

PEMBENTUKAN SISTEM GEODEMOGRAFI DI PULAU PINANG

KAMARUL BIN ISMAIL

UNIVERSITI SAINS MALAYSIA

2013

PEMBENTUKAN SISTEM GEODEMOGRAFI DI PULAU PINANG

oleh

KAMARUL BIN ISMAIL

**Tesis yang diserahkan untuk memenuhi keperluan bagi
Ijazah Doktor Falsafah**

DISEMBER 2013

PENGHARGAAN

Pertama-tamanya saya memanjatkan setinggi-tinggi kesyukuran ke hadrat Allah s.w.t kerana dengan limpah rahmat dan izinNya, saya berjaya menyiapkan kajian ini. Terima kasih yang tidak terhingga diucapkan kepada penyelia saya, Profesor Ruslan bin Rainis kerana idea yang dilontarkan dan memberi ruang seluas-luasnya kepada saya untuk meneroka bidang kajian ini. Terima kasih juga di atas segala teguran, motivasi dan tunjuk ajar yang dicurahkan setiap hari selama beberapa tahun ini tanpa jemu.

Tidak lupa juga kepada pensyarah-pensyarah geografi di Pusat Pengajian Ilmu Kemanusiaan, terutamanya Dr. Wan Muhammad Muhiyuddin, Prof. Madya Tarmiji Masron, Prof. Madya Narimah Samat dan Prof. Madya Usman Yaakob yang sering bertanya khabar, berkongsi ilmu, memberi teguran dan semangat untuk menyiapkan kajian ini. Terima kasih juga kepada semua pensyarah geografi lain yang turut membantu saya menyiapkan tesis ini sama ada secara langsung ataupun tidak langsung. Terima kasih yang tidak terhingga juga diucapkan kepada kedua ibu bapa saya yang menyokong dan yang menjadi pendorong kepada saya. Segala kasih sayang dan didikan yang diberikan selama ini hanya Allah saja yang tahu nilai dan harganya. Juga buat isteri Noraini Md Zin, anak-anak Nur Aina Zulaikha, Nur Alia Maisara dan Nur Anis Zahirah, terima kasih sebanyaknya diucapkan kerana tanpa jemu membantu dan menunggu.

Kepada rakan-rakan seperjuangan di USM, Abdul Manaf Bohari, Danggat anak Chabo, Adrian Shah dan Jahan, rakan-rakan di UPSI, UM dan UiTM, terima kasih sebanyak-banyaknya. Akhirnya setinggi-tinggi penghargaan kepada Universiti Sains Malaysia dan Universiti Perguruan Sultan Idris yang memberi sokongan masa dan kewangan kepada saya untuk menyiapkan kajian ini.

SUSUNAN KANDUNGAN

	Muka surat
PENGHARGAAN	ii
JADUAL KANDUNGAN	iii
SENARAI JADUAL	vii
SENARAI RAJAH	viii
SENARAI SINGKATAN	x
SENARAI LAMPIRAN	xi
ABSTRAK	xii
ABSTRACT	xiii
BAB SATU : PENGENALAN	
1.0 Latar Belakang Kajian	1
1.1 Pernyataan Masalah	4
1.2 Persoalan Kajian	7
1.3 Objektif Kajian	7
1.4 Kepentingan Kajian	8
1.5 Skop Kajian	12
1.6 Organisasi Penulisan Tesis	14
BAB DUA : TINJAUAN BACAAN	
2.0 Pengenalan	17
2.1 Data Banci Penduduk dan Perumahan	17
2.2 Kawasan Statistik Terkecil	26
2.3 Sistem Geodemografi (SG)	31
2.4 Komponen-komponen Asas Sistem Geodemografi	35
2.4.1 Komponen Data Digital	36
2.4.2 Komponen Perlombongan Data	39
2.4.3 Komponen Sistem Maklumat Geografi (GIS)	49
2.5 Peringkat-peringkat Perkembangan Sistem Geodemografi	54
2.5.1 Sistem Pengkelasan Konvensional	54
2.5.2 Peralihan Sistem Geodemografi Digital	59
2.5.3 Pembangunan Sistem Geodemografi Komersial	64

2.5.4	Sistem Geodemografi Moden	67
2.5.5	Sistem Geodemografi Terbuka (Open Geodemographics)	71
2.6	Sistem Geodemografi di Malaysia dan Singapura	75
2.7	Pemilihan pembolehubah	79
2.8	Kesimpulan	82

BAB TIGA : KERANGKA KONSEP DAN METODOLOGI KAJIAN

3.0	Pengenalan	84
3.1	Rangka Konsep Kajian	84
3.2	Metodologi Kajian	86
3.2.1	Mengira Kekuatan Hubungan Antara Pembolehubah	90
3.2.2	Menggabungkan Pembolehubah	92
3.2.3	Melihat Perbezaan Taburan Secara Ruangan	93
3.2.4	Melaksanakan Analisis Komponen Utama (PCA)	94
3.3	Kaedah Penyeragaman Data	97
3.3.1	Penyeragaman z-skor	98
3.3.2	Penyeragaman Berkadar	98
3.4	Menentukan Saiz Kluster	99
3.5	Melabel dan Merangka Profil Kluster	101
3.6	Latar Belakang Kawasan Kajian	103
3.7	Kesimpulan	105

BAB EMPAT : ANALISIS KAJIAN

4.0	Pengenalan	107
4.1	Objek Pengklusteran	107
4.2	Pembolehubah-pembolehubah banci	110
4.3	Prosedur Pemilihan Pembolehubah	111
4.3.1	Analisis Komponen Utama (PCA)	112
4.3.2	Penentuan saiz komponen	113
4.3.3	Jumlah muatan setiap pembolehubah dalam komponen	117
4.3.4	Analisis Pekali Korelasi	119
4.3.5	Analisis Sisihan Piawai	124

4.3.6	Senarai Pembolehubah Akhir	125
4.4	Pembentukan Kluster Menggunakan algoritma <i>k</i> -means	127
4.5	Pengubahan Data	131
4.6	Penilaian kaedah penyeragaman data	134
4.7	Pembentukan Kluster Hierarki Kedua	136
4.8	Kesimpulan	141

BAB LIMA : PEMBENTUKAN DAN PENGUJIAN SISTEM

5.0	Pengenalan	142
5.1	Penentuan Hierarki Sistem Pengkelasan	142
5.2	Profil Kluster Hierarki Pertama	146
5.2.1	Kluster pekerja kolar biru (K1)	146
5.2.2	Kluster peniaga kecil sederhana dan penghuni kota (K2)	148
5.2.3	Kluster professional pinggir bandar (K3)	149
5.2.4	Kluster golongan berada dan berkerjaya (K4)	151
5.2.5	Kluster kelas pertengahan dan pekerja industri (K5)	153
5.2.6	Kluster luar bandar dan perkampungan tradisi (K6)	154
5.3	Profil Kluster Hierarki Kedua	156
5.3.1	Kolar biru berbagai bangsa dan budaya (1A)	157
5.3.2	Kolar biru berkeluarga (1B)	158
5.3.3	Berusia dan kurang berada (1C)	160
5.3.4	Peniaga berkeluarga (2A)	161
5.3.5	Keluarga muda dan penghuni kota (2B)	162
5.3.6	Pekerja perkhidmatan dan separa professional (2C)	163
5.3.7	Professional muda pinggir kota (3A)	165
5.3.8	Generasi Y berteknologi (3B)	167
5.3.9	Berpendidikan tinggi dan sederhana (3C)	168
5.3.10	Pekerja kolar putih berbagai etnik (3D)	170
5.3.11	Muda berada dan berjaya (4A)	171
5.3.12	Berkeluarga dan berkerjaya (4B)	173
5.3.13	Muda dan bermaklumat (4C)	174
5.3.14	Pekerja sektor industri (5A)	176
5.3.15	Membina keluarga (5B)	177
5.3.16	Generasi baru kelas pertengahan (5C)	179
5.3.17	Petani dan pesawah padi (6A)	180

