) Untuk Prediksi Kelulusan Mahasiswa." *Neo Teknika* 1(1).[Google Scholar](#)
- Srianto, Didik, and Edy Mulyanto. 2016. "Perbandingan K-Nearest Neighbor Dan Naive Bayes Untuk Klasifikasi Tanah Layak Tanam Pohon Jati." *Techno. Com* 15(3): 241–45.[Google Scholar](#)

Devi Damayanti

Tinggi, Badan Akreditasi Nasional Perguruan. 2008. “Akreditasi Program Studi Sarjana.” *Jakarta: BAN PT.*[Google Scholar](#)

Witten, Ian H et al. 2005. “Kea: Practical Automated Keyphrase Extraction.” In *Design and Usability of Digital Libraries: Case Studies in the Asia Pacific*, IGI global, 129–52[Google Scholar](#).

Copyright holder:

Devi Damayanti (2021)

First publication right:

Jurnal Syntax Admiration

This article is licensed under:

