

<http://www.ftsm.ukm.my/apjitm>

Asia-Pacific Journal of Information Technology and Multimedia

Jurnal Teknologi Maklumat dan Multimedia Asia-Pasifik

Vol. 7 No. 2, December 2018: 29 - 45

e-ISSN: 2289-2192

DATA ANALYTICS IN MALAYSIAN EDUCATION SYSTEM: REVEALING THE SUCCESS OF SIJIL PELAJARAN MALAYSIA FROM UJIAN APTITUD SEKOLAH RENDAH

AZHAR MOHD KHAIRY
AFZAN ADAM
MOHD RIDZWAN YAAKUB

ABSTRACT

Information on the student's cognitive abilities can help teachers to identify the strengths or potential of a student to plan a learning strategy. These data are collected through Ujian Aptitud Sekolah Rendah in year six (UASR), where the student's potential can be detected five years earlier before they take their Sijil Pelajaran Malaysia (SPM). Unfortunately, these data have not yet used as a criterion in the student's development plan. Therefore, this research has been done to see the strength of the connection between the student's potential in UASR and their results in SPM. Through data visualization, a strong connection between a student's potential in UASR and their successfulness in SPM can be seen. 71.48% of students in the first cohort that have been proven to have high potential during their year six in 2011 and 73.63% of the next cohort in 2012 have obtained a good SPM result that they took in 2016 and 2017 respectively. Further analysis using the decision tree technique shows a few other factors for the high potential student success in SPM. A few of them are their results in PT3, their grades in the Maths and Science subjects, academic stream, and the school's achievement for that particular stream for the year before.

Keywords: data analytics, result tree, aptitude test, SPM

ANALITIK DATA DALAM SISTEM PENDIDIKAN MALAYSIA: MEMBONGKAR KEJAYAAN SIJIL PELAJARAN MALAYSIA DARI UJIAN APTITUD SEKOLAH RENDAH

ABSTRAK

Maklumat kemahiran kognitif murid boleh membantu para guru mengenalpasti kekuatan atau potensi yang ada pada murid bagi merancang strategi pembelajaran. Maklumat ini dikumpul melalui Ujian Aptitud Sekolah Rendah tahun enam (UASR), dimana potensi murid dapat dikesan lima tahun lebih awal sebelum mereka menduduki peperiksaan Sijil Pelajaran Malaysia (SPM). Malangnya, maklumat yang ada ini tidak dijadikan sebagai salah satu kriteria dalam pelan pembangunan pelajar. Oleh itu, kajian ini dibuat untuk melihat kekuatan hubungan faktor potensi pelajar dalam UASR dengan kejayaan SPM mereka. Melalui visualisasi data, hubungan kuat antara murid berpotensi tinggi dengan kejayaannya dalam SPM boleh dilihat. Sebanyak 71.48% murid bagi kohort pertama yang telah diuji berpotensi tinggi semasa tahun enam pada tahun 2011 serta 73.63% kohort berikutnya pada tahun 2012, telah memperolehi keputusan SPM yang lebih berjaya semasa mendudukinya pada tahun 2016 dan 2017 masing-masing. Analisis lanjut dengan teknik pokok keputusan menunjukkan beberapa faktor lain yang mempengaruhi tahap kejayaan murid dalam SPM, antaranya ialah tahap pencapaian semasa menduduki PT3, gred dalam mata pelajaran Matematik dan Sains, aliran akademik dan faktor tahap pencapaian aliran di sekolah pada tahun sebelumnya.

Kata kunci: analitik data, pokok keputusan, ujian aptitud, SPM

PENGENALAN

FAKTOR DAN PENCAPAIAN AKADEMIK MURID

Tahap pencapaian akademik murid yang berkualiti merupakan keutamaan yang diberi oleh pihak pendidik dalam menentukan kejayaan pendidikan di setiap peringkat. Sejak beberapa dekad, terdapat banyak penyelidikan telah dibuat untuk mencari atribut atau pembolehubah yang menyumbang secara efektif terhadap kualiti murid dalam pencapaian akademik. Faktor-faktor ini boleh diistilahkan sebagai faktor murid, faktor sekolah, faktor keluarga dan faktor rakan sebaya (Crosnoe et al. 2004).

Terdapat juga beberapa kajian yang telah dijalankan di luar negara menggunakan perlombongan data untuk membuat peramalan pencapaian murid di peringkat kolej dan universiti. Menurut Shahiri et al. (2015) dalam kajiannya menyatakan terdapat dua faktor utama dalam membuat peramalan pencapaian murid iaitu atribut dan teknik peramalan yang digunakan. Bagi faktor atribut, Purata Nilai Gred Kumulatif atau PNGK yang mengambil kira penilaian dalaman sering digunakan. Penilaian dalaman dikumpulkan daripada kuiz, kerja amali, ujian bulanan dan rekod kehadiran. Melalui tinjauannya daripada 30 kajian antara tahun 2003 hingga 2014, satu pertiga daripadanya menggunakan PNGK sebagai atribut utama dalam peramalan pencapaian pelajar. Antara sebab atribut ini dipilih adalah kerana ia mempunyai nilai yang ketara untuk pendidikan masa depan dan pergerakan karier pelajar serta sebagai petunjuk sebenar potensi akademik mereka. Shahiri et. al. juga mendapati bahawa atribut penilaian dalaman juga merupakan peramal yang baik. Ini sejajar dengan penemuan Ibrahim (2007) yang menyatakan bahawa menerusi keputusan analisis korelasi koefisien, PNGK merupakan pembolehubah paling signifikan dengan nilai 0.87 berbanding pembolehubah yang lain.

Fadhilah Ahmad dan Asif *et.al.* juga telah menggunakan data PNGK dan demografi pelajar untuk meramal pencapaian akademik pelajar tahun satu (Ahmad, 2015), dan menjelang tamat tempoh pengajian (Asif *et.al* 2017) di unversiti. Menggunakan beberapa teknik pembelajaran mesin seperti *Naïve bayes*, pokok keputusan dan petua, 71.3% ketepatan peramalan berjaya diperoleh. Kemudian, Dutt *et.al* (2017) membuat kajian sorotan terhadap kaedah pengelompokan dan kebergunaannya keatas data pendidikan dari tahun 1983 hingga 2016. Hasil sorotan terhadap 30 tahun kajian, Dutt *et.al.* menyimpulkan bahawa teknik pengelompokan boleh memberikan definisi baru terhadap atribut-atribut yang digunakan. Terdapat juga kajian peringkat universiti yang menggunakan data-data tingkahlaku dan pencapaian pelajar dalam kelas, untuk memahami gaya pembelajaran dan hubungan dengan faktor kejayaan murid (Amin dan Mustafa, 2014), serta meramal prestasi pelajar (Zainudin, 2012). Jambari *et.al.* (2015) pula melaporkan penggunaan data pelajar institusi pengajian tinggi bagi membangunkan model peningkatan pengajian pelajar.

Selain itu, kajian yang melibatkan atribut demografi seperti jantina, umur, latar belakang keluarga dan ketidakupayaan, serta penilaian luaran melibatkan markah peperiksaan bagi mata pelajaran tertentu juga sering digunakan. Terdapat bukti menunjukkan bahawa jantina merupakan atribut yang penting mempengaruhi pencapaian murid. Kebanyakan murid perempuan mempunyai kepelbagaian gaya yang positif dalam pembelajaran dan tingkah lakunya berbanding lelaki (Meit et al. 2004). Mereka juga mempunyai strategi pembelajaran yang lebih efektif (Simsek & Balaban 2010). Haist et al. (2000) dalam kajiannya terhadap pelajar perubatan mendapati pencapaian lelaki dan perempuan adalah bergantung kepada ketetapan yang berbeza iaitu dalam situasi tertentu perempuan lebih baik daripada lelaki, dan begitu juga sebaliknya.

Justeru, faktor psikometrik pernah dikaji untuk digunakan sebagai peramal dalam beberapa kajian. Sembiring et al. (2011) menghubungkan atribut tingkah laku pelajar universiti melalui faktor psikometrik dengan pencapaian akhir akademik mereka. Sembiring

membuktikan bahawa kondisi mental pelajar; iaitu cara dan masa murid belajar, minat dan sokongan keluarga mempunyai korelasi yang tinggi dengan pencapaian akhir akademik. Kajian terbaru yang dijalankan di Iran University of Science and Technology pula menggunakan teknik pokok keputusan dan *induction tree* untuk mencari punca pelajar mencetak bahan-bahan pembelajaran interaktif (Rostaminezhad, 2018).

