

**PENDEKATAN PENGELASAN BERGABUNG  
UNTUK MENENTUKAN KEPUTUSAN  
KELULUSAN BIASISWA**

**WIRAWATI DEWI BINTI AHMAD**

**UNIVERSITI KEBANGSAAN MALAYSIA**

PENDEKATAN PENGELASAN BERGABUNG UNTUK MENENTUKAN  
KEPUTUSAN BIASISWA

WIRAWATI DEWI BINTI AHMAD

DISERTASI YANG DIKEMUKAKAN UNTUK MEMENUHI SEBAHAGIAN  
DARIPADA SYARAT MEMPEROLEHI IJAZAH  
SARJANA TEKNOLOGI MAKLUMAT (SAINS MAKLUMAT)

FAKULTI TEKNOLOGI DAN SAINS MAKLUMAT  
UNIVERSITI KEBANGSAAN MALAYSIA  
BANGI

2019

**PENGAKUAN**

Saya akui karya ini adalah hasil kerja saya sendiri kecuali nukilan dan ringkasan yang tiap-tiap satunya telah saya jelaskan sumbernya.

19 Januari 2019

**WIRAWATI DEWI BINTI  
AHMAD  
P86693**

## PENGHARGAAN

Pertama sekali saya ingin mengucapkan syukur ke hadrat Ilahi dengan limpah kurnia-Nya dapat saya menyiapkan kajian dan tesis ini dalam tempoh yang diberikan. Ribuan terima kasih jua kepada ibu bapa tercinta dan keluarga yang tersayang yang banyak menyokong saya dalam melanjutkan pelajaran ke peringkat sarjana. Terutama kepada puteri-puteri tersayang Amylea Qistyna Dewi, Asfarina Qistyna Dewi dan Adelia Qistyna Dewi yang menjadi sumber inspirasi.

Terima kasih khas juga ditujukan kepada penyelia saya Profesor Dr. Azuraliza Abu Bakar yang banyak membimbing saya dari awal hingga akhir tesis ini disiapkan. Ilmu dan pandangan serta nasihat yang bernilai amat membantu saya menyiapkan tesis kajian ini dengan sempurna. Tidak lupa juga kepada semua pensyarah Fakulti Teknologi dan Sains Maklumat (FTSM) UKM yang turut memberi bimbingan dan pandangan yang bernas dalam membantu kajian ini.

Ucapan terima kasih tidak terhingga juga buat sahabat-sahabat seperjuangan yang sama-sama berkongsi ilmu dan pandangan dalam membantu saya menghasilkan tesis yang baik bagi melayakkan saya mendapat Ijazah Sarjana dalam bidang Teknologi Maklumat.

## ABSTRAK

Biasiswa pendidikan tinggi merupakan kemudahan kewangan yang diberi kepada pelajar-pelajar yang layak bagi melanjutkan pendidikan ke peringkat pengajian di Universiti. Sumber dana yang terhad dengan bilangan pemohon yang semakin bertambah memaksa pihak Kerajaan mencari solusi bagi membantu mempercepat dan memudahkan proses pemilihan pemohon yang layak. Pada masa kini, kajian pengelasan keputusan biasiswa menggunakan pendekatan data analitik terhad kepada kaedah pengelasan tunggal. Kaedah ini mempunyai kelemahan-kelemahan tertentu dalam menghasilkan model pengelasan terbaik. Melalui kajian kesusasteraan, kelemahan-kelemahan kaedah ini boleh diatasi dengan pelaksanaan teknik pengelasan bergabung. Dalam kajian ini, data penawaran biasiswa sedia ada digunakan sebagai data latihan dan ujian dalam semua uji kaji dengan tiga objektif utama iaitu: merekabentuk satu rangka kerja pengelasan bergabung penentuan keputusan biasiswa pendidikan tinggi, membangunkan model pengelasan tunggal dan bergabung bagi menentukan keputusan biasiswa; dan membangunkan model pengelasan petua bergabung untuk meningkatkan keupayaan mengelas keputusan permohonan biasiswa. Lima algoritma pengelasan tunggal digunakan iaitu J48, SVM, NB, ANN dan RT manakala tiga teknik pengelasan bergabung digunakan iaitu peningkatan, pembungkusan, dan undian. Teknik pengelasan petua bergabung juga digunakan dengan menggabungkan petua model peraturan sekutuan algoritma Apriori dan petua model J48 bagi melihat perbandingan prestasi antara model. Hasil keputusan mendapati model pengelasan tunggal terbaik adalah dari algoritma SVM dengan ketepatan tertinggi 86.45%. Bagi pengelasan bergabung pula, model J48(pembungkusan) dan model undian mencapai ketepatan yang sama tertinggi iaitu 86.90% manakala bagi teknik pengelasan petua bergabung menghasilkan ketepatan 85.9%. Ini membuktikan bahawa teknik pengelasan bergabung berupaya meningkatkan ketepatan model pengelasan tunggal yang lemah manakala teknik pengelasan petua bergabung mempunyai keupayaan yang boleh dikategorikan baik dengan perbezaan kurang 0.5% berbanding model pengelasan tunggal terbaik.

## **ENSEMBLE MODEL FOR HIGHER LEARNING SCHOLARSHIP AWARD DECISION**

### **ABSTRACT**

Higher education scholarship is a financial facility given to eligible students to pursue higher education at the University. Limited funding sources with the growing number of applicants forced the Government to find solutions to help speed up and facilitate the prose selection of eligible students. Currently, the study of classification of scholarship results using analytical data approach is limited to a single classification method. This method has certain weaknesses in producing the best classification model. Through literary studies, the weaknesses of these methods can be overcome by the implementation of ensemble techniques. In this study, the scholarship data were used as training and testing data in all experiments with three key objectives: designing an ensemble framework for the Higher Education scholarship decision, developing a single classification model and ensemble model to determine the decision of the scholarship; and develop a rules-based ensemble model to enhance the classification capabilities of the decision of the scholarship candidates. Five single classification algorithms are used: J48, SVM, NB, ANN and RT while three ensemble techniques are used for ensemble classification modelling namely boosting, bagging, and voting. The rules-based classification techniques are developed by ensemble the rules obtained from the Apriori algorithm and J48 model to compare the performance between models. The results show that, the best single classification model is from the SVM algorithm with the highest accuracy of 86.45%. For the ensemble classification, the J48 (bagging) model and the voting model reached the highest accuracy of 86.90% while for the rules-based ensemble technique achieved 85.9% of accuracy. These shows that, ensemble classification techniques are capable of enhancing the accuracy of a weak single classification model while the rules-based technique has a well-defined capability of less than 0.5% compared to the best single classification model.

## KANDUNGAN

		<b>Halaman</b>
<b>PENGAKUAN</b>		<b>ii</b>
<b>PENGHARGAAN</b>		<b>iii</b>
<b>ABSTRAK</b>		<b>iv</b>
<b>ABSTRACT</b>		<b>v</b>
<b>KANDUNGAN</b>		<b>vi</b>
<b>SENARAI JADUAL</b>		<b>x</b>
<b>SENARAI SINGKATAN</b>		<b>xv</b>
<b>BAB I</b>	<b>Pengenalan</b>	
1.1	Latar Belakang	1
	1.1.1 Program Biasiswa Sektor Pendidikan Tinggi	1
	1.1.2 Pangkalan Data Sektor Pendidikan Tinggi	3
	1.1.3 Analisis Data Menggunakan Pendekatan Perlombongan Data	4
1.2	Pernyataan Masalah	4
1.3	Objektif Kajian	5
1.4	Skop Kajian	6
1.5	Metodologi Kajian	7
	1.5.1 Fasa Input Kajian	8
	1.5.2 Fasa Kajian	8
	1.5.3 Fasa Output Kajian	9
1.6	Kepentingan Kajian	9
1.7	Organisasi Tesis	10
<b>BAB II</b>	<b>Kajian Kesusasteraan</b>	
2.1	Pengenalan	12
2.2	Analitik Data Raya Sektor Awam(DRSA)	12
2.3	Perlombongan Data	14
	2.3.1 Data Analitik Sektor Pendidikan	14
	2.3.2 Pengelasan dan Ramalan Sektor Pendidikan	15
	2.3.3 Model Ramalan Pengelasan Penentuan Keputusan Biasiswa	16
	2.3.4 Kelemahan Model Pengelasan Tunggal	23

2.4	Model Pengelasan Bergabung	24
	2.4.1 Pengenalan	24
	2.4.2 Teknik Peningkatan	25
	2.4.3 Teknik Pembungkusan	26
	2.4.4 Teknik Undian	26
	2.4.5 Kelebihan dan Kelemahan Pengelasan Bergabung	27
	2.4.6 Kajian-kajian Model Pengelasan Bergabung	28
2.5	Perlombongan Data Menggunakan Teknik Peraturan Sekutuan	32
	2.5.1 Pengenalan	32
	2.5.2 Algoritma Apriori	32
2.6	Model Pengelasan Petua Bergabung	34
2.7	Pemilihan ciri	35
2.8	Analisis Pengetahuan	36
2.9	Cadangan penyelidikan	37
	2.9.1 Strategi Penyelidikan	37
	2.9.2 Rangka Kerja Pengelasan Bergabung	38
2.10	Kesimpulan	41
<b>BAB III</b>	<b>METODOLOGI KAJIAN</b>	
3.1	Pengenalan	42
3.2	Pendekatan Penyelidikan	42
3.3	Penyediaan Data	43
	3.3.1 Hasil Pembersihan dan Intergrasi Data	43
	3.3.2 Pemilihan dan Transformasi Data	46
3.4	Analisis Deskriptif Data Penawaran Biasiswa Pendidikan Tinggi	49
	3.4.1 Kelas Label Layak Jawatankuasa	49
	3.4.2 Julat Umur	50
	3.4.3 Status Penaja	50
	3.4.4 Status Pekerjaan	51
	3.4.5 Universiti	52
	3.4.6 Struktur Pengajian	52
	3.4.7 Julat Tahun Ijazah	53
	3.4.8 Atribut lain-lain	53
3.5	Rangka kerja Pengelasan Bergabung	53
	3.5.1 Fasa Input Kajian	56
	3.5.2 Fasa Kajian	56
	3.5.3 Fasa Output Kajian	72
3.6	Kesimpulan	72



<b>BAB IV</b>	<b>MODEL PENGELASAN TUNGGAL DAN BERGABUNG PENENTUAN KEPUTUSAN BIASISWA</b>	
4.1	Pengenalan	73
4.2	Keputusan Uji Kaji Model Pengelasan Tunggal	73
	4.2.1 Jadual Kontigensi	73
	4.2.2 Prestasi Model	75
4.3	Keputusan Model Pengelasan Bergabung	78
	4.3.1 Teknik Peningkatan <i>Adaboost</i>	78
	4.3.2 Teknik Pembungkusan	82
	4.3.3 Teknik Undian	86
4.4	Perbincangan	88
<b>BAB V</b>	<b>ANALISIS PENGETAHUAN DAN MODEL PENGELASAN PETUA BERGABUNG</b>	
5.1	Pengenalan	90
5.2	Analisis Petua Model J48	90
5.3	Model Peraturan Sekutuan Menggunakan algoritma Apriori	95
5.4	Pengelasan Petua Bergabung	98
5.5	Kesimpulan	99
<b>BAB VI</b>	<b>RUMUSAN DAN PENUTUP</b>	
6.1	Pendahuluan	100
6.2	Rumusan dan Penemuan Penyelidikan	100
6.3	Sumbangan Kajian	103
6.4	Batasan Kajian	104
6.5	Cadangan Perluasan Kajian	104
6.6	Penutup	105
<b>RUJUKAN</b>		<b>106</b>
Lampiran A	Jadual Aktiviti dan Dapatan Kajian	116
Lampiran B	Contoh Surat Kebenaran Menggunakan Data	117
Lampiran C	Borang Penilaian Pakar	119
Lampiran D	Contoh Kod Aturcara Permodelan Pengelasan Bergabung	122
Lampiran E	Senarai Kod Univerisiti yang diiktiraf	124

Lampiran F	Petua Model J48(Layak)	127
Lampiran G	Petua Model J48(Tidak Layak)	131
Lampiran H	Petua Model Apriori(Layak)	135
Lampiran I	Petua Model Apriori(Tidak Layak)	136

## SENARAI JADUAL

<b>No. Jadual</b>		<b>Halaman</b>
Jadual 1-1	Program-program biasiswa di KPT Malaysia	2
Jadual 2-1	Algoritma-algoritma dalam kajian pengelasan dan ramalan sektor pendidikan	15
Jadual 2-2	Contoh-contoh kajian menggunakan lima algoritma pengelasan tunggal J48, RT, NB, SVM dan ANN	21
Jadual 2-3	Perbandingan algoritma pengelasan tunggal	23
Jadual 2-4	Kelebihan dan kelemahan Algoritma Adaboost	26
Jadual 2-5	Perbandingan kelebihan dan kelemahan teknik pengelasan bergabung	27
Jadual 2-6	Perbezaan ciri-ciri teknik pengelasan bergabung	28
Jadual 2-7	Ringkasan kajian pengelasan bergabung pelbagai bidang kajian	28
Jadual 2-8	Ringkasan kajian pengelasan bergabung terkini	29
Jadual 2-9	Kajian-kajian peraturan persatuan menggunakan algoritma Apriori	33
Jadual 2-10	Kajian pengelasan petua bergabung	34
Jadual 2-11	Sistem Pengurusan Pengetahuan Perlombongan data	36
Jadual 2-12	Perbezaan Kaedah Kuantitatif dan Kualitatif	37
Jadual 3-1	Senarai atribut yang dihapuskan	44
Jadual 3-2	Kedudukan Kepentingan Atribut Menggunakan Teknik Info Gain dan Gain Ratio	47
Jadual 3-3	Senarai atribut yang dikekalkan dan dicipta menghasilkan set data lengkap	48
Jadual 3-4	Taburan data atribut struktur pengajian dengan peratusan berjaya	53
Jadual 3-5	Jadual kontigensi	69
Jadual 4-1	Jadual kontigensi model pengelasan tunggal	75

Jadual 4-2	Keputusan parameter ketepatan model-model terbaik pengelasan tunggal	75
Jadual 4-3	Keputusan keseluruhan prestasi model terbaik pengelasan tunggal	77
Jadual 4-4	Perbandingan jadual kontigensi RT(tunggal) dan RT(Adaboost) dengan teknik pecahan data 90:10	79
Jadual 4-5	Perbandingan keputusan teknik Adaboost bagi model pengelasan algoritma RT	80
Jadual 4-6	Nilai parameter ralat model RT dan RT(Peningkatan)	82
Jadual 4-7	Ketepatan parameter ketepatan model pengelasan tunggal dan pengelasan bergabung	83
Jadual 4-8	Perbandingan keputusan model J48 dan ANN tunggal dan bergabung	84
Jadual 4-9	Keputusan model pengelasan bergabung teknik pembungkusan	85
Jadual 4-10	Keputusan keseluruhan model teknik undian	87
Jadual 4-11	Keputusan ketepatan keseluruhan model teknik undian dan pengelasan tunggal	87
Jadual 5-1	Contoh-contoh peraturan petua model J48	91
Jadual 5-2	Contoh peraturan petua bagi kes calon biasiswa Tidak Layak	94
Jadual 5-3	Contoh petua 2 jujukan dan 3 jujukan algoritma Apriori penentuan keputusan biasiswa	96
Jadual 5-4	Contoh petua 4 jujukan dan 5 jujukan petua algoritma Apriori penentuan keputusan biasiswa	96
Jadual 5-5	Hasil analisis petua model peraturan sekutuan Apriori	97
Jadual 5-6	Kekerapan kewujudan atribut dalam petua yang dihasilkan algoritma Apriori yang diterima pakar	97
Jadual 5-8	Set petua pengelasan petua bergabung	98
Jadual 5-9	Keputusan model pengelasan petua bergabung	99
Jadual 5-10	Perbandingan keputusan model pengelasan petua bergabung dan model pengelasan tunggal	99

Jadual 6-1 Hasil ketepatan model-model terbaik pengelasan tunggal dan  
mdel-model pengelasan bergabung 102

## SENARAI ILUSTRASI

<b>No. Rajah</b>		<b>Halaman</b>
Rajah 1-1	Metodologi Kajian	7
Rajah 2-1	Agensi-agensi terlibat dalam projek perintis inisiatif DRSA	13
Rajah 2-2	Langkah-langkah dalam data analitik menggunakan pendekatan perlombongan data	14
Rajah 2-3	Cadangan Rangka Kerja Pengelasan Bergabung Penentuan Keputusan Biasiswa Pendidikan Tinggi	39
Rajah 3-1	Contoh data senarai Kod Universiti	45
Rajah 3-2	Contoh data mentah yang mentah	46
Rajah 3-3	Statistik tawaran biasiswa mengikut jantina	49
Rajah 3-4	Taburan data tawaran biasiswa mengikut julat umur dari kategori 1 hingga 6.	50
Rajah 3-5	Status penajaan pengajian peringkat Sarjana Muda berbanding calon Layak dan Tidak Layak	51
Rajah 3-6	Taburan rekod tawaran biasiswa mengikut status pekerjaan	52
Rajah 3-8	Rangka Kerja Pengelasan Bergabung	55
Rajah 3-9	Rekabentuk Fasa kajian	56
Rajah 3-11	Rekabentuk uji kaji model pengelasan tunggal	60
Rajah 3-12	Rekabentuk uji kaji pengelasan bergabung teknik Adaboost	62
Rajah 3-13	Rekabentuk uji kaji pengelasan bergabung menggunakan teknik pembungkusan	64
Rajah 3-14	Rekabentuk uji kaji pengelasan bergabung teknik undian	66
Rajah 3-15	Rangka Kerja Pengelasan Petua Bergabung	67
Rajah 3-16	Proses kerja analisis pengetahuan	72
Rajah 4-1	Perbezaan ukuran ketepatan pengelasan model Pokok rawak asas dan dengan teknik Adaboost	81

Rajah 4-2 Perbandingan ukuran ketepatan model pengelasan tunggal dan pembungkusan

84

**SENARAI SINGKATAN**

AC	Pengelasan Kesatuan
ANN	Artificial Neuron Network
DBMS	Sistem Pengurusan Pangkalan Data
GR	Gain Ratio
GIS	Sistem Maklumat Geografi
IG	Info Gain
IPT	Institut Pendidikan Tinggi
KNN	K-Nearest Neighbour
LR	Regresi Logistik
MSE	Mean squared error
NB	Negatif Betul
NS	Negatif Salah
RMSE	Root Mean squared error
KPT	Kementerian Pendidikan Tinggi
PhD	Dotor Falsafah
PB	Positif Betul
PS	Positif Salah
RT	Pokok rawak
SPM	Sijil Pelajaran Malaysia
STPM	Sijil Tinggi Pelajaran Malaysia
SVM	Mesin Sokongan Vektor



WEKA      Waikato Environment for Knowledge Analysis

## **BAB I**

### **PENGENALAN**

#### **1.1 LATAR BELAKANG**

Meramal keputusan biasiswa dengan tepat merupakan satu proses penting bagi memastikan bantuan kewangan diberikan kepada pelajar yang benar-benar layak dan memerlukan. Menurut kajian oleh (Raharja 2014), pemberian biasiswa kepada calon yang tidak tepat akan menyebabkan masalah kekurangan dana bagi pembiayaan pendidikan sektor tinggi dikalangan pelajar yang kurang berkemampuan seterusnya menutup peluang mereka untuk melanjutkan pengajian ke tahap lebih tinggi. Ini juga dilihat akan menjejaskan matlamat utama pembiayaan sektor pendidikan tinggi iaitu bagi melahirkan kepakaran dalam bidang sains dan teknologi di masa hadapan.

Lima(5) sebab utama kepentingan pemberian biasiswa dan bantuan kewangan kepada yang layak iaitu kos pendidikan yang semakin tinggi, keadaan ekonomi yang tidak stabil menyebabkan kebanyakan golongan ibubapa tidak berkemampuan menyediakan dana pendidikan anak-anak mereka, kos hidup di institusi pendidikan tinggi semakin meningkat dan bajet pembiayaan pendidikan yang disediakan semakin berkurang. Oleh itu, adalah sangat penting bagi memastikan keputusan pemberian biasiswa diberi kepada pelajar-pelajar yang layak dan memerlukan (Alhassan, & Lawal 2015; Azuraliza, & Arshad 2013; Raharja 2014; Tun, & Aye 2014).

##### **1.1.1 Program Biasiswa Sektor Pendidikan Tinggi**

Program biasiswa sektor Pendidikan tinggi diwujudkan sejak tahun 1970 an bagi membuktikan keperihatinan dan komitmen yang tinggi pihak berwajib membantu pelajar yang layak menyambung pengajian ke peringkat lebih tinggi. Matlamat tajaan ini adalah untuk membantu pelajar yang berpotensi dan berkelayakan terutama yang

memerlukan bantuan kewangan untuk mengikuti pengajian di Institusi Pengajian Tinggi dalam bidang-bidang Sains, Teknologi, Profesional, Kepakaran dan Pengurusan Perniagaan serta bidang-bidang lain yang dikenalpasti kepentingannya dari masa ke semasa. Selain itu, tajaan ini bertujuan meningkatkan kualiti anggota akademik berkeelayakan PhD atau setaraf seterusnya memperkasakan pengajian tinggi negara bagi menjadikan Malaysia sebagai pusat pengajian tinggi serantau.

Sehingga tahun 2018 terdapat 16 program biasiswa yang ditawarkan di Kementerian Pendidikan Tinggi (KPT) Malaysia di bawah beberapa kelompok pemohon iaitu pemohon lepasan SPM, pemohon peringkat Ijazah Pertama, pemohon peringkat Ijazah Lanjutan, pemohon antarabangsa(bukan warganegara) dan pemohon khas. Program-program biasiswa yang ditawarkan mempunyai sasaran dan objektif tersendiri seperti ditunjukkan dalam Jadual 1-1.

Jadual 1-1 Program-program biasiswa di KPT Malaysia

<b>Kategori Pemohon</b>	<b>Nama Program</b>	<b>Objektif</b>
Lepasan SPM	Bantuan Kewangan Asasi IPTA	Memberi bantuan kewangan kepada pelajar asasi di IPTA yang layak sebanyak RM1500
Ijazah Pertama	MyBrainSc	Memberi biasiswa kepada pelajar jurusan Sains Tulen bagi melanjutkan pelajaran di peringkat Ijazah Sarjana Muda di IPTA
Ijazah Lanjutan	MyBrain15	Menawarkan tiga jenis biasiswa MyMaster, MyPhd dan MyPhd Industri kepada pelajar peringkat Sarjana dan doktor falsafah di IPTA tidak termasuk penjawat awam
	Skim Latihan Akademik Bumiputra (SLAB) /Skim Latihan Akademik IPTA (SLAI)	Skim bantuan biasiswa bagi penyelidik yang layak yang sedang bekerja di IPTA
	Hadiah Latihan Persekutua (HLP) Bagi Kakitangan Akademik IPTA	Menyediakan biasiswa kepada ahli akademik IPTA bagi melanjutkan pendidikan di peringkat ijazah lanjutan di dalam dan luar negara

Bersambung..

Sambungan...	Hadiah Latihan Persekutuan (HLP) Bagi Kakitangan Bukan Akademik IPTA	Menyediakan biasiswa kepada bukan ahli akademik IPTA bagi melanjutkan pendidikan di peringkat ijazah lanjutan di dalam dan luar negara
Antarabangsa	Biasiswa Antarabangsa	Terbuka kepada pelajar antarabangsa yang ditawarkan melanjutkan pelajaran di Malaysia dan tidak mempunyai sebarang bantuan kewangan pihak lain
	Commonwealth Scholarship And Fellowship Plan (CSFP)	Terbuka kepada pelajar antarabangsa dari negara-negara kومنwel yang ditawarkan melanjutkan pelajaran di Malaysia dan tidak mempunyai sebarang bantuan kewangan pihak lain
Khas	Bantuan OKU	Bantuan kewangan kepada pemegang kad OKU Jabatan Kebajikan yang sedang melanjutkan pengajian di IPTA
	Biasiswa Sukan	Biasiswa kepada atlet peringkat kebangsaan yang melanjutkan pendidikan di IPTA
	Bantuan Kewangan Pelajar Kolej Komuniti (BKPKK)	Bantuan kewangan pelajar di Kolej Komuniti(KK) di bawah KPT

Sumber Kementerian Pendidikan Malaysia, Pendidikan Tinggi

### 1.1.2 Pangkalan Data Sektor Pendidikan Tinggi

Sistem pangkalan data, juga dikenali sebagai sistem pengurusan pangkalan data (DBMS), terdiri daripada kumpulan data yang saling berkaitan, yang dikenali sebagai pangkalan data, dan satu set program perisian untuk mengurus dan mengakses data. Program perisian melibatkan mekanisme untuk menentukan struktur pangkalan data; untuk penyimpanan data; untuk akses data serentak, dikongsi atau diedarkan; dan untuk memastikan ketekalan dan keselamatan maklumat yang disimpan, walaupun terdapat kemalangan atau cubaan sistem pada akses yang tidak dibenarkan. Di KPT, sistem pangkalan data yang digunakan menggabungkan dua jenis pengurusan pangkalan data itu berlesan dan perisian terbuka. Antara kumpulan data yang ada di KPT adalah seperti data pelajar Pendidikan tinggi, data institusi Pendidikan tinggi, data permohonan

kemasukan pelajar ke IPT bagi lepasan SPM, STPM dan setaraf juga data pengurusan biasiswa anjuran KPT.

### **1.1.3 Analisis Data Menggunakan Pendekatan Perlombongan Data**

Bidang Analisis data sektor pendidikan menggunakan pendekatan perlombongan data telah mula wujud disebabkan peningkatan jumlah data yang semakin meningkat (Azuraliza, & Arshad 2013) (Paul Baepler & Cynthia, 2010). Pelbagai masalah yang wujud dalam sektor ini dilihat dapat diatasi menggunakan pendekatan ini seperti menganalisis keputusan akademik pelajar bagi menentukan kelayakan tawaran universiti, membantu pelajar dalam memilih kursus-kursus yang sesuai diambil pada setiap semester bagi mengelakkan kegagalan pelajar pada semester akan datang (Cristóbal Romero & Sebastián Ventura & Enrique García, 2007; C. Romero & S. Ventura, 2007), meramal keputusan sejumlah pelajar dengan menganalisis corak atau ciri-ciri pelajar yang berjaya dan gagal sedia ada (Chong Ho Yu, Samuel DiGangi & Angel Jannasch-Pennell & Charles Kaprolet, 2010; Ying Zhang & Samia Oussena & Tony Clark & Hyeonsook Kim, 2010 ) serta banyak lagi contoh-contoh yang ada dalam kajian yang lepas.

## **1.2 PERNYATAAN MASALAH**

Kajian awal dalam penentuan keputusan biasiswa pendidikan tinggi mendapati terdapat dua(2) kaedah digunakan dalam menentukan keputusan biasiswa iaitu kaedah manual dan pelaksanaan pendekatan perlombongan data melalui kaedah pengelasan tunggal. Kaedah manual didapati mempunyai dua masalah utama iaitu ketidaktentuan dan keputusan bias wujud kerana ianya bergantung kepada ahli jawatankuasa yang dilantik.

Selain daripada itu, kajian yang melibatkan penentuan keputusan biasiswa menggunakan pendekatan perlombongan data hanya menggunakan kaedah tradisional menggunakan kaedah pengelasan tunggal bagi memilih calon yang layak (Yosoa Putra Raharja, 2001; Khin Tuzar, 2014; J. K. Alhassan, 2015; Azuraliza Abu Bakar & Azlinda Arsyad, 2013). Kajian awal juga menunjukkan bahawa, kaedah pengelasan tunggal mempunyai dua kelemahan utama iaitu mengurangkan peluang meneroka pengetahuan

menggunakan teknik-teknik lain yang menyebabkan kehilangan maklumat-maklumat yang berkemungkinan penting bagi proses peramalan (Oza, & Tumer 2008). Selain risiko ini, peningkatan tahap kompleksiti data yang wujud juga menyebabkan pendekatan perlombongan data keadah pengelasan tunggal kurang sesuai digunakan dalam analisis ramalan keputusan biasiswa yang mempunyai bilangan atribut dan saiz data yang besar.

Berdasarkan isu-isu ini, pendekatan terbaik bagi penentuan keputusan biasiswa wajar dikaji dan ditambah baik. Ini adalah bagi memastikan keputusan yang dijana adalah lebih tepat dan cepat (Yosha, 2010).

Dalam kajian awal juga mendapati teknik pengelasan bergabung digunakan secara meluas bagi meningkatkan prestasi pengelasan model berbanding pengelasan tunggal termasuk pendekatan kaedah pengelasan bergabung petua. Melalui pendekatan teknik pengelasan bergabung, pelbagai teknik digunakan bagi mencari model pengelasan yang terbaik. Model pengelasan akhir yang dibina berupaya menggabungkan ciri-ciri daripada pengelasan-pengelasan tunggal yang mempunyai sama ada dari fungsi yang sama atau berbeza (Nikunj C. Oza, 2017). Ini dapat meningkatkan peluang menemukan pengetahuan baru yang sama penting dalam semua pengelasan tunggal yang diuji (Cao et al. 2015). Teknik pengelasan bergabung petua juga diuji bagi melihat perbandingan prestasi model pengelasan bergabung biasa dan model pengelasan bergabung menggunakan petua yang didapati daripada model pengelasan tunggal berasaskan petua. Teknik ini telah banyak digunakan dalam kajian pengelasan dalam meningkatkan keupayaan pengelasan tunggal berasaskan petua (D'Angelo, Rampone & Palmieri 2016; Zhu, Huang & Zhou 2018). Bagi mengkaji keupayaan model pengelasan bergabung dalam menentukan keputusan biasiswa pendidikan tinggi tiga(3) objektif utama kajian digariskan.

### **1.3 OBJEKTIF KAJIAN**

Kajian ini menggariskan tiga(3) objektif utama iaitu:

1. Merekabentuk satu rangka kerja pengelasan bergabung dalam menentukan keputusan biasiswa pendidikan tinggi;

2. Membangunkan model pengelasan tunggal dan bergabung penentuan keputusan biasiswa pendidikan tinggi; dan
3. Membangunkan model pengelasan petua bergabung untuk meningkatkan keupayaan mengelas keputusan permohonan biasiswa.

#### 1.4 SKOP KAJIAN

Kajian ini memfokuskan kepada cadangan model pengelasan bergabung penentuan keputusan biasiswa pendidikan tinggi yang sesuai dengan pendekatan perlombongan data berdasarkan rangka kerja pengelasan bergabung yang dibangunkan. Data penawaran biasiswa selama tiga tahun bermula 2013 sehingga 2016 digunakan bagi tujuan permodelan. Pensampelan data ini hanya meliputi program penawaran biasiswa di bawah Kementerian Pendidikan Tinggi (KPT) dan terhad yang dapat disesuaikan dengan kemudahan storan dan sumber infrastruktur yang ada.

Kajian ini merangkumi tiga uji kaji iaitu uji kaji pengelasan tunggal, uji kaji pengelasan bergabung dan uji kaji pengelasan petua bergabung yang melalui proses pengujian dan penilaian bagi melihat hasil perbandingan prestasi model. Uji kaji pengelasan tunggal merangkumi lima algoritma pengelasan iaitu J48, SVM, ANN, NB dan RT dengan teknik pensampelan data latihan dan data ujian menggunakan teknik pecahan strata dan k-fold. Uji kaji pengelasan bergabung pula merangkumi tiga kaedah pengelasan bergabung iaitu peningkatan, pembungkusan dan undian. Uji kaji pengelasan petua bergabung pula akan menggabungkan petua yang dijana oleh algoritma Apriori dengan pendekatan pengelasan peraturan persatuan dengan petua yang diperolehi oleh model pengelasan J48. Hasil keputusan setiap dinilai menggunakan parameter jadual kontigensi dan parameter lain seperti ketepatan, kejituan, dapatan semula dan F-measure.