5.3.18	Pekerja industri luar bandar (6B)	182
5.3.19	Perkampungan tradisi (6C)	183
5.4	Analisis Sensitiviti	185
5.5	Perbincangan	189
5.6	Kesimpulan	196
BAB ENAM : RUMUSAN		
6.0	Pendahuluan	197
6.1	Rumusan Penyelidikan	198
6.2	Kelemahan Kajian	205
6.3	Cadangan Kajian Lanjutan	206
SENARAI RUJUKAN		208
SENARAI LAMPIRAN		220

SENARAI JADUAL

	Muka surat	
1.1	Kaedah dan teknik perlombongan data	3
1.2	Kajian kegunaan sistem geodemografi dalam bidang pencegahan jenayah	10
2.1	Pelan pembangunan dan data banci yang digunakan	19
2.2	Perbandingan soalan banci tahun 1970, 1980, 1991 dan 2000	24
2.3	Senarai pembolehubah banci dan bukan banci dalam sistem geodemografi	37
2.4	Ringkasan algoritma <i>k</i> -means	47
2.5	Kumpulan dan kelas yang terdapat dalam sistem pengkelasan PRIZM	62
2.6	Gambaran umum struktur pengkelasan ACORN tahun 1980-an	65
2.7	Kumpulan dan kelas penduduk dalam sistem pengkelasan MOSAIC	66
2.8	SG di UK dan sumber data yang digunakan sebelum tahun 2001	68
2.9	SG di UK dan sumber data yang digunakan selepas tahun 2001	70
2.10	Susunan kumpulan dan kelas dalam Output Area Classification	74
2.11	Pengkelasan pengguna pengangkutan awam di Lembah Klang	77
2.12	Kumpulan dan kelas dalam sistem pengkelasan MOSAIC Singapore	78
2.13	Sistem pengkelasan yang digunakan sebagai perbandingan	80
2.14	Penggunaan pembolehubah dalam sistem pengkelasan terdahulu	81
3.1	Proses-proses yang terlibat pada peringkat penyediaan data	89
3.2	Nilai pekali korelasi dan penentuan hubungan	91
4.1	Analisis kerapian zon banci mengikut mukim	108
4.2	Penentuan saiz komponen dengan menggunakan scree plot	115
4.3	Saiz komponen dengan menggunakan kaedah kriteria Kaiser	116
4.4	Senarai pasangan pembolehubah yang berkorelasi	120
4.5	Matrik korelasi pembolehubah umur	122
4.6	Senarai pembolehubah dengan sisihan piawai yang kecil	125
4.7	Senarai 62 pembolehubah akhir	126
4.8	Taburan zon banci bagi penyelesaian 6 kluster	130
4.9	Nilai statistik kepencongan asal, kuasa dua dan logaritma	132
4.10	Pecahan zon menggunakan penyeragaman berkadar	135
4.11	Pengiraan kluster optimum dengan menggunakan jumlah objek	137
5.1	Jadual ANOVA 62 pembolehubah akhir	145
5.2	Label dan peratusan isi rumah kluster utama	146
5.3	Label kluster pada hierarki kedua	156
5.4	Pembolehubah yang menambah jarak objek dari pusat kluster	188
5.5	Pembolehubah yang tidak mengubah jarak objek dari pusat kluster	188
5.6	Pembolehubah yang mengurangkan jarak objek dari pusat kluster	189

SENARAI RAJAH

	Muka surat	
1.1	Asas-asas dalam pembentukan pengkelasan penduduk	13
2.1	Model data vektor GIS	21
2.2	Hierarki kawasan geografi di Malaysia	26
2.3	Proses pengutipan dan penyimpanan data banci	28
2.4	Pecahan zon banci mengikut daerah pentadbiran	29
2.5	Komponen-komponen utama sistem geodemografi	36
2.6	Proses-proses yang terlibat dalam analisis penerokaan data	40
2.7	Pengklusteran berasaskan jarak terdekat	42
2.8	Kaedah-kaedah utama dalam proses pengklusteran	45
2.9	Kaedah pengklusteran menggunakan algoritma k-means	48
2.10	Perwakilan data geografi	50
2.11	Peta gambaran kemiskinan di London tahun 1898	55
2.12	Model Zon Berpusat Park dan Burgess	57
3.1	Rangka konsep kajian	85
3.2	Proses-proses pembentukan sistem geodemografi	87
3.3	Carta aliran proses pembentukan sistem geodemografi	88
3.4	Peringkat-peringkat dalam PCA	96
3.5	Graf penentuan jumlah optimum bilangan kluster	101
3.6	Kawasan kajian	104
3.7	Komposisi kumpulan etnik di Pulau Pinang	105
4.1	Peratusan zon banci mengikut mukim dalam kawasan kajian	109
4.2	Scree plot	114
4.3	Jarak purata objek dari pusat kluster	129
4.4	Proses pembentukan hierarki menggunakan algoritma k-means	136
4.5	Jarak purata bagi kumpulan satu	138
4.6	Jarak purata bagi kumpulan dua	139
4.7	Jarak purata bagi kumpulan tiga	139
4.8	Jarak purata bagi kumpulan empat	139
4.9	Jarak purata bagi kumpulan lima	140

4.10	Jarak purata bagi kumpulan enam	140
5.1	Carta radar kluster 1	143
5.2	Carta radar kluster 2	143
5.3	Carta radar kluster 3	143
5.4	Carta radar kluster 4	143
5.5	Carta radar kluster 5	143
5.6	Carta radar kluster 6	143
5.7	Pekerja kolar biru	147
5.8	Peniaga kecil sederhana dan penghuni kota	149
5.9	Professional pinggir kota	150
5.10	Golongan berada dan berkerjaya	152
5.11	Kelas pertengahan dan pekerja industri	153
5.12	Luar bandar dan perkampungan tradisi	155
5.13	Kolar biru berbagai bangsa dan budaya	158
5.14	Kolar biru berkeluarga	159
5.15	Berusia dan kurang berada	160
5.16	Peniaga berkeluarga	161
5.17	Keluarga muda dan penghuni kota	163
5.18	Pekerja perkhidmatan dan separa professional	165
5.19	Professional muda pinggir kota	166
5.20	Generasi Y berteknologi	168
5.21	Berpendidikan tinggi dan sederhana	169
5.22	Pekerja kolar putih berbagai etnik	171
5.23	Muda berada dan berjaya	172
5.24	Berkeluarga dan berkerjaya	174
5.25	Muda dan bermaklumat	175
5.26	Pekerja sektor industri	177
5.27	Membina keluarga	178
5.28	Generasi baru kelas pertengahan	180
5.29	Petani dan pesawah padi	181
5.30	Pekerja industri luar bandar	183
5.31	Perkampungan tradisi	184
5.32	Hasil analisis sensitiviti 62 pembolehubah akhir	186

SENARAI SINGKATAN

JPM	Jabatan Perangkaan Malaysia
RM-1	Rancangan Malaya Pertama
RM-2	Rancangan Malaya Kedua
RMK-1	Rancangan Malaysia Pertama
RMK-2	Rancangan Malaysia Kedua
RMK-3	Rancangan Malaysia Ketiga
RMK-8	Rancangan Malaysia Kelapan
GIS	Sistem Maklumat Geografi
SG	Sistem Geodemografi
UK	United Kingdom
USA	Amerika Syarikat
USD	Dollar Amerika
NSMC	The National Social Marketing Centre for Excellence
DEO	Jabatan Alam Sekitar
CACI	Consolidated Analysis Centers Incorporated
ACORN	A Classification Of Residential Neighbourhoods
DBMS	SistemPengurusan Pangkalan Data
BRGD	Jabatan Pendaftaran Umum British
IBM	International Business Machines Corporation
OEO	Jabatan Pembangunan Ekonomi USA
ZIP	Zone Improvement Plan
PRIZM	Potential Rating Index for ZIP Markets
CES	Pusat Kajian Alam Sekitar
MRS	Persatuan Kajian Pemasaran British
TGI	Target Group Index
PCA	Analisis Komponen Utama
ONS	Jabatan Perangkaan Kebangsaan UK
OAC	Pengkelasan Kawasan Hasilan
GROS	General Register Office for Scotland
TNS Global	Taylor Nelson Sofres Global
TK	Tempat Kediaman
BP	Blok Perhitungan
PDDP	Principal direction divisive partitioning
DTL	Daerah Timur Laut
SPT	Seberang Perai Tengah
SPSS	Statistical Package for the Social Sciences
SSL	Rotation Sums of Squared Loadings