Rumusannya, pengkajian menerusi penggunaan analitik data dapat memberi cadangan kepada pihak Kementerian Pendidikan Malaysia (KPM) untuk menggunakannya sebagai pendekatan alternatif baharu yang boleh diaplikasikan dalam penyelidikan pendidikan serta membantu dalam membuat perancangan, keputusan dan menggubal sesuatu dasar pendidikan. Oleh itu, kajian ini menawarkan cadangan penyelidikan yang menggabungkan teknik perlombongan data dalam mengesan atau mengecam faktor-faktor signifikan yang boleh dijelaskan terhadap pencapaian akademik murid melalui data berskala besar di KPM.

KEMAHIRAN KOGNITIF DALAM UJIAN APTITUD SEKOLAH RENDAH (UASR)

UASR ialah ujian aptitud yang dilaksanakan dari tahun 2009 dan berakhir pada tahun 2015. Ujian ini dilaksanakan ke atas murid yang berumur 12 tahun setelah melalui alam persekolahan rendah selama 6 tahun. Hasil ujian ini memberi gambaran awal tentang kebolehan, minat dan kecenderungan pelajar. Tahun berikutnya, ujian ini digantikan dengan Ujian Aptitud Am Tahun 3 dan Ujian Aptitud Am Tahun 6 di bawah payung Pentaksiran Berasaskan Sekolah (PBS) iaitu melalui komponen Pentaksiran Psikometrik (PPsi). Perbezaannya ialah dari segi instrumen dan peringkat pentadbiran ujian di mana UASR ditadbirkan secara serentak di seluruh negara sejurus murid tahun enam selesai menduduki Ujian Pencapaian Sekolah Rendah (UPSR) manakala di bawah PPsi, Ujian Aptitud Am ini ditadbirkan oleh pihak sekolah di premis masing-masing mengikut kesesuaian masa dan tempat.

UPSR adalah peperiksaan am untuk menguji tahap pencapaian akademik murid di sekolah rendah yang merujuk kepada Kurikulum Bersepadu Sekolah Rendah (KBSR). Berbeza dengan UPSR, UASR tidak merujuk kepada kurikulum dan tidak terdedah kepada intervensi seperti yang berlaku kepada ujian pencapaian. UASR digunakan untuk meramal kejayaan seseorang di masa akan datang iaitu dengan memberi maklumat awal tentang tahap kemahiran kognitif murid. Pelbagai definisi mengenai aptitud ini bolehlah disintesis sebagai kebolehan seseorang individu mempelajari sesuatu dan melakukan sesuatu kemahiran dengan mudah dan pantas, (Lembaga Peperiksaan, 2014). Justeru, melalui respon yang diberi semasa UASR, potensi murid boleh dikenalpasti. Maklumat ini boleh digunakan oleh pihak sekolah untuk membuat persediaan rapi dalam merencanakan segala program dan aktiviti bersesuaian dengan kemahiran kognitif dan potensi murid. UASR juga boleh digunakan untuk pemilihan sekolah, menentukan aliran pembelajaran yang sesuai serta mengetahui bidang dan minat kerjaya mereka.

Dalam kajian ini, UASR dihubungkan dengan dengan pencapaian terakhir akademik murid di sekolah menengah iaitu keputusan peperiksaan Sijil Pelajaran Malaysia (SPM), untuk mengenalpasti corak yang signifikan supaya boleh dijadikan sebagai faktor peramal terawal. Sepanjang tempoh lima tahun daripada tarikh menduduki UASR ke tarikh SPM, murid telah melalui persekolahan menggunakan Kurikulum Bersepadu Sekolah Menengah (KBSM) dan ditaksir secara berterusan melalui PBS. Walau bagaimanapun, pencapaian akademik di tiga tahun pertama sekolah menengah diukur secara sumatif melalui Pentaksiran Tingkatan 3 (PT3).

FAKTOR PT3 DAN PENCAPAIAN SEKOLAH DALAM SPM TERHADAP KEJAYAAN MURID

PT3 adalah salah satu komponen daripada 4 komponen dalam PBS yang mentaksir pencapaian akademik murid secara sumatif di peringkat menengah rendah. Pentaksiran ini berbeza dengan peperiksaan pusat iaitu PT3 diurustadbirkan sepenuhnya di peringkat sekolah sahaja.

Keputusan PT3 ini dilaporkan bersama-sama dengan komponen PBS lain iaitu pelaporan Pentaksiran Psikometrik (PPsi), Pentaksiran Aktiviti Jasmani Sukan dan Kokurikulum (PAJSK) dan Pelaporan Perkembangan Pembelajaran Murid (PPPM). Rangkuman kesemua pelaporan ini adalah bersifat deskriptif dan holistik yang memberi maklumat terperinci tentang perkembangan murid daripada sudut akademik dan bukan akademik sepanjang tiga tahun peringkat menengah rendah.

PT3 telah mula dilaksanakan mulai tahun 2014 bagi murid sekolah menengah rendah dan kajian ini turut melibatkan kohort pertama berkenaan. PT3 juga digunakan sebagai salah satu rujukan kepada murid dan pihak sekolah untuk menyambung ke aliran bersesuaian di peringkat menengah atas selama dua tahun sebelum menduduki SPM. Pihak sekolah di menengah atas mempunyai pakej mata pelajaran yang disediakan untuk murid dialirkan ke bidang yang bersesuaian dengan minat dan kebolehan akademik masing-masing. Pihak sekolah turut mempunyai rekod lalu tentang pencapaian semua muridnya dari segi prestasi keseluruhan mata pelajaran dalam SPM untuk digunakan sebagai rujukan.

Oleh itu, setiap sekolah mempunyai skor prestasi mata pelajaran yang berbeza. Ini sepatutnya boleh digunakan untuk pelajar dan ibubapa memilih sekolah yang dirasakan sesuai dan terbaik untuk meningkatkan atau mengekalkan tahap pencapaian mereka. Walau bagaimanapun, senario sebegini adalah amat terhad memandangkan terdapatnya faktor lain seperti lokasi sekolah dan dasar pemilihan murid oleh sekolah berkenaan. Kesemua ini memberi petunjuk bahawa terdapat pelbagai faktor atau atribut yang terdapat dalam persekitaran pendidikan sekolah di Malaysia. Justeru, faktor penyumbang utama terhadap pencapaian murid sepanjang tempoh lima tahun dari UASR ke SPM itu perlu dikenalpasti bagi menaikkan lagi peratus kejayaan, dan mengurangkan bilangan murid yang keciciran.

KAEDAH KAJIAN

Antara persoalan yang ingin dijawab dalam kajian ini ialah:

1. Apakah hubungan antara murid berpotensi tinggi dengan tahap kejayaan mereka semasa SPM?
2. Apakah faktor-faktor yang mempengaruhi murid yang berpotensi tinggi mendapat pencapaian yang lebih rendah dalam SPM?

DATA DAN PRA-PEMROSESAN

Data telah diperoleh dengan kebenaran dari Lembaga Peperiksaan, Kementerian Pendidikan Malaysia. Dua kohort terlibat dalam kajian ialah kohort pertama (K1) yang menduduki UPSR & UASR pada tahun 2011, PT3 pada tahun 2014 dan SPM pada tahun 2016, manakala kohort kedua (K2) menduduki UPSR & UASR pada tahun 2012, PT3 pada tahun 2015 dan SPM pada tahun 2017.

Data untuk kedua-dua kohort dikecilkkan kepada negeri Selangor sahaja. Set data UPSR & UASR, PT3 dan SPM diperoleh secara berasingan dan padanan ketiga-tiganya dibuat bersama dengan pihak LP. Tujuannya supaya akur kepada peraturan Akta Rahsia Rasmi 1972 untuk menjaga kerahsiaan data individu murid. Atribut sulit yang menunjukkan identiti individu dan identiti sekolah juga dikeluarkan terus daripada set data sebelum diserahkan kepada pengkaji. Atribut yang dikaji adalah seperti Jadual 2. Jumlah murid yang berjaya dipadankan bagi ketiga-tiga peperiksaan dan pentaksiran ialah seramai 60,583 murid untuk K1 dan 64,096 murid bagi K2 seperti ditunjukkan pada Jadual 1.

JADUAL 1. Bilangan dan peratus murid berdasarkan potensi mengikut kohort

Kohort	Potensi		Jumlah (Set S)
	Tinggi (Set T)	Rendah (Set R)	
K1	23,989 (39.60%)	36,594 (60.40%)	60,583 (100%)
K2	24,082 (37.57%)	40,014 (62.43%)	64,096 (100%)

Bagi tujuan pengkelasan tahap pencapaian SPM, kumpulan murid dikategorikan kepada dua tahap pencapaian: 'lebih berjaya' dan 'kurang berjaya'. Ukuran bagi kelas tahap pencapaian adalah berdasarkan purata nilai mata yang diperoleh bagi keputusan kesemua mata pelajaran yang diduduki. Kaedah pengiraan yang fleksibel ini digunakan oleh LP untuk mendapatkan nilai indeks Gred Purata bagi melaporkan pencapaian calon dan prestasi mata pelajaran di semua peringkat seperti sekolah, daerah, negeri dan nasional. Setiap gred akan ditukar kepada nilai mata iaitu A+=0, A=1, A-=2, B+=3, B=4, C+=5, C=6, D=7, E=8 dan G=9. Kesemua gred yang diperoleh dijumlahkan seterusnya dipuratakan sehingga 2 nilai titik perpuluhan. Setiap mata pelajaran membawa nilai pemberat yang sama. Purata nilai mata yang lebih kecil iaitu sama atau lebih menghampiri 0.00 adalah menggambarkan tahap pencapaian yang lebih baik berbanding purata nilai mata yang lebih besar menghampiri atau sama dengan 9.00. Oleh itu, purata nilai mata 0.00 hingga 4.49 dikelaskan sebagai 'lebih berjaya' manakala 4.50 hingga 9.00 dikelaskan sebagai 'kurang berjaya' Purata nilai 4.50 dipilih berbanding 5.00 (nilai sama jarak antara 0.00 dan 9.00) adalah untuk mengurangkan ralat yang berada di sempadan antara kedua-dua kelas.

JADUAL 2. Senarai atribut dan subset

Bil	Atribut	Subset
1	Jantina	{L, P}
2	Tahap UPSR.	{Cemerlang, Baik, Sederhana, Lemah}
3	Potensi	{Tinggi, Rendah}
4	02-Bahasa Melayu (Pemahaman) PT3	{AB, CD, EFT}
5	03-Bahasa Melayu (Penulisan) PT3	{AB, CD, EFT}
6	12-Bahasa Inggeris (Pemahaman) PT3	{AB, CD, EFT}
7	13-Bahasa Inggeris (Penulisan) PT3	{AB, CD, EFT}
8	50-Matematik PT3	{AB, CD, EFT}
9	55-Sains PT3	{AB, CD, EFT}
10	Tahap PT3	{Cemerlang, Baik, Sederhana, Lemah}
11	Aliran	{Ag, SAMPEI, SAS, SMPEI, SMT, ST, STA, STMPEI, XSTEM}
12	Jum MP SPM	{6, 7, 8, 9, 10, 11, 12}
13	Tahap aliran sebelum	{Cemerlang, Baik, Sederhana, Lemah, Null}
14	Tahap kejayaan SPM	{Lebih berjaya, Kurang berjaya}

ATRIBUT YANG DIGUNAKAN

Sebanyak 14 atribut dikaji dan senarainya ditunjukkan pada Jadual 2. Tahap UPSR ialah atribut yang dibentuk daripada pencapaian murid secara keseluruhannya dan dikategorikan kepada 4 tahap. Seperti SPM, pelaporan pencapaian calon dan prestasi mata pelajaran UPSR juga menggunakan nilai indeks Gred Purata. Nilai yang diberi pada gred UPSR ialah A=1, B=2, C=3, D=4 dan E=5. Bagi tujuan kajian, pengurangan data dibuat dengan menukarkan nilai indeks tersebut daripada numerik kepada nominal mengikut 4 julat tahap pencapaian iaitu Cemerlang (1.00 hingga 1.49), Baik (1.50 hingga 2.99), Sederhana (3.00 hingga 4.49) dan Lemah (4.50 hingga 5.00). Atribut Tahap PT3 juga dibentuk mengikut kaedah sama seperti atribut Tahap UPSR. Perbezaannya ialah tambahan pada gred F=6 dan julat pada tahap pencapaian iaitu Cemerlang (1.00 hingga 1.99), Baik (2.00 hingga 3.49), Sederhana (3.50 hingga 4.99) dan Lemah (5.00 hingga 6.00).

Atribut aliran murid disediakan dengan memetakan secara manual pakej aliran STEM yang dikemukakan oleh Bahagian Pembangunan Kurikulum kepada maklumat pendaftaran mata pelajaran SPM. Pakej aliran STEM adalah seperti ditunjukkan dalam Jadual 3.

Atribut Aliran SPM Tahun Sebelum menggunakan gred purata aliran murid SPM di sekolah masing-masing pada tahun sebelumnya sebagai faktor peramal untuk tahun semasa. Gred purata tersebut dikurangkan kepada 4 tahap pencapaian dengan julat tahap pencapaian Cemerlang (0.00 hingga 1.99), Baik (2.00 hingga 4.49), Sederhana (4.50 hingga 6.99) dan Lemah (7.00 hingga 9.00). Set data SPM tahun 2016 menggunakan atribut SPM tahun 2015 dan set data SPM tahun 2017 menggunakan atribut bagi tahun 2016.

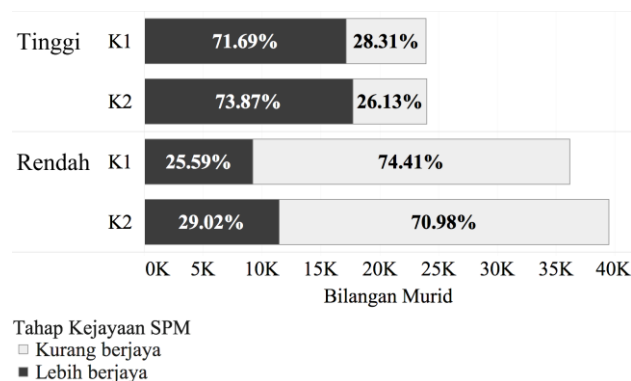
JADUAL 3. Aliran STEM dan pakej mata pelajaran SPM

Kod	Pakej Mata Pelajaran (MP)
ST	Sains Tulen
STA	Sains Tulen & Agama
STMPEI	Sains Tulen & MP Elektif Ikhtisas
SMT	Sains & Matematik Tambahan
SAS	Sains & Sains Tambahan
SAMPEI	Sains, Agama & MP Elektif Ikhtisas
SMPEI	Sains & MP Elektif Ikhtisas
XSTEM	Aliran Selain STEM

HUBUNGAN POTENSI MURID DAN PENCAPAIAN SPM

Potensi yang tinggi dalam kajian ini didefinisikan sebagai murid yang mendapat peringkat kombinasi band tertinggi bagi kemahiran berfikir, kemahiran menyelesaikan masalah dan kemahiran membuat keputusan yang terdapat dalam UASR. Jadual 1 menunjukkan sekitar 37.57% hingga 39.60% dari jumlah sampel kedua-dua kohort adalah murid berpotensi tinggi. Daripada murid berpotensi tinggi tersebut, 28.31% daripada K1 dan 26.13% daripada K2 mendapat keputusan yang 'kurang berjaya' dalam SPM seperti ditunjukkan dalam Rajah 1. Hubungan secara langsung antara murid berpotensi tinggi dengan tahap kejayaan SPM bagi kedua-dua kohort dapat diperhatikan berbanding murid berpotensi rendah yang berhubung secara songsang. Faktor-faktor yang menyebabkan murid ini tidak mencapai sasaran berdasarkan potensi yang ada memerlukan siasatan lanjut. Kajian ini akan menggunakan teknik

pembelajaran mesin menerusi pokok keputusan untuk mengenalpasti faktor-faktor yang mempengaruhi kejayaan pelajar berkenaan. Faktor-faktor yang terdapat adalah daripada atribut daripada persekitaran data PT3 dan SPM sahaja (rujuk Jadual 2).



RAJAH 1. Peratus murid berdasarkan tahap kejayaan SPM berdasarkan potensi

TEKNIK POKOK KEPUTUSAN UNTUK PENGENALPASTIAN FAKTOR

Terdapat banyak kajian mengaplikasikan teknik pembelajaran mesin untuk mencapai pelbagai tujuan dalam pendidikan dan kebanyakannya berfokus kepada kejayaan pelajar. Contohnya, antara kajian terawal iaitu aplikasi pembelajaran mesin dalam bidang akademik dan pendidikan dilakukan oleh Kotsiantis et al. (2004) iaitu dengan meramal keciciran pelajar. Bhardwaj dan Pal (2012) telah mengkaji pelbagai faktor melalui 17 atribut penting yang memberi kesan besar kepada pencapaian pelajar. Bakhshinategh et al. (2017) telah mengkategorikan semula pelbagai tugas dan pengaplikasian yang wujud dalam EDM sejak 10 tahun yang lalu. Terdapat 13 kategori aplikasi EDM iaitu meramal pencapaian murid, mengesan tingkah laku tidak diingini pada murid, pengelompokan murid, analisis jaringan sosial, penyediaan laporan, membuat amaran awal kepada pemegang taruh, perancangan dan penjadualan, penyediaan perisian kursus, membina peta konsep, menjana cadangan, sistem penyesuaian, penilaian (bantuan komputer) dan penyiasatan saintifik.