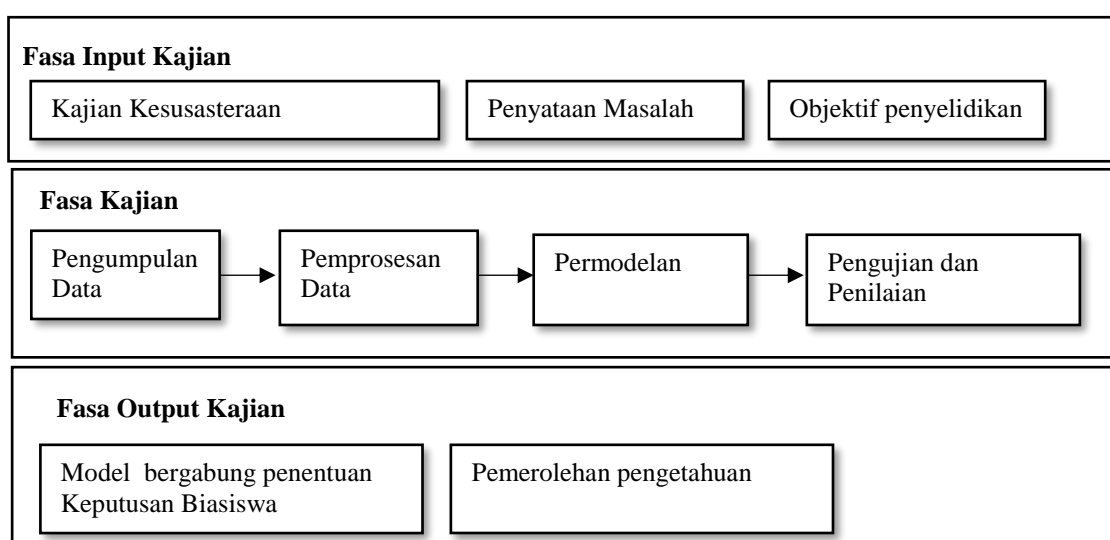
Proses analisis pengetahuan juga dilakukan merangkumi analisis teknikal dengan mengambil kira parameter batasan tertentu atau dengan pengetahuan pakar domain. Proses ini merangkumi mentafsir petua-petua yang dihasilkan oleh model tertentu bagi melihat kepentingan dan corak maklumat yang tersembunyi yang boleh digunakan pada masa hadapan.

## 1.5 METODOLOGI KAJIAN

Kajian ini dikendalikan dengan menggunakan pendekatan kuantitatif. Penggunaan pendekatan ini membolehkan kaedah kuantitatif digunakan bagi proses pembangunan dan penilaian model yang dibangunkan, disamping mengukur maklum balas penilaian pakar terhadap model yang dicadangkan.

Pendekatan kuantitatif dalam kajian perlombongan data telah digunakan secara meluas dalam menentukan prestasi dan keupayaan model pengelasan (Gutierrez et al. 2017; Kaiwen 2018; Yang et al. 2017). Proses pengumpulan dan pemilihan data dilakukan dengan menukar atribut-atribut kepada bentuk numerik bagi menambah kecekapan proses permodelan. Pengujian dan penilaian model juga dilakukan dengan kaedah kuantitatif dengan menggunakan jadual kontigensi dan parameter-parameter lain yang dikira daripada parameter jadual kontigensi sebagai metrik ukursan prestasi model.

Secara ringkas, bagi mencapai objektif yang ditentukan, rangka kerja permodelan pengelasan bergabung penentuan keputusan biasiswa pendidikan tinggi merangkumi tiga fasa utama iaitu fasa input kajian, fasa kajian dan fasa output kajian seperti ditunjukkan dalam Rajah 1-1.



Rajah 1-1 Metodologi Kajian



### **1.5.1 Fasa Input Kajian**

Fasa pertama kajian ini bermula dengan tinjauan literatur tentang kajian pendekatan perlombongan data dengan teknik pengelasan dan ramalan pelbagai domain. Ia melibatkan kajian berkenaan proses kerja yang terlibat dalam membangunkan model pengelasan, isu-isu dan cabaran, inisiatif dan pencapaian kajian lampau dan kajian terkini dalam bidang ini. Tinjauan ini dapat membantu merancang proses kerja kajian ini disamping memastikan model yang dibangunkan memenuhi kehendak semasa dan betul.

Berdasarkan kepada penemuan literatur tersebut, kajian ini diteruskan dengan definisi pernyataan masalah yang dikenalpasti dengan objektif kajian. Penerangan hasil tinjauan ini dihuraikan dalam Bab II kajian ini.

### **1.5.2 Fasa Kajian**

Fasa kajian pula merangkumi empat proses utama iaitu pengumpulan data, pemprosesan data, permodelan dan pengujian dan penilaian. Langkah-langkah dalam fasa ini dilakukan berdasarkan kajian literatur yang dilakukan dalam fasa input kajian.

Proses pengumpulan data merupakan proses awal permodelan. Data mentah yang di ambil dari pangkalan data berstruktur tawaran biasiswa pendidikan tinggi mulai tahun 2013 sehingga 2016 dikaji dan jenis atribut yang ada dikenalpasti. Data ini kemudian melalui proses saringan menggunakan pelbagai teknik seperti teknik saringan menapis kebisingan data, kelompongan data, data tidak konsisten dan menukar jenis atribut tersebut kepada bentuk yang tertentu bagi keperluan algoritma pengelasan dilakukan. Ini adalah bagi memastikan data yang lengkap dan sampel yang tepat digunakan dalam proses permodelan. Dalam peringkat ini, teknik pemilihan atribut juga dijalankan bagi memastikan atribut yang tidak penting dihapuskan bagi memastikan model yang dibina adalah terbaik (Rathore, & Gupta 2014).

Proses seterusnya dalam fasa ini adalah proses permodelan. Proses permodelan merupakan proses penting yang akan membina semua uji kaji yang dirangka dalam rangka kerja model pengelasan bergabung dalam kajian ini iaitu model pengelasan tunggal, model pengelasan bergabung dan model pengelasan petua bergabung penentuan keputusan biasiswa pendidikan tinggi. Bagi membuktikan hasil kajian yang lebih jitu, kajian ini akan melakukan uji kaji model pengelasan tunggal dengan menggunakan algoritma-algoritma pengelas yang ditentukan. Hasil uji kaji ini akan menjadi asas input kepada uji kaji pengelasan bergabung dengan teknik-teknik bergabung seperti peningkatan, pembungkusan dan undian. Setiap model yang dibina dalam uji kaji ini akan dinilai dan dibandingkan menggunakan metrik penilaian yang tertentu bagi mencari model pengelas keputusan biasiswa terbaik. Output model terbaik daripada uji kaji ini akan dianalisis, bagi mendapatkan corak pengetahuan yang tersembunyi, hubungan antara atribut-antribut dan petua-petua yang penting bagi membangunkan model pengelasan petua bergabung.

### **1.5.3 Fasa Output Kajian**

Fasa output kajian akan menentukan model akhir terbaik daripada semua uji kaji yang dijalankan semasa fasa permodelan. Dalam fasa ini, pengetahuan yang dihasilkan daripada proses permodelan ini akan dianalisis dan dibentangkan kepada pihak berkepentingan bagi mendapatkan maklum balas dan cadangan bagi tujuan penambahbaikan.

## **1.6 KEPENTINGAN KAJIAN**

Kajian ini mempunyai tiga kepentingan utama iaitu menyokong projek inisiatif Analitik Data Raya Sektor Awam(DSRA), menyumbang dalam kajian berkenaan pembelajaran mesin amnya dan membantu dalam proses pembuatan keputusan calon yang layak ditawarkan biasiswa pendidikan tinggi pada masa hadapan.

Analisis Data Raya Sektor Awam(DSRA) merupakan inisiatif yang diperkenalkan pada tahun 2014 sejajar dengan perkembangan data raya di seluruh

dunia. Kajian ini menggunakan kaedah analitik pintar menggunakan data sebenar tawaran biasiswa pendidikan tinggi negara yang kelak membantu dalam menentukan corak-corak calon yang pernah ditawarkan. Salah satu keunikan pendekatan analitik ini adalah menggunakan data sedia ada bagi mencari maklumat yang berpotensi bagi kegunaan pada masa hadapan terutama dalam proses pembuatan keputusan strategik Kerajaan.

Kajian ini juga menyumbang dari aspek kajian ilmiah dalam bidang pembelajaran mesin terutama kajian pengelasan bergabung. Kajian ini menghasilkan dua pendekatan teknik pengelasan bergabung bagi menentukan keputusan biasiswa pendidikan tinggi iaitu pengelasan bergabung menggunakan algoritma bergabung dan pengelasan petua bergabung menggunakan teknik majoriti mudah. Hasil kajian ini boleh menjadi rujukan bagi kajian akan datang dalam menambahbaik model pengelasan penentuan keputusan biasiswa pada masa hadapan khususnya dan data analitik amnya.

## **1.7 ORGANISASI TESIS**

Tesis ini merangkumi enam(6) bab iaitu:

Bab 1: Bab ini menrangkan secara ringkas berkenaan latar belakang kajian, objektif kajian, pernyataan masalah, skop kajian dan kepentingan kajian.

Bab 2 : Bab ini akan menerangkan secara lanjut berkenaan kajian kesusasteraan yang dijalankan bagi memastikan keperluan maklumat adalah cukup dan jelas. Kajian terdahulu berkaitan metodologi dan hasil kajian yang berkaitan kajian diterangkan dalam bab ini.

Bab 3 : Bab ini akan menerangkan secara visual aliran kerja keseluruhan kajian merangkumi fasa input, fasa kajian dan fasa output.

Bab 4: Model pengelasan tunggal dan bergabung akan menerangkan hasil ujikaji pengelasan tunggal dan pengelasan bergabung. Prestasi model-model yang dibangunkan akan dibandingkan antara satu sama lain bagi menunjukkan model pengelasan terbaik sebagai input uji kaji seterusnya.

Bab 5: Analisis pengetahuan, model peraturan persatuan dan model petua bergabung akan menerangkan contoh hasil pemerolehan maklumat daripada output yang dihasilkan oleh model pengelasan terbaik. Di samping itu, bab ini juga akan menerangkan hasil uji kaji permodelan peraturan persatuan dan pengelasan bergabung petua bagi melihat perbandingan teknik pengelasan bergabung terbaik.

