PERMODELAN RAMALAN PENCAPAIAN PELAJAR: APLIKASI KAEDAH PERLOMBONGAN DATA DI UNIVERSITI PENDIDIKAN SULTAN IDRIS

NORSYADZILA BINTI MUSTAFA

UNIVERSITI PENDIDIKAN SULTAN IDRIS

2011

PERMODELAN RAMALAN PENCAPAIAN PELAJAR: APLIKASI KAEDAH PERLOMBONGAN DATA DI UNIVERSITI PENDIDIKAN SULTAN IDRIS

NORSYADZILA BINTI MUSTAFA

DISERTASI DIKEMUKAKAN BAGI MEMENUHI SYARAT UNTUK MEMPEROLEHI IJAZAH SARJANA PENDIDIKAN MATEMATIK

FAKULTI SAINS DAN MATEMATIK
UNIVERSITI PENDIDIKAN SULTAN IDRIS

2011

PENGAKUAN

Saya mengaku disertasi ini adalah hasil kerja saya sendiri kecuali nukilan dan ringkasan yang setiap satunya saya jelaskan sumbernya.

07 MAC 2011

…….……..

NORSYADZILA BINTI MUSTAFA
M20091000564

PENGHARGAAN

Alhamdulillah, syukur ke hadrat Illahi kerana dengan limpah rahmat dan kurniaNya, kajian ini dapat disempurnakan mengikut jadual.

Terlebih dahulu, saya ingin mengucapkan ribuan terima kasih kepada Kementerian Pengajian Tinggi kerana biasiswa yang diberikan membolehkan saya menyambung pelajaran di peringkat sarjana (sepenuh masa) dalam bidang Pendidikan Matematik di Universiti Pendidikan Sultan Idris.

Kepada semua staf dan pensyarah di Fakulti Sains dan Matematik, saya sangat menghargai segala bantuan dan nasihat yang diberikan. Terima kasih diucapkan kepada Ketua Jabatan Matematik Fakulti ini, Dr. Zulkifley bin Mohamed kerana sokongan yang diberikan dalam sepanjang penyelidikan ini.

Sekalung penghargaan dan ucapan terima kasih yang tidak terhingga kepada Prof. Madya Dr. Abd Wahab bin Jusoh selaku penyelia pertama saya dan Pn Norsida binti Hasan kerana sedia membantu dan membimbing baik di dalam mahu pun di luar waktu kuliah. Tanpa kerjasama dan bantuan yang sepenuhnya diberikan, adalah mustahil kajian ini dapat disempurnakan.

Ucapan terima kasih juga dirakamkan kepada Dr. Mohd Uzi bin Dollah selaku penyelia kedua saya dan semua pensyarah yang telah mencurah ilmu di sepanjang tempoh pengajian saya disini, serta rakan-rakan yang telah memberi kerjasama yang penuh pengertian. Setinggi ucapan terima kasih juga kepada Bahagian Akademik Universiti Pendidikan Sultan Idris kerana memberi kebenaran kepada pengkaji untuk menjalankan kajian ini.

Setinggi penghargaan dan ucapan terima kasih dirakamkan kepada ibu bapa, suami dan keluarga tercinta atas segala pengorbanan, pengertian, sokongan dan dorongan yang diberikan terutama di sepanjang laluan program ini.

Akhir sekali, tidak ketinggalan juga jutaan terima kasih diucapkan kepada semua pihak yang telibat sama ada secara langsung mahupun tidak langsung dalam proses menyiapkan kajian ilmiah ini. Semoga kerjasama dan bakti yang dicurah akan mendapat ganjaran dari Yang Maha Kuasa dan semoga Allah memberkati kehidupan kita semua di dunia dan akhirat. Amin.

Sekian, salam hormat,

NORSYADZILA BINTI MUSTAFA

ABSTRAK

Pelajar merupakan hasil utama Pusat Pengajian Tinggi(IPT). Bagi memastikan kejayaan universiti, pihak pengurusan universiti harus memastikan semua pelajar mendapat bantuan pendidikan yang sewajarnya dan berjaya bergraduat dengan cemerlang. Salah satu cabaran besar perancangan IPT adalah untuk meramal pencapaian pelajar semasa mereka mendaftar di IPT lagi, misalnya pelajar mana yang berpotensi untuk cemerlang serta pelajar manakah yang perlu dibantu supaya boleh berijazah dengan cemerlang. Berdasarkan situasi yang dinyatakan, kajian ini mengemukakan satu model pohon regresi iaitu salah satu alat dalam kaedah perlombongan data untuk meramal pencapaian pelajar. Kajian ini menggunakan data profil pelajar yang diperolehi dari pangkalan data Bahagian Akademik UPSI (1997-2008), data tersebut dilombong dengan menggunakan bantuan pakej perisian *R Language*. Pencapaian pelajar diukur berdasarkan PNGK semasa pelajar bergraduat. Kajian menggunakan pendekatan kuantitatif, keseluruhan populasi kajian digunakan untuk kaedah analisis deskriptif dan permodelan ramalan yang melibatkan pembinaan model pohon regresi. Hasil kajian menunjukkan pohon regresi berkesan dalam memodelkan ramalan pencapaian pelajar semasa bergraduat.

ABSTRACT

Student is the main output for the higher education institution or *Pusat Pengajian Tinggi* (IPT). To maintain the successfull of the institution, university's administration must make sure that all students received guidance and can successfully graduated from this university. One major challenge for IPT's administration is to predict the student achievement as early as the admission stage, for example, how to determine which student would be an excellent student and which students need assistance in order to graduate with good grade. Based on this situation, this research present a regression model which is one of the tools in data mining to predict student performance. The research instrument consists of the student's profile data (1997-2008) obtained from Academic Department of UPSI database, the data is mined using *R Language* software packages. Student achievement was measured by final CGPA when graduated. The study uses a quantitative approach, the entire population will be used for descriptive analysis and modelling analysis involving the construction of Regression Tree models. The result obtained has shown that regression tree is able to model data set used.

KANDUNGAN

SENARAI JADUAL

SENARAI RAJAH

SENARAI SINGKATAN PERKATAAN

rpart	Pemisahan Rekursi dan Pohon Regresi (Recursive Partitioning and Regression Trees).
minsplit	Bilangan minimum pemerhatian yang perlu wujud didalam suatu nod bagi proses melakukan pecahan.
Cp	Parameter kompleks.
maxdepth	Menetapkan kedalaman maksimum nod bagi pohon terakhir, dengan nod akar dikira sebagai 0 kedalaman (lepas 30 *rpart* akan memberi keputusan yang tidak masuk akal pada mesin 32-bit). Default kepada bilangan kelas.
CART	Algoritmanya adalah berdasarkan Klasifikasi dan Pohon Regresi (Classification And Regression Tree)
C4.5	Algoritma yang digunakan untuk menjana pohon keputusan, dicipta oleh Ross Quinlan.
ID3	(Iterative Dichotomiser 3) adalah algoritma yang digunakan untuk menjana pohon keputusan yang dicipta oleh Ross Quinlan.

BAB 1

PENDAHULUAN

1.1 Pengenalan

Bab ini akan membincangkan secara keseluruhan kajian dan permasalahan kajian. Seterusnya akan melihat tujuan kajian dan persoalan kajian. Akhir sekali bab ini akan menjelaskan kepentingan dan batasan kajian yang dijalankan. Perbincangan ini adalah bertujuan untuk memberi gambaran secara umum tentang kajian yang dijalankan.

1.2 Latar Belakang Kajian

Kini, organisasi pengajian tinggi berada di dalam persekitaran yang berpersaingan tinggi dan mencapai kelebihan persaingan terhadap pesaing-pesaing perniagaan yang lain. Organisasi ini perlu memperbaiki kualiti perkhidmatan dan memenuhi kehendak pelanggan (industri dan kerajaan). Untuk kekal bersaing, organisasi ini memerlukan pengetahuan yang mendalam dan memadai untuk melakukan penaksiran, penilaian, perancangan dan pembuatan keputusan yang terbaik. Majoriti pengetahuan yang diperlukan tersimpan di dalam pangkalan data organisasi pendidikan dan ia boleh didapati daripada data rekod lama atau bersejarah dan data operasi.

Satu pendekatan untuk menghadapi cabaran pelajar dan pentadbiran dengan berkesan ialah melalui analisis dan persembahan data, atau perlombongan data (data mining). Perlombongan data membantu organisasi menggunakan keupayaan laporan semasa, untuk mendapatkan dan mengenalpasti pola tersembunyi di dalam pangkalan data. Pola yang telah dikenalpasti kemudiannya digunakan bagi membina model perlombongan data. Model yang terhasil boleh digunakan dalam meramal pencapaian dan kelakuan dengan lebih efektif.

1.2.1 Pengenalan Perlombongan Data

Tugas sainstis ialah menjadikan suatu maklumat atau data bermakna, mengenalpasti pola yang mengatur dunia fizikal bekerja dan seterusnya merumus pola tersebut ke dalam teori-teori yang boleh digunakan untuk melakukan andaian perkara yang akan

berlaku di dalam situasi baru. Tugas ahli industri ialah mengenalpasti peluang melalui pola pencapaian yang boleh ditukarkan kepada perniagaan yang menguntungkan.

Teknologi masa kini telah membolehkan kita mengumpul dan menyimpan sejumlah data yang banyak. Bagaimanapun, tugas mencari pola dan perbezaan dalam suatu set data serta merumuskannya kepada sebuah model kuantitatif ringkas merupakan suatu cabaran besar kerana proses menukar data kepada maklumat dan menukarkan maklumat kepada pengetahuan bukanlah suatu perkara yang mudah (Witten & Frank, 2005).

Jumlah data di dunia sentiasa meningkat dan tiada penghujungnya. Dalam bidang perlombongan data; data disimpan secara elektronik dan pencarian dilakukan secara automatik atau sekurang-kurangnya menggunakan komputer. Ahli ekonomi, saintis, peramal, dan jurutera komunikasi dan mereka yang telah lama bekerja dengan menggunakan data percaya bahawa maklumat di dalam data boleh di cari secara automatik, dikenalpasti, disahkan, serta digunakan untuk membuat ramalan.

Perlombongan data ialah mengenai penyelesaian masalah melalui analisis data sedia ada di dalam pangkalan data. Perlombongan data juga merupakan proses pencarian pola dalam suatu data. Proses tersebut mestilah automatik atau separa-automatik. Pola yang dikenalpasti hendaklah bermakna dan ini bermaksud pola tersebut akan memberikan suatu kebaikan. Data yang terlibat dalam penganalisisan

perlombongan data hendaklah sentiasa dalam kuantiti yang besar (Turban, *et al*., 2007; Williams, 2005; Hunt & Madhyastha, 2002).

Perlombongan data melibatkan tugas seperti mengeluarkan pengetahuan, arkeologi data, penjelajahan data, pemprosesan pola data, mengorek data dan penuaian maklumat. Semua aktiviti ini dilakukan secara automatik dan membenarkan penemuan dibuat walaupun oleh bukan seorang pengaturcara atau programmer. Menurut Turban, *et al.*, (2007) ciri-ciri utama dan objektif perlombongan data ialah pertama, data selalunya tertanam di dalam pangkalan data yang besar dan kadangkala mengandungi data yang bertahun lamanya. Dalam kebanyakan kes, data tersebut telah dibersihkan dan diperkukuhkan di dalam gudang data. Manakala, persekitaran perlombongan data pula kebiasaanya ialah pelanggan atau struktur perkhidmatan atau struktur berasaskan *Web*.

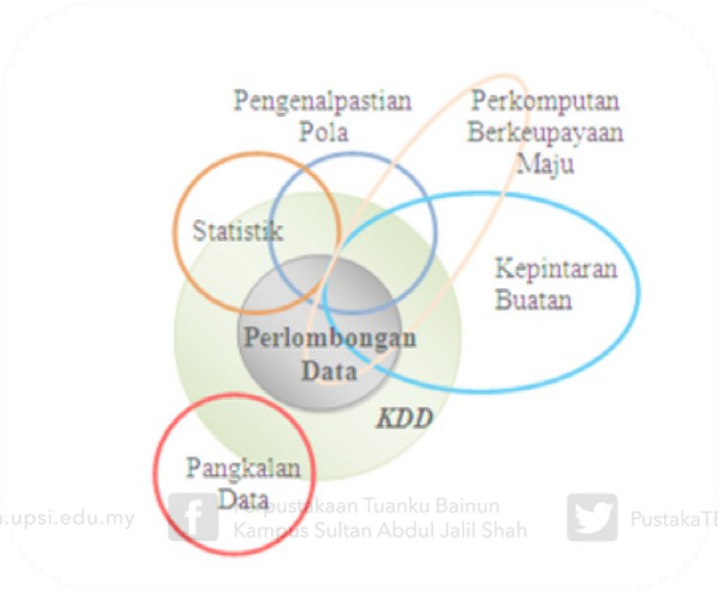
Peralatan-peralatan baru yang canggih, termasuklah peralatan persembahan visual terkini akan digunakan untuk membantu membuang maklumat yang tertanam di dalam fail koorporat atau rekod-rekod capaian umum. Pencarian maklumat pula melibatkan membaiki dan menyelaraskan data tersebut untuk mendapatkan keputusan yang tepat. Selain itu, proses melombong data juga melibatkan proses menyelidik data yang ringan (contohnya teks tak-berstruktur yang tersimpan di suatu tempat seperti pangkalan data *Lotus Notes*, fail teks dalam internet, atau perusahaan-luas rangkaian dalaman (enterprise-wide intranet).

Ciri-ciri seterusnya ialah, pelombong atau penyelidik biasanya adalah pengguna terakhir, dengan menggunakan data latihan dan alatan pertanyaan yang berteknologi tinggi untuk menyoal dan mendapatkan jawapan dengan cepat atau serta merta, dengan menggunakan sedikit atau tanpa kemahiran pengaturcaraan. Perlombongan data selalunya melibatkan penemuan keputusan yang tidak dijangka dan memerlukan kreativiti pemikiran pengguna terakhir.

Peralatan perlombongan data telah dilengkapi dengan hamparan elektronik (spreadsheet) dan peralatan pembangunan perisian yang lain. Perisian yang boleh digunakan ialah perisian *R Language, SAS Enterprise Miner, DBMiner, SPSS Version 17.0, Weka, Clenmentine, See5, E4ML* atau menggunakan hamparan elektronik *Microsoft Excel* (solver) (Tudor & Carbureanu, 2007; Witten & Frank, 2005; Walsh, 2002). Oleh itu, data yang dilombong boleh dianalisis dan diproses dengan lebih cepat dan mudah. Walaubagaimanapun, disebabkan jumlah data yang besar dan usaha pencarian yang rumit, ianya kadangkala perlu menggunakan proses perlombogan data yang serentak.

Penemuan pengetahuan dalam pangkalan data (Knowledge Discovery in Database) atau KDD dan perlombongan data digunakan secara sinonim oleh kebanyakan penyelidik dalam bidang ini. Bagaimanapun, terdapat sedikit perbezaan di antara kedua-duanya dari segi teori. KDD ialah satu proses keseluruhan dalam mendapatkan maklumat daripada data yang melibatkan beberapa langkah penting dalam memastikan pengetahuan yang diperoleh munasabah. Proses yang terdapat dalam KDD berfungsi untuk mengekstrak maklumat yang mungkin berguna, tersirat

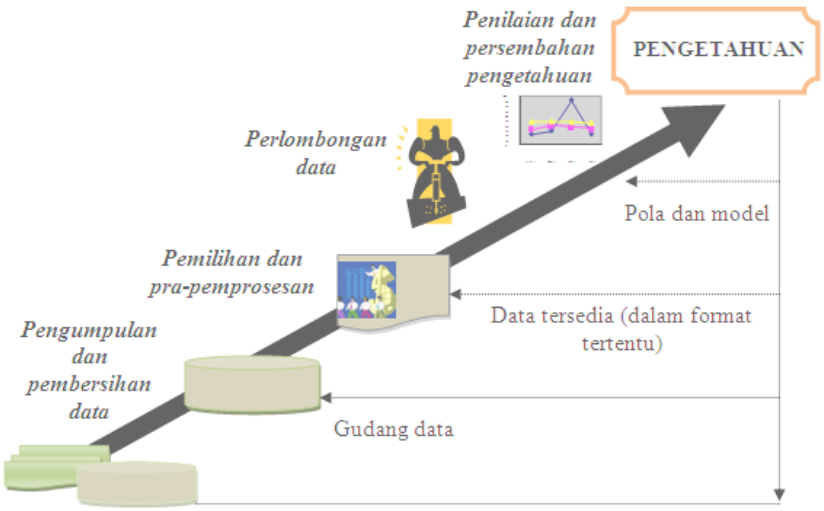
dan sebelum ini tidak diketahui dari suatu koleksi data atau pangkalan data yang besar (Walsh, 2002). Perlombongan data pula merupakan kaedah yang menerangkan perolehan pengetahuan dan pola berguna daripada data, serta merupakan salah satu proses penting untuk mendapatkan pengetahuan dalam proses penemuan pengetahuan seperti ditunjukkan dalam rajah 1.1.



Rajah 1.1: Pelbagai Bidang Penemuan Pengetahuan dalam Pangkalan Data, KDD.

Perlombongan data merupakan suatu langkah dalam proses penemuan pengetahuan yang terdiri daripada analisis data dan algoritma penemuan iaitu ia akan menghasilkan beberapa pola tertentu daripada data apabila wujudnya kekangan perkomputan yang boleh diterima. KDD ialah proses penggunaan pangkalan data bersama-sama keperluan pemilihan, pra-pemprosesan, persampelan dan penukaran data dengan menggunakan kaedah perlombongan data untuk mendapatkan pola dan seterusnya membuat penilaian hasil perlombongan data bagi mengenalpasti pola yang mempunyai pengetahuan. Lima langkah penting dalam proses KDD iaitu; I) Pengumpulan data, II) Pembersihan data, III) Pemilihan data serta pra pemprosesan,

IV) Penilaian pengetahuan dan V) Penggunaan pengetahuan. Rajah 1.2 menunujukkan proses yang terlibat dalam penemuan pengetahuan.



Rajah 1.2: Susunan proses dalam KDD.

I) Pengumpulan Data

Dalam dunia yang berteknologi serba maju ini, kita sebenarnya telah mengumpulkan terlalu banyak data. Data terdiri daripada pelbagai bentuk dan disimpan didalam pelbagai media penyimpanan yang didapati di pasaran. Berdasarkan teknologi pengumpulan data sedia ada, pelbagai sumber data telah kita kumpulkan contohnya data pelajar, data penduduk negara, data perlawanan, log tetamu, transaksi perniagaan, data sainstifik, video, gambar dan banyak lagi.

Data-data ini disimpan didalam pelbagai bentuk seperti Sistem Pengurusan Pangkalan Data (RDBMS), fail, teks dan sebagainya. Bagi sistem perlombongan data yang menyeluruh, sumber-sumber data yang disebutkan tadi dikumpulkan kepada satu bentuk data yang sangat besar dinamakan gudang data.

Gudang data ialah koleksi data yang besar dan dikendali secara berasingan daripada data operasi dalam sesebuah organisasi dan digunakan untuk menyelesaikan masalah dengan menyediakan data tertentu apabila diperlukan. Data merupakan subjek utama bagi perlombongan data, pengumpulan data penting kerana kualiti perlombongan data bergantung kepada data yang dikumpulkan tersebut.

II) Pemilihan dan Pra-pemprosesan

Domain suatu aplikasi termasuklah matlamat applikasi dan jenis pengetahuan yang hendak diketahui dari data yang diperolehi perlu difahami terlebih dahulu dan kemudian, data sasaran perlu dikenalpasti sebelum perlombongan data dilakukan. Pra-pemprosesan data adalah berkaitan kebersihan data. Data sebenar tidak bersih kerana beberapa faktor seperti mungkin terdapat data yang hilang semasa mengisi pangkalan data, attribut yang tidak mencukupi, ralat di dalam data dan data yang tidak konsisten. Disebabkan oleh faktor-faktor inilah pra-pemprosesan dianggap proses yang sangat penting dalam perlombongan data kerana ia memastikan kualiti pengetahuan yang diperoleh daripada data. Empat kaedah dalam pra-pemprosesan disenaraikan di dalam jadual 1.1;

Jadual 1.1
Kaedah Pra-pemprosesan.