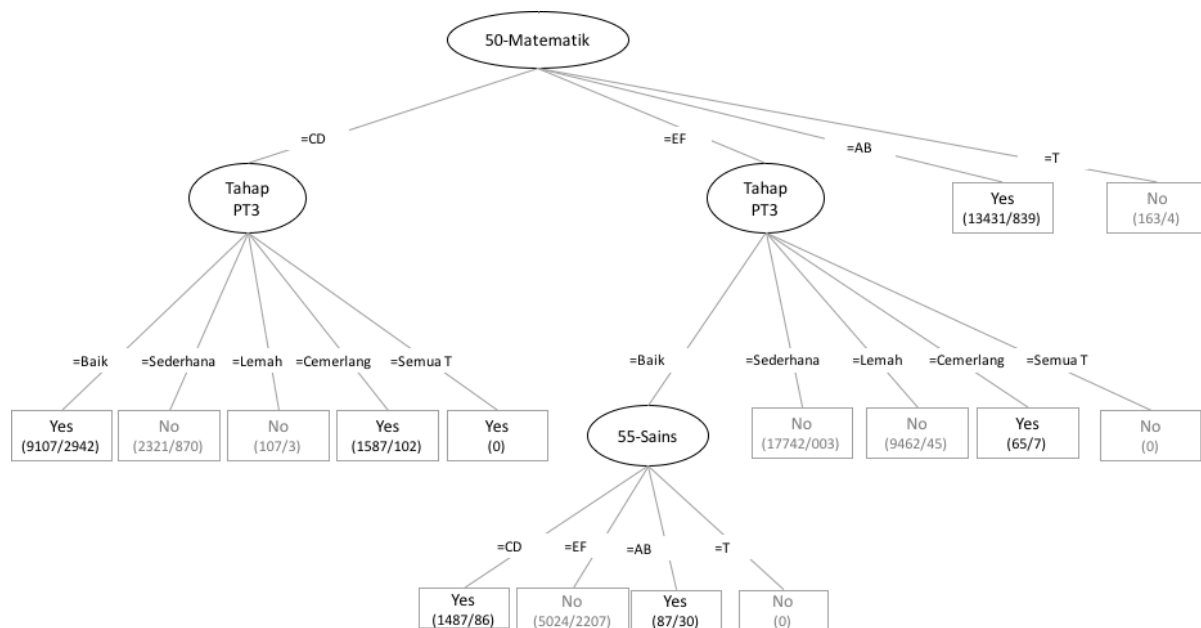
Dalam senario pendidikan di Malaysia, terdapat beberapa atribut penting yang boleh dikaji menggunakan pembelajaran mesin seperti demografi murid dan guru, pencapaian murid dan sekolah dalam peperiksaan awam dan pentaksiran berasaskan sekolah, psikometrik murid, penglibatan murid dalam aktiviti ko-kurikulum iaitu persatuan dan sukan, kehadiran murid, kompetensi guru dan banyak lagi. Namun sehingga kini, kajian berbentuk peramalan menggunakan pembelajaran mesin di peringkat sekolah menengah masih lagi kurang mendapat tempat atau diketengahkan.

Algoritma yang akan digunakan ialah C4.5 (J48) yang telah dibangunkan oleh Quinlan (1993) iaitu lanjutan penambahbaikan program asal *Iterative Dichotomiser 3* (ID3). Algoritma C4.5 membina pokok keputusan dengan memilih atribut yang terbaik dahulu untuk dijadikan akar dan kemudian meletakkan atribut-atribut lainnya pada ranting. Proses ini berterusan sehingga kesemua atribut diletakkan pada pokok tersebut. Kriteria pemilihan atribut yang digunakan ialah berdasarkan pengiraan entropi dan *information gain* yang terbesar untuk memaksimumkan kadar ramalan. Seterusnya, algoritma C4.5 akan menggunakan kaedah pantas untuk meminimumkan ralat ramalan dengan mengelak berlakunya *overfitting*. Selain ID3 dan C4.5, terdapat juga algorithm yang lain seperti CART (Breiman et al. 1984), SPRINT (Shafer et al. 1996) dan SLIQ (Mehta et al.1996).

Pokok keputusan adalah salah satu model yang popular dalam algoritma pembelajaran mesin. Model ini adalah lebih serasi dengan tahap kemampuan logik manusia untuk memahami

dan menginterpretasi data dengan mudah. Sesuai dengan namanya, model ini adalah seperti struktur pokok yang bermula dengan nod akar, kemudian bercabang kepada nod dalaman dan akhirnya membentuk nod daun. Setiap cabang adalah hasil yang datang daripada ujian di nod dalaman manakala nod daun mewakili label kelasnya (Han et al. 2012).

Martínez Abad dan Chaparro Caso López (2017) mengaplikasikan pokok keputusan untuk mengecah faktor yang mempengaruhi pencapaian akademik murid tahun pertama, kedua dan ketiga daripada 99 buah sekolah menengah di Mexico. Keputusan kajiannya mendapati faktor peribadi menunjukkan pengaruh yang paling kuat terhadap pencapaian akademik, diikuti oleh faktor berkaitan persekitaran sekolah dan seterusnya faktor sosial. Menurut beliau, terdapat sejumlah besar teknik analisis digunakan dalam mengesan faktor-faktor berkaitan akademik, namun penggunaan perlombongan data adalah jarang diterokai. Beliau membuktikan bahawa pengaplikasian algoritma pengelasan pokok keputusan adalah amat berguna untuk mengenalpasti faktor yang boleh menjelaskan pencapaian akademik. Bunkar et al. (2012) pula telah menggunakan ketiga-tiga algoritma pokok keputusan iaitu ID3, C4.5 dan CART serta membandingkan ketepatannya dalam meramal pencapaian akademik pelajar. Dapatan beliau bahawa melalui ramalan, pelajar yang lemah boleh dibantu untuk memperbaiki strategi pembelajarannya supaya mencapai keputusan yang lebih baik. Mishra et al. (2014) pula dalam kajiannya meneroka hubungan antara kemahiran emosional bersama sosioekonomi dan pencapaian akademik lampau dalam meramal pencapaian akademik pelajar. Beliau mendapati teknik algoritma J48 memberi ketepatan ramalan sehingga 88.4%.



RAJAH 2. Output pokok keputusan dihasilkan menggunakan Weka

Pokok keputusan juga dikenali sebagai teknik induksi petua atau *rule induction techniques* (Luan 2002). Pokok keputusan adalah teknik yang membolehkan penggunaan kedua-dua kriteria pembolehubah kategorikal dan kuantitatif, dan juga berupaya membuat pengecaman terhadap kesan interaksi dalam kes-kes sub kumpulan, iaitu di dalam peringkat atau lapisan pokok serta di beberapa kawasan pokok yang lain (Castro dan Lizasoain 2012).

Selain itu, teknik pokok keputusan dipilih adalah kerana lebih mudah difahami dan lebih mudah ditafsirkan secara visual atau menggunakan set petua. Pokok keputusan lebih mudah untuk dijelaskan kepada sesiapa sahaja terutamanya kepada eksekutif atau pembuat keputusan.

Pokok keputusan juga menyediakan indikator yang jelas daripada susunan kepentingan faktor untuk membuat keputusan, ramalan dan pengecaman.

Perisian atau alat analisis yang akan digunakan dalam membuat pengelasan ini ialah Weka (Waikato Environment for Knowledge) iaitu merupakan perisian sumber terbuka perlombongan data yang ditulis menggunakan Java dan boleh digunakan dalam pelbagai platform.

Secara ekperimental, set data dibahagikan kepada tiga kumpulan bagi setiap kohort iaitu keseluruhan set data (set S), kumpulan set data mengandungi hanya murid berpotensi tinggi (set T) dan set data hanya mengandungi murid berpotensi rendah (set R). Set data K1 setiap satu dijadikan set latihan (set S_L , T_L dan R_L) bagi mendapatkan model pengelasan untuk diuji ke atas set data K2 iaitu sebagai set ujian (set S_U , T_U , R_U). Ketepatan ujian bagi ketiga-tiga nya dibandingkan dan faktor-faktor yang mempengaruhi keputusan SPM boleh diperhatikan melalui kedudukannya dalam pokok tersebut.

