

Pemodelan Guru Cemerlang KPM Menggunakan Perlombongan Data

Excellence Teacher Modelling using Data Mining Techniques

Nik Haslinda Abdul Halim¹, Abdul Razak Hamdan², Zulaiha Ali Othman²,
Hamidah Jantan³

¹Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia; liendahalim@gmail.com

²Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia; {arh, zao}@ukm.edu.my

³Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA; hamidahjm@tganu.uitm.edu.my

Abstrak

Guru merupakan panggilan kepada seorang pendidik yang mengajar di sekolah. Kecemerlangan seseorang pelajar biasanya berkait rapat dengan kecemerlangan guru yang mengajar. Kewujudan guru cemerlang amat diperlukan di semua sekolah dan masalah yang biasa dihadapi oleh pihak pentadbiran sekolah ialah menjadikan seseorang guru itu guru yang cemerlang. Oleh itu tujuan kajian ini dilakukan adalah untuk mendapatkan model guru cemerlang yang terbaik menggunakan dua algoritma dalam teknik pepohon keputusan. Algoritma yang digunakan ialah C4.5 (J48) dan Hutan Rawak (RF). Kajian ini menggunakan data guru cemerlang di sekolah. Atribut yang digunakan merupakan 4 faktor dengan 26 kriteria pemilihan guru cemerlang sebagai input serta 1 output. Set data diambil daripada Sistem Pengurusan Latihan Kementerian Pelajaran Malaysia (SPL KPM) yang ditadbir oleh Sektor Latihan ICT, BPG. Keputusan kajian menunjukkan bahawa Model 2 algoritma J48 memperolehi keputusan yang lebih baik iaitu ketepatan sebanyak 98.86% dengan min kuasa dua ralat (RMSE) hanya 0.1061 berbanding dengan algoritma Hutan Rawak (RF).

Kata Kunci: Guru cemerlang, pepohon keputusan, J48, hutan rawak.

Abstract

Teacher is an educator who is teaching at the school. Student's performances are usually associated with excellence teacher. The existence of excellent teachers is needed in all schools and the common problems faced by the school administration is making an excellent teacher's teacher. Therefore, the purpose of this study is to get the excellence teachers using two algorithms from decision tree technique. The algorithm used is a C4.5 (J48) and Random Forests (RF). This study uses the data of excellence teachers in schools. The attributes used are 4 factors with 26 criteria for selection of outstanding teachers as input and one output. The data set used were taken from the Training Management System Ministry of Education (MOE SPL) which is administered by the ICT Training Sector, BPG. The results show that the Model 2 algorithm J48 was better than an accuracy of 98.86% with a mean square error (RMSE) is only 0.1061 compared with algorithm Random Forest (RF).

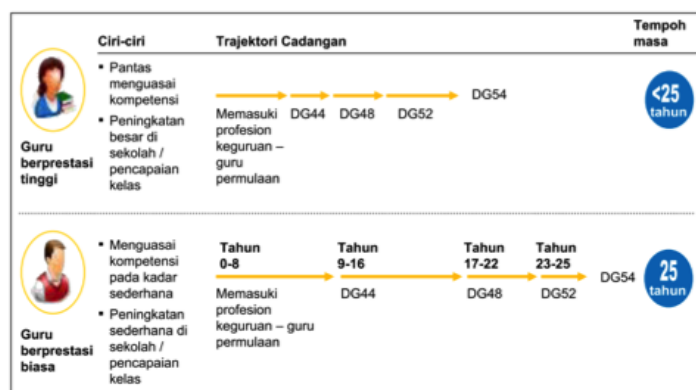
Keywords: excellence teacher, decision tree, J48, random forest.

PENGENALAN

Salah satu cabaran dalam bidang pengurusan bakat ialah mengenal pasti bakat-bakat yang ada di kalangan pekerja (Stavrou-Costea, E., 2005). Seorang pembuat keputusan kadangkala menghadapi banyak keadaan yang tidak menentu dan kesukaran untuk menyelesaikan sesuatu keputusan pengurusan. Ini merupakan permasalahan yang utama dalam pengurusan sumber manusia untuk memilih calon-calon yang cemerlang.

Dalam proses penilaian dan pengurusan bakat, ia amat bergantung kepada faktor seperti pengalaman, pengetahuan, penilaian dan kecenderungan seseorang pembuat keputusan (Hamidah, 2011). Faktor tersebut selalunya akan menghasilkan keputusan yang tidak konsisten, kurang tepat, kurang adil serta kadangkala di luar jangkaan (Martinsons, 1995). Apabila keadaan ini berlaku, ia kadangkala akan memberi kesan motivasi kepada individu yang dinilai (Delahaye, 2011). Kesan ini juga akan mempengaruhi produktiviti organisasi di masa hadapan. Pelan Pembangunan Pendidikan Malaysia 2013-2025 menyatakan bahawa kemajuan kerjaya guru bergantung kepada kepantasan seseorang guru menguasai kompetensi yang ditetapkan pada setiap peringkat. Guru yang juga dikenali sebagai Pegawai Perkhidmatan Pendidikan (PPP) terbahagi kepada dua kumpulan iaitu Pegawai Perkhidmatan Pendidikan Siswazah (PPPS) dan Pegawai Perkhidmatan Pendidikan Lepas Diploma (PPPLD). Manakala sistem pemangkuan dan kenaikan pangkat PPP membahagikan empat (4) laluan yang utama untuk PPP iaitu Laluan Hakiki, Laluan Cemerlang, Laluan Time-Based dan Laluan Khas bagi Guru Besar Ex-PPPLD.

