EQUILIBRIA PENDIDIKAN
Jurnal Ilmiah Pendidikan Ekonomi
Vol. 4, No. 2, 2019

http://journal.upgris.ac.id/index.php/equilibriapendidikan

STUDENT GRADUATION PREDICTION USING ALGORITMA K-MEANS WITH FITUR SELECTION CHI SQUARE

MF Darmaputra[1], Nugroho Dwi Saputro [2]
fajardharma@upgris.ac.id, *nugputra1@gmail.com*
Universitas PGRI Semarang

Abstrak

Prediksi masa studi sangatlah dibutuhkan oleh manajemen perguruan tinggi dalam menentukan kebijakan preventif terkait dengan pencegahan sejak awal kasus DO (drop out) Penelitian ini bertujuan untuk menentukan faktor akademis yang berpengaruh terhadap mahasiswa tersebut dapat lulus tepat waktu atau tidak dan membangun model prediksi terbaik dengan teknik data mining .Kriteria pemilihan model yang digunakan adalah dengan metode fitur selection chi square. Dengan Algoritma K-Means dihasilkan bahwa lama studi dipengaruhi oleh indeks prestasi per semester, jumlah mata kuliah mengulang , jumlah mata kuliah yang ditempuh dan jumlah pengambilan mata kuliah tertentu. Oleh sebab itu faktor - faktor tersebut dapat digunakan sebagai bahan evaluasi bagi pihak pengelola perguruan tinggi.

Kata Kunci: chi square,k-means, prediksi, mahasiswa, data mining.

Abstract

Prediction study period is required by the university management in determining the policy of preventive related to prevention since the beginning of the case DO (Drop Out) This study aims to determine the factors of academic influence on these students can graduate on time or not, and build predictive models best with data techniques .Criteria elections mining model used is the chi-square method selection feature. K-Means algorithms generated by that period of study is influenced by indeks prestasi of semester, the number of repeated courses, courses taken and the shots certain subjects. Therefore factors - these factors can be used as an evaluation for the manager of the college.

Keywords: Chi Square, K-Means, prediction, students,datamining..

ISSN
2548-6535 (print)
2615-6784 (online)

Pendahuluan

Dewasa ini bidang ilmu komputasi telah berkembang hingga mampu membantu aspek-aspek kebutuhan manusia yang lebih mendalam. *Data Mining* adalah merupakan satu cabang bidang ilmu dari ilmu komputasi yang memanfaatkan sekumpulan data yang sangat besar untuk digali dan di *extract* guna mendapatkan informasi yang lebih bermanfaat. Sejumlah cabang pengetahuan dari *Data Mining* banyak berkembang dengan variasi dan peruntukan yang berbeda-beda. Diantara cabang-cabang tersebut adalah prediksi, estimasi, asosiasi, *clustering*, klasifikasi dan lain-lain.

Sejumlah penelitian terkait prediksi telah banyak dilakukan peneliti yang berasal dari dalam negeri maupun luar negeri, beberapa diantaranya ada penelitian yang terkait dengan permasalahan prediksi kinerja akademik mahasiswa, beberapa metode juga telah diusulkan oleh peneliti untuk mendapatkan tingkat akurasi yang lebih besar.

Analisa Kinerja Mahasiswa pada sistem pembelajaran jauh dengan algoritma genetik dan pohon keputusan analisa ini di perguruan tinggi di *Hellenic Ope Univercity* melalui kemampuan pekerjaan rumah, yang pada akhirnya akan diperoleh hubungan dengan tingkat keberhasilan di ujian akhir (*final exam*) atau dengan kata lain bahwa kinerja akademik mahasiswa berkaitan langsung dengan tingkat kelulusan dari siswa