KEPUTUSAN

Kesemua 14 atribut (rujuk Jadual 2) diuji menggunakan J48 dengan Tahap Kejayaan SPM sebagai label kelas bagi ketiga-tiga set data K1. Jadual 4 berikut menunjukkan ketepatan ramalan secara silang bagi ketiga-tiga set data menggunakan set latihan dan set ujian pada set data K2 Ketepatan secara keseluruhan mengambil kira peratus jumlah data yang dikelaskan dengan betul berbanding jumlah keseluruhan data yang diuji, mengikut model pembelajaran yang digunakan.

JADUAL 4. Ketepatan ramalan menggunakan J48 bagi ketiga-tiga set data bagi K1 dan K2.

Set data (K1)	Ujian ketepatan pada set latihan	Ujian ketepatan pada set ujian (K2)		
		S_U	T_U	R_U
S_L	86.89%	82.93%	84.08%	82.24%
T_L	86.26%	82.30%	83.73%	81.43%
R_L	87.51%	83.06%	84.06%	82.46%

Nilai sensitiviti dan spesifisiti boleh dibandingkan dengan nilai ketepatan keseluruhan (Jadual 4) untuk memilih model pembelajaran yang terbaik. Weka memberi nilai sensitiviti dengan merujuk kepada perkadaran murid yang mempunyai tahap kurang berjaya diramalkan dengan betul kepada kelas 'kurang berjaya' manakala nilai spesifisiti merujuk kepada perkadaran murid yang mempunyai tahap lebih berjaya diramalkan dengan betul kepada kelas 'lebih berjaya'. Untuk kajian ini, kedua-dua nilai ini dijadikan rujukan perbandingan ketepatan pengelasan. Jadual 5 membandingkan kesemua nilai berkenaan bagi parameter *confidence factor*, $c=0.25$.

Berdasarkan Jadual 5, model pembelajaran set T_L - T_U dan set R_L - S_U dipilih bagi meramal kejayaan pencapaian SPM cara keseluruhan. Nilai sensitiviti dan spesifisiti bagi kedua-dua set tersebut mencatatkan nilai yang paling rendah perbezaannya berbanding set-set yang lain, di samping nilai ketepatan keseluruhannya juga antara yang tinggi. Saiz pokok yang semakin kecil dengan nilai peratus ketepatan yang tinggi serta konsisten adalah diutamakan dalam kajian ini. Oleh demikian, model pembelajaran set T_L dipilih sebagai model pembelajaran terbaik.

JADUAL 5. Nilai sensitiviti dan spesifisiti berbanding ketepatan keseluruhan

Model pembelajaran-Set data ujian	Sensitiviti	Spesifisiti	Ketepatan
S _L -S _U	0.879	0.781	0.829
S _L -T _U	0.757	0.862	0.841
S _L -R _U	0.902	0.663	0.822
T _L -S _U	0.759	0.898	0.823
T _L -T _U	0.850	0.779	0.837
T _L -R _U	0.636	0.918	0.814
R _L -S _U	0.867	0.793	0.831
R _L -T _U	0.728	0.874	0.841
R _L -R _U	0.895	0.674	0.825

Faktor-faktor yang mempengaruhi tahap kejayaan SPM murid dipilih menggunakan J48 berdasarkan kedudukan dari akar ke pucuk dalam pokok. Kedudukan atribut yang lebih dekat pada akar adalah peramal terbaik. Atribut pertama bagi set model pembelajaran S_L (keseluruhan data) ialah Matematik, diikuti Tahap PT3, Sains, Tahap Aliran Sebelum dan seterusnya. Atribut Tahap PT3 adalah atribut pertama bagi model pembelajaran T_L (data murid berpotensi tinggi), diikuti Matematik, Sains, Tahap Aliran Sebelum dan seterusnya. Atribut Sains pula dipilih sebagai peramal pertama bagi model pembelajaran R_L (data murid berpotensi rendah) diikuti Tahap PT3, Tahap Aliran Sebelum, Matematik dan seterusnya. Rumusannya, kesemua faktor tersebut adalah sama sebagai peramal terawal berbanding atribut-atribut yang lain dan perbezaannya adalah dari segi kedudukan.

PERBINCANGAN

INTERPRETASI KEJAYAAN SPM

Tahap kejayaan sebenar adalah terlalu subjektif memandangkan pengamalan sijil terbuka dalam peperiksaan SPM. Nilai mata pelajaran teras dan pilihan murid adalah sama penting dan tidak dibezakan. Mulai tahun 2000, pihak LP tidak lagi meletakkan sebarang nilai angka akhir kepada keputusan individu murid seperti pangkat, agregat, purata nilai mata, purata nilai gred kumulatif dan sebagainya bagi menginterpretasi kejayaan murid berkenaan. Sebaliknya, kejayaan murid dinilai berdasarkan gred yang diperoleh bagi setiap keputusan mata pelajaran yang diambil. Contohnya, calon dikategorikan Cemerlang apabila memperoleh gred A bagi semua mata pelajaran yang diduduki, bermula dari semua gred A+, kombinasi gred A+, A, A- hinggalah semua gred A-, manakala calon dikatakan lemah apabila kombinasi grednya ialah D, E dan G.

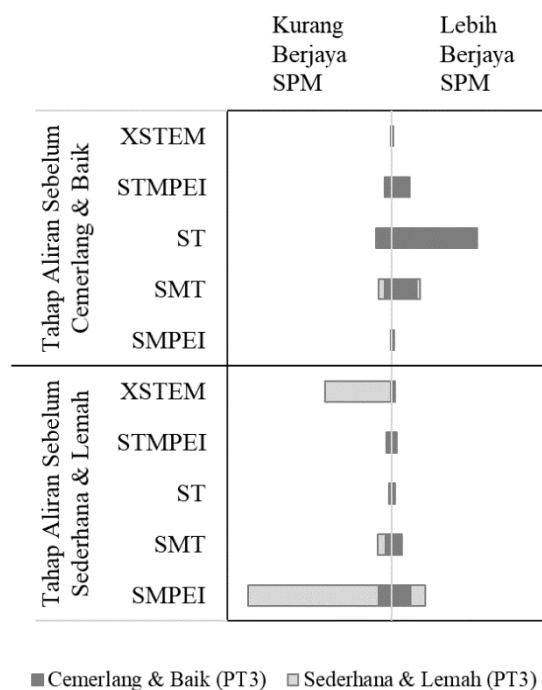
Oleh itu, calon yang lebih banyak mendapat gred yang Cemerlang dan Baik adalah lebih berjaya. Bermakna interpretasi gred setiap mata pelajaran dalam keputusan SPM yang diperoleh oleh murid adalah berbeza serta bergantung kepada fokus atau keperluan kesemua pemegang taruh yang berada di institusi pengajian tinggi serta pelbagai bidang pekerjaan berkaitan.

Dalam kajian ini, nilai kejayaan murid dalam SPM diukur dengan purata nilai gred yang diperoleh. Bagi tujuan pengelasan, nilai purata yang lebih kecil dari 4.50 menunjukkan murid tersebut cenderung mendapat kombinasi gred yang lebih baik dan dikelaskan sebagai 'lebih

berjaya' berbanding sebaliknya. Walau bagaimanapun, kemungkinan terdapat murid yang gagal mendapat sijil SPM secara realitinya, dikelaskan sebagai 'lebih berjaya' dalam kajian ini kerana syarat utama yang diletakkan oleh LP untuk lulus SPM ialah iaitu sekurang-kurangnya gred E (nilai gred 8.00) bagi kedua-dua mata pelajaran Bahasa Melayu dan Sejarah. Perkara ini bagaimanapun dilihat sebagai tidak relevan bagi objektif kajian.

FAKTOR TAHAP ALIRAN SEBELUM DAN TAHAP PT3

Tahap Aliran Sebelum adalah antara faktor yang baik untuk meramalkan kejayaan murid. Kedudukannya adalah selepas Matematik dalam pokok keputusan bagi model pembelajaran set TL. Faktor ini menggambarkan prestasi aliran yang ditawarkan di sesebuah sekolah iaitu setahun sebelum murid menduduki SPM. Faktor ini menjadi antara peramal yang baik ialah kerana sebilangan besar murid di tahap Cemerlang dan Baik dalam peperiksaan sebelumnya telah pun mendapat tempat di aliran SPM yang juga mempunyai latar belakang prestasi pada tahap Cemerlang dan Baik. Begitu juga berlaku pada murid yang yang mendapat keputusan pada tahap Sederhana dan Lemah, mereka diumpukkan kebanyakannya di aliran SPM yang juga berada di tahap Sederhana dan Lemah. Maka sudah tentu murid yang berada di aliran SPM yang tahapnya lebih baik adalah terdiri daripada murid berpencapaian PT3 yang lebih baik, dan juga sebaliknya.