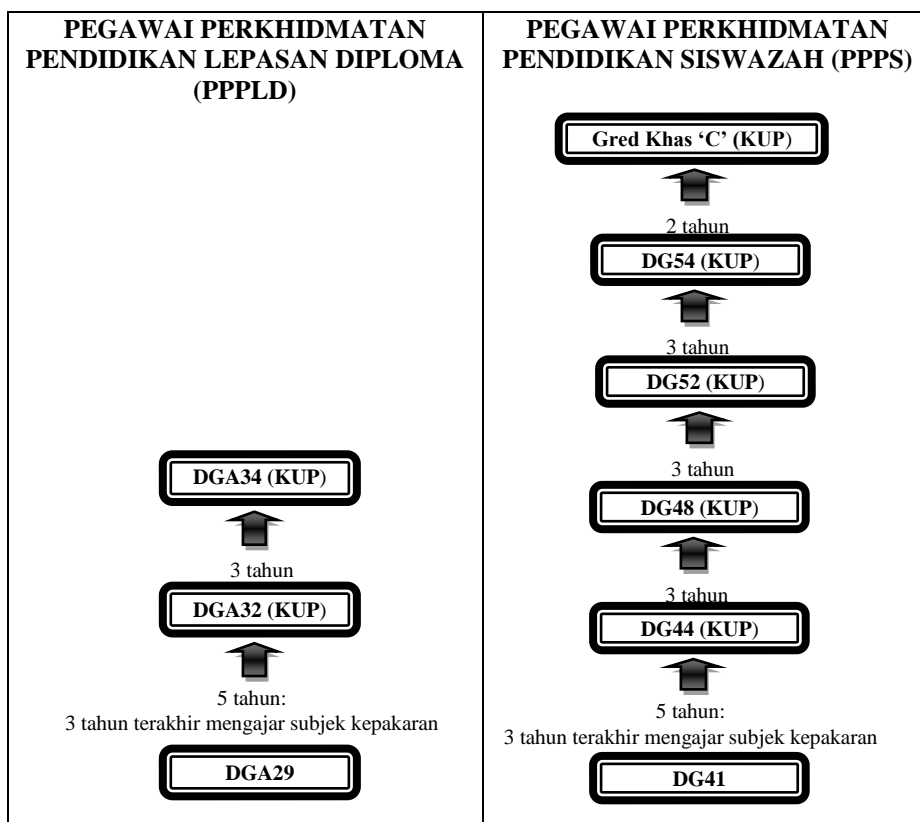
Objektif utama kajian ini ialah mencari model guru cemerlang yang akan diperolehi daripada pemodelan terbaik antara algoritma C4.5 (J48) dan Hutan Rawak (RF). Rajah 1 menunjukkan cadangan guru berprestasi tinggi dinaikkan gred dari DG41 ke DG54 dalam tempoh masa yang lebih pendek berbanding tempoh 25 tahun sekarang. Ia bertujuan memberikan ganjaran kepada guru yang cemerlang dan menjadi dorongan kepada guru yang terbaik untuk bertugas di sekolah berkenaan.



Rajah 1: Cadangan laluan bagi guru berprestasi tinggi.

Guru Cemerlang

Buku Terma Rujukan Konsep Guru Cemerlang (KPM, 2011) mendefinisikan bahawa Guru Cemerlang (GC) ialah guru yang mempunyai pengetahuan, kemahiran, dan kepakaran yang tinggi dalam mata pelajaran khususnya dalam bidang pengajaran dan pembelajaran. GC juga sentiasa berdedikasi dan bermotivasi dalam melaksanakan tugas dan tanggungjawab yang cemerlang secara berterusan. Seorang GC memiliki pengetahuan, kemahiran, dan kepakaran yang tinggi dalam mata pelajaran kepakarannya, bijak dalam mengatur strategi pengajaran yang berupaya menyampaikan ilmu dengan berkesan dan berupaya menggerakkan akal fikiran murid untuk terus memperoleh ilmu tambahan (UKK, 2010). Prestasi GC melebihi daripada prestasi guru biasa dari segi peribadi, pengetahuan dan kemahiran, hasil kerja, komunikasi, potensi dan sumbangan kepada kemajuan pendidikan negara.



Rajah 2: Carta Aliran Laluan Kerjaya Guru Cemerlang (KPM, 2011).

Rajah 2 menunjukkan carta aliran laluan kerjaya Guru Cemerlang bagi Pegawai Perkhidmatan Pendidikan Lepasn Diploma (PPPLD) dan Pegawai Perkhidmatan Pendidikan Siswazah (PPPS).

Laluan kerjaya GC berbeza daripada laluan kerjaya guru biasa seperti digambarkan dalam Rajah 2. GC diberi peluang untuk merancang pembangunan diri bagi merebut peluang peningkatan ke gred yang lebih tinggi tanpa meninggalkan tugas mengajar di sekolah asalkan mereka dapat membuktikan peningkatan kecemerlangan. Manakala guru biasa pula walaupun cemerlang dalam perkhidmatan, mereka lebih terikat dengan zon lantikan sebelum layak mengisi jawatan kenaikan pangkat. Kaedah kenaikan pangkat GC adalah secara Khas untuk Penyandang (KUP) manakala kaedah kenaikan pangkat untuk guru biasa adalah secara *Time-Based*.