Berdasarkan data dan penelitian di atas tampak bahwa dengan isu yang terkait dengan prediksi kelulusan masih menjadi trend penelitian di bidang komputasi. Tampak pula paparan diatas bahwa sejumlah algoritma telah diusulkan untuk mendapatkan hasil akurasi yang lebih baik. Namun demikian tingkat ideal masih dari jauh dari capaian, untuk itu riset terkait prediksi kelulusan perlu dilakukan sebagai usaha untuk menyelidiki permasalahan yang ada, penulis memilih menggunakan algoritma *K-Means* dalam penelitian ini karena ada kelebihan dari algoritma ini antara lain

a. Daerah pengambilan keputusan yang sebelumnya komplek dan global dapat diubah menjadi *simple* dan spesifik;
b. Sampel yang diuji berdasarkan kriteria atau kelas tertentu

Pernyataan Masalah

Berdasarkan latar belakang diatas algoritma *K-Means* banyak digunakan pada penelitian tentang bidang pendidikan khususnya tentang prediksi kelulusan mahasiswa. Dari komparasi yang dilakukan oleh beberapa peneliti, algoritma *K-Means* dianggap memiliki nilai yang baik dibanding algorima lain antara lain daerah pengambilan keputusan yang sebelumnya komplek, global dan luan dapat dirubah menjadi *simple* atau sederhana serta spesifik. Namun demikian algoritma *K-Means* ini *relative sensitive* terhadap dataset untuk sejumlah kasus algoritma *K-Means* memiliki tingkat akurasi yang lebih baik namun dalam kasus tertentu justru sebaliknya dan dataset ini sangat berpengaruh terhadap akurasi yang dihasilkan *K-Means*. Dengan demikian dapat disimpulkan algoritma *K-Means* masih kurang akurat.

Pertanyaan Masalah

Bagaimanakah tingkat akurasi algoritma *K-Means* untuk prediksi kelulusan mahasiswa apabila dikombinasikan dengan metode fitur seleksi *chi square* pada *dataset*?

Tujuan Penelitian

Tujuan dari penelitian ini adalah meningkatan akurasi algoritma *K-Means* dengan melakukan seleksi atribut terlebih dahulu dengan menggunakan *fitur selection chi square* untuk memprediksi kinerja akademik mahasiswa.

Manfaat Penulisan

a. Manfaat praktis dari dari penelitian ini adalah diharapkan agar dapat digunakan oleh para praktisi dunia pendidikan seperti dosen dan tenaga kependidikan untuk memprediksi kinerja akademik mahasiswa di sebuah perguruan tinggi.
b. Manfaat teoritis dari hasil penelitian ini diharapkan memberikan sumbangsih bagi pengembang teori yang berkaitan dengan prediksi kinerja akademik mahasiswa dengan menggunakan algoritma *K-Means* dan metode *chi square*

Penelitian Terkait

Beberapa penelitian yang tekait dengan prediksi kinerja akademik mahasiswa pernah dilakukan oleh beberapa peneliti. Kovacic pada tahun 2010 melakukan penelitian prediksi keberhasilan studi mahasiswa menggunakan data pendaftaran. Data yang digunakan berasal dari Politeknik New Zealand yang merupakan data dari 450 orang mahasiswa dari tahun 2006 hingga 2009. Dalam penelitian ini, digunakan algoritma *CHAID* dan *CART*. Algoritma *CHAID* menghasilkan nilai akurasi sebesar 59,4% sedangkan *CART* menghasilkan akurasi sebesar 60,5%.

Bharadwaj dan Pal melakukan penelitian terhadap kinerja mahasiswa menggunakan algoritma *Naïve Bayes*. *Naïve Bayes* dipilih karena dianggap sebagai algoritma yang mudah digunakan dan memerlukan data pelatihan dalam jumlah kecil untuk memperkirakan parameter yang diperlukan dalam proses klasifikasi. Data diambil dari 300 mahasiswa dari beberapa berguruan tinggi berbeda yang terbagi menjadi 17 atribut. Atribut yang nilai probabilitasnya lebih besar dari 0,50 dianggap memiliki pengaruh terhadap kinerja mahasiswa.