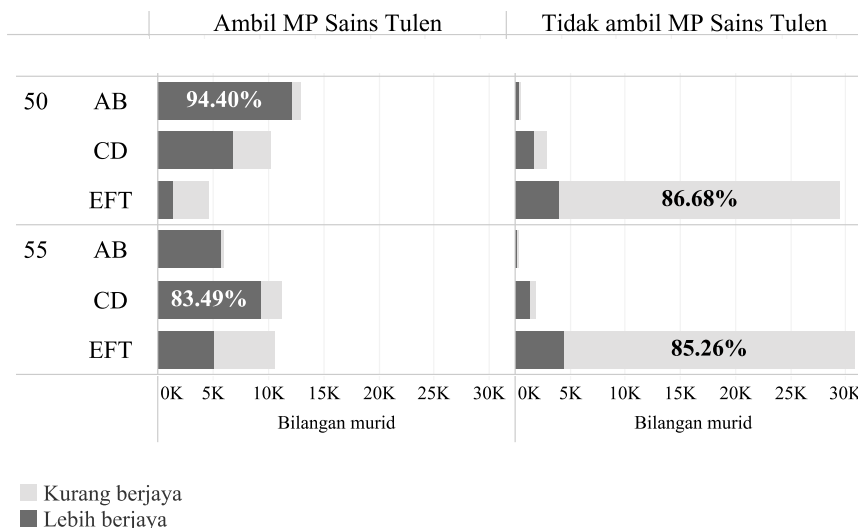
RAJAH 3. Hubungan Tahap Aliran Sebelum, Tahap PT3 dan Tahap Kejayaan SPM bagi aliran melebihi 4,000 murid bagi K1.

Rajah 3 mengilustrasikan hubungan antara faktor Tahap Aliran Sebelum bagi lima aliran terpilih (rujuk Jadual 2) dan dibezakan pada dua lajur, faktor Tahap PT3 dibezakan dengan kod warna pada bar (rujuk petunjuk Jadual 3) dan faktor Tahap Kejayaan SPM yang dibezakan pada dua kolumn. Panjang bar menggambarkan bilangan murid iaitu lebih panjang bar, bilangan murid lebih ramai. Data adalah murid daripada K1. Berdasarkan Rajah 3, jumlah terbesar murid yang lebih berjaya dalam SPM adalah berada dalam aliran ST iaitu kelompok murid yang telah menduduki di aliran yang mempunyai tahap Cemerlang dan Baik dalam SPM pada tahun sebelumnya dan sebilangan besarnya pula telah memperoleh keputusan PT3 juga pada tahap Cemerlang dan Baik (warna bar kelabu melebihi warna bar putih). Jumlah terbesar

murid yang kurang berjaya dalam SPM pula berada dalam aliran SMPEI iaitu kelompok murid yang telah menduduki di aliran yang mempunyai tahap Sederhana dan Lemah dalam SPM pada tahun sebelumnya dan sebilangan besar mereka telah memperoleh keputusan PT3 pada tahap Sederhana dan Lemah (warna bar putih melebihi warna bar kelabu).

FAKTOR MATEMATIK, SAINS DAN ALIRAN

Faktor Matematik atau pencapaian murid dalam mata pelajaran matematik di peringkat PT3; merupakan peramal terawal bagi model pembelajaran set data S_L dan R_L dalam menentukan tahap kejayaan SPM murid, manakala faktor Sains (tahap pencapaian mata pelajaran sains semasa PT3) dipilih selepas faktor Tahap PT3, diikuti pula dengan faktor Aliran dalam pokok keputusan. Kebanyakan murid dengan keputusan gred Matematik dan Sains yang lebih baik akan cenderung ke aliran yang mengandungi mata pelajaran Sains Tulen di peringkat menengah atas iaitu Matematik Tambahan, Biologi, Fizik dan Kimia. Aliran tersebut ialah ST, STA, STMPEI dan SMT. Rajah 4 menunjukkan gabungan gred AB, CD dan EFT bagi kedua-dua mata pelajaran berkenaan beserta peratus dan bilangan tahap kejayaan murid K1 dalam SPM berdasarkan mata pelajaran yang diambil atau kumpulan aliran. Dalam kajian ini, gabungan gred AB dijadikan sebagai gred tertinggi, CD sebagai gred sederhana, manakala EFT sebagai gred terendah. Gred T pada EFT ialah gred untuk murid yang tidak hadir mata pelajaran PT3 berkenaan dan jumlahnya kurang daripada 0.5% bagi setiap mata pelajaran yang dikaji. Secara rumusan, pola data menunjukkan murid yang mengambil mata pelajaran Sains Tulen mendominasi tahap pencapaian SPM yang lebih berjaya berbanding yang tidak mengambil mata pelajaran tersebut. Bagi Matematik (kod 50), 94.40% murid yang mendapat gred AB dan mengambil mata pelajaran Sains Tulen adalah lebih berjaya. Dari sudut saiz, bilangan murid yang mendapat gred AB bagi Matematik dan tidak mengambil mata pelajaran Sains Tulen adalah jauh lebih kecil, berbanding murid yang mendapat gred EFT dengan bilangan yang jauh lebih besar seperti pada Rajah 4. Situasi ini turut sama berlaku pada mata pelajaran Sains (kod 55).



RAJAH 4. Peratus murid K1 berdasarkan tahap kejayaan SPM, gred bagi Matematik dan Sains dan kumpulan aliran.

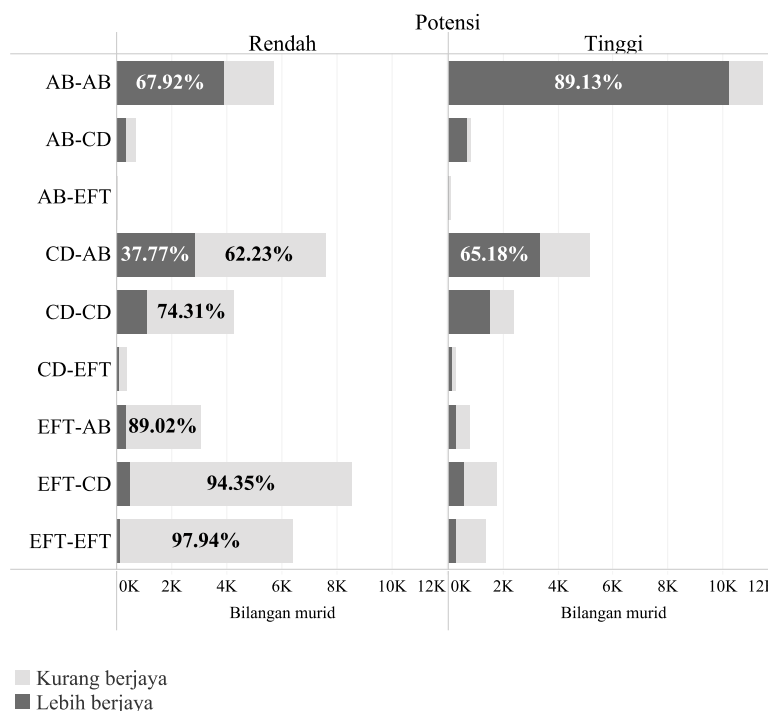
FAKTOR BAHASA MELAYU PEMAHAMAN & PENULISAN

Mata pelajaran Bahasa Melayu (BM) diuji dalam PT3 melalui komponen Pemahaman dan komponen Penulisan. Kedua-dua komponen tersebut dilaporkan secara gred yang berasingan

dalam keputusan PT3. Dalam hierarki pokok keputusan yang terhasil, kedudukan faktor BM Penulisan sebagai peramal adalah di bawah faktor Sains dan selari dengan faktor Aliran bagi model pembelajaran S_L dan R_L . Kedudukan bawahnya pula diikuti oleh faktor BM Pemahaman. Bagi set T_L , kedudukan BM Penulisan pula selari dengan faktor Aliran dan faktor Tahap Aliran Sebelum.