SENARAI LAMPIRAN

	Muka surat
1.1 Senarai pembolehubah awal dan sektor yang diwakili	220
1.2 Senarai 40 komponen yang diperolehi daripada analisis PCA	224
1.3 Senarai akhir daripada proses pemilihan pembolehubah	229
1.4 Pengiraan kluster optimum menggunakan jumlah objek	233

PEMBENTUKAN SISTEM GEODEMOGRAFI DI PULAU PINANG

ABSTRAK

Pangkalan data banci penduduk dan perumahan negara merekodkan maklumat semua orang yang tinggal disesebuah kawasan. Jumlah data yang besar dan dikutip secara berterusan menyebabkan maklumat-maklumat yang terdapat dalam pangkalan data ini sukar untuk dimanfaatkan sepenuhnya. Maklumat-maklumat ini digunakan oleh pihak kerajaan dalam merangka dasar, pembahagian sumber dan merupakan petunjuk penting kepada kemajuan sosioekonomi sesebuah negara. Oleh itu, kajian pembentukan sistem geodemografi yang mempunyai keupayaan untuk meringkaskan data-data ini dalam bentuk yang lebih mudah difahami telah dilaksanakan. Objektif-objektif kajian yang lain termasuklah untuk menguji kesesuaian penggunaan data banci, menilai kaedah pemilihan dan penyeragaman data dan membentuk serta menguji sistem geodemografi yang dibangunkan. Peringkat-peringkat metodologi kajian iaitu peringkat penyediaan data, penentuan jumlah kluster, pembentukan profil dan melabelkan kluster yang digunakan dalam penyelidikan ini adalah berasaskan kepada proses pembentukan sistem OAC (*Output Area Classifications*) yang dibuat oleh Jabatan Perangkaan Kebangsaan (*Office for National Statistics*) United Kingdom. Penggunaan kaedah perlombongan data dan analisis statistik yang lain seperti analisis pekali korelasi, analisis komponen utama (PCA), analisis sisihan piawai dan analisis kluster menggunakan algoritma *k*-means telah berjaya membentuk sistem geodemografi di Pulau Pinang yang mengandungi sejumlah 6 kluster utama dan 19 sub kluster bagi mewakili kumpulan penduduk berbeza. Kluster pada hierarki pertama dilabelkan dengan nama seperti Pekerja kolar biru (K1), Peniaga kecil sederhana dan penghuni kota (K2), Professional pinggir bandar (K3), Golongan berada dan berkerjaya (K4), Kelas pertengahan dan pekerja industri (K5) serta kluster Luar bandar dan perkampungan tradisi (K6). Manakala kluster pada hierarki kedua antaranya dilabelkan dengan nama-nama seperti Peniaga berkeluarga, Keluarga muda dan penghuni kota, Pekerja perkhidmatan dan separa professional, Professional muda pinggir kota, Generasi Y berteknologi serta kluster Berpendidikan tinggi dan sederhana yang memberi gambaran umum ciri-ciri penduduk di kawasan tersebut.

THE DEVELOPMENT OF GEODEMOGRAPHICS SYSTEM IN PENANG

ABSTRACT

The study of the development of geodemographics classification system that was conducted in Penang has fully utilised the data source from population and housing census that are gathered from the Department of Statistics. Huge amount of data collected continuously cause the information in the database is difficult to be fully utilized. Therefore, the study of the development of geodemographics system that has the ability to summarize data in a more easily understood form was implemented. The objectives of the other studies included to verify the usage of census data, to determine the appropriate methods of analysis and to create and verify the developed geodemographics system. Stages of the research methodology for data preparation, determination of the clusters and labeling profiles used in this research is based on the formation of Output Area Classifications system (OAC) made by the United Kingdom Office for National Statistics. The use of data mining and statistical analysis such as correlation analysis, Principal Component Analysis (PCA), the standard deviation analysis and cluster analysis using the *k*-means algorithm has been successfully formed Penang geodemographics system containing a total of 6 major clusters and 19 sub-clusters to represent the different population groups. The major groups in the hierarchy labeled with names such as Blue-collar workers (K1), Medium and small traders of the city (K2), Professional suburbs (K3), The prosperous group (K4), The middle class and industrial workers (K5), Rural and traditional village (K6). While both of the sub-clusters in the second hierarchy are labeled with names like Trader family, Young family and the city, Semi-professional and service workers, Young suburban professional and Generation Y groups which gives a general overview of the characteristics of the population in the area.

BAB SATU

PENGENALAN

1.0 Latar belakang

Pengkelasan secara umumnya merujuk kepada proses menyusun dan mengumpul sesuatu benda mengikut kumpulan, berasaskan kepada kesamaan dan ketidaksamaan. Hasil daripada proses ini dirujuk sebagai sistem pengkelasan (Batley, 2005). Kajian berkaitan dengan sistem pengkelasan, menurut Tilton (2009) berasal daripada bidang taksonomi yang melibatkan kajian tentang sains pengkelasan dengan tumpuan khusus kepada pengkelasan organisma hidup dan hubungan semulajadi yang wujud. Perintis kepada bidang taksonomi ialah Aristotle, iaitu seorang ahli falsafah Yunani terkenal yang hidup antara tahun 384 hingga 322 sebelum masihi.

Aristotle memperkenalkan dua konsep pengkelasan yang utama iaitu pengkelasan organisma mengikut jenis dan pengkelasan mengikut definisi binomial (kombinasi dua perkataan) (Tilton, 2009). Dalam kajian yang bertajuk "*The History of Animals*" (350 s.m.), Aristotle membezakan organisma hidup berdasarkan kepada persamaan yang dimiliki oleh organisma-organisma tersebut, antaranya termasuklah pengkelasan haiwan berdarah dan haiwan tidak berdarah, haiwan yang hidup di daratan dan haiwan yang hidup di dalam air. Aristotle mengandaikan bahawa organisma hidup boleh disusun mengikut hierarki dengan spesis manusia berada pada hierarki paling tinggi (Tilton, 2009; Avery, 2003).

Gordon (1999) melihat sistem pengkelasan sebagai satu alat yang boleh digunakan untuk menyiasat hubungan yang wujud dalam satu kumpulan objek, sama ada objek-

objek tersebut boleh diringkaskan atau tidak dalam kumpulan yang lebih kecil. Kumpulan-kumpulan kecil ini dibentuk dengan mengelaskan objek yang serupa dalam kelas yang sama. Gordon (1999) juga menyatakan teori-teori pembangunan pengkelasan yang dibincangkan oleh beberapa orang pengkaji terdahulu melibatkan penggunaan sistem pengkelasan dalam bidang kajian yang berbeza.

Walau bagaimanapun, matlamat utama penggunaan sistem pengkelasan menurut Gordon (1999) adalah untuk meringkaskan data dan membuat tanggapan terhadap corak yang wujud didalamnya. Kewujudan pangkalan data dengan jumlah data yang besar dan semakin bertambah setiap hari, telah melampaui keupayaan manusia untuk menganalisis data-data tersebut. Kumpulan data mentah ini tidak dapat memberi apa-apa manfaat sekiranya ia tidak diproses dengan menggunakan kaedah analisis yang bersesuaian. Analisis data dengan menggunakan analisis statistik konvensional menurut Miller dan Han (2001), mempunyai kelemahan yang jelas walaupun kaedah ini boleh memberi gambaran asas dan menerangkan fenomena yang berlaku di sesuatu kawasan.

