

RAMALAN PRESTASI AKHIR PELAJAR MELALUI KAEDAH PERLOMBONGAN DATA

[PREDICTING STUDENT FINAL PERFORMANCE THROUGH DATA MINING METHODS]

RIZALINA MOHD RADWAN¹

AIDA MUSTAPHA

BADARIAH ABDOLLAH

Abstrak

Cabaran besar yang dihadapi oleh pensyarah dalam merancang dan melaksanakan pengajaran dan pembelajaran (PdP) berkesan adalah mengenal pasti pelajar yang berprestasi lemah pada peringkat lebih awal sebelum mereka bergraduasi. Bagi menyoal cabaran ini, kajian ini dibangunkan bertujuan untuk meramal prestasi pelajar yang menyambung pengajian di Kolej Komuniti Kuala Langat berdasarkan keputusan kemasukan Sijil Pelajaran Malaysia (SPM) dengan menggunakan kaedah perlombongan data melalui teknik pengklasifikasian. Empat (4) algoritma pengklasifikasian telah digunakan untuk membina model ramalan keputusan akhir peperiksaan melalui Himpunan Purata Nilai Mata (HPNM) iaitu (1) Pepohon Keputusan (J48), *Multilayer Perceptron* (MP), Hutan Rawak (RF) dan ZeroR. Set data yang digunakan di dalam kajian ini diperolehi daripada *Sistem ePelajar* dan *Integrated Student Evaluation Management System* (iSEMS), Jabatan Pengajian Politeknik dan Kolej Komuniti (JPPKK). Keputusan kajian menunjukkan bahawa model pengklasifikasian J48 adalah model terbaik dengan nilai ketepatan 56.56% diikuti oleh algoritma ZeroR dan RF, kedua-dua dengan nilai ketepatan 50.67%. Manakala algoritma MP memperolehi ketepatan terendah dengan peratusan 44.34% sahaja. Selain daripada itu, didapati mata pelajaran Bahasa Melayu (BM) dan Sejarah (SJ) mempunyai pengaruh terhadap pengelasan status HPNM pelajar. Seramai 49/76 orang pelajar telah dikenal pasti mempunyai kebarangkalian HPNM yang rendah di semester akhir. Kajian ini boleh membantu institusi dalam mengenal pasti dari semester awal pelajar yang diramalkan berprestasi rendah supaya program bersasar boleh dianjurkan untuk melonjakkan prestasi pelajar.

Kata Kunci: Prestasi pelajar, Perlombongan data, pelajar TVET, J48, ZeroR, Random Forest, Multilayer Perceptron.

¹ Kolej Komuniti Kuala Langat, Jalan Sultan Suleiman Shah, 42700, Banting, Selangor, MALAYSIA.

² Fakulti Sains Komputer dan Teknologi Maklumat, Universiti Tun Hussein Onn Malaysia, Parit Raja, 86400 Batu Pahat, Johor, Malaysia

Corresponding Author:

RIZALINA MOHD RADWAN, Kolej Komuniti Kuala Langat, Jalan Sultan Suleiman Shah, 42700, Banting, Selangor, Malaysia.

E-mail: rizalina@kkkla.edu.my

Abstract

The biggest challenge faced by lecturers in implementing effective teaching and learning (T&L) is to identify students who perform poorly at an earlier stage for planning an intervention. To address this challenge, this study is set to predict the performance of students at Kuala Langat Community College based on the results of Malaysian Certificate of Education (MCE) using classification-based data mining approach. Four (4) classification algorithms are used to construct the results prediction model via the Cumulative Grade Point Average (CGPA), which are Decision Tree (J48), Multilayer Perceptron (MP), Random Forest (RF), and ZeroR. The dataset used in this study is sourced from the ePelajar System and Integrated Student Evaluation Management System (iSEMS), Department of Polytechnic and Community Colleges (DPCCE). The results showed that the J48 classification model is the best model with an accuracy value of 56.56% followed by ZeroR and RF algorithms, where both produced the same accuracy value of 50.67%. Meanwhile, the MP algorithm obtained the lowest accuracy with a percentage of only 44.34%. In addition, it was found that Malay Language and History has an influence on the student's CGPA classification status. Overall, a total 49/76 of students have been identified to have potentially low CGPA in the final semester. The findings from this study is able to help institutions in identifying weak students in their early semesters so intervention and further development programs can be organized to boost student performance.

Keywords: Student performance, Data mining, student TVET, J48, ZeroR, Random Forest, Multilayer Perceptron.

Cite This Article:

Rizalina Mohd Radwan, Aida Mustapha & Badariah Abdollah. (2020). Ramalan prestasi akhir pelajar melalui kaedah perlombongan data. *Asian Journal of Civilizational Studies (AJOCS)*, 2(3), 1-16.

PENGENALAN

Perlombongan data digunakan untuk melatih data tidak berstruktur agar maklumat penting dapat diekstrak daripada himpunan data yang besar. Pelbagai sektor mula menggunakan kaedah perlombongan data termasuklah sektor pendidikan di Malaysia. Pendidikan merupakan salah satu aspek penting yang menyumbang kepada kualiti kehidupan masyarakat. Ia bermula seawal pendidikan pra-sekolah sehinggalah kepada pendidikan tinggi. Pelajar lepasan menengah mempunyai pilihan sama ada untuk berkerja atau memohon untuk menyambung pengajian. Bagi mereka yang ingin menyambung pengajian, terdapat institusi pengajian tinggi yang menawarkan peluang kepada mereka dengan syarat keputusan peperiksaan yang minimum. Pelajar yang berjaya melanjutkan pelajaran ke institusi pendidikan tersebut akan melalui proses pembelajaran yang berbeza daripada proses pembelajaran di sekolah. Kecemerlangan akademik di sekolah tidak semestinya menjamin kejayaan yang cemerlang di peringkat pendidikan tinggi tetapi boleh diramal berdasarkan kelangsungan sikap atau disiplin diri pelajar-pelajar bermula dari sekolah.