Pada tahun 2014, Ogunde dan Ajibade membuat sebuah sistem yang berfungsi untuk memprediksi kelulusan mahasiswa menggunakan algoritma ID3. Data yang digunakan pada penelitian ini berasal dari Universitas Redeemer Nigeria. Dari penelitian ini, didapatkan akurasi ID3 sebesar 79,556%.

Landasan Teori

Penerimaan Mahasiswa Baru dan wisuda merupakan kegiatan yang selalu dilaksanakan dalam sebuah perguruan tinggi. Pada kegiatan Penerimaan Mahasiswa Baru, calon mahasiswa mendaftar pada program studi yang diminati dengan mengikuti berbagai tes yang menjadi persyaratan. Setelah menempuh seluruh mata kuliah dan mengerjakan skripsi atau tugas akhir, mahasiswa dinyatakan lulus dan dapat diwisuda.

Permasalahan yang selama ini muncul dalam sebuah perguruan tinggi adalah ketidak seimbangan atara jumlah mahasiswa yang dalam suatu angkatan dengan jumlah mahasiswa yang lulus. Ketidak seimbangan ini muncul karena adanya keterlambatan kelulusan mahasiswa. Keterlambatan kelulusan akan menjadi masalah bagi mahasiswa karena selain harus menambah biaya administrasi kuliah juga menjadi hambatan dalam meraih kesempatan di dunia kerja. Bagi perguruan tinggi sendiri, keterlambatan kelulusan akan membawa dampak menurunnya tingkat kepercayaan masyarakat terhadap perguruan tinggi tersebut serta akan berdampak terhadap akreditasi yang dikeluarkan oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) serta dapat menjadi beban bagi Universitas karena rasio antara dosen dan mahasiswa menjadi sangat besar. Apabila Universitas di dalam rasio antara dosen dan mahasiswa tidak ideal maka akan mendapat teguran dari pemerintah dalam hal ini Direktorat Pendidikan Tinggi di bawah Kementrian Riset dan Teknologi

Pendidikan Tinggi. Teguran ini sangat berpengaruh sekali terhadap kehidupan di kampus, antara lain dihentikannya pelayanan penelitian dan pelayanan online lainnya sampai yang paling ekstrim adalah pencabutan ijin penyelenggaraan perguruan tinggi.

Data Mining

Proses yang menggunakan teknik statistik, matematika kecerdasan buatan dan *machine learning* untuk mengektraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar.

Kemajuan luar biasa yang terus berlanjut dalam bidang *data mining* didorong oleh beberapa faktor, antara lain:

1. Pertumbuhan yang cepat dalam kumpulan data;
2. Penyimpanan data dalam data *warehouse* sehingga seluruh perusahaan memiliki akses ke dalam *database* yang andal;
3. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi;
4. Perkembangan teknologi perangkat lunak untuk *data mining* (ketersediaan teknologi);
5. Adanya peningkatan akses data melalui navigasi web dan internet.

Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan

Metode *Chi Square*

Data berdimensi besar menjadi salah satu masalah dalam *data mining*. Ribuan fitur atau variabel dalam dataset berisi informasi yang tidak berkontribusi terhadap kemampuan memprediksi dan menurunkan kinerja algoritma.

Teknik reduksi dimensi dibagi menjadi dua kategori yaitu ekstraksi fitur dan seleksi fitur Ekstrasi fitur mengurangi dimensi dengan tranformasi aljabar linear. Pada seleksi fitur dilakukan pemilihan pada variabel berdasarkan hubungan antara variabel sehingga dimensi fitur terkurangi, data yang *redundant*, *noise* dan tidak relevan dihapus. Selain mengurangi fitur, peningkatan akurasi juga akan terjadi.