Kaedah	Kegunaan	Contoh
Pembersihan	Menyingkirkan atau membetulkan data yang mempunyai ralat, hilang, tidak konsisten dan rekod bertindih.	Penggunaan kaedah Binning, Pengelompokan, Regresi.
Integrasi data	Menganalisis data yang mempunyai masalah seperti jenis data, nilai data, rekod yang bertindih, atribut yang bertindih dan skema penamaan attribut yang tak sama.	Jika STUDID = Sudent_ID = Stud_IC Maka berikan satu skema penamaan yang jelas.
Penukaran data	Menukar, meringkas atau penskalaan semula data kepada bentuk atau wakil tertentu. Kaedah ini penting kerana kaedah perlombongan data yang tertentu mungkin memerlukan jenis data yang telah diringkas atau ditukar kepada skala yang lebih kecil. Terdiri daripada, *Smooting, Aggregation, Generalisation, Normalisation* dan *Discretization*.	Kaedah binning, Min-max Normalisasi, Normalisasi Z-skor, Pengskalaan-perpuluhan.
Pengurangan data	Penggunaan keseluruhan data daripada gudang data yang besar memakan masa yang lama untuk mendapatkan keputusan. Pengurangan data perlu dilakukan supaya set data yang digunakan mampu menghasilkan keputusan analitikal yang sama dengan set data yang lebih besar.	Penurunan kiub data, Penurunan dimensi, Diskritisasi, Pohon keputusan (mendapatkan atribut yang perlu sahaja)

Sumber: Mohd Shamrie Sainin (2003).

III) Perlombongan Data

Seperti yang telah diterangkan sebelum ini, perlombongan data ialah kaedah yang digunakan untuk mendapatkan pola daripada data yang boleh dijadikan pengetahuan dan melibatkan proses pembinaan model secara berulang menggunakan kaedah tertentu (kaedah pintar). Kaedah pintar melibatkan algoritma pembelajaran terhadap data. Proses perlombongan data menggunakan konsep pembelajaran mesin, statistik dan teknik persembahan keputusan supaya hasil kajian mudah difahami.

Tiga perkara penting dalam perlombongan data ialah; 1) menentukan jenis model yang hendak dibina, 2) Pemilihan algoritma atau fungsi perlombongan data, dan 3) Pembinaan model menggunakan data latihan, data ujian dan data pengesahan. Model yang hendak dibina daripada data adalah berdasarkan tugas perlombongan data. Tugas-tugas perlombongan data yang popular ialah ramalan, pengkelasan, pengelompokan, hubungan dan korelasi.

Pemilihan algoritma perlombongan data berdasarkan kepada dua faktor iaitu pertama pembelajaran terselia iaitu kaedah yang mementingkan sasaran dalam data, konsep sasaran tersebut dijadikan panduan untuk mengetahui hubungan diantara atribut yang terlibat dalam konsep sasaran tersebut. Klasifikasi dan peramalan adalah contoh tugas perlombongan data yang bersifat pembelajaran terselia. Manakala faktor kedua ialah pembelajaran tanpa seliaan iaitu kaedah perlombongan data yang tidak mempunyai konsep sasaran dalam data. Kaedah ini akan melihat kepada corak data

dan kaitan antara data supaya ianya dapat diletakkan dalam kumpulan tertentu contohnya pengelompokan.

Teknik perlombongan data yang popular ialah pohon keputusan, rangkaian neural, Bayesian dan regresi. Namun, tidak semua algoritma pernemuan pengetahuan tersebut sesuai untuk sesuatu data, kerana setiap algoritma memerlukan kaedah penyediaan data tertentu yang mungkin berbeza. Pemilihan pembelajaran sama ada terselia atau tidak bergantung kepada data yang telah disediakan dan jenis pengetahuan yang akan dicapai.

Teknik yang sesuai harus diapplikasikan kepada data. Data dibahagikan kepada dua atau tiga bahagian yang terdiri daripada data latihan, data ujian dan data pengesahan. Data yang di bahagi dua biasanya dibahagikan mengikut nisbah 3:2 iaitu 60% data latihan dan 40% data ujian. Data latihan digunakan untuk menjana model dan data ujian digunakan untuk memastikan ketepatan model yang diperolehi.

IV) Penilaian dan Persembahan Pengetahuan

Suatu model atau pengetahuan yang diperolehi perlu ditentusahkan bagi menguji sejauh manakah model tersebut menghasilkan ketepatan dalam membuat keputusan atau klasifikasi. Ralat model diambil kira dalam menentukan ketepatan. Ralat data latihan boleh dikira dengan menggunakan data ujian atau data latihan. Kebanyakan pengujian terhadap ralat ini adalah berdasarkan kepada *matrik konfusi*.