Intervensi awal kecemerlangan di peringkat pendidikan tinggi adalah penting bagi institusi pendidikan menangani masalah kegagalan atau keciciran pelajar bagi mana-mana kursus yang ditawarkan sepanjang pengajian. Pelbagai langkah diambil oleh pihak institusi

pendidikan untuk menangani masalah ini. Walau bagaimanapun, sekiranya program pengukuhan dijalankan tanpa mengkaji latar belakang pelajar di peringkat menengah, dikhuatiri program-program tersebut tidak mendapat impak yang maksimum kerana tidak menepati profil pelajar yang disasarkan. Oleh itu seharusnya institusi mengenal pasti terlebih dahulu latar belakang dan susur galur pendidikan kelompok pelajar lemah sebelum sesuatu program peningkatan akademik dilaksanakan. Objektif kajian ini adalah untuk mengenal pasti kelompok pelajar-pelajar yang diramalkan lemah dalam akademik untuk dibimbing secara tersusun melalui sesi kaunseling dan penglibatan dalam program-program pembangunan kecemerlangan dan aktiviti ilmiah yang lain.

Bagi mencapai objektif tersebut, kajian ini dihasilkan menggunakan metodologi perlombongan data iaitu teknik pengklasifikasian (*Classification*) bagi meramal keputusan akhir peperiksaan melalui Himpunan Purata Nilai Mata (HPNM) para pelajar. Ia bertujuan mengenal pasti samada pelajar bakal berada di dalam kelas Cemerlang, Kepujian atau Baik. Teknik pengklasifikasian adalah teknik pembelajaran terselia yang dapat membantu pihak pengurusan untuk membuat keputusan. Kelebihan teknik pengklasifikasian berbanding teknik lain seperti pengelompokan (*Clustering*) atau perlombongan petua (*Association Rules Mining*) ialah ramalan yang diperolehi adalah berbentuk objek. Ini bermaksud, pelajar yang diramal mendapat keputusan HPNM selain daripada Cemerlang dapat dikenal pasti dan diberi pengukuhan melalui program-program motivasi dan kaunseling, serta program akademik seperti Bengkel *Effective Learning*, Program Pembangunan Kecemerlangan dan Program *Creative Optimistic Leadership*.

KAJIAN KESUSASTERAAN

Perlombongan data merupakan satu teknologi yang menggabungkan kaedah pemprosesan data secara tradisional dengan algoritma berbeza, bertujuan untuk menganalisis jenis data baru dan mengekstrak maklumat (Weiping, 2013). Tujuan utama perlombongan data ada dua iaitu analisis data (deskriptif) dan ramalan data (prediktif). Menurut (Lewichi, 2005), tiga peringkat proses perlombongan data ialah penerokaan awal, pembangunan model (*pattern*) dan penjaanaan ramalan ke atas data baru.

Lambakan data yang terkumpul di institusi pengajian tinggi boleh dilombong bagi mendapatkan corak tertentu yang konsisten terhadap pemboleh ubah di sesebuah institusi pendidikan. Data pelajar bergraduat Kolej Komuniti Kuala Langat pernah digunakan di dalam kajian sebelumnya namun ia digunakan untuk tujuan berbeza. Di dalam kajian Sudyatno (2014), Pepohon Keputusan J48 dibuktikan memiliki ketepatan yang lebih tinggi berbanding dengan Regresi Berganda. Atribut pemboleh ubah yang digunakan oleh penyelidik ini ialah status sosial, ekonomi keluarga, motivasi pelajar, disiplin pelajar dan prestasi lepas.

Jalil et al. (2013) membuktikan bahawa algoritma Naïve Bayesian merupakan algoritma yang mempunyai ketepatan paling tinggi berbanding algoritma Pepohon Keputusan (Decision Tree) dan Multilayer Perceptron. Berbeza pula hasil penyelidikan Permodelan Guru Cemerlang KPM (Halim et al., 2017) yang menunjukkan algoritma Pepohon Keputusan J48 lebih baik berbanding algoritma Hutan Rawak (*Random Forest*).

Kajian yang dibuat oleh Saxena (2015) adalah unik kerana beliau menggunakan dua teknik berbeza ke atas set data yang sama iaitu teknik pengklasifikasian dan pengelompokan.

Hasil daripada dapatan kajiannya mendapati algorithma pengklasifikasian Pepohon Keputusan J48 lebih sesuai digunakan berbanding algorithma pengelompokan K-means. Pada tahun 2018, satu kajian berkaitan data set biasiswa KPT telah dijalankan oleh Ahmad dan Bakar (2018) bertujuan membandingkan lima algorithma teknik *classification* yang memiliki ketepatan tertinggi untuk meramal bakal pemegang biasiswa KPT. Kajian tersebut mendapati algorithma SVM memberikan nilai ketepatan tertinggi manakala Random Forest adalah algorithma berketepatan terendah.

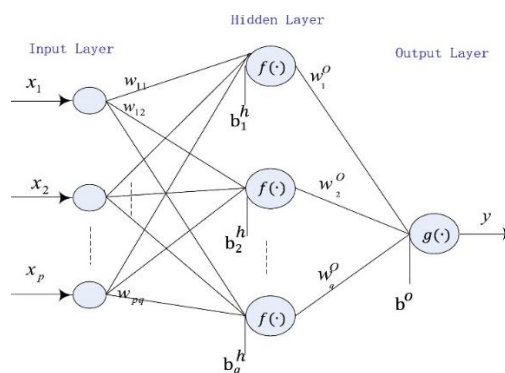
PEPOHON KEPUTUSAN (J48)

Hasil daripada kajian yang dijalankan oleh Fayyad dan Irani (1993) berjaya memperkenalkan kaedah *discretization* yang dikenali sebagai *entropy* dalam algorithma ini. Pepohon Keputusan (*Decision Trees*), J48 merupakan algorithma yang amat popular dikalangan penyelidik teknik klasifikasi kerana ia mudah untuk ditafsirkan dan difahami. Model peramalan yang dihasilkan merupakan pepohon bercabang yang ditampilkan dalam bentuk hierarki. Rumus 1 menunjukkan rumus *entropy* yang diperkenalkan oleh Fayyad dan Irani (1993) berdasarkan jadual kekerapan terhadap setiap satu atribut.

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i \quad (1)$$

Multilayer Perceptron (MP)

Multilayer Perceptron (MP) merupakan rangkaian saraf tiruan. Ia terdiri lebih dari satu *perceptron*. Mereka terdiri daripada lapisan input yang bersedia untuk menerima signal dan lapisan output yang akan meramal mengenai maklumat tersembunyi yang tersimpan didalam lapisan tersembunyi. Rajah 4 menunjukkan rangkaian MP.