Keseluruhan rumusan kajian seterusnya akan diterangkan dalam Bab 6: merangkumi batasan kajian, sumbangan kajian dan cadangan kajian akan datang.

## **BAB II**

### **KAJIAN KESUSASTERAAN**

#### **2.1 PENGENALAN**

Bab ini membincangkan berkenaan kajian literatur berkenaan kajian pendekatan perlombongan data menggunakan teknik pengelasan dan ramalan dalam pelbagai domain terutama domain pendidikan. Ini bagi melihat keupayaan teknik pengelasan tunggal dan bergabung yang telah banyak digunakan dalam kajian sedia ada dalam dunia sebenar.

Menyedari tentang kelemahan-kelemahan model pengelasan tunggal dalam mengelas dan meramal, bab ini seterusnya menumpu kepada perbincangan tentang keupayaan pendekatan perlombongan data menggunakan teknik pengelasan bergabung dalam pelbagai domain bagi melihat signifikan kajian-kajian tersebut dengan kajian ini.

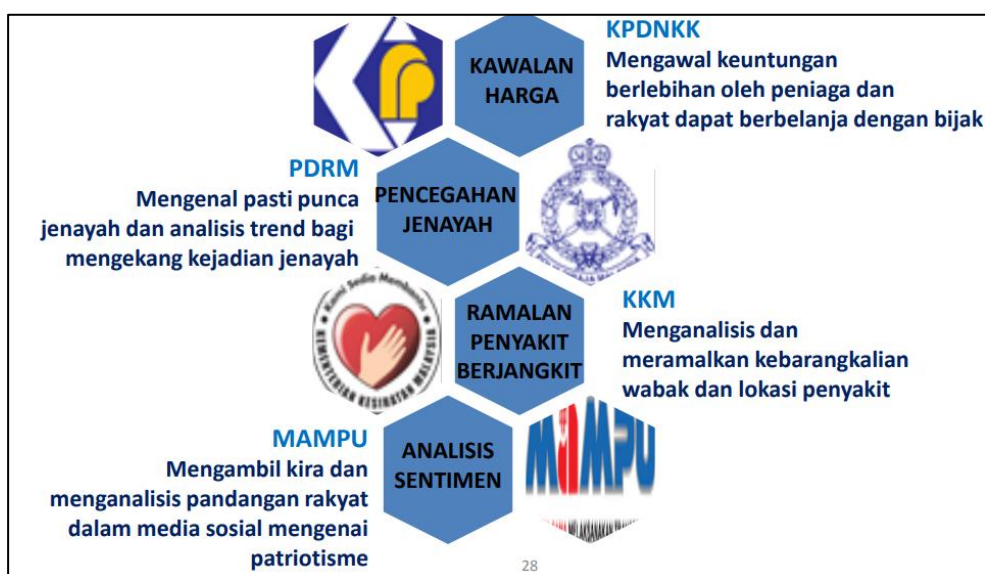
Berdasarkan kepada literatur yang ditinjau, bab ini mengemukakan rangka kerja pengelasan dan ramalan, algoritma pengelasan dan ramalan popular yang digunakan dalam kajian, dan analisa pengetahuan hasil daripada model pengelasan dan ramalan.

#### **2.2 ANALITIK DATA RAYA SEKTOR AWAM(DRSA)**

Analitik Data Raya Sektor Awam (DSRA) merupakan inisiatif Kerajaan dalam menambah baik kualiti perkhidmatan Kerajaan kepada rakyat. Inisiatif yang diperkenal pada tahun 2014 adalah sangat berkait rapat dengan perkembangan Data Raya di semua industri. Menurut Gartner, data raya adalah aset maklumat yang bersaiz besar (high-volume), perubahan pantas (high-velocity) dan pelbagai (high-variety) yang memerlukan kos yang berkesan, pemprosesan maklumat yang berinovasi bagi meningkatkan pengetahuan dan pemahaman (insight) dan membuat keputusan. Dalam

sektor awam, usaha melaksanakan inisiatif DSRA merupakan usaha yang baik dalam memastikan data yang disimpan mempunyai nilai dan berupaya menambah nilai produktiviti sektor awam negara.

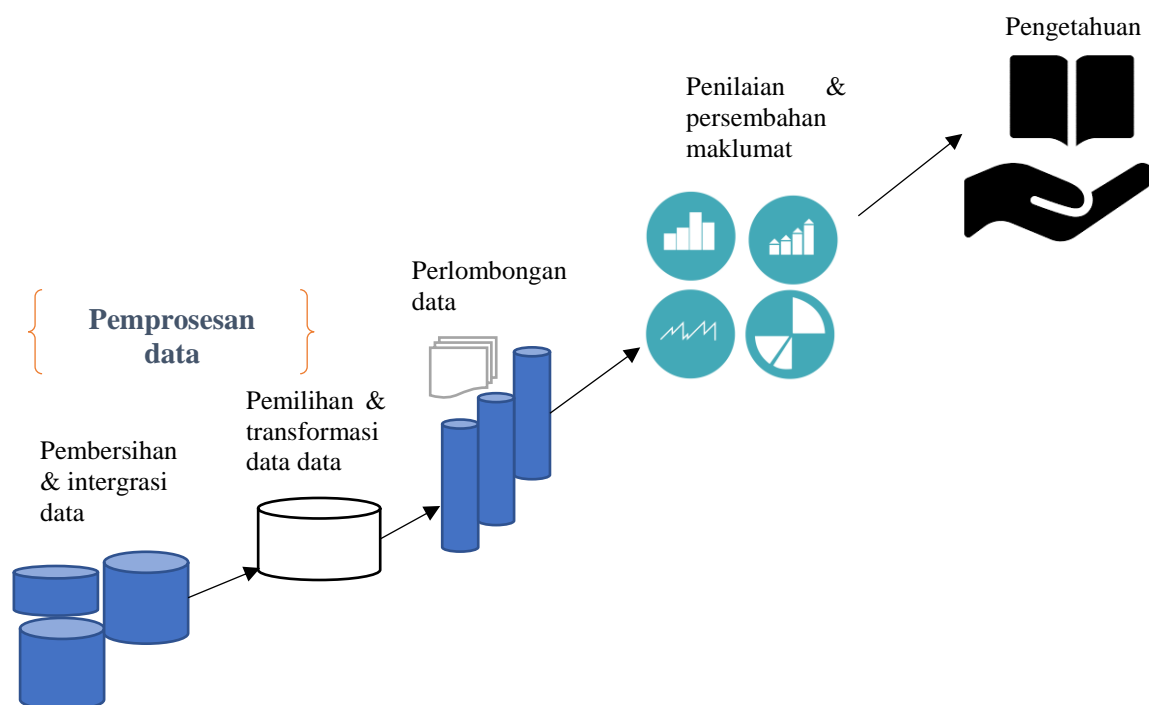
DSRA adalah proses bagi menilai data yang bersaiz besar dan pelbagai jenis untuk menemui corak tersembunyi, korelasi yang tidak diketahui dan maklumat berguna yang lain berkaitan data sektor awam. Inisiatif ini telah diperkenalkan sejak tahun 2014 dan mempunyai lima tujuan utama iaitu bagi menyokong proses pembuatan keputusan dalam pelbagai domain pengurusan sektor awam, membantu perancangan strategi yang lebih baik, mewujudkan hubungan lebih baik dengan pelanggan, pengesanan risiko, memantau prestasi kewangan Kerajaan dan lain-lain. Dalam rangka kerja analitik DSRA, terdapat tujuh kluster utama iaitu sosio ekonomi infra asas luar bandar, jenayah, rasuah, pendidikan, pengangkutan dan penjagaan kesihatan. Kesemua tujuh kluster ini merupakan sasaran utama analitik DSRA yang berkesan dan hasil tinjauan mendapati antara faedah yang didapati daripada inisiatif ini adalah ramalan wabak penyakit, ramalan pencegahan jenayah, maklumat pintar kesesakan jalan raya, ramalan cuaca atau bencana, keselamatan cyber pertahanan negara, dan banyak lagi. Dalam projek perintis inisiatif ini terdapat beberapa agensi Kerajaan yang dipertanggungjawabkan melaksanakan projek analitik DSRA dengan tujuan berbeza seperti Rajah 2-1.



Rajah 2-1 Agensi-agensi terlibat dalam projek perintis inisiatif DSRA

## 2.3 PERLOMBONGAN DATA

Perlombongan data merupakan pendekatan yang menggunakan pelbagai teknik dan algoritma bagi mencari corak tersembunyi, hubungkait antara atribut, dan pemerolehan pengetahuan yang mungkin berguna berdasarkan set data sedia ada. Rajah 2-2 menunjukkan secara visual langkah-langkah ini yang mana output setiap proses menjadi input bagi proses yang lain. Ini bermakna setiap proses harus dilakukan dengan sistematik dan tepat agar output di peringkat akhir iaitu maklumat dan pengetahuan yang diperoleh adalah maklumat yang berguna serta betul.



Rajah 2-2 Langkah-langkah dalam data analitik menggunakan pendekatan perlombongan data

### 2.3.1 Data Analitik Sektor Pendidikan

Kajian pendekatan perlombongan data dalam sektor pendidikan telah berkembang dengan meluas. Kajian-kajian ini banyak menyumbang dalam memberi solusi bagi menyelesaikan masalah-masalah berkaitan sektor pendidikan di sesebuah negara menggunakan kaedah pengelasan dan ramalan (Dutt, Ismail & Herawan 2017). Pemerolehan maklumat berguna melalui kaedah ini membantu dalam proses pembuatan

keputusan di masa hadapan. Pertumbuhan pesat dalam saiz data pendidikan juga menunjukkan bahawa model pengelasan dan ramalan yang ingin digunakan dalam proses pemerolehan maklumat ini, memerlukan set algoritma yang lebih efisien dalam memberi keputusan yang terbaik (Dutt et al. 2017).

### 2.3.2 Pengelasan dan Ramalan Sektor Pendidikan

Berdasarkan kajian literatur, kajian pelaksanaan pendekatan kaedah pengelasan dan ramalan dalam sektor pendidikan semakin berkembang dan meluas dengan pelbagai kaedah telah dicadangkan. Penerokaan penyelidik dalam menggunakan pendekatan ini tidak terbatas pelaksanaannya dan sehingga 2017, banyak kajian-kajian baru diterbitkan. Dua elemen utama dalam sektor pendidikan yang sering difokus adalah pengelasan dan ramalan bagi permasalahan kumpulan pelajar dan pengelasan dan ramalan berkaitan permasalahan institusi pendidikan itu sendiri (Adejo, & Connolly 2017; Baepler, & Murdoch 2010). Ini termasuklah kajian membangunkan model pengelasan dan ramalan dalam menentukan prestasi akademik pelajar berdasarkan rekod pencapaian pelajar sedia ada (Adejo, & Connolly 2017; Mesarić, & Šebalj 2016; Shahiri, Husain & Rashid 2015), menentukan senarai kursus yang sesuai diajar di organisasi akademik (Mesarić, & Šebalj 2016) dan menentukan keputusan pengambilan pelajar yang sesuai dengan program bagi memastikan pencapaian kelulusan yang baik bagi menjaga prestasi graduan organisasi akademik secara keseluruhan (Yu et al. 2010).