Berdasarkan Rajah 5, kedua-dua gred setiap komponen BM tersebut digabungkan (BM Pemahaman-BM Penulisan) bagi tujuan analisis secara sekaligus. Hasil gabungan gred tersebut menunjukkan bahawa murid K1 daripada kalangan berpotensi tinggi dan mendapat gabungan gred AB-AB bagi kedua-dua kertas BM mempunyai tahap kejayaan SPM lebih berjaya yang tertinggi iaitu sebanyak 89.13% berbanding kombinasi lain. Murid berpotensi rendah dan mendapat kombinasi gred EFT-EFT pula menunjukkan sebaliknya dengan nilai sebanyak 97.94%.

Kesimpulannya, murid yang berpotensi tinggi diramal atau mempunyai kebarangkalian yang lebih tinggi untuk berjaya dalam SPM sekiranya gred BM PT3 bagi kedua-dua komponennya adalah yang terbaik iaitu AB. Walau bagaimanapun, murid yang berpotensi rendah tetapi mendapat kombinasi gred AB-AB bagi kedua-dua komponen BM juga mempunyai kebarangkalian dan peluang sebanyak 67.92% untuk berjaya dalam SPM. Ini menunjukkan bahawa murid yang berpotensi rendah masih lagi berpeluang mendapat keputusan yang lebih baik dalam SPM sekiranya mendapat keputusan yang baik dalam BM PT3. Faktor-faktor lain yang berkait rapat dengan kejayaannya juga boleh dikaji.



RAJAH 5. Hubungan kombinasi Bahasa Melayu Pemahaman-Penulisan dengan potensi dan tahap kejayaan SPM bagi K1.

HUBUNGAN POTENSI, TAHAP PT3 DAN SPM

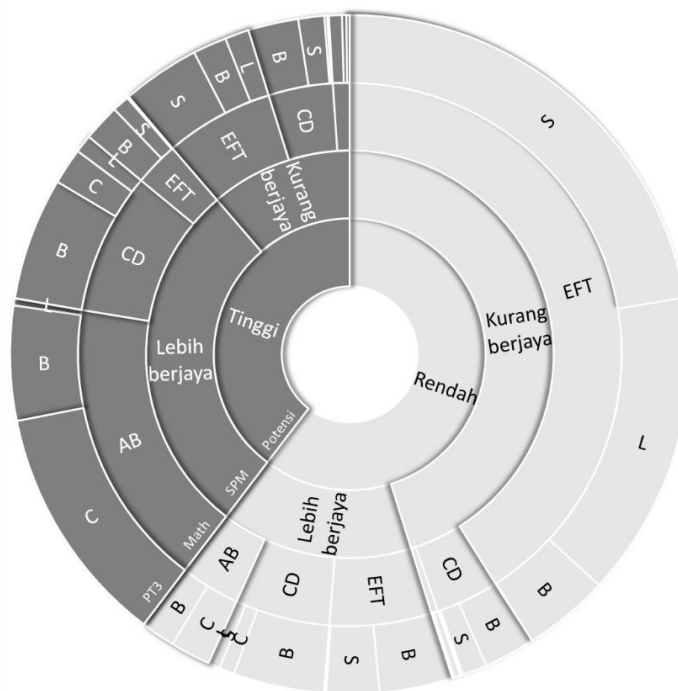
Murid yang berpotensi tinggi diramal akan berjaya dalam mana-mana peringkat pendidikan sekiranya murid tersebut mengekalkan momentum prestasinya yang baik dalam faktor-faktor yang dibincangkan sebelum ini. Seperti digambarkan pada Rajah 6, hubungan yang kuat antara potensi, Matematik, tahap PT3 dan tahap kejayaan SPM dapat dibuktikan.

Interpretasi Rajah 6 adalah melalui kaedah pemvisualan dengan memerhatikan pada saiz bahagian yang melingkungi subset bagi beberapa atribut. Saiz bahagian yang lebih besar

menunjukkan bilangan atau peratusan murid yang lebih besar. Bahagian pertama iaitu bermula dari tengah ialah atribut Potensi, diikuti atribut Tahap Kejayaan SPM, atribut Matematik dan terakhir ialah Tahap PT3. Subset bagi atribut Tahap PT3 diringkaskan kepada C iaitu Cemerlang, B iaitu Baik, S iaitu Sederhana dan L iaitu Lemah. Atribut Potensi menggambarkan pembahagian dua kumpulan murid iaitu murid berpotensi tinggi dan murid berpotensi rendah.

Kedua-dua kumpulan murid diasingkan diperingkat awal sebagai menunjukkan potensi mereka yang telah dikesan lebih awal menggunakan UASR semasa di Tahun Enam. Setiap kumpulan Potensi tersebut masing-masing melingkungi Tahap Kejayaan SPM semasa di Tingkatan Lima iaitu ukuran kejayaan terakhir di peringkat sekolah menengah. Murid berpotensi tinggi melingkungi saiz murid yang lebih berjaya dalam SPM yang lebih besar berbanding kurang berjaya, manakala murid berpotensi rendah melingkungi saiz murid yang kurang berjaya dalam SPM lebih besar berbanding lebih berjaya. Atribut seterusnya iaitu Matematik dan Tahap PT3 pula menggambarkan faktor kepada kejayaan murid dalam SPM berdasarkan pembahagian saiz bagi setiap subset kedua-dua atribut berkenaan.

Murid daripada kumpulan potensi tinggi dilihat mempunyai kebarangkalian yang lebih tinggi (saiz bahagian yang lebih besar) atau diramalkan akan lebih berjaya dalam SPM. Mereka diramal lebih berjaya berdasarkan saiz lebih besar pada faktor Tahap PT3 iaitu tahap Cemerlang atau Baik, serta faktor Matematik dengan gred AB dan CD. Kebarangkalian untuk kurang berjaya dalam SPM adalah tinggi bagi murid yang mendapat gred EFT bagi Matematik walaupun berpotensi tinggi sebelum itu.



RAJAH 6. Hubungan potensi, Matematik PT3, Tahap PT3 dan SPM bagi K1.
Nota: C: Cemerlang, B: Baik, S: Sederhana & L: Lemah

KESIMPULAN DAN CADANGAN UNTUK KAJIAN MASA HADAPAN

Limitasi dalam kajian ini ialah faktor yang digunakan amat tertakluk kepada data pencapaian dan prestasi akademik murid. Kajian selanjutnya dicadangkan lebih melihat kepada demografi lain yang boleh memberikan maklumat komprehensif. Contohnya seperti status sosiekonomi murid, tahap pendidikan ibu bapa, pekerjaan ibu bapa, lokasi tempat tinggal, bilangan adik beradik, hubungan dengan rakan sebaya dan sebagainya. Selain itu, latar belakang sekolah merupakan peluang yang juga boleh memberi kesan atau perubahan kepada personaliti murid,

minat dan motivasi dalam pendidikannya. Contohnya ialah pengalaman mengajar guru, bilangan murid dalam kelas, penglibatan aktiviti dalam persatuan, sukan dan ko-kurikulum, pergaulan rakan sebaya dan banyak lagi. Penglibatan murid dalam interaksi jaringan sosial juga boleh dikaji. Walau bagaimanapun, pengumpulan data melalui faktor-faktor lain seperti dicadangkan ini kemungkinan perlu melalui proses yang agak sukar dan memerlukan kerjasama pelbagai pihak. Antara solusi yang boleh digunakan ialah menggunakan faktor sedia ada pada data PBS seperti Pentaksiran Psikometrik dan Pentaksiran Aktiviti Sukan & Ko-kurikulum.

Selain itu, hanya dua atribut kemahiran yang diambil daripada UASR lebih kepada mengukur keupayaan kognitif dan logik murid. Maklumat lain yang juga diukur seperti minat dan kecenderungannya terhadap pelbagai bidang turut boleh diberi perhatian dalam kajian akan datang. Sebagai alternatif, kajian menggunakan dapatan daripada Ujian Aptitud Am Tahun 3 boleh diaplikasikan menggantikan UASR. Ujian tersebut mengandungi dua domain untuk menguji aptitud murid. Pertama, domain Bahasa Melayu dan Bahasa Inggeris menguji kemahiran menaakul seperti kemahiran perkataan, konsep verbal dan aplikasi kritikal. Kedua, domain Matematik pula menguji sistem numerik, penyelesaian masalah, penaakulan matematik untuk tugas logik nombor dan membuat persepsi.

Modifikasi pengkaji kepada ketiga-tiga tahap pencapaian iaitu UPSR, PT3 dan SPM juga boleh mempengaruhi keputusan kajian. Contohnya ialah bagi tahap kejayaan SPM, kaedah lain menggantikan pengiraan purata nilai mata boleh digunakan seperti pengiraan kombinasi gred bagi murid cemerlang, sederhana dan lemah. Faktor tambahan yang sedia ada seperti daerah, lokasi dan gred purata sekolah boleh digunakan sebagai atribut tambahan.