Rajah 1: *Multilayer Perceptron*

Menurut Saed Sayad di dalam artikel *artificial neural network* menyatakan *Multilayer Perceptron* mempunyai struktur yang sama dengan *Single Layer Perceptron* di mana perbezaannya adalah MP mempunyai satu atau lebih lapisan. Ia turut menjelaskan bahawa algorithma ini memiliki dua fasa iaitu fasa maju dan fasa mundur. Fasa maju merujuk kepada

aktiviti perlombongan yang disebarkan daripada input hingga ke lapisan output. Fasa mundur pula merujuk kepada apabila berlaku ralat diantara pemerhatian sebenar dan permintaan nilai nominal.

Hutan Rawak (RF)

Hutan Rawak (*Random Forest*) merupakan algoritma pembelajaran terarah. Ia merupakan kombinasi pokok-pokok peramal sehingga setiap pokok akan bergantung kepada nilai vektor rawak yang diambil sampelnya secara bebas dengan taburan yang sama untuk semua pokok. Kesalahan generalisasi yang berlaku bergantung kepada kekuatan pokok secara individu dan juga hubungan diantara setiap pokok (Breiman, 2001). Rumus 2 menunjukkan teorem yang digunakan untuk membuktikan ketepatan algoritma RF.

$$mg(X, Y) = av_k I(h_k(X) = Y) - \max_{j \neq Y} av_k I(h_k(X) = j) \quad (2)$$

ZeroR

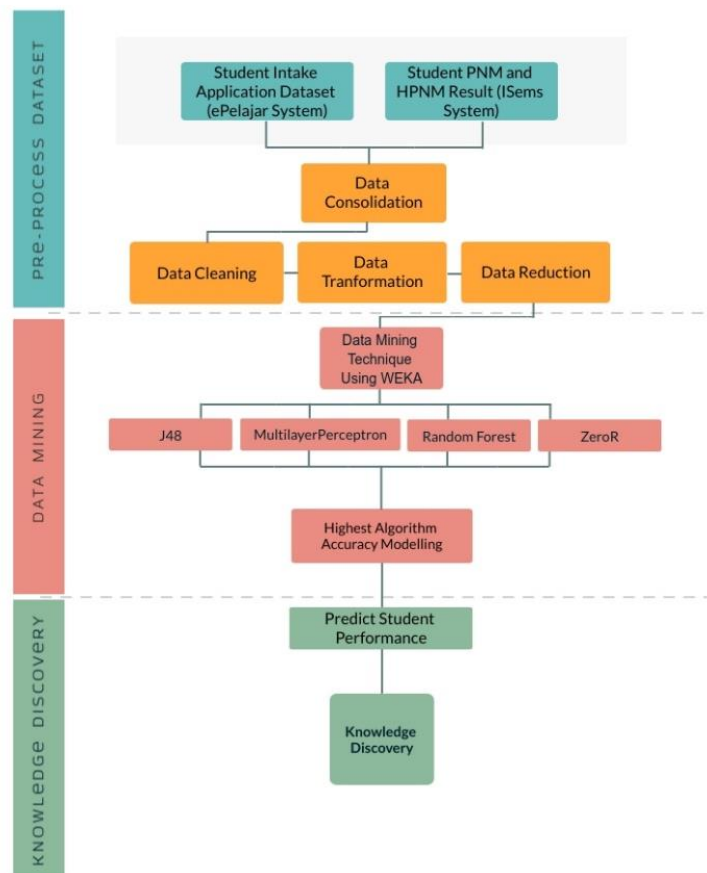
ZeroR merupakan singkatan daripada perkataan *Zero Rules*. Saed Sayad menyatakan bahawa ZeroR merupakan kaedah klasifikasi termudah di mana ia tertumpu sepenuhnya pada sasaran dan mengabaikan peramal. Ramalan dapat dihasilkan dengan mudah menggunakan kaedah ini. Cara ZeroR bertindak ialah dengan mengira nilai paling kerap muncul di dalam kelas dan membuat ramalan menggunakan kekerapan tertinggi di dalam kelas dan mengabaikan nilai-nilai lain. Rajah 5 menunjukkan matrik kekeliruan (*confusion matrix*) bagi ZeroR.

Jadual 1: Metrik Kekeliruan bagi Algoritma ZeroR

Confusion Matrix	Play Golf			
	Yes	No		
ZeroR	Yes	9	5	Positive Predictive Value 0.64
	No	0	0	Negative Predictive Value 0.00
		Sensitivity 1.00	Specificity 0.00	Accuracy = 0.64

METODOLOGI

Kajian ini dijalankan menggunakan teknik pengklasifikasian berserta empat (4) algoritma perlombongan data yang dipilih iaitu Pepohon Keputusan (*Decision Trees*) (J48), *Multilayer Perceptron* (MP), Hutan Rawak (*Random Forest*) (RF) dan ZeroR. Ujilari dilakukan menggunakan kaedah pengesahan silang 10 lipat untuk set data latihan dan pengujian. Di dalam kajian ini, perisian WEKA (Hall et al., 2009) digunakan untuk melombong data-data keputusan kemasukan program melalui teknik pengklasifikasian. Kerangka utama yang digunakan bagi menjalankan kajian ini adalah seperti Rajah 1.



Rajah 2: Kerangka Utama Penyelidikan

Pemrosesan Set Data

Pada peringkat pre-pemrosesan data, dataset diperolehi daripada Jabatan Pengajian Politeknik dan Kolej Komuniti (JPPKK) melalui sistem penilaian *Sistem Integrated Student Evaluation Management System* (iSEMS) dan Sistem ePelajar Kolej Komuniti. Sebanyak 9,527 data diperolehi daripada sistem iSEMS yang merupakan data penilaian pelajar Kolej Komuniti Kuala Langat (KKKL) ambilan tahun 2014 hingga tahun 2019 manakala 4,769 data diperolehi daripada Sistem ePelajar yang terdiri daripada maklumat SPM pelajar bermula tahun 2013 hingga 2019. Data yang digunakan untuk ramalan keputusan peperiksaan akhir ialah pelajar Program Sijil Teknologi Maklumat Semester 1, Semester 2 dan Semester 3 yang akan bergraduasi pada tahun 2021-2023.

Teknologi Maklumat pengambilan tahun 2015 sehingga 2018. Oleh itu, hanya 221 data sahaja yang digunakan dalam kajian ini.