Seleksi fitur dibagi menjadi tiga kelas yaitu *filter method*, *wrapper method* dan *embedded method*. Algoritma *chi square* merupakan salah satu algoritma yang termasuk dalam *filter method* selain *Information Gain* dan *Correlation Coefficient Scores*. *Chi square* sangat umum digunakan untuk mengevaluasi nilai dari fitur dengan menghitung nilai statistik *chi square* untuk mengetahui hubungannya dengan kelas.

Persamaan *chi square* adalah sebagai berikut :

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \quad (1)$$

Oi = frekuensi yang diperoleh
Ei = frekuensi yang diharapkan

Uji *chi-square* merupakan uji non parameter yang paling banyak digunakan. Namun perlu diketahui syarat-syarat uji ini adalah: frekuensi responden atau sampel yang digunakan besar, sebab ada beberapa syarat di mana *chi square* dapat digunakan yaitu:

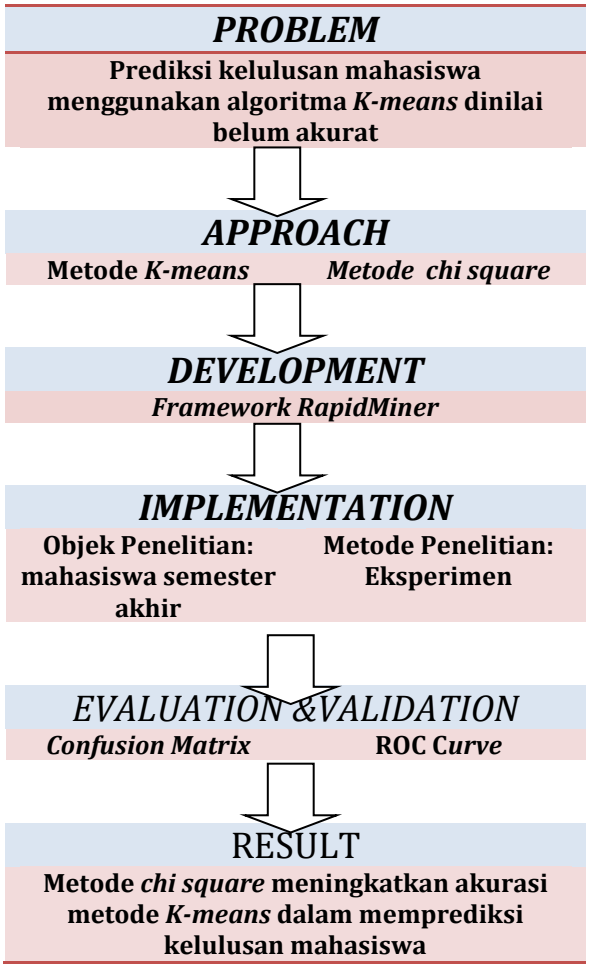
1. Tidak ada *cell* dengan nilai frekuensi kenyataan atau disebut juga *Actual Count* (F0) sebesar 0 (Nol).
2. Apabila bentuk tabel kontingensi 2X2, maka tidak boleh ada 1 *cell* saja yang memiliki frekuensi harapan atau disebut juga *expected count* (Fh) kurang dari 5.
   Apabila bentuk tabel lebih dari 2x2, misal 2x3, maka jumlah *cell* dengan frekuensi harapan yang kurang dari 5 tidak boleh lebih dari 20%.

Apabila tabel kontingensi 2x2 seperti di atas, tetapi tidak memenuhi syarat seperti di atas, yaitu ada *cell* dengan frekuensi harapan kurang dari 5, maka rumus harus diganti dengan rumus *Fisher Exact Test*.

Algoritma *K-Means*

K-means merupakan salah satu metode clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster. Metode ini mempartisi data ke dalam cluster sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karateristik yang berbeda di kelompokan ke dalam cluster yang lain.

Kerangka Pemikiran



Metode

Metode yang diusulkan pada penelitian ini adalah menggunakan metode *K-means* dan metode *K-means* dengan metode *chi square* sebagai algoritma penyeleksi atribut

Menunjukan proses yang dilakukan dalam tahap modeling untuk menyelesaikan prediksi kelulusan mahasiswa dengan menggunakan dua metode yaitu algoritma *K-means* dan algoritma *K-means* berbasis *chi square*.