Masalah ini mendorong para pengkaji mencari kaedah analisis yang berupaya untuk mengubah data-data ini kepada bentuk lain sehingga membawa kepada pengenalan bidang perlombongan data (*data mining*) pada tahun 1950-an (Ayre, 2006). Istilah perlombongan data menurut Han dan Kamber (2006) dan Larose (2005) merujuk kepada proses mencari maklumat yang berguna dalam satu pangkalan data yang besar. Ia melibatkan penggunaan algoritma khusus yang dibangunkan melalui gabungan pelbagai bidang kajian lain seperti statistik, kebijaksanaan tiruan, mesin belajar (*machine learning*), sains pangkalan data dan pemerolehan maklumat. Kamath (2002) pula melihat proses perlombongan data sebagai satu proses interaktif yang saling

berkitaran. Proses-proses penting yang perlu dilaksanakan untuk memperolehi maklumat semasa menjalankan analisis perlombongan data termasuklah pengumpulan data, pengurangan dimensi, pengubahan dan penyusunan data serta pengenalan corak. Jadual 1.1 menunjukkan susunan kaedah perlombongan data, keterangan kaedah dan teknik yang lazimnya digunakan untuk melaksanakan analisis dengan menggunakan kaedah-kaedah tersebut.

Jadual 1.1: Kaedah dan teknik perlombongan data

Kaedah	Keterangan	Teknik
Segmentasi	Kluster - menentukan saiz sebenar kelas yang terdapat dalam data Pengkelasan - mengumpulkan data dalam kumpulan kelas yang ditentukan terlebih dahulu	Analisis kluster, Pengkelasan Bayesian <i>Classification Tree</i> , Rangkaian saraf tiruan
Analisis Kebergantungan	Mencari aturan tertentu untuk meramalkan nilai sesuatu atribut berdasarkan nilai atribut yang lain	Rangkaian bayesian, Aturan kesatuan
Analisis taburan dan nilai melampau	Meneliti taburan data yang tidak berada dalam lingkungan jangkaan	Analisis kluster, Pengesanan nilai melampau
Penelitian Kecenderungan	Meneliti garisan dan kecondongan, mengkelaskan pangkalan data, lazimnya mengikut masa	Regresi, Pencerapan corak berperingkat
Generalisasi dan pencirian	Gambaran secara tepat sesuatu data	Aturan ringkasan, Pengenalan atribut bertumpu

Sumber: Miller dan Han (2001)

Berdasarkan kepada jadual yang ditunjukkan, secara keseluruhannya bidang perlombongan data boleh dibahagikan kepada lima kaedah yang utama iaitu segmentasi (*segmentation*), analisis kebergantungan, analisis taburan dan nilai melampau (*outliers*), penelitian kecenderungan serta generalisasi dan pencirian. Teknik-teknik yang sering digunakan bagi setiap kaedah perlombongan data ini antaranya termasuklah analisis kluster, pengelasan bayesian, pohon pengelasan, rangkaian saraf tiruan, aturan kesatuan dan regresi (Miller dan Han, 2001).

Kaedah pengklusteran dengan menggunakan analisis kluster merupakan salah satu daripada kaedah perlombongan data yang popular. Kaedah ini menurut Miller dan Han (2009) melibatkan proses penyusunan data ke dalam kelas atau kumpulan yang mempunyai makna. Proses ini biasanya melibatkan penggunaan algoritma analisis kluster yang boleh melihat hubungan di antara data, menentukan bilangan kelas dan memastikan data-data tersebut mempunyai persamaan yang maksimum di antara kelas (*intra*class) serta mempunyai perbezaan yang minimum di dalam kelas (*inter*class). Tegasnya, analisis kluster yang digunakan dalam bidang perlombongan data menentukan bilangan kelas dan mengagihkan data ke dalam kelas yang berbeza berdasarkan kepada kekuatan hubungan data-data tersebut dalam ruang maklumat (Miller dan Han, 2009).

Dramovicz (2004) menyatakan bahawa hasil daripada analisis kluster ini dirujuk sebagai sistem segmentasi (*segmentation system*), segmen-segmen dalam sistem pengkelasan ini mempunyai perbezaan yang sangat ketara antara satu dengan yang lain. Objek-objek yang sering digunakan untuk membentuk segmen ini termasuklah pelanggan (pengkelasan pelanggan), pasaran (pengkelasan pasaran) ataupun kejiranan (pengkelasan geodemografi). Salah satu daripada sumber data penting yang digunakan untuk membentuk sistem-sistem pengkelasan ini ialah data banci penduduk dan perumahan negara.

1.1 Pernyataan Masalah

Penyelidikan ini dijalankan berasaskan kepada pernyataan bahawa pengguna pada hari ini tidak lagi berhadapan dengan masalah kekurangan data, sebaliknya mereka

berhadapan dengan masalah data yang berlebihan. Keadaan ini jelas ditunjukkan melalui kajian yang dijalankan oleh Naisbitt melalui pernyataan “*we are drowning in information but starved for knowledge*” pada tahun 1982. Pernyataan ini jelas membuktikan masalah sebenar yang sedang dihadapi oleh pengguna bukan ketiadaan data tetapi masalah jumlah data yang semakin bertambah sehingga menimbulkan kesukaran untuk memanfaatkan data-data tersebut (Larose, 2005).

Data banci penduduk dan perumahan negara merupakan sumber data utama dalam proses pembentukan sistem geodemografi bersifat rencam dan dikutip secara berterusan dalam selang masa yang telah ditetapkan. Bagi negara-negara maju, proses pengutipan data banci dilaksanakan dalam tempoh masa lima tahun, manakala di negara-negara sedang membangun kerja-kerja merekod maklumat mengenai penduduk ini dijalankan pada setiap sepuluh tahun. Data banci penduduk dan perumahan mengandungi pelbagai maklumat mengenai ciri-ciri populasi seperti umur, jantina, status perkahwinan, pendapatan dan pendidikan. Data-data ini disimpan dalam pangkalan data banci dan merupakan sumber data yang lengkap dan paling komprehensif bagi kebanyakan negara sedang membangun (Smith, 2003).

Proses pengutipan data banci yang berterusan ini menimbulkan masalah kepada pengguna kerana jumlah data yang semakin bertambah setiap hari menyukarkan pencarian maklumat dilakukan. Untuk menangani masalah ini, kaedah analisis yang boleh meringkaskan maklumat demografi dan perumahan dalam pangkalan data banci kepada bentuk yang lebih mudah difahami sangat diperlukan. Oleh itu, pembentukan sistem geodemografi yang menggunakan algoritma pengklusteran dalam penyelidikan ini merupakan kaedah yang paling berkesan untuk mengurangkan jumlah data dan

menyusun data-data tersebut ke dalam kumpulan kecil yang lebih mudah untuk diurus dan difahami. Dalam konteks pembentukan sistem geodemografi di Pulau Pinang, sejumlah 195 pembolehubah berpotensi dan 1138 unit zon banci digunakan sebagai input dalam pembentukan sistem ini.

Selain itu, penyelidikan ini juga dijalankan disebabkan oleh kesukaran memperolehi maklumat lengkap mengenai individu dalam sesebuah kawasan. Dalam dunia nyata, proses pembentukan pangkalan data yang menyimpan maklumat individu hampir mustahil untuk direalisasikan. Ini kerana proses menghitung setiap orang dalam sesebuah kawasan sangat sukar dilaksanakan. Kerja-kerja pengutipan data di lapangan bukan sahaja memerlukan perbelanjaan yang besar, malah data-data yang dikutip ini perlu disimpan dan diuruskan secara sistematik. Masalah ini menyebabkan kesukaran untuk mewujudkan pangkalan data yang lengkap walaupun sebilangan kecil maklumat mengenai penduduk boleh direkodkan dalam pangkalan data (Vickers, 2007).