Jadual 2: Senarai Atribut dan Nilai

Bil.	Atribut	Nilai
1	Nama	{ }
2	No. Matrik	{ }
3	BM	{A, A-, B+, B, C+, C, D, E, F}
4	BI	{A, A-, B+, B, C+, C, D, E, F}
5	MT	{A, A-, B+, B, C+, C, D, E, F}
6	SJ	{A, A-, B+, B, C+, C, D, E, F}
7	Kategori HPNM	{Cemerlang, Kepujian, Baik}

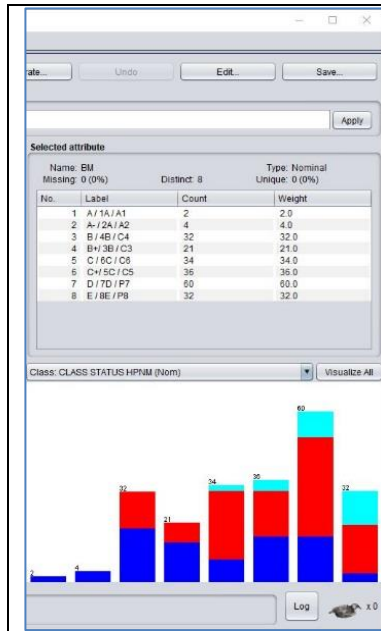
Mata pelajaran selain dari senarai di atas tidak diambil kira di dalam kajian ini kerana hanya empat (4) mata pelajaran ini yang menjadi syarat kemasukan pelajar ke kolej komuniti.

DAPATAN DAN ANALISIS

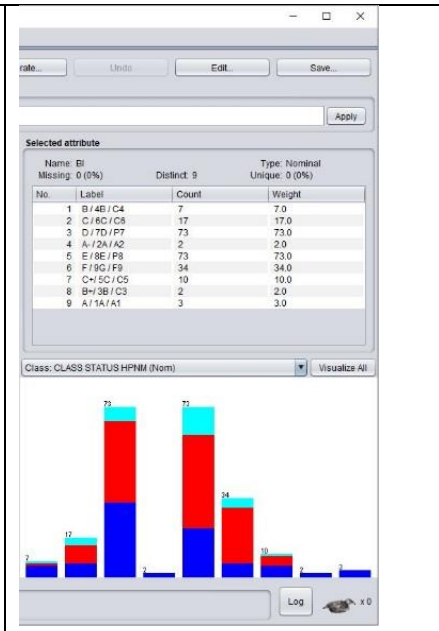
Data yang diperolehi terlebih dahulu dianalisis untuk menunjukkan kekerapan Gred Keputusan di dalam Peperiksaan SPM dan kekerapan Status HPNM pelajar. Empat model pengklasifikasian telah dibina berdasarkan empat algoritma perlombongan data yang telah dipilih. Rajah 6 menunjukkan set data yang dimasukkan ke dalam perisian WEKA sebagai input eksperimen pengklasifikasian. Hanya atribut BM, BI, MT, SJ dan atribut kelas Status HPNM dipilih dan diujilari menggunakan perisian WEKA. Hasil ujilari adalah seperti Rajah 7-10.

NO	1: BM	2: BI	3: MT	4: SJ	5: CLASS STATUS HPNM
1	A+	B+	A+	A+	CEMERLANG
2	A+	C+	C+	C+	CEMERLANG
3	A+	D+	D+	A+	CEMERLANG
4	A+	A+	D+	B+	CEMERLANG
5	A+	B+	C+	B+	CEMERLANG
6	A+	A+	D+	C+	CEMERLANG
7	B+	B+	D+	B+	CEMERLANG
8	B+	B+	D+	C+	CEMERLANG
9	B+	F	D+	C+	CEMERLANG
10	B+	D+	C+	B+	KEPUJIAN
11	B+	D+	C+	C+	CEMERLANG
12	B+	F	C+	B+	CEMERLANG
13	B+	F	B+	D+	KEPUJIAN
14	B+	B+	F	C+	KEPUJIAN
15	B+	C+	D+	C+	CEMERLANG
16	B+	B+	B+	D+	CEMERLANG
17	B+	E	D+	C+	CEMERLANG
18	B+	D+	D+	C+	CEMERLANG
19	B+	D+	D+	C+	CEMERLANG
20	B+	C+	D+	D+	KEPUJIAN
21	B+	F	B+	C+	KEPUJIAN
22	B+	B+	D+	C+	KEPUJIAN
23	B+	F	D+	C+	KEPUJIAN
24	B+	D+	B+	D+	KEPUJIAN
25	B+	D+	F	C+	KEPUJIAN
26	B+	E	C+	C+	CEMERLANG
27	B+	F	D+	C+	CEMERLANG
28	B+	B+	D+	C+	CEMERLANG
29	B+	D+	F	C+	CEMERLANG
30	B+	B+	D+	C+	KEPUJIAN
31	B+	D+	C+	C+	KEPUJIAN
32	B+	F	B+	D+	KEPUJIAN
33	B+	B+	B+	D+	CEMERLANG
34	B+	B+	B+	D+	CEMERLANG
35	B+	B+	C+	C+	CEMERLANG
36	B+	F	C+	D+	CEMERLANG
37	B+	F	D+	B+	CEMERLANG
38	B+	D+	C+	B+	CEMERLANG
39	B+	D+	B+	C+	CEMERLANG
40	B+	C+	B+	C+	CEMERLANG
41	B+	C+	B+	D+	CEMERLANG
42	B+	D+	B+	C+	CEMERLANG
43	B+	D+	C+	C+	CEMERLANG
44	B+	D+	C+	B+	KEPUJIAN
45	B+	D+	C+	C+	CEMERLANG
46	B+	C+	D+	C+	CEMERLANG
47	B+	D+	C+	B+	CEMERLANG
48	B+	D+	C+	C+	KEPUJIAN
49	B+	D+	F	C+	KEPUJIAN
50	B+	A+	A+	E	CEMERLANG
51	B+	C+	D+	D+	KEPUJIAN
52	B+	E	D+	C+	KEPUJIAN
53	B+	D+	D+	A+	CEMERLANG
54	B+	C+	F	C+	KEPUJIAN
55	B+	D+	C+	C+	CEMERLANG
56	B+	D+	E	C+	KEPUJIAN