1. Algoritma *K-means* digunakan untuk membangun sebuah pohon keputusan yang mudah dimengerti, fleksibel, dan menarik karena dapat divisualisasikan dalam bentuk gambar.
2. *Seleksi fitur chi square yaitu* metode yang dilakukan pemilihan pada variabel berdasarkan hubungan antara variabel sehingga dimensi fitur terkurangi, data yang *redundant*, *noise* dan tidak relevan dihapus [*14]*

Hasil dan Pembahasan

Tujuan dari penelitina ini yaitu untuk mengetahui atau membandingkan tingkat atau prediksi kelulusan mahasiswa yang termuat atau ada di dalam kinerja akademik mahasiswa dengan Algoritma *K-means* dengan Algoritma *K-means* yang sebelumnya diseleksi dengan *Chi Square*. Hasilnya nanti akan diperoleh tingkat akurasi dan AUC yang tertinggi, atribut awal yang digunakan adalah 17 atribut yang antara lain adalah no, npm, nama, jurusan, SKS semester 1, IP Semester 1, SKS semester 2, IP Semester 2, SKS Semester 3, IP Semester 3, SKS Semester 4, IP Semester 4. SKS semester 5, IP Semester 5, SKS Semester 6, IP Semester

6.sedangkan setelah dilakukan metode seleksi fitur *chi square* menjadi 9 atribut dan yang akan di gunakan yaitu jurusan, SKS semester 1, IP Semester 1, SKS semester 2, IP Semester 2, SKS Semester 3, IP Semester 3, SKS Semester 4, IP Semester 4.

Algoritma *K-means* untuk model yang pertama dilakukan. Berikut langkah-langkah yang akan dilakukan sebagai berikut:

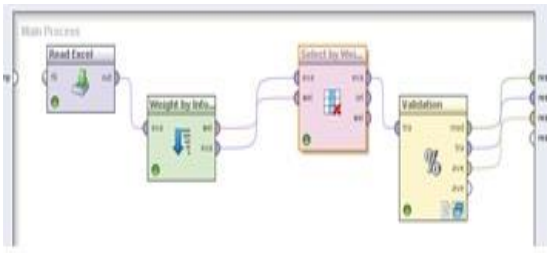
1. Menghitung jumlah kelas Tepat dan kelas Terlambat serta nilai *Entropy* dari semua kasus. Kasus dibagi berdasarkan atribut dengan jumlah kasus 200 *record*, kelas Tepat ada 180 *record* dan kelas Terlambat sebanyak 20 *record* sehingga didapat *entropy*.

$$Entropy(i) = \sum_{j=1}^{m} \mathrm{f}(i,j).2\ \mathrm{f}[(i,j)]$$

= (-20/200.log2(20/200)) + (-180/200.log2(180/200))
= 0.332193 + 0.136803
= 0.468996