Penilaian calon GC adalah salah satu kaedah penilaian pencapaian menggunakan instrumen yang mengukur tahap pengurusan mata pelajaran, sumbangan kepimpinan, khidmat kepakaran, sumbangan profesional serta pengajaran dan pembelajaran (PDP) dalam bilik darjah (KPM, 2011).

Kaedah kenaikan pangkat GC dikeluarkan khas untuk guru-guru yang memang cemerlang dalam mata pelajaran yang diajar atau dalam bidang masing-masing selari dengan usaha kerajaan memartabatkan profesion keguruan.

Faktor dan Kriteria Guru Cemerlang

Secara umumnya pada masa ini, terdapat enam individu atau Bahagian yang terlibat dalam pemilihan GC ini. Selain Pegawai Perkhidmatan Pendidikan (PPP-guru) yang merupakan calon GC, Bahagian yang terlibat ialah Ketua Jabatan (Pengetua/Guru Besar), Jemaah Nazir Negeri, Jemaah Nazir dan Jaminan Kualiti (JNJK-Putrajaya), Bahagian Pengurusan Sumber Manusia (BPSM) dan search committee KPM.

Pada umumnya, pemilihan GC adalah bergantung kepada permohonan yang dilakukan oleh guru-guru yang layak mengikut syarat-syarat yang telah ditetapkan seperti Jadual 1.

Jadual 1: Syarat pemilihan guru cemerlang: Tapisan Ketua Jabatan (Sumber: Surat Siaran bertariikh 18 Mac 2014).

Syarat Umum	Syarat Khusus
<ul style="list-style-type: none">• Disahkan dalam perkhidmatan• Markah Laporan Nilai Prestasi Tahunan (LNPT) tiga tahun berturut-turut sekurang-kurangnya 80%• Bebas daripada tindakan tatatertib• Lulus laporan pengisytiharan harta• Lulus tapisan keutuhan Susruhanjaya Pencegahan Rasuah Malaysia (SPRM)• Bebas daripada senarai peminjam tegar pinjaman Pendidikan• Diperakui oleh Ketua Jabatan	<ul style="list-style-type: none">• Tempoh perkhidmatan tidak kurang 5 tahun untuk gred sekarang.• Disahkan dalam jawatan• Mata pelajaran kepakaran merupakan mata pelajaran utama dalam bidang pengkhususan.• Mengajar mata pelajaran kepakaran 3 tahun berturut-turut.

Jadual 1 menunjukkan kriteria yang ditapis oleh Ketua Jabatan terhadap pemohon GC. Seseorang guru yang memohon gelaran GC harus melepasi semua syarat umum dan syarat khusus yang telah ditetapkan. Setelah melalui proses saringan calon, proses pemantauan dan pemilihan calon dilakukan. Empat faktor yang digunakan dalam proses tersebut iaitu ketrampilan profesional, pengajaran dan pembelajaran (PDP), peranan dalam aktiviti bukan akademik dan keperibadian. Bagi memudahkan proses pemantauan, instrumen Standard 4 dalam Standard Kualiti Pendidikan Malaysia (SKPM) digunakan sebagai instrumen pemantauan untuk faktor pengajaran dan pembelajaran.

Setelah pemantauan dan penilaian dilaksanakan oleh JNJK, senarai calon-calon yang layak akan dikemukakan kepada Bahagian Pengurusan Sumber Manusia (BPSM). Tapisan seterusnya dilakukan di peringkat BPSM dengan mengambil kira calon-calon yang masih belum dinaikkan pangkat secara KUP ataupun hakiki ke gred yang dipohon serta memenuhi syarat-syarat pemangkuan yang telah ditetapkan oleh Lembaga Kenaikan Pangkat Perkhidmatan Pelajaran (LKPPP).

Memandangkan proses pemangkuan GC melibatkan pemantauan dan penilaian oleh JNJK, hanya calon-calon yang memenuhi syarat-syarat kenaikan pangkat akan dipertimbangkan bagi urusan pemangkuan berdasarkan kekosongan kuota yang ada (UKK, 2009). Seseorang GC yang dipilih haruslah mempunyai ciri-ciri seorang pendidik yang cemerlang dari segi peribadi, pengetahuan dan kemahiran, hasil kerja, komunikasi, potensi dan sumbangan kepada kemajuan pendidikan Negara. Jadual 2 menunjukkan ciri-ciri GC dari segi kualiti.

Jadual 2: Ciri-ciri guru cemerlang dari segi kualiti (KPM, 2010).