Pembentukan sistem geodemografi seperti yang dicadangkan dalam penyelidikan ini mampu menyelesaikan masalah ketiadaan pangkalan data yang menyimpan maklumat demografi dan sosioekonomi penduduk di sesebuah kawasan. Pembentukan sistem geodemografi membolehkan pengguna membuat andaian mengenai ciri-ciri individu sekiranya tempat mereka tinggal diketahui. Secara umumnya sempadan kawasan di Malaysia terbahagi kepada dua bahagian iaitu, sempadan pentadbiran dan sempadan statistik. Sempadan pentadbiran merujuk kepada negeri, daerah, mukim, bandar atau kampung. Manakala sempadan statistik pula merupakan kawasan yang dihasilkan oleh Jabatan Perangkaan bagi memudahkan kerja-kerja pengutipan data banci dijalankan. Sempadan-sempadan kawasan statistik terkecil ini termasuklah daerah banci, zon banci

dan blok penghitungan. Sistem geodemografi menggunakan sempadan- sempadan kawasan terkecil ini untuk membuat ringkasan ciri-ciri demografi bagi semua penduduk dalam sesebuah kawasan secara purata (Harris et al., 2005).

1.2 Persoalan Kajian

Berdasarkan kepada perbincangan tentang pernyataan masalah dalam bahagian sebelum ini, beberapa persoalan penting yang perlu dijelaskan dalam penyelidikan ini. Persoalan-persoalan kajian dalam penyelidikan ini adalah seperti berikut:

- i) Adakah data banci penduduk dan perumahan negara sesuai digunakan sebagai input dalam proses pembentukan sistem geodemografi?
- ii) Bagaimanakah kaedah perlombongan data dan analisis statistik yang lain membantu proses pencarian maklumat yang terdapat dalam pangkalan data banci penduduk dan perumahan negara?
- iii) Sejauhmanakah pembentukan sistem geodemografi memudahkan pengguna memahami gambaran asas ciri-ciri penduduk di Pulau Pinang?

1.3 Objektif Kajian

Matlamat utama kajian ini adalah untuk membentuk satu sistem geodemografi (SG) dengan menggunakan data banci penduduk dan perumahan tahun 2000. Untuk mencapai matlamat pembentukan sistem pengkelasan ini dan bagi membolehkan kajian dijalankan dengan lebih bersistematik, objektif-objektif berikut turut digariskan.

- i) Menguji kesesuaian penggunaan data banci penduduk dan perumahan negara yang diperolehi daripada Jabatan Perangkaan Negara untuk membentuk sistem geodemografi di Pulau Pinang.

- ii) Menilai kaedah pemilihan dan penyeragaman data yang paling sesuai bagi pembentukan sistem geodemografi di Pulau Pinang.
- iii) Membentuk dan menguji sistem geodemografi yang dibangunkan di Pulau Pinang.

1.4 Kepentingan Kajian

Data yang dikutip secara berterusan dan disimpan dalam pangkalan data yang besar menimbulkan masalah kepada pengguna untuk memperolehi maklumat daripada pangkalan data tersebut. Oleh itu, rasional kajian pembentukan sistem geodemografi dijalankan adalah untuk membantu pengguna memanfaatkan sumber data sedia ada ini sepenuhnya. Metodologi yang digunakan dalam penyelidikan ini adalah berasaskan kepada kajian pembangunan sistem pengkelasan OAC (*Output Area Classification*) yang dijalankan oleh Vickers (2006). Walaupun terdapat pelbagai jenis sistem geodemografi di pasaran, sistem OAC merupakan satu-satunya sistem pengkelasan sumber terbuka yang memberikan penjelasan terperinci bagaimana sistem tersebut dihasilkan. Kajian pembentukan sistem geodemografi di Pulau Pinang merupakan sumbangan penting penyelidikan ini dalam perkembangan bidang pengkelasan kawasan terutamanya dalam analisis statistik kawasan terkecil (*small area statistics*).

Pangkalan data banci penduduk dan perumahan negara yang kompleks menyukarkan pengguna untuk memperolehi maklumat yang terlindung di dalam kumpulan data-data tersebut. Data-data banci ini diperolehi melalui proses pengutipan data statistik terbesar yang dijalankan oleh Jabatan Perangkaan Malaysia (JPM). Maklumat-maklumat yang terdapat dalam pangkalan data ini sangat penting untuk digunakan dalam proses penggubalan dasar negara, perancangan pembangunan sosioekonomi, menjadi input kepada pentadbiran kerajaan dan kegunaan penyelidikan (Abdul Rahman dan Norfariza,

2007). Bagi membolehkan semua maklumat yang tersembunyi dalam pangkalan data banci diperolehi, penggunaan sistem geodemografi sebagai alat bantu memudahkan analisis dilaksanakan disebabkan oleh keupayaan sistem ini untuk meringkaskan pangkalan data yang rencam dan menyusun data-data tersebut dalam bentuk yang lebih mudah difahami.

Rasional penyelidikan ini dijalankan adalah juga berasaskan kepada perancangan strategik pengagihan data untuk memaksimumkan penggunaan data banci oleh JPM (Rabieyah dan Siti Asiah, 2000). Pangkalan data banci penduduk dan perumahan negara yang semakin bertambah data banci penduduk yang dikutip setiap sepuluh tahun dengan melibatkan peruntukan perbelanjaan yang besar. Banci pada tahun 2000 sebagai contoh melibatkan penggunaan lebih daripada RM200 juta untuk menampung keseluruhan kos pengutipan data di lapangan. Objektif utama banci pada tahun 2000 ialah untuk mencapai matlamat penggunaan data yang tidak terhad kepada pihak kerajaan sahaja, tetapi data-data ini juga perlu digunakan oleh sektor-sektor lain seperti penganalisis industri, usahawan, ahli akademik, penyelidik dan ahli politik (Malaysia, 2001). Kajian pembentukan sistem geodemografi yang dijalankan ini selaras dengan matlamat JPM memaksimumkan penggunaan data banci dan merupakan langkah proaktif bagi menangani masalah pembaziran data yang serius.

Selain itu, penyelidikan ini juga dijalankan adalah disebabkan oleh kepentingan nilai ekonomi sistem geodemografi itu sendiri. Ini kerana sistem geodemografi boleh digunakan sebagai alat bantu dalam pelbagai bidang kajian (Feng dan Flowerdew, 1999). Peningkatan permintaan terhadap penggunaan sistem geodemografi menurut Brown (1991) menjadi bukti keberkesanan sistem ini untuk meramalkan perlakuan

pengguna dan seterusnya menyokong proses menganalisis pasaran dengan lebih berkesan. Pada tahun 1992 nilai pasaran sistem pengkelasan ini hanya berjumlah 25 juta pound tetapi telah meningkat kepada 54 juta pound pada tahun 1995 (Harris et al., 2005). Peningkatan permintaan sistem geodemografi di pasaran ini menjadi rasional kajian pembentukan sistem geodemografi di Pulau Pinang perlu dibuat dan dimajukan.

Sistem geodemografi juga merupakan alat bantu yang penting dalam pelbagai bidang kajian lain, ini termasuklah kepentingan dalam bidang sosial dan pencegahan jenayah (Breetzke dan Horn, 2009). Jadual 1.2 menunjukkan beberapa contoh ringkasan kajian penggunaan sistem geodemografi untuk menangani masalah sosial dan jenayah.