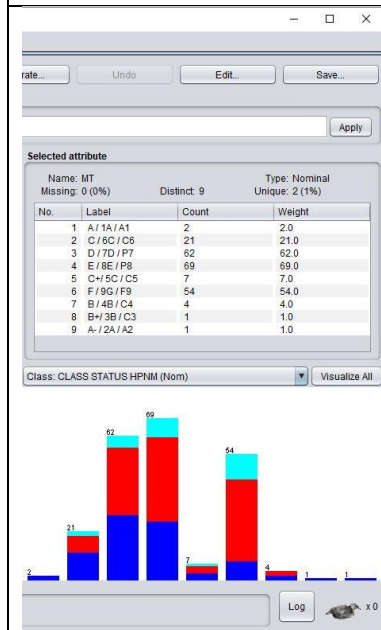
Rajah 6: Set Data yang Diujilari Menggunakan WEKA



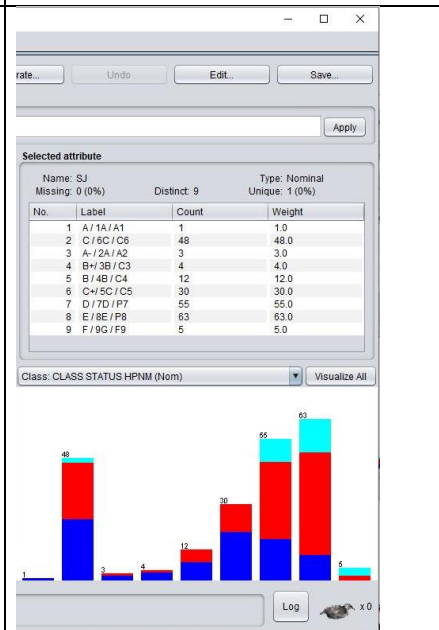
Rajah 7: Pre-pemprosesan Atribut BM



Rajah 8: Pre-pemprosesan Atribut BI



Rajah 9: Pre-pemprosesan Atribut MT



Rajah 10: Pre-pemprosesan Atribut SJ

Rajah 11 menunjukkan kelas Status Gred HPNM pelajar Sijil Teknologi Maklumat pada semester akhir pengajian. Jadual rumusan bagi kelas status HPNM adalah ditunjukkan di dalam Jadual 2.



Rajah 11: Pre-pemprosesan HPNM

Jadual 3: Bilangan Pelajar Sijil Teknologi Maklumat Mengikut HPNM

Bil	Kategori HPNM	Bilangan Pelajar
1	Cemerlang	82
2	Kepujian	112
3	Baik	27

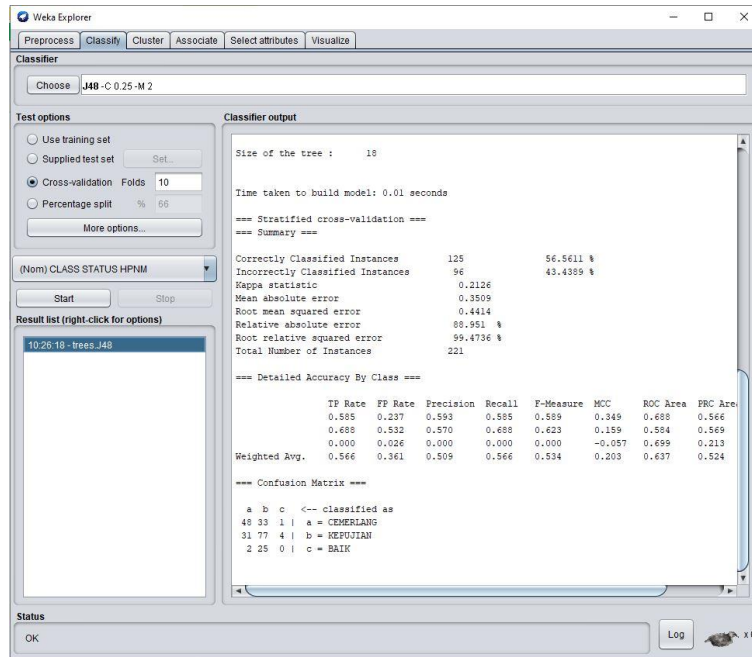
Berdasarkan analisis tersebut, dapat disimpulkan bahawa maklumat keputusan peperiksaan bagi 221 pelajar Program Sijil Teknologi Maklumat (2015-2018) di Kolej Komuniti Kuala Langat adalah seperti Jadual 3.

Jadual 4: Gred Keputusan SPM pelajar Teknologi Maklumat

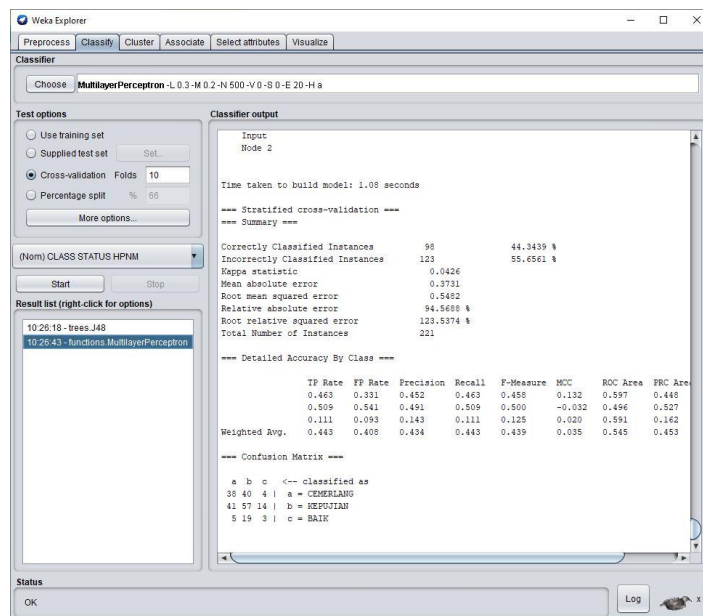
Gred	BM	BI	MT	SJ
A	2	3	2	1
A-	4	2	1	3
B+	21	2	1	4
B	32	7	4	12
C+	36	10	7	30
C	34	17	21	48
D	60	73	62	55
E	32	73	69	39
F	0	34	54	5

Bagi membandingkan dan menentukan algoritma berketepatan tinggi, kesemua data – data ini dilarikan ke atas empat algoritma berbeza iaitu algoritma Pepohon Keputusan J48, *Multilayer Perceptron*, Hutan Rawak (*Random Forest*) dan ZeroR dengan menggunakan set data pelajar Program Teknologi Maklumat, Kolej Komuniti Kuala Langat. Rajah 12 sehingga