Selain itu, konsep terlebih-latih (overtraining) juga boleh berlaku sehingga menghasilkan keputusan yang tidak tepat. Teknik pengesahan silang (cross-validation) boleh digunakan untuk menyelesaikan masalah ini bagi kajian yang menggunakan sampel data yang sedikit. Kaedah pengesahan silang-k (k-Cross Validation) bertujuan memastikan setiap data digunakan untuk latihan dan ujian.

Hasil perlombongan data boleh dipersembahkan menggunakan teknik visual, simulasi dan sebagainya. Model keputusan seperti pohon keputusan dan peraturan (rules) boleh digambarkan kepada peraturan atau dengan menggunakan visual yang menunjukkan pohon klasifikasi. Keputusan model pengelompokan pula boleh di plot kedalam graf menggunakan paparan visual dua atau tiga dimensi.

1.3 Pernyataan Masalah

Pada awal abad ke-21, beberapa perubahan dan perkembangan dalam sistem pendidikan Malaysia telah berlaku disebabkan oleh cabaran yang dihadapi akibat kesan globalisasi, liberalisasi, dan perkembangan teknologi maklumat dan komunikasi (Jamil Ahmad & Norlia Goolamally, 2008). Cabaran globalisasi kepada dunia pendidikan membawa kepada persaingan dalam bidang pendidikan iaitu persaingan antara institusi-institusi pengajian tinggi semakin meningkat demi mencapai matlamat negara tersebut. Untuk kekal bersaing, perancang universiti mestilah bijak dalam

menggunakan peluang dan maklumat semasa bagi penambahbaikan mutu perkhidmatan universiti.

Institusi pengajian tinggi mempunyai sistem untuk menyimpan maklumat pelajar dan telah mengumpulkan jumlah data maklumat pelajar yang besar. Bagaimanapun, data ini lazimnya tidak diletakkan ke dalam bentuk memperbaiki kegunaannya. Kini, universiti telah "kaya-data" tetapi "miskin maklumat". Kebanyakan universiti tidak memanfaatkan perlombongan data dalam menganalisis dan mendedahkan maklumat tersembunyi didalam data kemasukan pelajar.

Usaha mendedahkan maklumat tersembunyi sangat berguna untuk menghasilkan pengetahuan yang berfungsi dalam penambahbaikan pembuatan keputusan oleh pihak pengurusan dan perancangan. Oleh itu, dengan analisis yang mendalam dan berstruktur ke atas pangkalan data institusi sebagai contoh, data dari pangkalan data sistem maklumat pelajar di institusi pengajian tinggi, boleh membantu memberi maklumat bermanfaat berkaitan pencapaian pelajar. Dalam kajian ini analisis deskriptif digunakan untuk mengenal pasti pola yang terdapat dalam data.

Hasil utama pusat pengajian tinggi adalah graduan, oleh itu menjadi tanggungjawab pihak pengurusan universiti untuk memastikan UPSI dapat melahirkan jumlah pelajar yang boleh bergraduat dengan cemerlang dan boleh berkhidmat sebagai tenaga pendidik yang berkualiti di sekolah serta mampu menjadi seorang kakitangan

yang berkredibiliti di tempat kerja sejajar dengan Tujuh Teras Asas Pendidikan Malaysia (KPTM, 2010); 1) Meluaskan peluang dan meningkatkan ekuiti, 2) Memperbaiki kualiti pengajaran dan pembalajaran, 3) Meningkatkan penyelidikan dan inovasi, 4) Memperkuatkan institusi pengajian tinggi, 5) Meningkatkan kebangsaan, 6) Membudayakan pengajian sepanjang masa dan 7) Memperkukuhkan sistem pengagihan Kementerian Pengajian Tinggi.

Fakta bahawa latar belakang siswazah seperti umur, jantina, bangsa, kewarganegaraan, saluran kemasukan ke institusi pengajian tinggi, pencapaian pelajar sebelum memasuki universiti, keputusan semester pertama di universiti dan bidang pengkhususan yang diambil dikatakan mempengaruhi pencapaian pelajar semasa bergraduat (Shulruf *et al*. (2010); Jennifer E. G. dan Bryndl H. M. (2007); Hayes *et a*l. (1997); Graham (1991)). Mengikut kajian lampau, faktor sosio-demografi dan persekitaran pelajar sedikit sebanyak mempengaruhi pencapaian pelajar semasa menuntut di sesebuah institusi pengajian tinggi sama ada dari segi isprirasional ataupun spiritual.

Sebagai langkah awal untuk memastikan kejayaan universiti, pihak pentadbiran universiti juga perlu berkeupayaan untuk meramal jumlah pelajar yang akan berjaya serta pelajar bermasalah supaya tindakan susulan dapat diambil. Seterusnya, hasil keputusan permodelan ramalan juga dapat membantu pihak pengurusan UPSI mengenalpasti pelajar yang harus diberi perhatian dan bantuan

berdasarkan kriteria-kriteria pelajar yang telah dikenalpasti seawal selepas semester pertama pelajar di universiti lagi.