Ciri-ciri	Penerangan
Peribadi	Mempunyai sahsiah yang terpuji, berketrampilan dan berpegang teguh kepada etika perguruan dan etika perkhidmatan awam serta peka dan prihatin terhadap keperluan pelajar yang dapat menjadi <i>role model</i> kepada warga pendidik.
Pengetahuan dan kemahiran	Menguasai dan menghayati kandungan mata pelajaran atau bidang pengkhususan dan kepakarannya, menjadi pakar rujuk pengajaran dan pembelajaran, bijak mengurus masa, bahan sumber, alat pengajaran dan pembelajaran, mampu mengenal pasti keperluan dan masalah pembelajaran pelajar serta menyediakan program tindakan susulan, berpengetahuan dan berkemahiran dalam bidang teknologi maklumat dan komunikasi
Hasil Kerja	Berupaya meningkatkan hasil pembelajaran (<i>learning outcomes</i>) ke arah membangunkan potensi pelajar selaras dengan Falsafah Pendidikan Kebangsaan.
Komunikasi	Memiliki kemahiran komunikasi yang tinggi serta berkemampuan menyampaikan idea dan mesej dengan berkesan dalam pelbagai situasi
Potensi	Berwawasan, proaktif, mempunyai inisiatif, bersifat responsif dan inovatif dalam rangka menjalankan tugas sebagai pendidik.
Sumbangan kepada kemajuan pendidikan negara	Berupaya menjana idea baru dalam mata pelajaran atau bidang pengkhususan masing-masing dan berkongsi idea tersebut bagi meningkatkan mutu pendidikan negara

Ketrampilan profesional merangkumi keupayaan dalam mengurus, mengguna dan memanfaatkan pengetahuan dan pengalaman dalam bidang mata pelajaran kepakaran serta prinsip pedagogi untuk meningkatkan kualiti pencapaian pelajar. Pengajaran dan Pembelajaran melibatkan keupayaan menjalankan pengajaran dan pembelajaran (PdP), sumbangan kepakaran dalam dan luar sekolah dalam bidang dan keupayaan mengurus rekod serta maklumat pencapaian pelajar dengan baik untuk diakses bagi tujuan tindakan susulan.

Guru yang berperanan dalam aktiviti bukan akademik juga akan cuba memperlihatkan pengetahuan dan kemahiran serta melibatkan diri dalam aktiviti kokurikulum, hal ehwal pelajar dan tugas-tugas bukan mengajar yang lain. Faktor keperibadian pula melibatkan cara berpakaian, menepati masa, mesra, sopan, bertanggungjawab, berkeyakinan diri, amanah dan mempunyai kemahiran komunikasi yang baik (KPM, 2011).

KAJIAN LITERATUR

Perlombongan data adalah satu proses analisis untuk meninjau data dalam mencari corak perhubungan yang konsisten dan sistematik antara pembolehubah, dan kemudian untuk mengesahkan penemuan itu, corak yang di kesan tersebut akan digunakan ke atas data baru. Matlamat utama dalam perlombongan data adalah peramalan. Proses perlombongan data dalam peramalan terdiri daripada tiga peringkat iaitu penerokaan awal, pembangunan model atau corak pengenalan dengan pengesahan dan penjana (iaitu, penggunaan model untuk data baru bagi menjana ramalan) (Lewichi, 2005).

Data mining for forecasting offers the opportunity to leverage the numerous sources of time series data, internal and external, now readily available to the business decision maker, into actionable strategies that can directly impact profitability (Bhardwaj et al, 2011).

Perlombongan data adalah satu proses mencari maklumat yang penting daripada sejumlah data yang besar (Kova et al, 2012). Ia merupakan usaha kerjasama antara manusia dan komputer. Manusia mereka bentuk pangkalan data, menggambarkan masalah dan menetapkan matlamat. Manakala komputer pula menapis melalui data serta mencari corak yang sesuai dengan matlamat yang ditetapkan. Peramalan dalam perlombongan data ialah pencarian suatu corak yang sangat kukuh dalam data besar yang boleh menjana untuk mendapatkan satu keputusan masa depan yang tepat.

Masalah dalam perlombongan data terbahagi kepada dua kategori iaitu peramalan dan penemuan pengetahuan. Masalah peramalan digambarkan dari segi matlamat yang telah ditentukan yang mana berhubungkait dengan rekod lama yang sudah diketahui jawapannya. Manakala masalah penemuan pengetahuan biasanya digambarkan peringkat sebelum peramalan yang mana maklumat yang ada tidak mencukupi untuk membuat ramalan. Terdapat dua jenis tugas yang umum dalam permasalahan ramalan iaitu pengelasan dan regresi. Sampel daripada pengalaman lama yang sudah diketahui

jawabannya akan diuji dan dijana untuk mendapat kes yang akan datang (Kova et al, 2012).

Dalam perlombongan data terdapat beberapa metod (task) iaitu *classification* (pengelasan), *Association Rules Mining*, *Prediction*, *Sequential rules*, *Clustering* dan *Deviation Detection*. Setiap metod mempunyai beberapa teknik yang biasa digunakan mengikut keperluan kajian.

Pengelasan ialah tugas perlombongan data yang paling popular bagi penemuan pengetahuan dan rancangan masa depan. Ia yang yang biasa digunakan dalam banyak bidang juga boleh dilaksanakan untuk bakat manusia (Hamidah et al, 2010). Antara teknik pengelasan (*classification*) dalam perlombongan data yang digunakan dalam kebanyakan kajian ialah seperti teknik Pepohon Keputusan, Rangkaian Neural, Set Teori yang kasar, teori Bayesian dan logik kabur. Jadual 3 menunjukkan kajian sebelum yang menggunakan teknik dalam kaedah pengelasan yang melibatkan prestasi insan.

Jadual 3: Teknik untuk meramal prestasi insan.