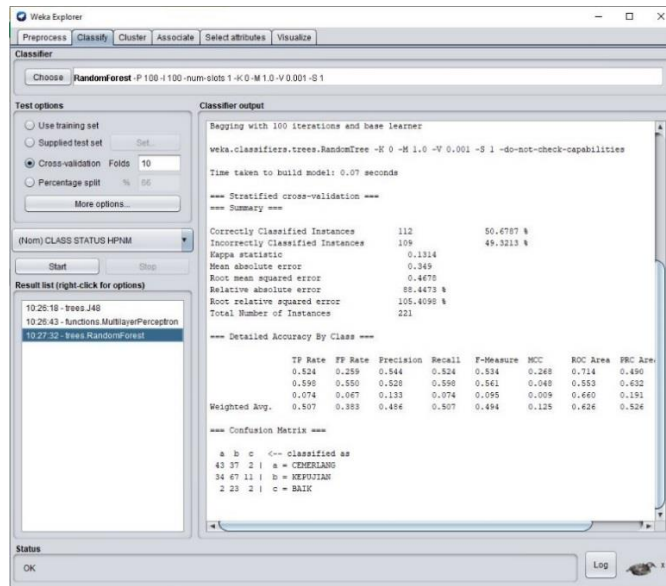
Rajah 15 memaparkan gambarajah output bagi keempat-empat ujian yang dijalankan menggunakan algoritma J48, MP, RF ZeroR.



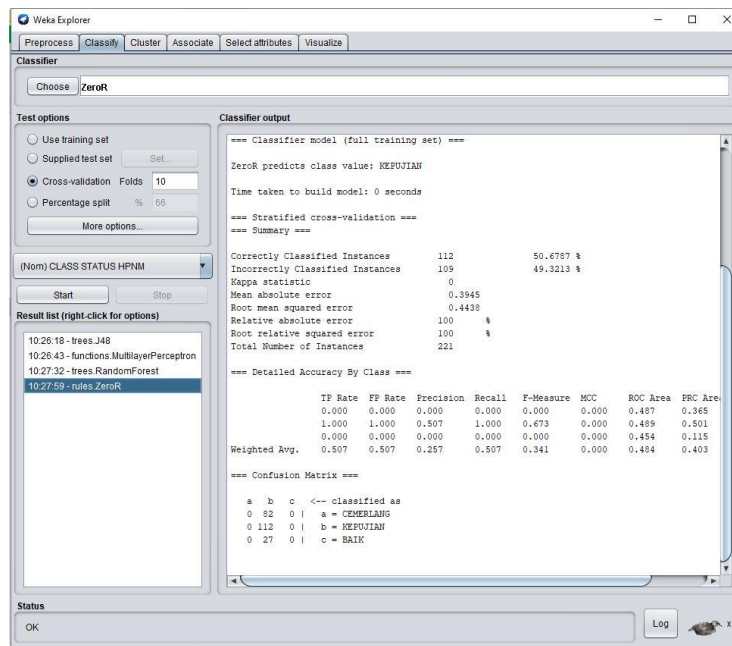
Rajah 12: Output Pepohon Keputusan J48



Rajah 13: Output Multilayer Perceptron



Rajah 14: Output Hutan Rawak



Rajah 15 : Output ZeroR

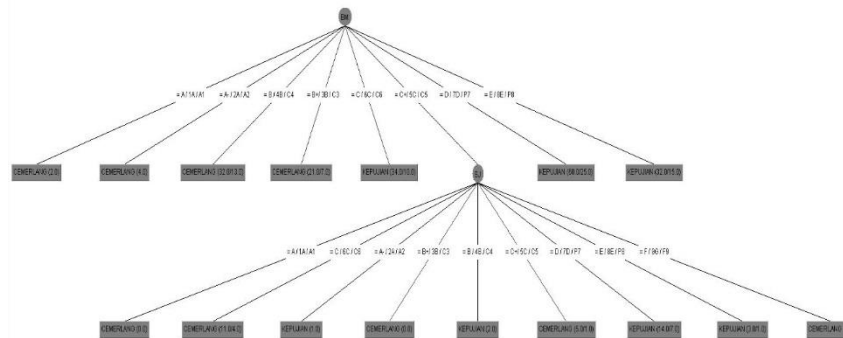
Output bagi ujian tersebut dirumuskan melalui Jadual 4. Jadual tersebut menunjukkan hasil output bagi ketepatan ramalan (*correctly classified instance*) bagi setiap ujilari.

Jadual 4: Perbandingan Keputusan Ketepatan, Kejituan dan Dapatan Semula

Algoritma	Ketepatan (%)	Kejituan (%)	Dapatan Semula (%)
J48	56.56	0.509	0.566
Multilayer Perceptron	44.34	0.434	0.443
Random Forest	50.67	0.486	0.507
ZeroR	50.67	0.257	0.507

Berdasarkan jadual keputusan tersebut, keputusan eksperimen menunjukkan algoritma Pepohon Keputusan J48 mempunyai ketepatan ramalan yang paling tinggi (56.56%) berbanding algoritma *Random Forest* (50.67%), *ZeroR* (50.67%) dan *Multilayer Perceptron* (44.34%). Algoritma Pepohon Keputusan J48 merupakan algoritma ramalan terbaik di dalam kajian ini berkemungkinan kerana jenis data yang dilombong adalah dari jenis kategorikal. Selain dari itu kaedah pengesahan silang 10 lipat Pepohon Keputusan J48 memberikan pengesahan analisis data yang lebih general terhadap set data bebas yang digunakan.

Berdasarkan perbandingan algoritma, Pepohon Keputusan J48 digunakan untuk mengenal pasti pengaruh diantara atribut terhadap status HPNM. Rajah 16 menunjukkan atribut yang mempunyai pengaruh terhadap pengelasan status HPNM. Keputusan SPM BM didapati mempunyai pengaruh terhadap kelas HPNM diikuti oleh keputusan SPM SJ. Manakala keputusan SPM BI dan SPM MT didapati tidak mempengaruhi keputusan HPNM pelajar.