2. Hitung nilai Gain dari masing-masing atribut sebagai contoh untuk PPKn.

Berikut adalah gambar pengujian menggunakan motode K-*Fold Cross Validation*



Jurusan = PPKN
| IP smstr 2 = Baik: Tepat {Tepat=185, Terlambat=2}
| | IP smstr 2 = Cukup
| | | IP smstr 1 = Baik
| | | | SKS smster 5 = Lebih: Tepat {Tepat=6, Terlambat=0}
| | | | SKS smster 5 = Paket
| | | | | IP smstr 6 = Baik: Tepat {Tepat=2, Terlambat=0}
| | | | | IP smstr 6 = Cukup: Terlambat {Tepat=3, Terlambat=3}
| | | | SKS smster 5 = Paket Kurang: Terlambat {Tepat=2, Terlambat=2}
| | | IP smstr 1 = Cukup
| | | | IP smstr 3 = Baik: Tepat {Tepat=11, Terlambat=1}
| | | | IP smstr 3 = Cukup
| | | | | SKS smster 6 = Lebih: Tepat {Tepat=5, Terlambat=1}
| | | | | SKS smster 6 = Paket: Terlambat {Tepat=1, Terlambat=1}
| | | | | SKS smster 6 = Paket Kurang: Terlambat {Tepat=1, Terlambat=1}
| | | | IP smstr 3 = Kurang: Tepat {Tepat=2, Terlambat=0}
| | | IP smstr 1 = Kurang: Terlambat {Tepat=1, Terlambat=3}
| | IP smstr 2 = Kurang
| | | IP smstr 1 = Baik: Terlambat {Tepat=0, Terlambat=3}
| | | IP smstr 1 = Cukup: Tepat {Tepat=2, Terlambat=1}
| | | IP smstr 1 = Kurang: Terlambat {Tepat=0, Terlambat=2}

Jumlah *True Positive* (TP) adalah 117 *record* diklasifikasikan sebagai *Tepat* terpilih dan *False Negative* (FN) sebanyak 16 *record* diklasifikasikan sebagai *Tepat* terpilih tetapi *Terlambat* terpilih. Berikutnya 4 *record* untuk *True Negative* (TN) diklasifikasikan sebagai Terlambat terpilih, dan 4 *record False Positive* (FP) diklasifikasin sebagai *Terlambat* terpilih ternyata *Tepat* Berdasarkan Gambar 4.1 tersebut menunjukan bahwa, tingkat akurasi dengan menggunakan algoritma *K Means* adalah sebesar 85.81%, dan dapat dihitung untuk mencari nilai *accuracy*, *sensitivity*, *specificity*, *ppv*, dan *npv* pada persamaan dibawah ini:

$$acc = \frac{tp + tn}{tp + tn + fp + fn} \qquad acc = \frac{117 + 4}{117 + 4 + 4 + 16}$$