Pelbagai algoritma turut dikaji bertujuan melihat perbandingan keupayaan model yang digunakan seperti di Jadual 2-1.

Jadual 2-1 Algoritma-algoritma dalam kajian pengelasan dan ramalan sektor pendidikan

<b>Penulis, Tahun</b>	<b>Algoritma</b>
(Shahiri et al. 2015)	J48, SVM, ANN
(Mesarić, & Šebalj 2016)	J48, C4.5, RT, Hutan Rawak
(Ong, & Chandrasekaran 2015)	NB
(Yu et al. 2010)	Pokok Keputusan
(Daud et al. 2017)	NB, SVM, C4.5
(Mittal, Aggarwal & Mahajan 2017)	Apriori
(Maniar 2017)	Apriori Algorithm And Linear Regression
(Anuradha, & Velmurugan 2015)	J48, Naïve Bayes, KNearest Neighbor



Secara keseluruhan, perkembangan kajian permodelan pengelasan dan ramalan dalam sektor pendidikan sewajarnya digiat dalam memastikan kesinambungan sumbangan teknologi dalam bidang kritikal iaitu pendidikan dalam sesebuah negara.

### **2.3.3 Model Ramalan Pengelasan Penentuan Keputusan Biasiswa**

Selain daripada bidang masalah di atas, kajian literatur kajian ini juga menfokuskan dalam mengkaji kajian-kajian sedia ada dalam membangun pengelasan dan ramalan dalam proses penganugerahan biasiswa pendidikan. Ini kerana penganugerahan biasiswa juga merupakan salah satu bidang kajian yang penting dalam sektor pendidikan. Biasiswa didefinisikan sebagai wang bantuan yang diberikan kepada pelajar atau kepada mahasiswa untuk membantu menanggung keperluan yuran pelajaran dan keperluan asasi mereka sepanjang tempoh belajar.

Melalui tinjauan literatur, kajian-kajian pengelasan dan ramalan ini merangkumi kepada kajian pengelasan tunggal atau pengelasan tunggal yang mengkaji algoritma-algoritma yang sesuai digunakan dengan menggunakan set data penawaran biasiswa terdahulu. Melalui kajian ini, pemilihan ciri yang pelbagai bergantung kepada set data yang disimpan oleh pihak pemberi biasiswa. Bagi pihak pemberi biasiswa ini, mekanisme sistematik diperlukan bagi tujuan ini kerana perkembangan saiz data permohonan dan sedia ada yang besar serta proses penganugerahan perlu dilakukan dalam tempoh yang singkat (Azuraliza, & Arshad 2013; Raharja 2014).

Menurut kajian (Raharja 2014) pemberian biasiswa kepada calon yang tidak tepat akan menyebabkan masalah kekurangan dana bagi pembiayaan pendidikan tinggi dikalangan pelajar yang kurang berkemampuan seterusnya menutup peluang mereka untuk melanjutkan pengajian ke tahap lebih tinggi. Ini juga dilihat akan menjejaskan matlamat utama pembiayaan sektor pendidikan tinggi iaitu bagi melahirkan kepakaran dalam bidang sains dan teknologi di masa hadapan.

Kos pendidikan yang semakin membebankan juga merupakan faktor peningkatan kepada keperluan biasiswa pendidikan dalam sesebuah negara. Ini turut memberi kesan kepada golongan ibu bapa yang berppendapatan rendah bagi menyediakan dana pendidikan anak-anak mereka. Oleh itu, adalah sangat penting bagi

memastikan pemberian biasiswa diberi kepada pelajar-pelajar secara teratur dan sistematik (Alhassan, & Lawal 2015; Azuraliza, & Arshad 2013; Raharja 2014).

Perlombongan data merupakan satu pendekatan di dalam proses penemuan pengetahuan (Yang et al. 2017) yang terdiri daripada pelbagai fungsi dengan teknik-teknik berbeza. Dalam mengelas dan meramal keputusan biasiswa, beberapa kajian telah menggunakan pendekatan ini dengan membangunkan model-model melalui algoritma pengelasan tunggal yang popular seperti algoritma pepohon keputusan (Alhassan, & Lawal 2015; Raharja 2014), KNN (Tun & Aye 2014) dan Set Kasar (Azuraliza, & Arshad 2013). Kajian-kajian ini telah membuktikan bahawa teknik-teknik ini adalah sesuai digunakan dalam menyokong proses ramalan keputusan biasiswa. Walaubagaimanapun, kajian-kajian ini hanya terhad mengkaji pelaksanaan pendekatan perlombongan data menggunakan teknik pengelas tunggal yang mempunyai banyak kelemahan dan kekangan.

Antara algoritma biasa digunakan dalam model pengelasan tunggal adalah seperti berikut:

**a. Pokok keputusan**

Pokok keputusan adalah serasi dengan kedua-dua jenis data sama ada numerik atau nominal. Selain sesuai dengan apa pun saiz data, ketepatan yang lebih tinggi dalam teknik klasifikasi pokok keputusan menggambarkan bahawa teknik itu berjaya mengelas dengan baik. Pokok keputusan dapat menangani dan mengendalikan data input kuantiti yang banyak seperti teks dengan data berangka hanya teks atau nominal. Ia adalah pendekatan pembelajaran yang diselia yang mempunyai keupayaan mengekstrak maklumat dari sejumlah besar data berdasarkan peraturan atau petua keputusan. Dalam kajian ini, dua jenis algoritma asas di bawah kategori pokok keputusan digunakan iaitu J48 dan RT.

**i. J48**

Algoritma J48 dibangunkan oleh Ross Quinlan pada tahun 1993. Ia adalah pengelas berbentuk pokok yang digunakan untuk mewakili pengelasan model dengan

hubungkait atribut dalam sesuatu set data. Peraturan rekod yang dikelaskan dengan betul akan menentukan sama ada model yang dijana berprestasi baik atau tidak. Peraturan dalam bentuk *IF-THEN* boleh diekstrak daripada algoritma ini. Setiap laluan dari akar ke nod daun boleh ditulis sebagai satu peraturan. Peraturan yang terdahulu (bahagian *IF*) dibentuk dengan menggabungkan kriteria pemisahan di sepanjang laluan yang diberikan menggunakan sambungan *AND*. Nod daun yang mengandungi ramalan kelas membentuk peraturan yang berlaku (bahagian *THEN*). Langkah-langkah asas dalam algoritma ini adalah seperti di bawah (G. Kaur, & Chhabra 2014):

Langkah 1 : Nilai atribut sebagai kelas label ditetapkan;

Langkah 2 : Maklumat tahap kepentingan atribut dikira untuk setiap atribut, dengan ujian pada atribut. Jika atribut mencapai nilai yang tinggi dalam maklumat yang dikira, atribut tersebut akan dijadikan nod akar;

Langkah 3 : Kemudian atribut terbaik ditemui berdasarkan kriteria pemilihan semasa dan atribut yang dipilih;

Langkah 4: Sekiranya kes tergolong dalam kelas yang sama pokok mewakili daun supaya daun dikembalikan oleh pelabelan dengan kelas yang sama.

## ii. RT

RT adalah salah satu pendekatan pokok di mana atribut 'K' dipilih secara rawak untuk mengklasifikasikan data. Ia tidak mengandungi sebarang teknik pemangkasan untuk mengurangkan ralat. Algoritma pokok rawak mempunyai pilihan untuk menganggarkan kebarangkalian kelas untuk klasifikasi. Langkah asas pokok rawak adalah seperti berikut:

Langkah 1: Muatkan data latihan pada akarnya;

Langkah 2 : Data input dibahagikan berdasarkan sifat 'K' secara rawak;

Langkah 3: Membina pokok keputusan dengan perpecahan secara rawak

Langkah 4: Algoritma berjalan secara rekursif pada nod bukan daun sehingga semua data diklasifikasikan

## b. SVM

SVM adalah teknik pembelajaran mesin yang diperkenalkan oleh Boser, Guyon & Vapnik dan telah digunakan secara meluas dalam pelbagai masalah pengelasan sehingga kini (Statnikov et al. 2005). SVM secara amnya boleh dilihat sebagai satu satah hiper yang memisahkan objek (mata) kepunyaan kelas (objek positif) daripada yang tidak termasuk kelas itu (objek negatif). Pemisahan ini dilaksanakan oleh algoritma SVM semasa langkah pembelajaran di mana pelan hiper diperoleh untuk membahagikan objek positif dan negatif dengan margin maksimum. Margin menunjukkan ruang dari ruang hyper ke objek positif dan negatif terdekat. Contoh algoritma SVM adalah seperti di bawah (Nurfarahin, Shaikh & Shapi 2017):

Langkah 1:

- Tentukan L, merupakan pengelas
- sampel X
- Kelas label y (atribut kelas label)

Langkah 2:

- Bagi setiap k dalam sampel X;
- Bina label vektor yang baru, z dimana  $z_i=1$  jika  $y_i=k$  dan  $z_i = 0$  atau
- Gunakan L kepada X,z untuk mendapatkan  $f_k$

## c. Naive Bayes (NB)

Selain itu, algoritma asas lain yang biasa digunakan dalam uji kaji model pengelasan adalah *Naive Bayes* (NB). NB adalah teknik klasifikasi berdasarkan teori kebarangkalian sesuatu data atau atribut dengan kekerapan data atau atribut yang sama dalam satu set data menggunakan teori Bayes (Fatima, & Pasha 2017). Model NB mudah digunakan untuk set data yang sangat besar. Secara ringkas, NB mengandaikan bahawa nilai ciri tertentu tidak berkaitan dengan kehadiran atau ketiadaan sebarang ciri lain, yang diberikan dalam pembolehubah kelas. Langkah-langkah algoritma NB adalah seperti berikut:

Langkah 1: Ekstrak, bersih dan klasifikasi label keputusan.

Langkah 2 : Keluarkan tanda baca yang besar dan berpecahnya.

Langkah 3 : Menghitung Token dan mengira kebarangkalian. Kebarangkalian ini dipanggil sebagai kebarangkalian posterior yang dikira oleh formula yang dijelaskan dalam.

Langkah 4: Menambah kebarangkalian dan kemudian membungkus.

