

MODEL ALGORITMA *K-NEAREST NEIGHBOR* (K-NN) UNTUK PREDIKSI KELULUSAN MAHASISWA

Abdul Rohman

Dosen Jurusan Elektronika Fakultas Teknik Universitas Pandanaran Semarang

Abstrak

Dalam sistem pendidikan mahasiswa adalah aset penting bagi sebuah institusi pendidikan dan untuk itu perlu diperhatikan tingkat kelulusan mahasiswa tepat pada waktunya. Prosentase naik turunnya kemampuan mahasiswa untuk menyelesaikan studi tepat waktu merupakan salah satu elemen penilaian akreditasi universitas. Untuk itu perlu adalah pemantauan dan evaluasi terhadap kelulusan mahasiswa dengan menggunakan klasifikasi data mining. Dengan mengolah data mahasiswa akan memperoleh hal penting untuk keperluan perguruan tinggi terutama untuk prediksi kelulusan mahasiswa. Dalam penelitian ini dalam memprediksi kelulusan mahasiswa dengan menggunakan algoritma klasifikasi data mining *K-Nearest Neighbor* dengan mengkluster data $k=1$, $k=2$, $k=3$, $k=4$, dan $k=5$. Hasil yang diperoleh dengan cluster data $k=5$ *accuracy* adalah 85,15% dan nilai *AUC* adalah 0.888 adalah akurasi paling tinggi.

Kata kunci: mahasiswa, *K-Nearest Neighbor*

PENDAHULUAN

Latar Belakang

Lembaga pendidikan tertinggi adalah perguruan tinggi yang merupakan penyelenggara pendidikan akademik bagi mahasiswa. Mahasiswa sering disebut kelompok masyarakat yang memiliki ciri intelektualitas yang lebih luas dibandingkan dengan kelompok usia 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 (Azwar, 2004).

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.

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

adanya pemantauan maupun evaluasi terhadap kecenderungan mahasiswa lulus tepat waktu atau tidak.

Pemantauan atau evaluasi terhadap kinerja akan menghasilkan informasi yang bermanfaat untuk membantu mahasiswa, dosen, administrator, dan pembuat kebijakan (Ogor, 2007). Dengan demikian jelas bahwa memprediksi kelulusan mahasiswa adalah hal yang penting bagi penyelenggara pendidikan untuk menentukan strategi bagi institusinya.

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 (Ogor, 2007). Dengan menerapkan data mining dapat memecahkan masalah dengan menganalisis data yang dimiliki perguruan tinggi.

Dalam mengolah data mahasiswa untuk prediksi telah diselesaikan telah oleh Karamouiz dan Vrettos dengan menggunakan metode *neural network* (Vrettos, 2009), Qudri dan Kalyankar dengan metode *decision tree* (Kalyankar, 2010), Suhartini dan Ernastuti dengan metode *C4.5* dan *naïve bayes*, (Ernastuti, 2010), Hastuti dengan komparasi metode *Logistic Regression*, *Decision Tree*, *Naïve Bayes*, *Neural Network* (Hastuti, 2012) dan

Tahyudin, Utami dan Amborowati dengan mengkomparasi algoritma *decision tree*, *naïve bayes*, *ANN*, *Support Vector Machine* (SVM) dan *Logistic Regression* (LR) (Tahyudin, 2013). Tetapi belum ada yang menggunakan metode klasifikasi data mining lainnya seperti *K-Nearest Neighbour*, *ID3*, *CHAID* dan *Learning Discriminant Analysis* (LDA).

Untuk itu dalam penelitian ini akan dilakukan prediksi kelulusan mahasiswa dengan metode *K-Nearest Neighbour*

Rumusan masalah

Seberapa besar tingkat akurasi algoritma *K-Nearest Neighbor* dalam memprediksi ketepatan kelulusan mahasiswa

Tujuan dan Manfaat

Tujuan penelitian ini adalah mengetahui tingkat akurasi algoritma *K-Nearest Neighbour* dalam memprediksi ketepatan kelulusan mahasiswa.