Rajah 16: Pepohon Keputusan J48

Kajian ini juga menggunakan algoritma ramalan yang telah dikenal pasti berketepatan tinggi iaitu Pepohon Keputusan J48 untuk meramal keputusan status HPNM pelajar pengambilan tahun 2019 dan 2020. Dataset pelajar Sijil Teknologi Maklumat ambilan 2019 hingga 2020 diperolehi daripada Sistem ePelajar. Seramai 76 pelajar semester 1, 2 dan 3 telah diujilarkan dengan ramalan algoritma tersebut.

Rajah 17 merupakan senarai 76 pelajar Sijil Teknologi Maklumat yang masih belum menamatkan pengajian ketika ini. Mereka terdiri daripada pelajar semester 1, 2 dan 3.

	A	B	C	D	E	F
1	BM	BI	MT	SJ	HPNM	
2	D / 7D / P7	D / 7D / P7	F / 9G / F9	D / 7D / P7		
3	D / 7D / P7	D / 7D / P7	E / 8E / P8	E / 8E / P8		
4	E / 8E / P8	F / 9G / F9	E / 8E / P8	C / 6C / C6		
5	C+ / 5C / C5	D / 7D / P7	B / 4B / C4	C / 6C / C6		
6	D / 7D / P7	E / 8E / P8	F / 9G / F9	E / 8E / P8		
7	D / 7D / P7	C+ / 5C / C5	F / 9G / F9	E / 8E / P8		
8	E / 8E / P8	F / 9G / F9	F / 9G / F9	E / 8E / P8		
9	C+ / 5C / C5	D / 7D / P7	D / 7D / P7	D / 7D / P7		
10	C / 6C / C6	D / 7D / P7	D / 7D / P7	E / 8E / P8		
11	A- / 2A / A2	C / 6C / C6	A / 1A / A1	B / 4B / C4		
12	D / 7D / P7	E / 8E / P8	F / 9G / F9	E / 8E / P8		
13	B / 4B / C4	C+ / 5C / C5	F / 9G / F9	E / 8E / P8		
14	C+ / 5C / C5	D / 7D / P7	F / 9G / F9	D / 7D / P7		
15	D / 7D / P7	E / 8E / P8	F / 9G / F9	D / 7D / P7		
16	B / 4B / C4	F / 9G / F9	F / 9G / F9	C+ / 5C / C5		
17	D / 7D / P7	E / 8E / P8	E / 8E / P8	E / 8E / P8		
18	E / 8E / P8	F / 9G / F9	D / 7D / P7	E / 8E / P8		
19	B / 4B / C4	E / 8E / P8	E / 8E / P8	C / 6C / C6		
20	E / 8E / P8	E / 8E / P8	E / 8E / P8	E / 8E / P8		
21	C / 6C / C6	E / 8E / P8	E / 8E / P8	E / 8E / P8		
22	D / 7D / P7	E / 8E / P8	C / 6C / C6	D / 7D / P7		
23	B / 4B / C4	C / 6C / C6	E / 8E / P8	D / 7D / P7		
24	E / 8E / P8	D / 7D / P7	E / 8E / P8	D / 7D / P7		
25	B / 4B / C4	C+ / 5C / C5	D / 7D / P7	C+ / 5C / C5		
26	A- / 2A / A2	C+ / 5C / C5	E / 8E / P8	C / 6C / C6		
27	C+ / 5C / C5	D / 7D / P7	F / 9G / F9	C / 6C / C6		
28	B / 4B / C4	C+ / 5C / C5	D / 7D / P7	C+ / 5C / C5		
29	C+ / 5C / C5	F / 9G / F9	E / 8E / P8	D / 7D / P7		
30	B+ / 3B / C3	B+ / 3B / C3	B / 4B / C4	C / 6C / C6		
31	E / 8E / P8	F / 9G / F9	E / 8E / P8	E / 8E / P8		
32	C+ / 5C / C5	E / 8E / P8	D / 7D / P7	C / 6C / C6		
33	C / 6C / C6	E / 8E / P8	E / 8E / P8	D / 7D / P7		
34	C+ / 5C / C5	E / 8E / P8	E / 8E / P8	C / 6C / C6		

Rajah 17: Dataset Pelajar Ambilan 2019-2020

Rajah 18 merupakan keputusan ramalan yang telah dijalankan ke atas 76 pelajar tersebut. Rumusan hasil daripada ujilari ramalan tersebut adalah seperti Jadual 5.

	A	B	C	D	E	F	G	H	I	J
1	Bil	Nama Pelajar	No Matrik	BM	BI	MT	SJ	'prediction margin'	'predicted CLASS STATUS HPNM'	'CLASS STATUS HPNM'
2				D / 7D / P7	D / 7D / P7	F / 9G / F9	D / 7D / P7	-0.316667	KEPUJIAN	?
3				D / 7D / P7	D / 7D / P7	E / 8E / P8	E / 8E / P8	-0.316667	KEPUJIAN	?
4				E / 8E / P8	F / 9G / F9	E / 8E / P8	C / 6C / C6	-0.4375	KEPUJIAN	?
5				C+ / 5C / C5	D / 7D / P7	B / 4B / C4	C / 6C / C6	0.363636	CEMERLANG	?
6				D / 7D / P7	E / 8E / P8	F / 9G / F9	E / 8E / P8	-0.316667	KEPUJIAN	?
7				D / 7D / P7	C+ / 5C / C5	F / 9G / F9	E / 8E / P8	-0.316667	KEPUJIAN	?
8				E / 8E / P8	F / 9G / F9	F / 9G / F9	E / 8E / P8	-0.4375	KEPUJIAN	?
9				C+ / 5C / C5	D / 7D / P7	D / 7D / P7	D / 7D / P7	-0.142857	KEPUJIAN	?
10				C / 6C / C6	D / 7D / P7	D / 7D / P7	E / 8E / P8	-0.470588	KEPUJIAN	?
11				A- / 2A / A2	C / 6C / C6	A / 1A / A1	B / 4B / C4	1	CEMERLANG	?
12				D / 7D / P7	E / 8E / P8	F / 9G / F9	E / 8E / P8	-0.316667	KEPUJIAN	?