Teknik	Aplikasi
Pohon Keputusan	Pengurusan Bakat (Hamidah, 2011) Meramal prestasi pekerja (Stavrou, 2005; KPM, 2012) Meramal prestasi pelajar (Pandey et. al., 2011; Kabkchieva et al, 2009; Rey et. al., 2013; UKK, 2009)
Rangkaian Neural	Pengurusan Bakat (Hamidah, 2011) Meramal prestasi pekerja (Stavrou, 2005; KPM, 2012) Meramal prestasi pelajar (Stavrou, 2005; Kabkchieva et. al., 2009)
Jiran Terdekat-K	Pengurusan Bakat dan meramal prestasi pekerja (Hamidah, 2011)
Bayesian classifiers	Meramal prestasi pekerja (Al-Radaideh et. al., 2012) Meramal prestasi pelajar (Pandey et. al., 2011; Hamidah, 2011; Ibrahim, 2014; UKK, 2009)
Nearest Neighbor	Meramal prestasi pelajar (Pandey et. al., 2011; UKK, 2009)
Parzen-Window	Meramal prestasi pelajar (Pandey et. al., 2011)
Genetic Algorithm	Meramal prestasi pelajar (Pandey et. al., 2011)
SSVM	Meramal prestasi akademik pelajar kolej (Karahoca et. al., 2008)
K-Mean Algorithm	Meramal prestasi pelajar (Kabkchieva et. al., 2009)
Regression	Meramal prestasi pelajar (Rey et. al., 2013)
Rule learners	Meramal prestasi pelajar (Kabakchieva & KI, 2009)

Teknik pepohon keputusan adalah antara teknik pengelasan yang boleh menghasilkan kaedah-kaedah yang boleh ditafsirkan atau pernyataan logik. Kaedah-kaedah yang dihasilkan daripada teknik yang dipilih boleh digunakan untuk ramalan masa depan.

Terdapat banyak algoritma daripada teknik pepohon keputusan seperti C4.5/J48, Hutan Rawak (RF), NBTree, SimpleCart, REPTree, BFTree dan lain-lain. Jadual 4 menunjukkan pendekatan perlombongan data yang digunakan dalam peramalan prestasi manusia.

Algoritma klasifikasi C4.5/J48 adalah mudah difahami sebagai peraturan yang diperolehi mempunyai tafsiran yang sangat mudah. Ia merupakan pepohon keputusan yang induksi yang mana kelasnya ada yang nominal dan input data boleh nominal atau selang. Kadang-kadang saiz pepohon yang diinduksikan berkurangan apabila strategi cantasan berbeza digunakan.

Manakala algoritma Hutan Rawak (RF) pula, pemilihan ujian adalah berasaskan pada bilangan ciri-ciri rawak pada setiap nod dan perlaksanaannya adalah tanpa catasan. RF membina pepohon rawak dengan mengurangkan kumpulan pepohon rawak.

Jadual 4: Pendekatan perlombongan data.

Kajian	Jenis Model	Teknik	Algoritma	Hasil Kajian
Kerangka Kerja Sistem Sokongan Keputusan Cerdas Untuk Pengurusan Bakat (Hamidah, 2011)	Pengelasan	Pohon Keputusan	C4.5 (J48) Hutan Rawak (RF)	Pohon Keputusan - C4.5 (J48)
		Rangkaian Neural	Rambatan Multi-Lapisan (MLP) Rangkaian Fungsi Asas Radial (RBFN)	
		Jiran Terdekat-K	K-Star	
Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance (Al-Radaideh et. al., 2012)	Pengelasan	Pohon Keputusan	C4.5 (J48) ID3	Pohon Keputusan
		Naïve Bayes	Naïve Bayes	
Applying Data Mining Classification Techniques for Employee's Performance Prediction (Hamidah et. al., 2010)	Pengelasan	Pohon Keputusan	C4.5 (J48) Hutan Rawak (RF)	Pohon Keputusan - C4.5 (J48)
		Rangkaian Neural	Rambatan Multi-Lapisan (MLP) Rangkaian Fungsi Asas Radial (RBFN)	
		Jiran Terdekat-K	K-Star	
Data Mining to Cluster Human Performance by Using Online Self-Regulating Clustering Method (Lewichi et. al., 2005)	Pengelompokkan	Pengawalan Diri (Self Regulating)	X-mean	Optimum clusters

METODOLOGI KAJIAN

Pemodelan GC KPM bermula dengan pengumpulan dan penyediaan data GC di sekolah. Kemudian pra pemprosesan data dan seterusnya pembangunan model bagi menentukan model terbaik.

Pengumpulan dan Penyediaan Data

Data yang digunakan diperolehi secara terus daripada Bahagian Pendidikan Guru (BPG) melalui Sistem Pengurusan Latihan Kementerian Pelajaran Malaysia (SPL KPM). Sistem ini ditadbir oleh

Sektor Latihan ICT, BPG yang berperanan untuk mengumpul semua maklumat mengenai profil dan latihan semua pendidik di bawah Kementerian Pelajaran Malaysia. Oleh kerana salah satu faktor pemilihan guru cemerlang ialah pemantauan pengajaran dan pembelajaran (PdP), maka maklumat taraf PdP harus diperolehi daripada Jemaah Nazir dan Jaminan Kualiti yang mengendalikan proses pemantauan dan pemilihan tersebut. Pada akhir tahun 2013, Jemaah Nazir dan Jaminan Kualiti telah menyerahkan maklumat mengenai pemantauan PdP kepada BPG untuk melengkapkan profil seseorang guru. Jadual 5 menunjukkan data mentah yang diperolehi daripada Bahagian Pendidikan Guru, KPM.

Jadual 5: Data mentah.