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Memberikan kontribusi keilmuan pada penelitian bidang klasifikasi data mining khususnya untuk prediksi kelulusan mahasiswa.
2. Membantu administrasi perguruan tinggi untuk memberikan peringatan dini dan pembimbingan awal bagi mahasiswa yang kemungkinan tidak lulus tepat waktu.
3. Membantu perguruan tinggi dalam membuat kebijakan untuk bisa meningkatkan kelulusan mahasiswa

TINJAUAN PUSTAKA

Penelitian terkait

Penelitian yang dilakukan oleh Karamouiz dan Vretoz pada tahun 2009 dengan judul *Sentivity Analysis of Neural Network for Identifying the Factors for Collage Students Success*. Masalah yang dikaji adalah tingkat kelulusan dianggap sebagai indikator efektivitas suatu lembaga institusi, Metode yang digunakan

adalah NN (*Neural Network*). Dari hasil data *training* yang dilakukan diperoleh kategori yang lulus adalah 86.04% dan *training* data untuk kategori yang tidak sukses adalah 68.21%, dan *error* yang diperoleh untuk kedua kategori tersebut adalah 0.18% (Vrettos, 2009).

Penelitian yang dilakukan oleh Qudri dan Kalanyar pada tahun 2010 dengan judul *Drop Out Feature of Student Data for Academic Performance Using Decision Tree techniques*. Masalah dalam penelitiannya adalah prestasi akademik siswa sangat penting bagi lembaga pendidikan karena program-program strategis dapat direncanakan untuk meningkatkan atau mempertahankan prestasi siswa selama periode mereka studi di lembaga. Metode yang digunakan adalah *Decision Tree*, yakni algoritma J4.8. Hasil penelitian ini adalah sebuah pohon keputusan yang dapat dijadikan *rule* bagi prediksi siswa yang putus sekolah (Kalyankar, 2010).

Penelitian yang dilakukan oleh Suhartina dan Ernastuti pada tahun 2010 dengan judul dengan judul *Graduation Prediction of Gunadarma University Students Using Algorithm and Naive Bayes C4.5 Algorithm*. Permasalahannya adalah banyaknya mahasiswa yang tidak lulus tepat waktu. Untuk mengetahui tingkat kelulusan mahasiswa dalam satu tahun ajaran dapat dilakukan suatu prediksi berdasarkan data-data mahasiswa pada tahun ajaran pertama. Algoritma yang digunakan adalah C45 dan *naïve bayes*. Hasil dari penelitian ini adalah akurasi untuk metode *naïve bayes* adalah 80,85% dengan presentasi kesalahan 19,05% Akurasi ketepatan hasil prediksi C4.5 85.7%, dan presentasi kesalahannya adalah 14,3% (Ernastuti, 2010).

Penelitian yang dilakukan oleh Hastuti pada tahun 2012 dengan judul analisis komparasi algoritma klasifikasi data mining untuk prediksi mahasiswa non aktif. Permasalahannya adalah mahasiswa non aktif adalah mahasiswa yang berhenti studi dan tidak melakukan registrasi administratif. Mahasiswa yang memiliki status non aktif memiliki kecenderungan untuk drop out.

Metode yang digunakan adalah *Logistic Regression*, *Decision Tree*, *Naïve Bayes*, *Neural Network*. Hasilnya adalah akurasi *Logistic Regression* 81,64%, *Decision Tree* 95,29%, *Naïve Bayes* 93,47%, dan *Neural Network* 94,59%.

Selanjutnya penelitian yang dilakukan oleh Tahyudin, Utami dan Amborowati pada tahun 2013 dengan judul *Comparing Clasification Algorithm Of Data Mining to Predict the Graduation Students on Time*. Permasalahannya adalah persentase mahasiswa yang lulus tepat waktu adalah salah satu unsur yang mempengaruhi akreditasi program studi. Metode yang digunakan adalah mengkomparasi algoritma *decision tree*, *naïve bayes*, *ANN*, *Support Vector Machine* (SVM) dan *Logistic Regression* (LR). Hasilnya adalah akurasi algoritma *decision tree* 80,01%, *naïve bayes* 75,16%, *ANN* 100%, *SVM* 100%, dan *LR* 100%. (Tahyudin, 2013)

Selain algoritma klasifikasi data mining diatas, ada algoritma klasifikasi yang lainnya untuk melakukan suatu prediksi seperti *K-Nearest Neighbour*, *ID3*, *CHAID* dan *Learning Discriminant Analysis* (LDA) (Gorunescu, 2011).