Dapatan data menunjukkan bahawa mata pelajaran Matematik & Sains PT3 dan Tahap PT3 merupakan peramal terpenting dalam menentukan tahap kejayaan murid dalam SPM. Maksudnya, jika seseorang murid berjaya dalam ketiga-tiga atribut berkenaan, peluangnya untuk berjaya dalam SPM pula adalah lebih cerah, manakala murid yang kurang berjaya adalah sebaliknya. Korelasi yang sangat rapat ini membolehkan kajian akan datang dicadangkan untuk mengenalpasti apakah faktor-faktor yang mempengaruhi kejayaan murid bagi ketiga-tiga atribut berkenaan.

Kajian ini telah membuktikan dengan jelas bahawa pengaplikasian perlombongan data adalah merangkumi penyelidikan pelbagai bidang yang penting termasuklah bidang pendidikan di Malaysia khususnya. Ini menyimpulkan bahawa teknik perlombongan data yang digunakan dalam menjelaskan dan menginterpretasi faktor berhubung pencapaian akademik melalui pokok yang terhasil daripada algoritma pembelajaran merupakan keupayaan terkini melibatkan penggunaan data yang berskala besar secara optimum. Rantainya kajian dengan penambahbaikan dengan mempelbagaikan teknik pembelajaran mesin boleh diaplikasikan selepas ini untuk mencungkil kaedah baharu dari aspek teknik digunakan dalam membuat pengecaman faktor disamping meluaskan aplikasi dan replikasi kajian dalam penyelidikan yang berlainan. Melalui perlombongan data, pelbagai isu pendidikan dari segenap lapisan bermula daripada dasar sehinggalah produk pendidikan boleh dikaji dengan mengoptimalkan penggunaan rekod-rekod data sedia ada di KPM supaya pengetahuan tersembunyi dapat diteroka dan dimanfaatkan dalam menangani isu berkaitan. Diharapkan penyelidikan menerusi perlombongan data sebegini dijadikan sebagai batu loncatan untuk memperkembangkan serta menyemarakkan bidang sains data di KPM dalam memacu transformasi pendidikan. Salah satu contoh ialah melalui pembudayaan kemahiran komunikasi data secara visual yang merupakan sebahagian elemen sains data disamping mencetuskan lagi idea-idea baharu dalam teknik pemvisualan.

PENGHARGAAN

Penerbitan kertas kerja ini dibiayai oleh geran penyelidikan RCRP-2016-002 dan Universiti Kebangsaan Malaysia.

RUJUKAN

- Ahmad, F., Ismail, N.H. and Aziz, A.A. 2015. The Prediction of Student's Academic Performance Using Classification Data Mining Techniques, *Applied Mathematical Sciences* 9(129): 6415-6426.
- Amin, H dan Mustafa, N.H. 2014. Gaya Pembelajaran dan Pencapaian Pelajar dalam Kursus Reka Bentuk Aturcara. *Asis-Pacific Journal of Information Technology and Multimedia*. 3(2):1-12.
- Asif, R., Merceron, A., Ali, S.A. and Haider, N.G.2017. Analyzing undergraduate student;s performance using educational data mining. *Computers & Education* 111:117-194.
- Bakhshinategh, B., Zaiane, O. R., ElAtia, S. & Ipperciel, D. 2018. Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies*. 23(1):537-553. doi:10.1007/s10639-017-9616-z.
- Bhardwaj, B. & Pal, S. 2011. Data Mining: A prediction for performance improvement using classification. *International Journal of Computer Science and Information Security*. 9(4):136-140.
- Bunkar, K. ., Singh, U. K. ., Pandya, B. . & Bunkar, R. . 2012. Data mining: Prediction for performance improvement of graduate students using classification. *International Conference on Wireless and Optical Communications Networks, WOCN*. Indore, 2012, 1-5. doi:10.1109/WOCN.2012.6335530.
- Castro, M. & Lizasoain, L. 2012. Las técnicas de modelización estadística en la investigación educativa: minería de datos, modelos de ecuaciones estructurales y modelos jerárquicos lineales. (Spanish). *Statistical Modeling Techniques in Educational Research: Data Mining, Structural Equation Models and Hierarchical Linear Models*. (English). *Revista Espanola de Pedagogia* 70(251):131-148.
- Crosnoe, R., Johnson, M. K. & Elder, G. H. 2004. School size and the interpersonal side of education: An examination of race/ethnicity and organizational context. *Social Science Quarterly*. 85(5):1259-1274. doi:10.1111/j.0038-4941.2004.00275.x.
- Dutt, A., Ismail, M.A., and Herawan, T. 2017. A systematic Review on Educational Data Mining. *IEEEAccess* 5:15991-16005.
- Haist, S. a., Wilson, J. F., Elam, C. L., Blue, A. V. & Fosson, S. E. 2000. The Effect of Gender and Age on Medical School Performance: An Important Interaction. *Advances in health sciences education : theory and practice* 5(3): 197–205. doi:10.1023/A:1009829611335.
- Han, J., Kamber, M. & Pei, J. 2012. Data Mining: Concepts and Techniques. The Morgagan Kaufmann Series in Data management Systems) Edisi ke3. . doi:10.1016/C2009-0-61819-5.
- Ibrahim Z, R. D. 2007. Predicting Students' Academic Performance: Comparing Artificial Neural Network, Decision Tree and Linear Regression, 2007. *21st Annual SAS Malaysia Forum, 5th September 2007, Shangri-La Hotel, Kuala Lumpur*.
- Jambari, D.I., Mokhtar, U.A., Ishak, H.Y. dan Yaakub, M.R. 2015. Information Technology Graduate Employability Enhancement Model: Case Study in the FTSM, UKM. *Asia-Pacific Journal of Information Technology and Multimedia*, 4(2):57-68.
- Kotsiantis, S., Pierrakeas, C. & Pintelas, P. 2004. Predicting students' performance in distance learning using machine learning techniques. *Applied Artificial Intelligence* 18(5): 411–426. doi:10.1080/08839510490442058.
- Luan, J. 2002. Data Mining and Its Applications in Higher Education. *New Directions for Institutional Research*. . p: 17-36doi:10.1002/ir.35.
- Martínez Abad, F. & Chaparro Caso López, A. A. 2017. Data-mining techniques in detecting factors linked to academic achievement. *School Effectiveness and School Improvement*. doi:10.1080/09243453.2016.1235591.
- Meit, S. S., Borges, N. J., Cubic, B. A. & Seibel, H. R. 2007. Personality Differences in Incoming Male and Female Medical Students. Result oo a Multi-site 9-year Study. *Medical Education Online* 12(7):1-6.
- Mishra, T., Kumar, D. & Gupta, S. 2014. Mining students' data for prediction performance. *International Conference on Advanced Computing and Communication Technologies, ACCT*, hlm. 255–262. doi:10.1109/ACCT.2014.105.
- Rostaminezhad, M.A. 2018. To interact or not to interact: Why students print interactive instructional multimedia? Problem of reading or reviewing? *Asia-Pacific Journal of Information Technology*

and Multimedia. 7(1) :19-28.

- Sembiring, S., Zarlis, M., Hartama, D., Ramliana, S. & Wani, E. 2011. Prediction of Student Academic Performance By an Application of Data Mining Techniques. International Conference on Management and Artificial Intelligence (IPEDR), hlm. Vol. 6, 110–114. Retrieved from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Prediction+of+student+academic+performance+by+an+application+of+data+mining+techniques+1+1#0>.
- Shahiri, A. M., Husain, W. & Rashid, N. A. 2015. A review on predicting student's performance using data mining techniques. 3rd Information Systems International Conference, 2015. Vol. 72, 414–422. doi:10.1016/j.procs.2015.12.157.
- Simsek, A. & Balaban, J. 2010. Learning Strategies of Successful and Unsuccessful University Students. *CONTEMPORARY EDUCATIONAL TECHNOLOGY* 1(1): 36–45.
- Zainudin, S. and Aziz, M.A.A. 2015. Perlombongan data bagi Meramal Prestasi Akademik Pelajar. *Prosiding Kongress Pengajaran dan Pembelajaran UKM 2012*. 1(11): 96-103.

Azhar Mohd Khairy

Afzan Adam

Mohd Ridzwan Yaakub

Fakulti Teknologi & Sains Maklumat,

Universiti Kebangsaan Malaysia.

azharkhairy81@gmail.com, afzan@ukm.edu.my, ridzwanyaakub@ukm.edu.my

Received: 2 Mei 2018

Accepted: 1 July 2018

Published: 6 December 2018