Maklumat Data	Keterangan
Pemantauan PdP	Taraf pemantauan yang menggunakan Standard 4 SKPM
Pengalaman	Pengalaman mengajar dan kemahiran yang dimiliki
Kepakaran	Kepakaran yang ada
Sumbangan	Sumbangan dalaman, luaran dan kokurikulum
Status Kecemerlangan	Guru Cemerlang atau tidak

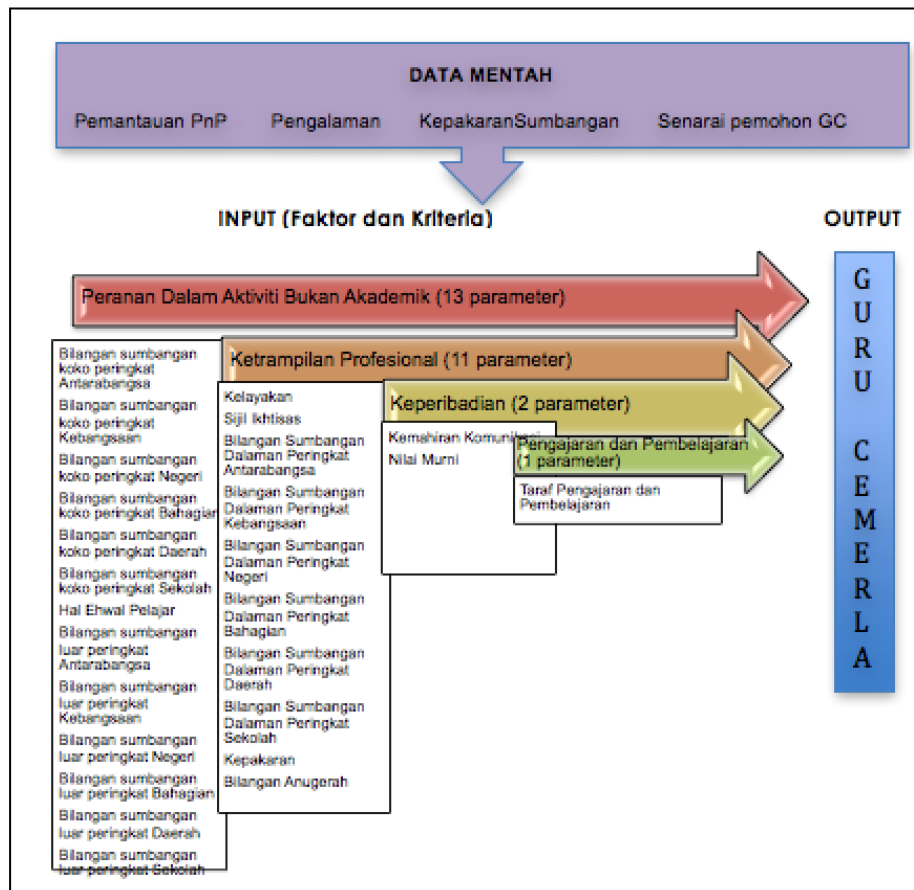
Data mentah yang terdapat dalam Jadual 5 dianalisis iaitu memahami apa yang terdapat disebalik semua data tersebut, mengelompokannya, meringkasnya menjadi suatu senang difahami, serta menemukan pola umum yang timbul dari data tersebut. Analisis data disebut juga pengolahan data dan penafsiran data.

Semasa menganalisis, data akan dikelompokkan berdasarkan parameter atau atribut, menyediakan data setiap parameter dan melakukan kajian untuk menjawab semua masalah yang dikemukakan. Data yang berupa kualitatif (kata-kata) akan dikuantifikasikan terlebih dahulu kemudian baru dianalisis.

Pra Pemprosesan Data

Data mentah yang diperolehi mempunyai banyak maklumat yang tidak wajar. Ini mungkin kesalahan memasukkan maklumat. Maka data-data tersebut dibersihkan dan bagi data yang tidak sempurna dan tidak konsisten juga dibuang. Oleh itu data akan semakin berkurangan.

Selepas itu data mentah tersebut digabungkan menjadi satu ketika pembangunan data warehouse dengan key nombor ID guru. Proses penggabungan data ini dilakukan ketika proses ETL (*extract, transform, and load*). Kemudian dipecahkan menjadi 4 faktor kecemerlangan guru seperti Rajah 3. Kriteria setiap faktor akan melalui pra pemprosesan bagi mendapat data yang bersih dan sesuai digunakan untuk perlombongan data. Oleh itu dalam gudang data terdapat dua set data iaitu data gabungan 4 faktor kecemerlangan guru dan data setiap faktor kecemerlangan guru.



Rajah 3: Padanan data.

Sebelum pembangunan model, data yang telah dipadankan akan melalui proses transformasi ke format yang sesuai untuk diproses dalam perlombongan data. Nilai setiap atribut akan dikategorikan dan diubah mengikut Jadual 6. Manakala nilai data yang nominal akan didiskritkan seperti Jadual 7.

Jadual 6: Perubahan nilai data atribut.

Nilai asal	Nilai diubah
0-3	1
4-6	2
7-9	3
10-12	4
13-15	5
>15	6

Jadual 7: Perubahan nilai data nominal ke diskrit.