Jadi dalam penelitian ini akan dilakukan analisis prediksi menggunakan metode klasifikasi data mining *K-Nearest Neighbour* (KNN) untuk mengetahui seberapa besar tingkat akurasi dalam prediksi ketepatan kelulusan mahasiswa.

Landasan teori

Dalam penelitian ini menggunakan tinjauan pustaka didasarkan pada beberapa buku, jurnal dan paper yang mendukung teoritis dari penelitian ini. Diantaranya:

Kelulusan Mahasiswa

Mahasiswa sering disebut kelompok masyarakat yang memiliki ciri intelektualitas yang lebih luas dibandingkan dengan kelompok usia 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 (Azwar, 2004)

Kelulusan mahasiswa adalah hal yang penting diperhatikan, karena persentase jumlah kelulusan mempengaruhi penilaian pemerintah serta mempengaruhi status akreditasi program studi (Vrettos, 2009). Faktor-faktor yang dapat mempengaruhi kelulusan mahasiswa antara lain adalah nilai akhir SMA, Indeks Prestasi Semester (IPS), gaji orang tua dan pekerjaan orang tua (Ernastuti, 2010). Indeks prestasi sering digunakan sebagai indikator penilaian akademik, banyak perguruan tinggi memberi standar minimum yang sulit di peroleh mahasiswa (Oyelade, 2010). Adapun variabel yang dapat digunakan dalam prediksi kelulusan mahasiswa seperti umur, status pernikahan, jumlah saudara (Yingkuachat, 2007).

Pada penelitian ini parameter yang digunakan adalah usia, jenis kelamin, indeks prestasi semester satu sampai dengan indeks prestasi semester empat.

Data Mining

Data mining adalah serangkaian proses mendapatkan pengetahuan atau pola dari kumpulan data (Ian H. Witten, 2011). *Data mining* akan memecahkan masalah dengan menganalisis data yang telah ada dalam basis data. *Data mining*, sering juga disebut *knowledge discovery in database* (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan pola keteraturan, pola hubungan dalam set data berukuran besar (Santoso, 2007). Hasil keluaran dari data mining ini dapat dijadikan untuk memperbaiki pengambilan keputusan di masa depan.

Dalam penelitian ini akan memanfaatkan *data mining* untuk mengklasifikasi data mahasiswa sehingga hasil keluarannya akan dimanfaatkan untuk keperluan prediksi kelulusan mahasiswa.

Metode Klasifikasi Data Mining

Klasifikasi adalah proses penemuan model (atau fungsi) yang menggambarkan

dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Kamber, 2006).

Klasifikasi data terdiri dari 2 langkah proses. Pertama adalah *learning* (fase *training*), dimana algoritma klasifikasi dibuat untuk menganalisa data *training* lalu direpresentasikan dalam bentuk *rule* klasifikasi. Proses kedua adalah klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari *rule* klasifikasi (Kamber, 2006).

Proses klasifikasi didasarkan pada empat komponen (Gorunescu, 2011):

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

Berikut ini adalah algoritma klasifikasi data mining yang paling populer yaitu adalah (Gorunescu, 2011):

- Decision/classification trees*
- Bayesian classifiers/Naive Bayes classifiers*
- Neural networks*
- Statistical analysis*
- Rough sets*
- k-nearest neighbor classifier*
- Rule-based methods*
- Memory based reasoning*
- Support vector machines*.

Algoritma *K-Nearest Neighbor* (KNN)

Algoritma *K-Nearest Neighbor* merupakan metode klasifikasi yang mengelompokan data baru berdasarkan jarak data baru itu kebeberapa data/tetangga (neighbor) terdekat (Santoso, 2007).