Rajah 18: Keputusan Ramalan Keputusan HPNM Pelajar SPM 2019- 2020

Jadual 5: Rumusan Ramalan Keputusan HPNM Pelajar SPM 2019- 2020

Bil	Kategori HPNM	Bilangan Pelajar
1	Cemerlang	27
2	Kepujian	49
3	Baik	0

KESIMPULAN DAN CADANGAN

Kajian berkaitan ramalan prestasi pelajar mula menarik perhatian ramai penyelidik untuk diterokai secara lebih meluas. Pelbagai program yang lebih strategik dan berfokus boleh dirancang untuk pelajar-pelajar yang lemah sekiranya mereka dapat dikenal pasti ketika berada

di semester pertama lagi. Dapatan kajian ini menunjukkan algoritma terbaik bagi meramal prestasi keputusan semester akhir melalui Himpunan Purata Nilai Mata (HPNM) pelajar Program Teknologi Maklumat di Kolej Komuniti Kuala Langat adalah algoritma Pepohon Keputusan J48. Selain daripada itu, kajian ini juga mendapati keputusan Sijil Pelajaran Malaysia (SPM) mata pelajaran Bahasa Melayu (BM) mempunyai pengaruh yang lebih tinggi terhadap kelas HPNM berbanding dengan keputusan mata pelajaran Sejarah. Manakala keputusan mata pelajaran Bahasa Inggeris (BI) dan Matematik (MT) didapati tidak mempengaruhi keputusan HPNM pelajar. Penguasaan BM yang tinggi di kalangan pelajar adalah penting kerana medium pengantaraan pengajaran dan pembelajaran (PdP) di Kolej Komuniti Kuala Langat adalah dalam Bahasa Melayu.

Sementara itu, keputusan ramalan HPNM menunjukkan daripada 76 orang pelajar Program Teknologi Maklumat yang belum tamat pengajian, 27 orang pelajar diramal bakal mendapat keputusan Cemerlang dan bakinya 49 orang bakal memperolehi keputusan berstatus Kejujangan. Dapatan ini dapat membantu pihak pengurusan Kolej Komuniti Kuala Langat terutamanya para pensyarah, Unit Pembangunan Pelajar dan kaunselor untuk memberi lebih tumpuan kepada pelajar Kejujangan tersebut supaya mendapat keputusan Cemerlang, seterusnya meningkatkan prestasi keseluruhan kolej.

Bagi kajian akan datang, dicadangkan agar kajian ini diperluaskan bagi meramal prestasi pelajar-pelajar yang bakal bergraduat melalui pelbagai program akademik yang lain. Faktor-faktor lain selain daripada keputusan SPM sebagai syarat kemasukan juga perlu dikaji, contohnya hubungkait di antara penempatan asrama di kolej atau latar belakang keluarga berserta maklumat demografik lain. Kajian ini diharapkan dapat membantu pihak pengurusan Kolej Komuniti untuk lebih memahami keperluan semasa para pelajar dalam usaha memaksimumkan prestasi pembelajaran mereka sepanjang pengajian.

RUJUKAN

- Abdul Rohman, Muhammad Rochcham (2019). Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Mahasiswa, *Jurnal Neo Teknika*,5(1): 22-33.
- Azura Abd Jalil, Aida Mustapha, Dzulizah Santa, Nurul Zaiha Zainal, Rizalina Radwan (2013). Prediction Of College Student Academic Performance Using Data Mining Techniques, *International Conference On Engineering Education 2013*, Madinah, Kingdom of Saudi Arabia
- Fang, Weiping & Wang, Yuming (2013). The Development of Data Mining, *International Journal of Business and Social Science*,4((16): MukasuratLewicki, P. & Hill, T. (2005). *Statistics: Methods and Applications*. (StatSoft, Ed.) (1st ed.). Tulsa: Statistica Data Analysis Software and Services.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. & Witten, I. H.. "The WEKA data mining software: an update." *ACM SIGKDD explorations newsletter* 11(1): 10-18.
- Heri Susanto, Sudiyatno (2014). Data Mining Untuk Memprediksi Prestasi Siswa, *Jurnal Pendidikan Vokasi*. 4(2): 222-231.
- Jing, Luan (2002). Datamining And Knowledge Management In Higher Education Potential Applications. *Annual Forum For The Association, For Institutional Research* (42nd, Toronto, Ontario, Canada, June 2-5, 2002)

- Kaur, P. Singh, M. & Singh Josan, G. (2015), Classification And Prediction Based Data Mining Algorithms To Predict Slow Learners In Education Sector, *3rd International Conference On Recent Trends In Computing 2015 (ICRTC-2015)*
- Leo Breiman (2001). Random forests. *Machine Learning*, 45: 5–32.
- Nik Haslinda Abdul Halim, Abdul Razak Hamdan, Zulaiha Ali Othman, Hamidah Jantan (2017). Pemodelan Guru Cemerlang KPM Menggunakan Perlombongan Data, *Journal Of ICT In Education (JICTIE)* 4: 21-34.
- Ritika Saxena (2015). Educational Data Mining: Performance Evaluation Of Decision Tree And Clustering Techniques Using WEKA Platform, *International Journal Of Computer Science And Business Informatics*, 15(2): 230-240.
- Sayad, Saed. “Artificial Neural Network - Perceptron.” *Perceptron*, saedsayad.com/artificial_neural_network_bkp.htm.
- Sayad, Saed. “ZeroR – Model Evaluation.” *ZeroR*, saedsayad.com/ZeroR.htm.
- Sen, B. Ucar, U. & Delen, D. (2012). Predicting And Analyzing Secondary Education Placement-Test Scores: A Data Mining Approach. *Expert Systems With Applications* 39: 9468–9476.
- Siti Rohaidah Ahmad and Azuraliza Abu Bakar.(2005). Perlombongan Data Prestasi Pelajar Siswazah Menggunakan Kaedah Aruhan Berasaskan Atribut, *Jurnal Teknologi Maklumat & Multimedia* 2: 1-17.
- Usama M. Fayyad and Keki B. Irani.(1993). Multi-interval discretization of continuous valued attributes for classification learning, *13 International Joint Conference on Artificial Intelligence*: 1022– 1027.
- Viloria, A., Bonerge, O., Lezama, P. & Varela, N. (2019). Bayesian Classifier Applied To Higher Education Dropout, *6th International Symposium on Emerging Information, Communication And Networks (EICN 2019)* November 4-7, 2019, Coimbra, Portugal
- Wirawati Dewi Ahmad, Azuraliza Abu Bakar (2018). Classification Models For Higher Learning Scholarship Award Decisions, *Asia-Pacific Journal of Information Technology and Multimedia Jurnal Teknologi Maklumat dan Multimedia Asia-Pasifik*.7(2): 131-145.