Kesalahan atau kelewatan dalam memberi bantuan bukan sahaja akan merugikan pihak universiti, bahkan tidak akan bermakna sekiranya bantuan tersebut disalurkan kepada pelajar yang salah. Sebagai contoh dalam kajian ini, bantuan yang dimaksudkan adalah bantuan seperti kursus-kursus intensif dan kelas tambahan perlu diberikan kepada pelajar yang bermasalah dalam pelajaran, pihak universiti akan kerugian jika bantuan tersebut diberikan kepada pelajar yang cemerlang.

Kajian ini menjelaskan penggunaan dan keupayaan perlombongan data dan aplikasinya terhadap pangkalan data institusi pengajian tinggi terutama dalam memahami data kemasukan pelajar sarjana muda di Universiti Pendidikan Sultan Idris, UPSI. Kajian ini memanfaatkan pendekatan perlombongan data deskriptif dan peramalan dalam mengenalpasti maklumat tersembunyi di dalam pangkalan data universiti. Kajian ini juga akan menguji sejauh manakah faktor latar belakang siswazah mempengaruhi pencapaian pelajar di UPSI. Oleh sebab itu, kajian ini membangunkan model ramalan bagi mengenalpasti dan membezakan kriteria pelajar yang cemerlang dan juga pelajar yang bermasalah untuk bergraduat di UPSI secara keseluruhan dan dengan lebih spesifik mengikut fakulti.

1.4 Objektif Kajian

1) Menganalisis serta menentukan pola atau corak tersembunyi bagi data pelajar UPSI.
2) Mengkaji hubungan antara latar belakang pelajar terhadap pencapaian akademik melalui kaedah analisis deskriptif dan peramalan.
3) Membangunkan dan menilai keberkesanan model ramalan yang dibangunkan serta mengenalpasti faktor yang membezakan pelajar yang berpotensi untuk cemerlang dan pelajar yang perlu dibantu untuk cemerlang.

1.5 Soalan Kajian

Kajian ini dijalankan adalah untuk menilai keberkesanan aplikasi model dalam kaedah perlombongan data untuk meramal pencapaian pelajar-pelajar UPSI. Hal ini berguna sebagai sumber rujukan pihak pengurusan universiti bagi menguruskan masalah pemilihan pelajar, peruntukan dana atau bantuan akedemik pelajar serta meramal pelajar-pelajar yang begraduat dengan cemerlang menggunakan kaedah perlombongan data yang sistematik.

1.5.1 Persoalan kajian

1) Bagaimanakah pola atau corak tersembunyi yang didapati dari data pelajar UPSI?

2) Adakah latar belakang pelajar mempengaruhi pencapaian pelajar semasa bergraduat?

   i. Adakah umur, jantina, bangsa, dan kewarganegaraan mempengaruhi pencapaian mereka di UPSI?

   ii. Adakah terdapat perbezaan antara pencapaian yang dicapai di UPSI dengan program pengajian yang dipilih dan saluran kemasukan ke UPSI?

   iii. Adakah terdapat perbezaan pencapaian pelajar semasa bergraduat dengan keputusan semester pertama pelajar di UPSI?

3) Sejauhmanakah keberkesanan model matematik yang dibangunkan dalam meramal pencapaian pelajar di UPSI?

1.5.2 Hipotesis Kajian

Berikut disenaraikan hipotesis kajian bagi soalan kajian 1 yang telah dikenalpasti;

a) H_0: Tiada hubungan yang signifikan di antara pencapaian pelajar dengan umur pelajar di UPSI.

H_1: Terdapat hubungan yang signifikan di antara pencapaian pelajar dengan umur pelajar di UPSI.

b) H_0: Tiada hubungan yang signifikan di antara pencapaian pelajar dengan jantina pelajar di UPSI.

H_1: Terdapat hubungan yang signifikan di antara pencapaian pelajar dengan jantina pelajar di UPSI.

c) H_0: Tiada hubungan yang signifikan di antara pencapaian pelajar dengan bangsa pelajar di UPSI.

H_1: Terdapat hubungan yang signifikan di antara pencapaian pelajar dengan bangsa pelajar di UPSI.

d) H_0: Tiada hubungan yang signifikan di antara pencapaian pelajar dengan kewarganegaraan pelajar di UPSI.

H_1: Terdapat hubungan yang signifikan di antara pencapaian pelajar dengan kewarganegaraan pelajar di UPSI.

e) H_0: Tiada hubungan yang signifikan di antara pencapaian pelajar yang dicapai di UPSI dengan saluran kemasukan mereka ke UPSI.

H_1: Terdapat perbezaan yang signifikan di antara pencapaian pelajar yang dicapai di UPSI dengan saluran kemasukan mereka ke UPSI.

f) H_0: Tiada hubungan yang signifikan di antara pencapaian pelajar yang dicapai di UPSI dengan program pengajian yang dipilih di UPSI.

H_1: Terdapat perbezaan yang signifikan di antara pencapaian pelajar yang dicapai di UPSI dengan program pengajian yang dipilih di UPSI.

g) H_0: Tiada hubungan yang signifikan di antara pencapaian pelajar dengan keputusan semester pertama di UPSI.