Teknik *K-Nearest Neighbor* dengan melakukan langkah-langkah yaitu (Santoso, 2007), mulai input: Data training, label data training, *k*, data testing

- Untuk semua data testing, hitung jaraknya ke setiap data training
- Tentukan *k* data training yang jaraknya paling dekat dengan data
- Testing
- Periksa label dari *k* data ini
- Tentukan label yang frekuensinya paling banyak
- Masukan data testing ke kelas dengan frekuensi paling banyak
- Berhenti

Label untuk semua data testing didapat.

Untuk menghitung jarak antara dua titik *x* dan *y* bisa digunakan jarak Euclidean sebagai berikut:

$$d(X_1, Y_2) = \sum_i \left| \frac{n_{1i}}{n_1} - \frac{n_{2i}}{n_2} \right|$$

Yang mana X_1 , $l = 1, 2$, adalah atribut kategori, dan n_{lj} , n_l mewakili frekuensi yang sesuai.

Kerangka pemikiran

Sesuai dengan permasalahan diatas, maka kerangka pemikiran yang dilakukan adalah:



Gambar 2.1 Kerangka Pemikiran

Pengolahan awal data

Jumlah data awal yang diperoleh dari pengumpulan data yaitu sebanyak 1.633 data, namun tidak semua data dapat digunakan dan tidak semua atribut digunakan karena harus melalui beberapa tahap pengolahan awal data (*preparation data*). Untuk mendapatkan data yang berkualitas, beberapa teknik yang dilakukan adalah sebagai berikut (Vercellis, 2009)

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
2. *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.

Metode dan Evaluasi

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 menggunakan *k-fold cross validation*, kemudian dilakukan evaluasi dan validasi hasil dengan *confusion matrix* dan kurva *ROC*.

EKSPERIMEN DAN EVALUASI

Eksperimen

Setelah melakukan pengolahan data awal, data menjadi 1582 siswa dengan jumlah kasus yang “tepat” waktu berjumlah

671 siswa sedangkan kasus yang “terlambat” berjumlah 911 siswa. Dan atribut yang dipakai dalam eksperimen yaitu; fakultas, jenis kelamin, umur, IPK sari semester 1 sampai 4.

Model gambar dibawah ini adalah hasil dari eksekusi data diatas yaitu:

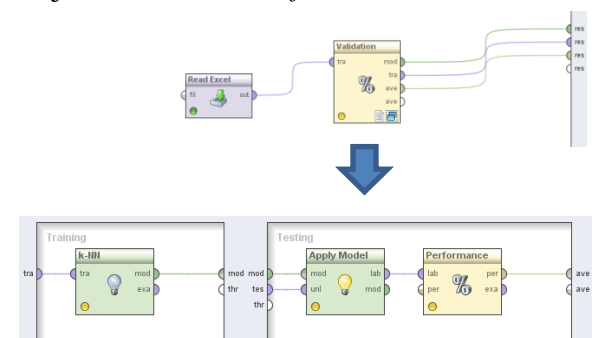
```
KNNearestClassification
1-Nearest Neighbour model for classification.
The model contains 1582 examples with 8 dimensions of the following classes:
TEPAT
TERLAMBAT
```

Gambar 4.1 Model KNN Clasification dengan k=1

Dengan mengklaster k=1 dari data 1582 mahasiswa, 8 dimensi terdiri dari 2 kelas yaitu “TEPAT” dan “TERLAMBAT”.

Pengujian Model

Pengujian model dalam penelitian ini menggunakan *Cross Validation* adalah teknik validasi dengan membagi data secara acak kedalam *k* bagian dan masing-masing bagian akan dilakukan proses klasifikasi (Han & Kamber, 2006). Dengan menggunakan *cross validation* akan dilakukan percobaan sebanyak *k*. Data yang digunakan dalam percobaan ini adalah data *training* untuk mencari nilai *error rate* secara keseluruhan. Secara umum pengujian nilai *k* dilakukan sebanyak 10 kali untuk memperkirakan akurasi estimasi. Dalam penelitian ini nilai *k* yang digunakan berjumlah 10 atau *10-fold Cross Validation*.