Nama Atribut	Nilai Asal	Nilai diubah
Hal Ehwal Pelajar	Ya, Tidak	1, 2
Kelayakan	Diploma, Ijazah, Sarjana, PhD	1, 2, 3, 4
Sijil Ikhtisas	Ada, Tiada	1, 2
Kepakaran dalam bidang	Bilangan kepakaran	1, 2, 3
Taraf PdP	Sangat Lemah, Lemah, Memuaskan, Harapan, Baik, Cemerlang	1, 2, 3, 4, 5, 6

Contohnya pada Jadual 7, apabila nilai asal adalah Ya atau Tidak maka nilai tersebut akan diubah menjadi nilai diskrit iaitu 1 dan 2. Begitu juga apabila nilai asal adalah ‘Bilangan kepakaran’ maka jumlah bilangan itu perlu diubah dulu menggunakan Jadual 6 kemudian baru diubah menggunakan Jadual 7.

Ekspirimen

Data bersih yang digunakan ialah sebanyak 5516 data. Set data yang digunakan melibatkan 26 atribut input dengan 1 atribut output. Untuk pemodelan ini, teknik pepohon keputusan dalam metod pengelasan digunakan kerana ia biasa diaplikasikan untuk pemodelan bakat manusia. Manakala dua algoritma yang popular daripada teknik pepohon keputusan ini digunakan iaitu algoritma C4.5 (J48) dan Hutan Rawak (RF). Bagi menentukan ketepatan model semasa pembangunan model, sebanyak 5 set data rawak digunakan dan diimplimentasikan dengan kaedah 10 lipatan silang (*10-fold cross validation*) untuk percubaan dan pengujian. Perisian WEKA 3.7.0 digunakan sebagai alat untuk pemodelan.

Petua yang dihasilkan melalui algoritma C4.5 dan RF mempunyai ciri-ciri seperti mudah difahami, mudah dibaca serta mempunyai pentafsiran yang jelas. Ia juga boleh menghasilkan petua dalam bentuk pohon dan set petua. Prestasi model dinilai berdasarkan peratus ketepatan dan ralat min kuasa dua (RMSE) setiap model.

DAPATAN KAJIAN

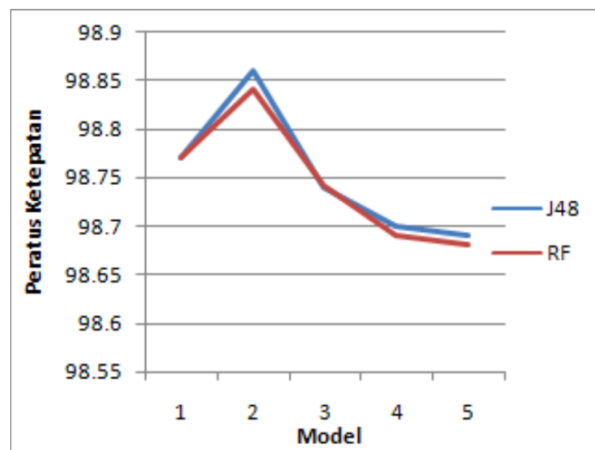
Jadual 8 menunjukkan peratus ketepatan dan ralat min kuasa dua model yang menggunakan algoritma J48 dan Hutan Rawak. Model 2 J48 merupakan model terbaik dengan peratus ketepatan 98.86%.

Walau bagaimanapun, algoritma RF juga mempunyai peratus ketepatan yang lebih kurang sama dengan J48 iaitu 98.84%. Model terbaik bagi RF juga model 2. Model terbaik bagi kedua-dua algoritma ini menjana 80% data untuk percubaan dan 20% untuk pengujian. Graf peratus ketepatan model dan ralat min kuasa dua model dapat dilihat seperti dalam Rajah 4 dan Rajah 5.

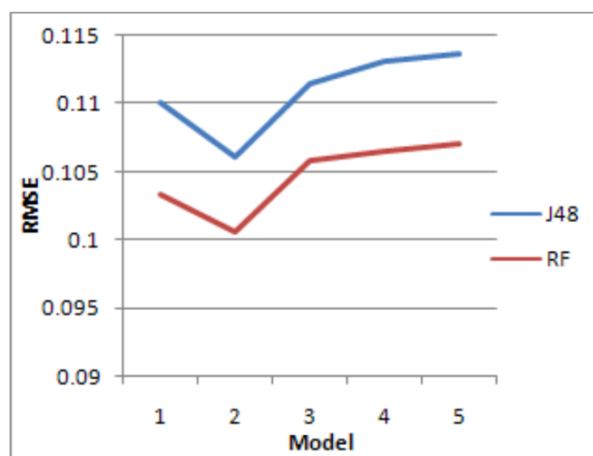
Jadual 8: Keputusan pemodelan algoritma J48 dan Hutan Rawak (RF).

Model	Model	J48		RF	
		Ketepatan (%)	RMSE	Ketepatan (%)	RMSE
1	90:10	98.77	0.1101	98.77	0.1033
2	80:20	98.86	0.1061	98.84	0.1006
3	70:30	98.74	0.1114	98.74	0.1058
4	60:40	98.70	0.1131	98.69	0.1064
5	50:50	98.69	0.1136	98.68	0.1070

Rajah 4: Peratus ketepatan model.



Rajah 5: Ralat Min Kuasa Dua (RMSE) Model.