#### **d. ANN**

ANN ialah model klasifikasi yang diwakili oleh nod yang saling berkaitan. Ia boleh dilihat sebagai nod bulat yang diwakili sebagai neuron buatan yang mewakili output satu neuron kepada input yang lain. Model ANN berfungsi dalam menunjukkan hubungan tersembunyi di dalam data sejarah, dengan itu memudahkan pengelasan dan ramalan data baru. Model ANN cukup tepat untuk membuat keputusan yang jelas dan relevan mengenai corak data sedia ada. Walaubagaimanapun, algoritma ini mengambil masa lebih lama berbanding algoritma pengelasan yang lain memandangkan ia membina nod-nod yang saling berkaitan.

Contoh-contoh kajian yang menggunakan kelima-lima algoritma ini adalah seperti di Jadual 2-2.

Jadual 2-2 Contoh-contoh kajian menggunakan lima algoritma pengelasan tunggal J48, RT, NB, SVM dan ANN

<b>Nama Penulis</b>	<b>Objektif</b>	<b>Tahun</b>	<b>Teknik</b>	<b>Dapatan</b>	<b>kekangan</b>
(Hamsagayathri, & Sampath 2017)	Menjalankan kajian perbandingan pengelasan pesakit kanser payu dara menggunakan empat algoritma pepohon keputusan	2017	J48, RT, <i>REPTree</i>	Kajian mendapati algoritma REPTree merupakan algoritma terbaik bagi pengelasan set data pesakit kanser payudara dengan ketepatan paling tinggi berbanding tiga algoritma pepohon keputusan yang lain.	Kajian meliputi hanya algoritma di bawah kategori Pepohon keputusan
(Ayub, & Karnalim 2017)	Menilai hubungan atribut terpenting bagi menentukan keputusan pelajar menggunakan teknik pengelasan algoritma J48	2017	J48	Menggunakan Teknik pengelasan J48 mendapati subjek matematik sangat mempengaruhi keputusan pelajar bagi kedua-kedua jenis subjek.	gabungan faktor untuk meramal kejayaan atau kegagalan boleh meningkatkan prestasi pengelasan
(Parthiban 2012)	Diagnostik penyakit menggunakan pendekatan pengelasan pembelajaran mesin	2012	SVM, NB	Kajian mendapati algoritma SVM berupaya mengelas dengan baik penyakit berkaitan hati berbanding algoritma NB.	Kajian hanya meliputi set data pesakit penyakit hati sahaja. Kajian lebih lanjut menggunakan pendekatan yang sama ke atas set data pesakit penyakit lain adalah dicadangkan.
(Deo, & Şahin 2017)	Peramalan radiasi solar global jangka panjang dengan algoritma ANN ditambah dengan suhu permukaan tanah yang diperolehi satelit (MODIS) (LST) untuk lokasi serantau di Queensland	2017	ANN	Model bagi pengelasan dan ramalan suhu permukaan tanah dibangunkan. Model ini dinilai dengan menggunakan algoritma sesuai bagi melihat prestasi model. Kajian membuktikan bahawa model ramalan yang baik bagi menentukan suhu permukaan tanah dengan atribut tertentu boleh dibangunkan dengan menggunakan algoritma ANN.	Kajian lanjut berkenaan tambahan atribut baru boleh dilakukan dalam mengelas dan meramal suhu permukaan tanah pada masa hadapan menggunakan teknik yang sama.

bersambung...

---

sambungan...					
(Afram et al. 2017)	Kawalan Prediktif Model (MPC) berasaskan ANN dan Pengoptimuman Sistem HVAC:	2017	ANN	Model ANN yang ditambah baik bagi mengurangkan pengurangan elektrik dan meningkatkan keselamatan kediaman dibangunkan.	Kajian lanjut dengan menambah atribut baru menggunakan kaedah yang sama dicadangkan
(Roy, & Garg 2017)	Menganalisa Prestasi Pelajar Dengan Menggunakan Teknik Perlombongan Data	2017	ANN,SVM,KNN	Kajian semula permodelan pengelasan menggunakan algoritma-algoritma berkaitan dalam melihat prestasi pelajar	Kajian perlombongan data memantau prestasi pelajar menggunakan teknik lain dicadangkan pada masa hadapan

---

Selain itu, hasil kajian literatur menunjukkan bahawa setiap algoritma yang digunakan mempunyai kelebihan dan kelemahan tersendiri. Jadual 2-3 menerangkan secara ringkas kelebihan dan kelemahan lima algoritma pengelasan tunggal yang biasa digunakan (Deo, & Şahin 2017; Fatima, & Pasha 2017; R. Kaur, & Gangwar 2017).

Jadual 2-3 Perbandingan algoritma pengelasan tunggal

<b>Algoritma</b>	<b>Kelebihan</b>	<b>Kekurangan</b>
J48	Mudah digunakan; Output dalam bentuk peraturan petua mudah ditafsir.	Kadar kesilapan klasifikasi yang tinggi semasa latihan ditetapkan kecil berbanding dengan bilangan kelas
SVM	Algoritma SVM mempunyai ketepatan yang tinggi jika dataset kecil	Kurang berkesan bagi set data dengan tiada nilai, Tidak menyokong pembolehubah bercampur
NB	Ia meningkatkan prestasi klasifikasi dengan menghapuskan ciri-ciri yang tidak berkaitan, ia memerlukan kurang masa pengiraan, mudah dilaksanakan, meramalkan hasil yang tepat untuk kebanyakan masalah pengelasan dan ramalan	Algoritma ini memerlukan saiz data yang besar untuk mencapai hasil yang baik. Kurang berkesan kerana mereka menyimpan keseluruhan contoh latihan.
RT	Mudah difahami, proses pengelasan dan ramalan adalah cepat	Pengiraan adalah rumit terutamanya jika nilai tidak pasti atau jika beberapa hasil dihubungkan
ANN	Ia sangat mudah digunakan, ia berfungsi dengan data yang tidak lengkap. Ia belajar dengan cepat dan pelaksanaannya agak mudah	Ia memerlukan masa pemprosesan yang tinggi jika rangkaian saraf besar

#### **2.3.4 Kelemahan Model Pengelasan Tunggal**

Seterusnya, kajian literatur ini akan melihat kepada kelemahan-kelemahan model pengelasan tunggal bagi memastikan kajian model pengelasan bergabung dalam kajian ini dibuat dengan pemilihan teknik terbaik. Teknik pengelasan tunggal mempunyai dua kelemahan utama iaitu tidak menyediakan penyelesaian terbaik dalam semua kes atau kelas data dan penggunaan hanya satu pengelas akan menyebabkan maklumat berpotensi pada pengelas lain hilang (Adejo & Connolly 2017; Tun & Aye 2014). Risiko-risiko ini wujud kerana, pengelas lain yang mungkin mempunyai pengetahuan lain yang mungkin sama penting tidak diambilkira dalam pengelas yang terpilih.



Selain itu, bagi data yang kompleks dan mempunyai bilangan atribut yang banyak, pengelas tunggal juga dikatakan tidak sesuai digunakan kerana pengelas tunggal tidak mempunyai fungsi nilai purata dan hanya menumpukan kepada satu-satu algoritma sahaja (Adejo, & Connolly 2017). Berbanding kaedah intergrasi atau gabungan, algoritma-algoritma ini masing-masing mempunyai kekuatan dan kelemahan tersendiri yang boleh saling melengkapi dalam membangunkan satu model tunggal yang lebih baik.

## **2.4 MODEL PENGELASAN BERGABUNG**

### **2.4.1 Pengenalan**

Teknik menggabungkan model-model dari pelbagai algoritma dipanggil pengelasan bergabung yang mana merupakan solusi bagi mengatasi kelemahan-kelemahan pengelas tunggal ini. Banyak kajian dilihat dapat mengatasi kelemahan-kelemahan ini dengan melaksanakan teknik pengelasan bergabung yang menghasilkan model akhir yang lebih teguh (Dietterich 2000; Oza & Tumer 2008; Rokach 2010; Stapel et al. 2016) .

Ini kerana melalui pendekatan teknik pengelasan bergabung, model pengelas akhir yang dibina berupaya menggabungkan ciri-ciri daripada pengelas-pengelas tunggal yang mempunyai samaada dari fungsi yang sama atau berbeza (Nikunj C. Oza, 2017) dan memberi keputusan lebih baik berbanding pengelas tunggal (Nikunj C. Oza & Kagan Tumer, 2017; Martin Stapel & Zhilin Zheng & Niels Pinkwart, 2006; Lior Rokach, 2010; Thomas G. Dietterich, 2000).

Teknik menggabungkan ramalan dari pelbagai model dengan dua matlamat utama iaitu: untuk meningkatkan ketepatan ramalan keseluruhan berbanding dengan pengelas tunggal dan meningkatkan kadar generalisasi yang lebih baik kerana pengukuran khusus cirinya yang berbeza. Hasilnya, pengelasan bergabung boleh mengatasi masalah yang tidak dapat diatasi dengan model ramalan tunggal.(Stapel, Zheng & Pinkwart 2016)

Teknik-teknik yang digunakan dalam kajian pengelasan bergabung adalah teknik Peningkatan, teknik Pembungkusan dan teknik undian.

#### 2.4.2 Teknik Peningkatan

Peningkatan merupakan teknik asas yang digunakan dalam pengelasan bergabung bagi meningkatkan keupayaan sesebuah model pengelasan. Teknik ini memfokus kepada pengelas-pengelas lemah dengan menguji semula item yang dikelaskan dengan tidak tepat sehingga dikelaskan dengan tepat pada lelaran berikutnya. Ia diperkenalkan oleh Schapire pada tahun 1990 dan telah digunakan secara meluas sehingga kini. Proses pengelasan bergabung teknik ini sama seperti teknik Pembungkusan iaitu dengan melakukan pensampelan data pada setiap lelaran dengan menggunakan teknik undian majoriti. Salah satu algoritma yang biasa digunakan teknik ini adalah *AdaBoost*.

Algoritma AdaBoost diperkenalkan oleh Freund and Schapire pada tahun 1997 (Deng 2007). Ia menjana satu set hipotesis, dan menggabungkan mereka melalui pengundian majoriti selari dengan kelas yang dicapai oleh setiap hipotesis. Hipotesis dihasilkan dengan melatih pengelas yang lemah, dengan menggunakan rekod-rekod atau contoh yang diperoleh semasa proses lelaran menggunakan data latihan secara teratur. Kitaran ini memastikan pengelasan rekod yang salah dikelaskan oleh pengelas sebelum ini lebih cenderung dimasukkan dalam data latihan lelaran seterusnya. Oleh itu, data latihan yang dibina semakin menjadi data yang sukar untuk dikelaskan tetapi kebarangkalian berjaya dikelaskan dengan tepat adalah tinggi bergantung kepada corak pengelasan sebelum. Ia juga berupaya menambahbaik ketepatan pengelasan dan biasanya digunakan dalam banyak bidang. Ia adalah mudah untuk dilaksanakan dan tidak terdedah kepada masalah *over-fitting*. (S. Palaniappan&T.V. Rajinikanth&A. Govardhan, 2017). Ia juga menghasilkan urutan model asas dengan pengagihan berat yang berbeza di atas set latihan (Nikunj C. Oza a,\*, Kagan Tumer A, 2007). Kelebihan dan kelemahan algoritma ini boleh dilihat di Jadual 2-4 (Palaniappan, Rajinikanth & Govardhan 2017).