Gambar 4.2 Ilustrasi 10 Fold Cross Validation Algoritma *K-Nearest Neighbor*

Evaluasi dan Validasi

Confusion matrix

Confusion matrix memberikan keputusan yang diperoleh dalam *training* dan

testing (Bramer, 2006). *confusion matrix* memberikan penilaian *performance* klasifikasi berdasarkan objek dengan benar atau salah (Gorunescu, 2011).

accuracy: 82.24% +/- 2.07% (micro: 82.24%)			
	true TEPAT	true TERLAMBAT	class precision
pred. TEPAT	535	145	78.86%
pred. TERLAMBAT	136	766	84.92%
class recall	79.73%	84.09%	

Tabel 4.1 Tingkat akurasi Algoritma *K-Nearest Neighbor*

Jumlah *True Positive* (TP) adalah 671 *record* diklasifikasikan sebagai TEPAT terpilih dan *False Negative* (FN) sebanyak 145 *record* diklasifikasikan sebagai TEPAT terpilih tetapi TERLAMBAT terpilih. Berikutnya 766 *record* untuk *True Negative* (TN) diklasifikasikan sebagai TERLAMBAT terpilih, dan 136 *record* *False Positive* (FP) diklasifikasikan sebagai TERLAMBAT terpilih ternyata TEPAT. Berdasarkan Gambar 4.2 tersebut menunjukkan bahwa, tingkat akurasi dengan menggunakan algoritma C4.5 adalah sebesar 85,15% dengan kondisi $k=5$.

Dibawah ini adalah perbandingan tingkat akurasi dengan kondisi k yang berbeda-beda:

Tabel 4.2 Perbandingan jumlah cluster dan akurasi

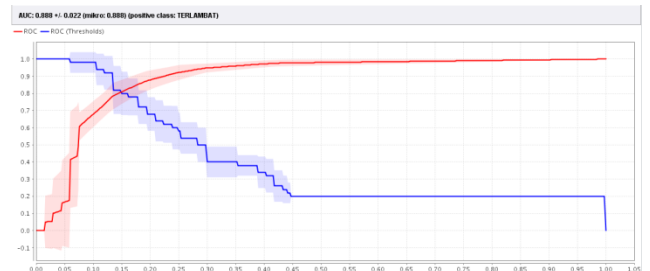
Cluster	Akurasi
<i>K1</i>	82,25%
<i>K2</i>	79,45%
<i>K3</i>	83,95%
<i>K4</i>	82,62%
<i>K5</i>	85,15%

Dari perbandingan diatas, tingkat akurasi dapat dipengaruhi oleh jumlah kalstering data.

Evaluasi *ROC curve*

Untuk dapat melihat akurasi secara manual dilakukan perbandingan klasifikasi menggunakan *curva ROC* hasil ekspresi dari *confusion matrix*. Kurva *ROC* (*Receiver Operating Characteristic*) adalah cara lain untuk mengevaluasi akurasi dari klasifikasi secara visual (Vercellis, 2009).

Dalam penelitian ini, setelah di evaluasi dengan *ROC curve*, menghasilkan nilai *AUC* (*Area Under Curve*) sebesar 0,888 dengan klatering data $k=5$.



Gambar 4.3 Grafik *ROC Curve* Algoritma *K-Nearest Neighbor*

Tabel 4.3 Perbandingan jumlah cluster dan akurasi dengan *ROC Curve*

Cluster	Akurasi
<i>K1</i>	0,500
<i>K2</i>	0,826
<i>K3</i>	0,853
<i>K4</i>	0,874
<i>K5</i>	0,888

Tingkat akurasi dapat di diagnosa sebagai berikut (Gorunescu, 2011):

- Akurasi 0.90 – 1.00 = *Excellent classification*
- Akurasi 0.80 – 0.90 = *Good classification*
- Akurasi 0.70 – 0.80 = *Fair classification*
- Akurasi 0.60 – 0.70 = *Poor classification*
- Akurasi 0.50 – 0.60 = *Failure*

Sesuai dengan grafik dan tabel diatas, tingkat akurasi dengan *ROC Curve* menunjukkan akurasi yang tergolong *Good Classification* yaitu nilai *AUC* (*Area Under Curve*) sebesar 0,888 dengan klastering data $K=5$.