PENUTUP DAN RUMUSAN

Dapatan kajian menunjukkan bahawa model yang diperolehi daripada pemodelan GC menggunakan teknik perlombongan data dan model pemilihan GC yang sedia ada mempunyai persamaan lebih kurang 98%. Maka untuk kajian yang akan datang, pemodelan boleh dilakukan selepas proses pemilihan atribut bagi mendapatkan atribut yang sepatutnya. Selain itu, data kajian juga boleh dilakukan proses pendiskritan atau normalisasi terlebih dahulu sebelum dimodelkan. Mungkin hasil yang diperolehi lebih baik dan terdapat banyak perbezaan di antara teknik.

Data yang digunakan untuk kajian ini juga merupakan data yang mentah yang belum digunakan untuk eksperimen yang lain. Oleh itu data ini boleh juga diuji menggunakan teknik dan algoritma pengelasan yang lain. Perbandingan boleh dilakukan di antara teknik bagi menentukan teknik yang terbaik.

RUJUKAN

- Al-Radaideh, Q. A., & Al Naqi, E. (2012). Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance. (IJACSA) *International Journal of Advanced Computer Science and Applications*, 3(2), 144–151. Retrieved from www.ijacsa.thesai.org.
- Bhardwaj, B. K., & Pal, S. (2011). Data Mining: A prediction for performance improvement using classification. *International Journal of Computer Science and Information Security*, 9(4).
- Delahaye, B. (2011). *Human Resource Development Managing Learning and Knowledge Capital* (Third Edit). Australia: Tilde University Press.
- Hamidah Jantan. (2011). *Kerangka Kerja Sistem Sokongan Keputusan Cerdas Untuk Pengurusan Bakat*. Universiti Kebangsaan Malaysia.
- Hamidah Jantan, Hamdan, A. R., & Othman, Z. A. (2010). Human Talent Prediction in HRM using C4.5 Classification Algorithm. (IJCE) *International Journal of Computer Science and Engineering*, 02(08), 2526–2534.
- Hamidah, Mazidah, Razak, A., & Ali, Z. (2010). *Applying Data Mining Classification Techniques for Employee's Performance Prediction*, 601–607. Retrieved from <http://www.kmice.cms.net.my/ProcKMICE/KMICE2010/Paper/PG601-607.pdf>
- Kabakchieva, D., & Kl, S. (2009). *Analyzing University Data for Determining Student Profiles and Predicting Performance*. In Proceedings of the 4th International Conference on Educational Data mining, Eindhoven. The Netherlands.
- Karahoca, A., Karahoca, D., & Kaya, O. (2008). *Data Mining To Cluster Human Performance By Using Online Self Regulating Clustering Method*. In 1st WSEAS International Conference on Multivariate Analysis and its Application in Science and Engineering (MAASE) '08. Istanbul (pp. 198–203).
- Kova, Z. J. (2012). Predicting student success by mining enrolment data. *Research in Higher Education Journal*, 15(iii), 1–20.
- KPM. (2010). *Terma Rujukan Konsep Guru Cemerlang*.
- KPM. (2012). *Laporan Awal Pelan Pembangunan Pendidikan Malaysia 2013-2025*.
- Lewichi, P., & Hill, T. (2005). *Statistics : Methods and Applications*. (StatSoft, Ed.) (1st ed.). Tulsa: Statistica Data Analysis Software and Services.
- Martinsons, M. G. (1995). Knowledge-based systems leverage human resource management expertise. *International Journal of Manpower*, 16(2), 17–34.
- Minaei-bidgoli, B., Kashy, D. A., Kortemeyer, G., & Punch, W. F. (2003). *Predicting student performance: an application of data mining methods with an educational Web-based system*. In FIE 2003 33rd Annual (p. T2A–13). Frontiers in Education.
- Pandey, U. K., & Pal, S. (2011). Data Mining: A prediction of performer or underperformer using classification. (IJCSIT) *International Journal of Computer Science and Information Technologies*. 2(2), 686–690.

- Rey, T. D., Dow, T., Company, C., Wells, C., Kahl, J., & Services, T. C. (2013). *Using Data Mining in Forecasting Problems*.
- Sembiring, S., Zarlis, M., Hartama, D., Ramliana, S., & Wani, E. (2011). *Prediction Of Student Academic Performance By An Application Of Data Mining Techniques*. In International Conference on Management and Artificial Intelligence IPEDR vil.6 (2011). Bali, Indonesia.
- Stavrou-Costea, E. (2005). The challenges of human resource management towards organizational effectiveness: A comparative study in Southern EU. *Journal of European Industrial Training*, 29(2), 112–134.
- Unit Komunikasi Korporat. (2009, June 23). *Laluan Naik Pangkat Pegawai Perkhidmatan Pendidikan "istimewa."* Utusan Online. Putrajaya. Retrieved from http://ww1.utusan.com.my/utusan/info.asp?y=2009&dt=0623&pub=Utusan_Malaysia&sec=Forum&pg=fo_03.htm
- Unit Komunikasi Korporat. (2010, September 28). *Guru Cemerlang Sedia Dipantau*. Berita Harian.
- Weiss, S. M., & Indurkha, N. (1998). *Predictive Data Mining A Practical Guide*. (M. B. Morgan, Ed.). San Francisco: Morgan Kaufman Publishers.
- Wook, M., Yahaya, Y. H., Wahab, N., Mohd Isa, M. R., Awang, N. F., & Seong, H. Y. (2009). *Predicting NDUM Student's Academic Performance Using Data Mining Techniques*. In 2009 Second International Conference on Computer and Electrical Engineering (pp. 357–361). Dubai.