Jadual 2-4 Kelebihan dan kelemahan Algoritma Adaboost

Kelebihan	Kelemahan
<ul style="list-style-type: none"> <li>▪ Sangat mudah dilaksanakan</li> <li>▪ Proses pemilihan ciri menghasilkan pengelasan mudah yang relatif</li> <li>▪ Generalisasi yang baik</li> </ul>	<ul style="list-style-type: none"> <li>Penyelesaian suboptimal</li> <li>Sensitif kepada data yang bising dan <i>outlier</i></li> </ul>

### 2.4.3 Teknik Pembungkusan

Selain teknik Peningkatan, teknik Pembungkusan merupakan teknik mudah yang digunakan dalam pengelasan bergabung. Ia boleh digabungkan dengan algoritma pengelasan yang lain seperti J48, SVM, ANN dan NB. Mengikut kajian (Goyal, Thakur & Chowdhury 2016) algoritma J48 tunggal dan J48(Pembungkusan) akan menghasilkan output pohon yang berbeza. Pohon yang berbeza terbentuk apabila terdapat perubahan dalam kedudukan titik mula data latihan yang menyebabkan penurunan kestabilan. Ia boleh digunakan untuk mana-mana kaedah pengelasan lain dalam pendekatan perlombongan data. Teknik ini juga sesuai digunakan dalam mencari model yang sesuai untuk data bersaiz besar kerana pengelasan menjadi lebih mudah kerana set data latihan dipecah mengikut teknik tertentu (Deng 2007). Output teknik ini adalah output akhir dengan teknik undian majoriti yang memilih hasil pengelasan terbaik di kalangan model yang baik.

### 2.4.4 Teknik Undian

Teknik undian juga merupakan teknik pengelasan bergabung yang popular (Ahmed et al. 2017; Hamsagayathri, & Sampath 2017; Pham, Tien Bui & Prakash 2017). Ia menggabungkan banyak pengelasan dan melakukan proses undian semasa pengelasan. Banyak kaedah digunakan bagi memilih undi pengelasan terbaik di kalangan pengelasan yang ada. Teknik purata, minimum dan maksimum adalah antara yang digunakan. Pengelasan menggunakan teknik majoriti sering dipilih kerana ia menghasilkan output yang model terbaik yang adil dan seimbang. Sebagai contoh, satu set data yang mempunyai 2 kelas label dengan 3 pengelasan terbaik akan menentukan majoriti pengelasan yang memberi label yang paling banyak.

### 2.4.5 Kelebihan dan Kelemahan Pengelasan Bergabung

Ketiga-tiga teknik pengelasan bergabung ini mempunyai kelebihan dan kelemahan masing-masing yang akan mempengaruhi hasil model yang dibangunkan. Kelebihan dan kelemahan ini boleh di lihat di Jadual 2-5 di bawah (Choi et al. 2016; Deng 2007; Goyal et al. 2016; Lewandowski, Co-investigator & Lewandowski 2015; Polikar 2006; Tibshirani 2013).

Jadual 2-5 Perbandingan kelebihan dan kelemahan teknik pengelasan bergabung

Teknik	Kelebihan	Kelemahan
Peningkatan	<ul style="list-style-type: none"> <li>▪ Meningkatkan keupayaan model yang lemah dengan secara berulang melakukan latihan menggunakan sampel data yang berbeza.</li> <li>▪ Menumpukan kepada contoh-contoh yang dikelaskan tidak tepat pada lelaran model seterusnya menghasilkan ketepatan model yang lebih tinggi</li> </ul>	<ul style="list-style-type: none"> <li>▪ Masa permodelan yang lebih panjang dan kos tinggi</li> <li>▪ Model pengelasan menjadi lebih kompleks dan sukar dilaksanakan menggunakan platform masa nyata</li> <li>▪ Kurang sesuai digunakan jika data mempunyai kadar kebisingan dan outliers yang tinggi</li> </ul>
Pembungkusan	<ul style="list-style-type: none"> <li>▪ meningkatkan ketepatan dan mengurangkan varians, dengan itu mengurangkan masalah <i>overfitting</i></li> <li>▪ Mengambil sampel contoh data berbeza bagi meningkatkan ketepatan model</li> <li>▪ Kaedah termudah berbanding teknik peningkatan</li> <li>▪ Sesuai bagi set data dengan walau dengan kadar kebisingan yang tinggi</li> </ul>	<ul style="list-style-type: none"> <li>▪ Meningkatkan kerumitan komputasi model</li> <li>▪ Berkesan hanya bagi contoh-contoh rekod dengan kekerapan yang tinggi</li> </ul>
Undian	<ul style="list-style-type: none"> <li>▪ peningkatan prestasi untuk bilangan kelas yang banyak.</li> <li>▪ Boleh digunakan pada semua jenis data</li> </ul>	<ul style="list-style-type: none"> <li>▪ Memerlukan lebih banyak pengelasan yang baik adalah lebih baik kerana kiraan pengelasan adalah berdasarkan majoriti.</li> </ul>

Pemilihan teknik pengelasan bergabung adalah bergantung kepada matlamat dan ciri-ciri data latihan yang ada. Selain itu, terdapat tiga ciri yang membezakan teknik-teknik ini antara satu sama lain (Cao et al. 2015). Perbezaan ini ditunjukkan dalam Jadual 2-6.

Jadual 2-6 Perbezaan ciri-ciri teknik pengelasan bergabung

Aspek	Peningkatan	Pembungkusan	Undian
Pembahagian data ke dalam subset	Sampel data dibina berdasarkan pemberat	Rawak	Rawak
Objektif Teknik	Meningkat keupayaan mengelas dengan tepat kes positif	Mengurangkan ralat	Teknik mudah menentukan pengelas terbaik
Kaedah untuk menggabungkan model tunggal	Pemberat tinggi bagi contoh-contoh yang dikelas salah	Pemberat purata di beri kepada setiap model pengelas	Undian Majoriti, Undian Minimum, Undian Maksimum, Undian pemberat

#### 2.4.6 Kajian-kajian Model Pengelasan Bergabung

Terdapat banyak kajian yang berjaya membuktikan kejayaan teknik ini dalam memberi cadangan penyelesaian dalam pelbagai masalah dalam domain berbeza. Jadual 2-7 menunjukkan ringkasan kajian-kajian yang menggunakan teknik-teknik ini dalam pelbagai domain kajian seperti bioinformatik, sensor jarak jauh, telekomunikasi dan pendidikan. Kajian-kajian ini berjaya membuktikan bahawa pendekatan pengelasan bergabung adalah lebih baik berbanding pengelasan tunggal.

Jadual 2-7 Ringkasan kajian pengelasan bergabung pelbagai bidang kajian

Bidang	Sistem	Penulis
Bioinformatik	Model pergantungan bergabung untuk pengelasan dan ramalan kanser dan normal gen	Peng Qiu Z. Jane Wang K. J. Ray Liu, 2005
	Ramalan protein membran menggunakan asid amino perpecahan dan pengelasan bergabung	Maqsood Hayat & Asifullah Khan & Mohammed Yeasin, 2011
	Model Ramalan Protein menggunakan Multi-label Pengelasan bergabung	Guoxian Yu & Beibei & Huzefa Rangwala & George Mason & Carlotta Domenicon & George Mason & Guoji Zhang & Zhiwen Yu 2013
	Sistem pengelasan leukemia akut oleh zarah menggunakan kaedah pengelasan bergabung bagi memilih model terbaik	Hugo Jair Escalantea & Manuel Montes & Jesús A. González & Pilar Gómez-Gila & Leopoldo Altamiranoa & Carlos bersambung...

sambungan...		A. Reyesa & Carolina Retaa & Alejandro Rosalesa 2012
Sensor jarak jauh	Sistem Ramalan Cuaca dan Bencana menggunakan teknik Pembungkusan/Peningkatan dan pengelasan Hutan Rawak	Fachao Qin, Jiming Guo & Weidong Sun, 2016
Telekomunikasi	Sistem Ramalan Rangkaian untuk Telekomunikasi menggunakan <i>Filter-Wrapper</i> dan Pengelasan Bergabung	Adnan Idris & Asifullah Khan,, 2017
Pendidikan	Sistem ramalan prestasi pelajar menggunakan kaedah peningkatan pengelasan bergabung	S. Kotsiantis & K. Patriarcheas & M. Xenos, 2010
	Analisis perbandingan kaedah pembelajaran mesin bagi menentukan ramalan prestasi pelajar menggunakan kaedah pengelasan bergabung	Dursun Delen, 2010
	Sistem saringan pelajar: Analisis risiko dan visual data menggunakan model ramalan pengelasan bergabung	Alfred Essa & Hanan Ayad, 2012
	Pendekatan perlombongan data bagi menyokong pelajar memilih program kursus berdasarkan prestasi akademik	César Vialard & Jorge Chue & Juan Pablo Peche & Gustavo Alvarado & Bruno Vinatea & Jhonny Estrella & Álvaro Ortigosa & 2011
	Sistem bantuan keputusan bagi meramal gred pelajar menggunakan pendekatan pembelajaran mesin	S. B. Kotsiantis, 2011

Sehingga 2017, banyak kajian pengelasan bergabung dijalankan dengan pelbagai teknik pengelasan bergabung digunakan. Perkembangan kajian ini dapat diringkaskan dengan ringkasan kajian pengelasan bergabung terkini seperti di Jadual 2-8 di bawah.

Jadual 2-8 Ringkasan kajian pengelasan bergabung terkini

<b>Penulis, Tahun</b>	<b>Algoritma</b>	<b>Model pengelasan bergabung</b>
(Manaf et al. 2017)	Majority undian	Pemantauan zon pantai untuk memberikan maklumat penting mengenai keadaan semasa di kawasan pantai negara dengan mengkaji perubahan yang sedang berlaku.
(Ahmed et al. 2017)	ANN, SVM, K-NN, J48, Hutan Rawak	Meningkatkan klasifikasi jantung berasaskan tulisan tangan menggunakan pengelasan bergabung
(Thomas, & Vidal 2017)	Regresi Logistik (LR), Pembungkusan(LR)	Ke Arah Mengesan Kemalangan dengan Sedia Ada Maklumat Lalulintas Pasif

bersambung...