KESIMPULAN DAN SARAN

Dalam penelitian ini dilakukan pengujian model dengan menggunakan algoritma *K-Nearest Neighbor* (*K-NN*) dengan menggunakan data kelulusan mahasiswa yang tepat dan terlambat. Model yang dihasilkan diuji untuk mendapatkan nilai *accuracy* dan *AUC* dari algoritma klasifikasi data mining sehingga didapat pengujian data mahasiswa dengan klatering data $k=1$ dengan menggunakan *K-Nearest*

Neighbor (K-NN) didapat nilai *accuracy* adalah 82,25% dan nilai *AUC* adalah 0.500, dengan cluster data $k=2$ *accuracy* adalah 79,45% dan nilai *AUC* adalah 0.826, dengan cluster data $k=3$ *accuracy* adalah 83,95% dan nilai *AUC* adalah 0.853, dengan cluster data $k=4$ *accuracy* adalah 82,62% dan nilai *AUC* adalah 0.874, dengan cluster data $k=5$ *accuracy* adalah 85,15% dan nilai *AUC* adalah 0.888.

Maka dengan demikian dapat disimpulkan tingkat akurasi pengujian model kelulusan mahasiswa dengan menggunakan algoritma *K-Nearest Neighbor (K-NN)* dipengaruhi oleh jumlah klustering data. Akurasi dan nilai *AUC* paling tinggi adalah dengan mengklaster data $k=5$

Dari hasil pengujian yang telah dilakukan dan hasil kesimpulan yang diberikan maka ada saran atau usul yang di berikan antara lain:

1. Untuk penelitian selanjutnya dapat menggunakan algoritma lain seperti *ID3*, *CHAID* dan *Learning Discriminant Analysis (LDA)* atau dengan mengkomparasinya sehingga diketahui akurasi yang tertinggi.
2. Hasil penelitian ini diharapkan bisa digunakan untuk keperluan perguruan tinggi khususnya prediksi kelulusan mahasiswa.

DAFTAR PUSTAKA

Buku VI Matriks Penilaian Instrumen Akreditasi Program Studi Badan Akreditasi Nasional Perguruan Tinggi. (2008).

Azwar. (2004). *Penyusunan Skala Psikologi.* Yogyakarta: Pustaka Pelajar.

Bramer, M. (2006). *Principles of Data Mining.* London: Springer-Verlag.

Ernastuti, S. &. (2010). Graduation Prediction of Gunadarma University Students Using Algorithm and Naive Bayes C4.5 Algoritmh.

Gorunescu, F. (2011). *Data Mining Concepts Models and Techniques.* Craiova: Springer.

Han, & Kamber. (2006). *Data Mining Concepts and technique.* San Francisco: Diane Cerra.

Hastuti, K. (2012, Juni). ANALISIS KOMPARASI ALGORITMA KLASIFIKASI DATA MINING V. *Seminar Nasional Teknologi Informasi & Komunikasi Terapan*(979 - 26 - 0255 - 0), 241-249.

Ian H. Witten, f. E. (2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3 ed.). (A. S. Burlington, Ed.) United States of America: Morgan Kaufmann.

Kalyankar, Q. &. (2010). Drop Out Feature of Student Data for Academic Performance Using Decision Tree techniques. *Global Journal of Computer Science and Technology*, 2-4.

Kamber, H. &. (2006). *Data Mining Concept and Tehniques.* San Fransisco: Morgan Kauffman.

Ogor. (2007). *Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques.*

Oyelade, O. &. (2010). Application of kmeans Clustering algorithm for predicting of Students Academic Performace. *International Journal of Computer Science and Information Security*, 292-295.

Riduwan. (2008). *Metode dan Teknik Menyusun Tesis.* Bandung: Alfabeta.

Santoso, B. (2007). *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis* (1 ed.). Yogyakarta: Graha Ilmu.

Tahyudin, I. (2013, December). Comparing Clasification Algorithm Of Data Mining to Predict the Graduation Students on Time. *Information Systems International Conference (ISICO).*

Vercellis. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making Decision Making.*

John Willey & Sons Inc: Southern Gate.

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).

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.