H_1: Terdapat hubungan yang signifikan di antara pencapaian pelajar dengan keputusan semester pertama di UPSI.

1.6 Kepentingan Kajian

Perlombongan data bukanlah suatu kaedah yang asing bagi penyelidik di luar Malaysia dan sering di aplikasikan dalam pentadbiran bagi mengenalpasti pola serta maklumat berguna kepada pentadbiran. Kaedah ini dapat membantu pentadbir dalam membuat keputusan dan memberi gambaran yang tepat bagi suatu taburan data. Keputusan hasil analisis deskriptif keatas data kemasukan pelajar dari pangkalan data universiti berupaya memberi maklumat berguna kepada pembuat keputusan. Manakala, hasil permodelan pencapaian pelajar menggunakan kaedah perlombongan data berketepatan tinggi memberi idea bagi penambahbaikan perkhidmatan pentadbiran universiti.

Berdasarkan pemahaman ini, institusi pengajian tinggi akan berupaya mengagihkan sumber dengan lebih berkesan. Maklumat yang diperolehi dari kaedah ini sangat berguna kepada pihak pengurusan universiti terutama dalam mengawal pergerakan pencapaian pelajar berdasarkan kriteria ramalan yang telah dikenalpasti. Jawatankuasa kurikulum boleh menggunakan hasil ramalan untuk mengawal perubahan yang berlaku di dalam kurikulum dan menilai kesan kepada perubahan tersebut. Seorang penasihat akademik dapat merujuk kepada hasil penyelidikan apabila memberi nasihat kepada pelajar yang bermasaalah dalam pelajaran mereka, dengan itu langkah pencegahan boleh diambil dengan lebih awal.

Tambahan, seorang tenaga pengajar dapat memperbaiki pendekatan pengajaran dan pembelajaran yang biasa digunakan kepada pendekatan yang lebih

baik. Begitu juga dengan plan pembelajaran dan khidmat bantuan dapat diberikan kepada pelajar yang lemah. Perlombongan data berupaya memberi maklumat yang berguna kepada pihak pengurusan universiti bagi mengambil tindakan sebelum seseorang pelajar diberhentikan, atau boleh mengagihkan sumber dengan lebih efisyen dengan jangkaan tepat mengenai jumlah pelajar lelaki dan perempuan yang mendaftar sesuatu program pengajian.

Selain daripada penyelidik dan kakitangan, kejayaan sebuah pusat pengajian tinggi juga dinilai melalui bilangan siswazah yang berjaya bergraduat dalam tempoh yang ditetapkan dan mendapat pekerjaan pada kadar yang segera sama ada disektor kerajaan mahupun swasta. Fokus utama UPSI adalah untuk menghasilkan bakal-bakal pendidik yang berkualiti bagi mendidik generasi pada masa akan datang. Oleh itu, menjadi tanggungjawab pihak pengurusan universiti untuk memastikan bakal-bakal graduan UPSI mendapat pendidikan perguruan secukupnya dari institusi ini. Dengan keputusan kajian ini, diharap dapat membantu pembuat keputusan universiti dalam proses penambahbaikan perkhidmatan yang disediakan kepada pelajar.

Disamping itu juga, pihak pengurusan UPSI terutama dari Bahagian Akademik UPSI dapat menjadikan model ramalan yang dihasilkan dalam kajian ini sebagai model asas dalam membangunkan model ramalan keatas pencapaian akademik pelajar universiti bagi mengenalpasti pelajar yang berpotensi untuk cemerlang semasa bergraduat seawal semester pertama pelajar di universiti.

1.7 Batasan Kajian

Kajian ini dijalankan di Universiti Pendidikan Sultan Idris, Tanjong Malim. Peserta kajian ini adalah terhad dan terdiri daripada pelajar-pelajar universiti berkenaan sahaja. Kajian menggunakan populasi pelajar universiti yang telah bergraduat dari tahun 1997 sehingga tahun 2008 diperolehi dari pangkalan data Bahagian Akademik universiti. Oleh itu, hasil dan kesimpulan kajian ini mungkin tidak sesuai digunakan dalam situasi uiniversiti yang berlainan.

Kajian ini dijalankan dengan menggunakan kaedah kuantitatif yang terbahagi kepada dua bahagian iaitu analisis deskriptif dan analisis ramalan. Bagi analisis ramalan, sebuah model dalam kaedah perlombongan data telah dipilih, iaitu pohon regresi. Dengan ini, segala maklumat mengenai responden adalah berdasarkan analisis data yang diperoleh dari Bahagian Akademik UPSI. Kebolehpercayaan dapatan kajian bergantung kepada ketepatan data yang diperolehi. Untuk memastikan ketepatan dalam keputusan, setiap sampel dipilih secara rawak dari pengkalan data sedia ada.

Selain itu, kualiti data ditentusahkan terlebih dahulu sebelum data tersebut diuji. Masalah data tak lengkap, data terganggu (*noisy*) dan data yang tak konsisten adalah merupakan perkara biasa. Data tersebut juga dibersihkan dan akhirnya dijelmakan kedalam format yang bersesuaian dengan perlombongan data sebelum diproses.