$$Sensitivity = \frac{tp}{tp + fn}$$

$$Specitivity = \frac{tn}{tn + fp}$$

$$ppv = \frac{tp}{tp + fp}$$

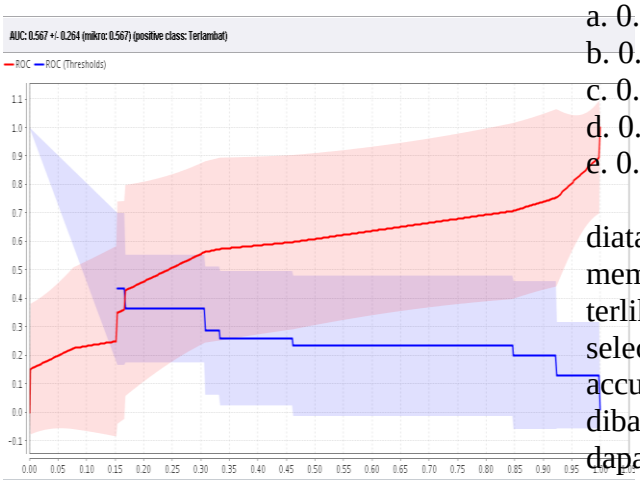
$$npv = \frac{tn}{tn + fn}$$

$$Sensitivity = \frac{117}{117 + 16}$$

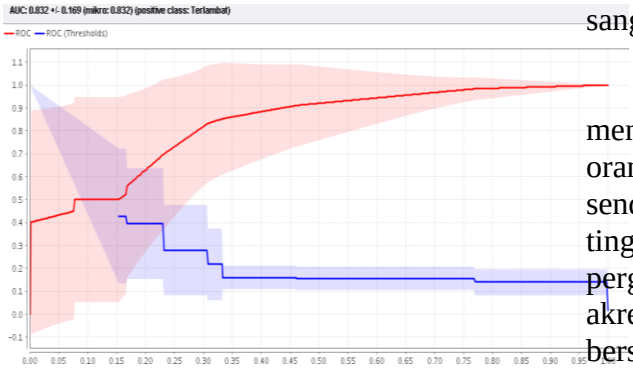
$$Specitivity = \frac{4}{4 + 4}$$

$$ppv = \frac{117}{117 + 4}$$

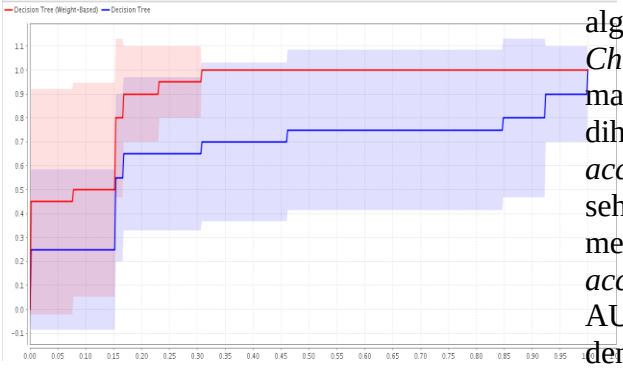
$$npv = \frac{4}{4 + 16}$$



Nilai AUC dalam grafik ROC algoritma *K-Means*



Nilai AUC dalam grafik ROC algoritma *K-means* fitur selection ChiSquare



Komparasi *ROC Curve* pada algoritma *K-means* dan algoritma *K-means* selection ChiSquare

Dari hasil perhitungan model pada gambar 6 dengan penerapan klasifikasi *performance* keakurasian AUC maka dapat diklasifikasikan menjadi lima kelompok [12], antara lain:

a. 0.50 – 0.60 = klasifikasi salah
b. 0.60 – 0.70 = klasifikasi buruk
c. 0.70 – 0.80 = klasifikasi cukup
d. 0.80 – 0.90 = klasifikasi baik
e. 0.90 – 1.00 = klasifikasi sangat baik

Jadi berdasarkan pengelompokan diatas pada table 4.1 dengan membandingkan nilai *accuracy* dan AUC terlihat bahwa algoritma *K-means* selection ChiSquare memiliki nilai accuracy dan nilai AUC yang lebih baik dibandingkan algoritma *K-means* dan dapat disimpulkan bahwa nilai AUC *K-means* dan *K-means* dengan fitur selection *Chi Square.*

antara 0.90 – 1.00 termasuk klasifikasi sangat baik

Simpulan dan Saran

Karena keterlambatan dalam kelulusan menjadi suatu masalah bagi mahasiswa, orang tua bahkan perguruan tinggi itu sendiri, karena dapat mempengaruhi tingkat kepercayaan masyarakat terhadap perguruan tinggi tersebut maupun akreditasi dari perguruan tinggi yang bersangkutan. Dalam penelitian ini dilakukan pengujian model dengan menggunakan algoritma *K-means* dan algoritma *K-means* dengan *fitur selection Chi Square* dengan menggunakan data mahasiswa UPGRIS. Model yang dihasilkan diuji untuk mendapatkan nilai *accuracy*, dan AUC dari setiap algoritma sehingga didapat pengujian dengan menggunakan *K-means* didapat nilai *accuracy* adalah 85.81 % dengan nilai AUC adalah 0.567. Sedangkan pengujian dengan mengunakan *K-means 5 fitur selection Chi Square* didapatkan nilai

accuracy 86.57 % dengan nilai AUC adalah. 0.832 % Maka dapat disimpulkan pengujian model prediksi kelulusan pada kinerja akademik Mahasiswa dengan menggunakan algoritma *K-means* dengan *fitur selection Chi Square* lebih baik dari pada *K-means* sendiri, dengan peningkatan akurasi sebesar 8.85 % dan peningkatan nilai AUC sebesar 0.467 %.

Dengan demikian dari hasil pengujian model diatas dapat disimpulkan bahwa *K-means* dengan *fitur selection Chi Square* memberikan hasil yang lebih akurat.dalam mengevaluasi kinerja akademik.