Jadual 1.2: Kajian kegunaan sistem geodemografi dalam bidang pencegahan jenayah

Kegunaan sistem geodemografi	Pengkaji
Menyusun sumber polis mengikut kepelbagaian skala ruangan di England dan Wales	Ashby dan Longley (2005b)
Meramal tahap sosial dan keberkesanan pemantauan secara kolektif	Williamson et al. (2006)
Menganalisis variasi jenayah mengikut lokasi dan khidmat polis	Ashby (2004)
Menunjukkan tahap dan corak perlakuan jenayah golongan remaja mengikut perbezaan kejiranan	Williamson et al. (2005)
Mengenalpasti kawasan kejiranan berisiko tinggi berlaku jenayah di UK	Mayhew et al. (1993)

Sumber: Breetzke dan Horn (2009)

Kajian yang dijalankan oleh Ashby dan Longley (2005a) menunjukkan kepentingan sistem geodemografi dalam proses menentukan lokasi kadar jenayah yang tinggi dan memerlukan lebih ramai anggota keselamatan berbanding dengan kawasan lain. Begitu juga dengan kajian Williamson et al. (2005) yang mengkaji perlakuan jenayah remaja mengikut kawasan kejiranan dan Mayhew et al. (1993) yang menggunakan

sistem geodemografi untuk mengenalpasti kawasan berisiko tinggi berlaku kejadian jenayah. Penggunaan sistem geodemografi membolehkan pembahagian kawasan mengikut kekerapan kejadian jenayah dilakukan. Maklumat yang diperolehi membantu proses pembuatan keputusan untuk menambahkan bilangan anggota keselamatan di kawasan-kawasan yang mempunyai kadar jenayah yang tinggi.

Kajian penggunaan sistem geodemografi dalam mengenalpasti kawasan berisiko berlakunya kejadian jenayah membuktikan bahawa sistem ini merupakan alat bantu yang sangat penting untuk menangani pelbagai masalah sosial dan jenayah dalam sempadan kawasan yang lebih kecil. Pembentukan sistem geodemografi di Pulau Pinang ini membantu pihak berkuasa menentukan kawasan-kawasan berisiko tinggi berlaku masalah sosial dan seterusnya membuat pembahagian anggota keselamatan berasaskan kepada keperluan kawasan yang telah dikenalpasti.

Sistem geodemografi juga telah terbukti menjadi alat yang penting untuk menyelesaikan masalah-masalah pentadbiran dan perkhidmatan awam. Badan kerajaan bukan sahaja boleh membandingkan pencapaian dan kejayaan dasar yang telah dilaksanakan, tetapi maklumat yang diperolehi daripada sistem pengelasan ini juga membolehkan pihak kerajaan membuat perancangan dan menyelesaikan masalah penduduk mengikut kumpulan sasaran yang khusus. Pembentukan sistem pengelasan berasaskan kepada kawasan kecil statistik di UK sebagai contoh adalah bertujuan untuk membantu pihak berkuasa untuk membuat perancangan dan mengagihkan peruntukan mengikut keperluan sesebuah kawasan (Webber dan Farr, 2001).

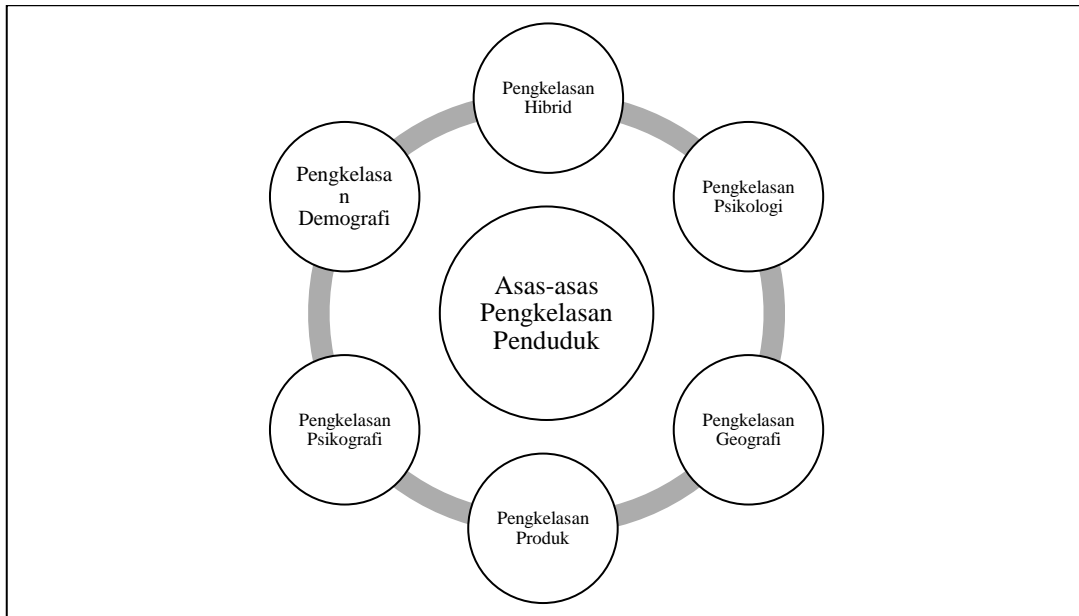
Pembentukan sistem geodemografi di Pulau Pinang, membolehkan pihak berkuasa tempatan membandingkan kawasan dan penduduk yang memerlukan kemudahan infrastruktur yang khusus. Sistem yang dibangunkan ini membolehkan penyediaan prasarana dan bantuan-bantuan lain disalurkan kepada kumpulan penduduk yang benar-benar memerlukan. Sistem ini bukan sahaja boleh digunakan oleh para perancang dan pembuat dasar, tetapi keupayaan sistem geodemografi sebagai alat bantu proses membuat keputusan membolehkan pihak berkuasa tempatan memahami ciri-ciri penduduk dalam kawasan pentadbiran mereka dan menyediakan kemudahan kepada kumpulan penduduk yang benar-benar memerlukan.

1.5 Skop Kajian

Terdapat beberapa batasan yang telah dikenalpasti dalam menjalankan penyelidikan ini terutamanya dari sudut penggunaan data, pemilihan kawasan kajian dan sumbangan kepada bidang pengkelasan kawasan. Sumber utama data yang digunakan dalam penyelidikan ini ialah data banci penduduk dan perumahan negara tahun 2000. Data-data yang diperolehi daripada pangkalan data banci JPM ini mengandungi sejumlah 195 pembolehubah dan lebih daripada 1000 unit zon banci yang digunakan sebagai objek pengklusteran. Hasil analisis yang diperolehi dengan menggunakan data-data ini berbeza sekiranya pembentukan sistem geodemografi dilakukan dengan menggunakan sumber data yang lain.

Profil penduduk yang dihasilkan daripada proses pembentukan sistem geodemografi juga terhad dalam kawasan kajian sahaja. Sekiranya sistem pengkelasan ini digunakan di kawasan lain, ia perlu disesuaikan dengan data dan sempadan kawasan yang baru. Proses-proses seperti pemilihan pembolehubah, pengurangan data, analisis kluster,

melabel dan membentuk profil kluster perlu dilaksanakan semula. Namun begitu, penyelidikan ini menjadi asas kepada pembentukan sistem geodemografi di kawasan-kawasan lain kerana proses-proses yang digunakan dalam pembentukan tersebut dinyatakan dengan jelas dalam kajian ini.



Rajah 1.1 : Asas-asas dalam pembentukan pengkelasan penduduk
Sumber: Schiffman dan Kanuk (2000)

Secara umumnya, terdapat enam asas utama yang oleh digunakan untuk mengkelaskan penduduk seperti yang ditunjukkan dalam rajah 1.1. Antara asas-asas pengkelasan tersebut ialah pengkelasan geografi, pengkelasan demografi, pengkelasan psikografi dan pengkelasan berasaskan produk. Pengkelasan geografi membahagikan penduduk berdasarkan kepada di mana lokasi mereka berada, manakala pengkelasan demografi pula membahagikan penduduk mengikut persamaan yang terdapat pada ciri-ciri seperti jantina, umur, pendapatan, pekerjaan, pendidikan, saiz isi rumah dan kitaran hidup keluarga. Pengkelasan psikografi mengelompokkan populasi mengikut ciri-ciri psikologi yang sepadan, nilai sosio budaya dan cara hidup. Pengkelasan berasaskan

produk pula mengelaskan pengguna berdasarkan persamaan yang terdapat dalam aspek pemilihan jenama, kecenderungan penggunaan dan penilaian pulangan kepuasan daripada produk tersebut.