Dari hasil penelitian yang telah dilakukan ini penulis merangkum ada beberapa saran yang bisa dilakukan agar penelitian menghasilkan prediksi yang akurat, saran itu antara lain :

a. Pemilihan data *testing* awal untuk penelitian ini sebaiknya harus yang sama karakteristiknya agar dalam penelitian hasilnya akan lebih baik.
b. Dalam penelitian kedepan dapat dikembangkan dengan menggunakan berbagai macam algoritma yang ada dalam *Data Mining***.**

Daftar Pustaka

B & Pal, S Bharadwaj. 2011 *"Data Mining : A Prediction for Performance Improvement Using Classificaion," International Journal of Computer* Science *and Information Security*, vol. 9,

Chandrima & Sarkar. 2013 *"Correlation Based Feature Selection Using Rank Aggregation for an Improved Prediction of Potentially Preventable Events,"*

Chen, Su, Lin Tsai. 2012 *"An Aplication of PSO Algorithm & Decision Tree for Medical Problem," 2n International Conference on Intelligent Computational System (ICS'2012)*, pp. 124 - 126

C. Vercellis. 2009. *Business Intelligence:* Data Mining and Optimization for Decision Making Decision Making. Southern Gate, Chichester, West Sussex, United Kingdom: John Wiley & Sons Ltd,

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

Jasmina & Novakovic2012. "*Toward Optimal Feature Selection Using Ranking Method and Classification Algorithm,"*

Ke & Wu. 2007. "*A Probabilistic Approach to Feature Selection for Multi-Class Text Categorization*," pp. 1310 - 1317

Kovacic. 2010. "*Early Prediction of Student Success : Mining Students Enrolment Data*," Proceedings of Informing Science & IT Education Conference (InSITE)

Lee Hyeseon, Jun Hyuk-Chi Ju Cho-Yun. 2011. "*Optimization of Decision Tree for Classification Using a Particle Swarm*," vol. 10, pp. 272 - 278

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

M.N & Kalyankar, N.V Quadril. 2010 "*Drop Out Feature of Student Data for Academic Performance Using Decision Tree Technique,"* Global Journal of Computer Science and Technology, pp. 1 – 5.

Ogunde & D.A, Ajibade A.O. 2014. "*A Data Mining System for Predicting University Students' Graduation Grades Using ID3 Decision Tree Algorithm*," Journal of Computer Science and Information Technology, vol. 2, pp. 21 – 46.

Paokanta P. 2012 "B- *Thalassemia Knowledge Elicitation Using Data Engineering : PCA, Pearson's Chi Square and Machine Learning*," International

Journal of Computer Theory and Engineering, vol. 2, pp. 702 – 706

Engineering & Technology, vol. 4, no. 2, pp. 173 - 178

Sonal & Khanna Vineet Saxena. 2014. "*Evaluation of Prediction Models for Tumor Class Detection with Classification Algorithm Employing Chi Square and Information Gain Feature Selection*," International Journal of Advanced Reaserch in Computer Science and Software Engineering, vol. 4, pp. 821 - 825

Tahyudin & Imam. 2013. "*Comparing Classification Algorithm of Data Mining to Predict the Graduation Students on Time*," Information System International Conference

Tsai Cheng Fa Jantawan Bangsuk. 2013. "*The Application of Data Mining to Build Classification Model for Predicying Graduate Employment*," International Journal of Computer Science and Information Security, vol. 11

V Sharanthi Vijay, Devi Gayathri M Kumar Varun. 2012. "*Hepatitis Prediction Model based on Data Mining Algorithm and Optimal Feature Selection to Improve Predictive Accuracy*," International Journal of Computer Application , vol. 51

V Sundaram T Smitha. 2012. "*Comparative Study of Data Mining Algorithm for High Dimensional Data Analysis*," International Journal of Advance in