Asas pengkelasan penduduk yang digunakan dalam penyelidikan ini ialah pengkelasan secara hibrid yang merupakan gabungan daripada pengkelasan demografi dan geografi. Penyelidikan ini tidak menggunakan kesemua asas pengkelasan seperti yang dikemukakan oleh Schiffman dan Kanuk (2000), sebaliknya membataskan kajian ini kepada asas pembentukan pengkelasan geodemografi. Sistem yang dibentuk dalam kajian ini menumpukan kepada pembentukan sistem geodemografi yang menggunakan pembolehubah-pembolehubah geografi dan demografi sahaja. Sumber utama data yang digunakan pula ialah data banci penduduk dan perumahan Malaysia tahun 2000 yang diperolehi daripada Jabatan Perangkaan Malaysia (JPM). Data-data ini dianalisis dengan menggunakan kaedah perlombongan data yang melibatkan beberapa proses utama seperti pembersihan, pengurangan, penyeragaman dan pemilihan data. Proses-proses ini perlu dilaksanakan dengan teliti kerana kualiti sistem geodemografi yang dibangunkan dalam kajian ini bergantung sepenuhnya kepada input yang digunakan dalam analisis kluster.

1.6 Organisasi Penulisan Tesis

Tesis ini mengandungi enam bab yang mempunyai kesinambungan dari bab satu hingga bab enam. Bab satu merupakan bab pengenalan yang memuatkan perbincangan tentang latar belakang kajian, isu dan masalah, objektif dan persoalan kajian, skop dan kepentingan kajian. Pengenalan secara ringkas latar belakang kajian dalam bahagian ini dibincangkan dengan lebih terperinci dalam bab dua. Bab dua adalah merupakan

semakan literatur. Dalam bab ini pengenalan kepada sistem geodemografi, konsep-konsep asas dan sejarah perkembangan sistem geodemografi dibincangkan dengan lebih terperinci. Bab ini membahagikan perkembangan sistem geodemografi kepada beberapa peringkat yang bermula sebelum tahun 1960, pengenalan kepada sistem geodemografi komersial dan revolusi sistem geodemografi yang bermula pada tahun 1990-an. Selain itu, perbandingan tentang jumlah pembolehubah, sumber data dan hierarki kluster juga dimasukkan dalam bahagian ini. Bab ini juga membincangkan sepintas lalu mengenai data banci penduduk dan perumahan yang merupakan input penting dalam proses pembangunan sistem geodemografi dan penggunaan sistem maklumat geografi (GIS) sebagai alat utama untuk menyimpan, mengubah dan memaparkan data secara ruangan.

Bab tiga pula mengandungi kerangka konsep dan metodologi kajian yang memberikan gambaran umum bagaimana keseluruhan kajian dijalankan. Kerangka konsep kajian menjelaskan secara teoritikal bagaimana kajian ini dilakukan dan mencapai matlamat untuk membentuk satu sistem geodemografi di dalam kawasan kajian. Metodologi kajian yang digunakan pula lebih tertumpu kepada penggunaan kaedah perlombongan data dengan tumpuan khusus kepada kaedah pemilihan data dan pembentukan kluster kejiranan.

Bab empat merupakan bab analisis yang merupakan perincian kepada perbincangan yang dilakukan dalam bab tiga. Bab ini mengandungi perbincangan terperinci sumber data yang digunakan, pemilihan pembolehubah yang bersesuaian dengan menggunakan analisis komponen utama dan proses-proses yang berlaku dalam menjalankan analisis kluster dengan menggunakan algoritma *k*-means.

Bab lima pula membincangkan tentang hasil yang diperolehi daripada analisis yang telah dijalankan dalam bahagian sebelumnya. Kluster-kluster yang dihasilkan ini dilabelkan dengan nama yang khusus. Proses melabelkan kluster ini dilakukan dengan meneliti ciri-ciri yang terdapat dalam pembolehubah yang mewakilinya dan juga melalui penelitian ciri-ciri sebenar kawasan di lapangan. Bab enam merupakan bab terakhir tesis ini yang merupakan bab perbincangan dan kesimpulan. Bab ini membincangkan tentang keseluruhan tesis dan meneliti semula objektif-objektif yang digariskan dalam bab pertama. Penilaian pencapaian objektif dan jawapan kepada persoalan kajian turut dimuatkan dalam bahagian ini.

BAB DUA

TINJAUAN BACAAN

2.0 Pengenalan

Konsep-konsep asas sistem pengkelasan, sistem geodemografi (SG), isu dan masalah serta kepentingan kajian telah dibincangkan secara sepintas lalu dalam bahagian pengenalan. Bab ini pula membincangkan tentang kajian-kajian terdahulu yang berkaitan konsep-konsep asas sistem geodemografi, peringkat-peringkat pembangunan sistem pengkelasan, kaedah perlombongan data dan pembentukan kluster kejiranan. Dalam bahagian ini, pengkaji turut membincangkan peringkat-peringkat perkembangan sistem geodemografi yang bermula pada tahun 1880-an hingga membawa kepada pengenalan sistem geodemografi terbuka dalam bidang akademik. Bahagian ini juga membincangkan dengan terperinci tiga komponen utama yang menjadi kerangka asas sistem geodemografi.

2.1 Data Banci Penduduk dan Perumahan

Banci penduduk dan perumahan merupakan kaedah yang paling berkesan untuk mengutip maklumat terperinci tentang ciri-ciri penduduk, persekitaran hidup dan keadaan perumahan (Smith, 2003; Nair dan Norhayati, 2006). Ia merupakan proses pengutipan data statistik terbesar yang dijalankan oleh Jabatan Perangkaan Malaysia (JPM). Data yang diperolehi melalui banci penduduk dan perumahan ini sangat berguna kepada penggubalan dasar dan perancangan pembangunan di Malaysia selain daripada memberi input kepada pentadbiran kerajaan dan kegunaan penyelidikan (Abdul Rahman dan Norfariza, 2007). Banci penduduk dan perumahan menurut Rabieyah dan Siti Asiah

(2000) secara prinsipnya membawa maksud pengumpulan data yang bertujuan untuk menyediakan maklumat statistik terperinci populasi penduduk dan perumahan. Prinsip pengutipan data banci ini menjadikan data banci penduduk dan perumahan Malaysia satu-satunya banci yang menyediakan maklumat bilangan penduduk dan isi rumah berserta dengan pelbagai maklumat ciri-ciri demografi, sosial dan ekonomi. Selain itu, maklumat keseluruhan tempat kediaman mengikut jenis, ciri-ciri struktur binaan dan ketersediaan kemudahan asas bagi setiap tempat kediaman juga turut dikutip dalam banci yang dijalankan.

Banci penduduk dan perumahan di negara ini telah mula dikutip sejak dari awal abad ke 20. Banci terperinci pertama yang dilakukan ialah pada tahun 1891 yang meliputi beberapa buah negeri seperti Pulau Pinang, Melaka, Singapura, Perak, Pahang, Negeri Sembilan dan Selangor. Banci kedua pula dilakukan pada tahun 1901 dan diikuti dengan banci ketiga pula dijalankan pada tahun 1911. Banci tahun 1931, 1947 dan 1957 pula dijalankan oleh kerajaan British. Banci tahun 1957 merupakan banci terakhir dijalankan oleh kerajaan British sebelum Persekutuan Tanah Melayu mencapai kemerdekaan pada 31 Ogos 1957. Selepas penubuhan negara Malaysia pada tahun 1963, kerajaan Malaysia telah menjalankan banci pertama pada tahun 1970 dan diikuti dengan banci tahun 1980, 1991 dan tahun 2000. Banci tahun 2000 merupakan banci penduduk dan perumahan ke empat yang dijalankan oleh kerajaan Malaysia (Abdul Rahman dan Norfariza, 2007).

Menurut Rabieyah dan Siti Asiah (2000) dan Rabieyah dan Roszaini (2002), tujuan utama data banci penduduk dan perumahan yang dikutip oleh JPM ialah untuk menyediakan input bagi pembentukan rancangan lima tahun pembangunan Malaysia.

Maklumat yang diperolehi daripada data banci ini menjadi input penting untuk menilai dan membentuk polisi, projek dan program bagi dimasukkan kedalam pelan pembangunan lima tahun (Malaysia, 2001). Jadual 1.2 menunjukkan senarai rancangan pembangunan lima tahun Malaysia dan sumber data banci penduduk yang digunakan sebagai input perancangan.

Jadual 2.1 : Pelan pembangunan dan data banci yang digunakan

Pelan Pembangunan	Tempoh	Tahun
Rancangan Malaya Pertama	1956 - 1960	1957
Rancangan Malaya Kedua	1961 - 1965	1957
Rancangan Malaysia Pertama	1966 - 1970	1957
Rancangan Malaysia Kedua	1971 - 1975	1970
Rancangan Malaysia Ketiga	1976 - 1980	1970
Rancangan Malaysia Keempat	1981 - 1985	1980
Rancangan Malaysia Kelima	1986 - 1990	1980
Rancangan Malaysia Keenam	1991 - 1995	1991
Rancangan Malaysia Ketujuh	1996 - 2000	1991
Rancangan Malaysia Kelapan	2001 - 2005	2000

Sumber: Rabieyah dan Siti Asiah (2000)

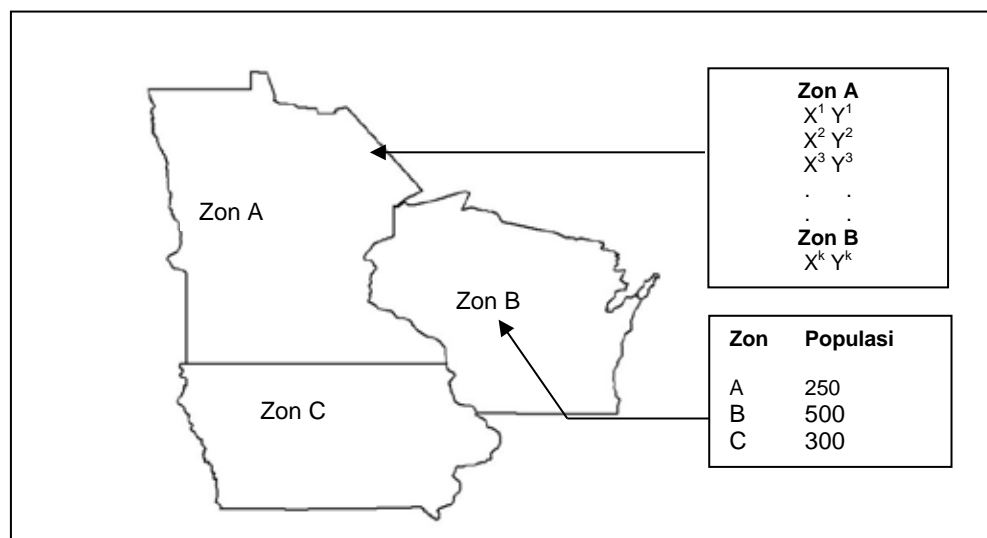
Jadual 2.1 menunjukkan penggunaan data banci sebagai input untuk merangka pelan pembangunan di negara ini telah bermula sejak dari tahun 1956. Pelan pembangunan yang pertama dibangunkan di negara ini dikenali sebagai Rancangan Malaya Pertama (RM-1) yang bermula pada tahun 1956 hingga 1960. Tumpuan utama pembangunan dalam RM-1 menjurus kepada pertumbuhan ekonomi dan menyediakan kemudahan asas seperti pendidikan, kesihatan, pembentungan, bekalan air dan elektrik. Pelan pembangunan kedua sebelum pembentukan negara Malaysia juga turut menggunakan data banci tahun 1957 sebagai input perancangan, pelan pembangunan ini dikenali sebagai Rancangan Malaya Kedua (RM-2) yang bermula dari tahun 1961 hingga 1965.

Rancangan Malaysia Pertama (RMK-1) yang bermula pada tahun 1966 hingga 1970 juga menggunakan data banci tahun 1957 untuk membuat perancangan pembangunan, tumpuan utama RMK-1 ialah dalam aspek pembangunan pendidikan, kesihatan awam dan penambahan peluang pekerjaan. Walau bagaimanapun, data banci tahun 1957 hanya boleh menyediakan maklumat mengenai taburan, saiz dan ciri-ciri populasi sahaja kerana maklumat mengenai perumahan hanya mula dikutip pada banci tahun 1970 (Rabieyah dan Siti Asiah, 2000). Pelan pembangunan yang menggunakan data banci tahun 1970 ialah Rancangan Malaysia Kedua (RMK-2) dan Rancangan Malaysia Ketiga (RMK-3).

Data banci penduduk dan perumahan tahun 2000 pula menurut Jabatan Perangkaan Malaysia (2001) digunakan sepenuhnya untuk menyediakan dan memantau Rancangan Malaysia Kelapan (RMK-8) yang bermula pada tahun 2001 hingga tahun 2005. Objektif utama banci tahun 2000 ini ialah untuk menyediakan maklumat terperinci ciri-ciri demografi, sosial dan ekonomi pada peringkat kawasan statistik terkecil (*small area statistics*). Objektif banci tahun 2000 ini juga cuba mencapai matlamat penggunaan data yang tidak terhad kepada pihak kerajaan sahaja, tetapi turut dimanfaatkan oleh sektor-sektor lain seperti usahawan, penganalisis industri, ahli akademik, penyelidik dan ahli politik (Malaysia, 2001). Abdul Rahman dan Norfariza (2007) menyatakan bahawa pengutipan data statistik dalam jumlah yang besar memerlukan program pengagihan yang teliti. Perancangan pengagihan dan memanfaatkan sepenuhnya data banci tahun 2000 telah bermula sejak dari tahun 1997.

Antara perancangan strategik pengagihan data banci penduduk dan perumahan Malaysia tahun 2000 menurut Abdul Rahman dan Norfariza (2007) juga termasuklah melalui

penerbitan laporan, penggunaan internet, produk pemetaan dan laporan analisis khusus dalam bidang-bidang tertentu seperti perbandaran, migrasi, populasi orang asli dan golongan kurang upaya. Matlamat Jabatan Perangkaan Malaysia untuk memaksimumkan penggunaan data banci menurut Rabieyah dan Siti Asiah (2000) pula turut disokong oleh perkembangan yang pesat dalam bidang teknologi komputer. Oleh kerana banci penduduk dan perumahan melibatkan penggunaan data dalam jumlah yang besar, penggunaan sistem maklumat geografi (GIS) yang berupaya untuk merekod, menyimpan, mencapai semula, menganalisis dan memaparkan semula maklumat ruangan membolehkan data banci dimanfaatkan dengan lebih mudah.



Rajah 2.1 : Model data vektor GIS
 Sumber: Martin dan Longley (1995)

Penggunaan GIS membolehkan penghasilan peta tematik, dan carian ruangan bagi tujuan paparan atau analisis lanjutan dilakukan pada pelbagai peringkat geografi. Selain itu keupayaan GIS untuk memaparkan data secara ruangan juga membolehkan taburan populasi, isi rumah dan tempat kediaman tunjukkan dengan lebih berkesan (Rabieyah dan Siti Asiah, 2000). Seperti contoh yang ditunjukkan dalam rajah 2.1, setiap

maklumat koordinat yang mewakili sempadan zon banci dihubungkan dengan ciri-ciri populasi yang terdapat dalam zon tersebut. Objek-objek ruangan dan bukan ruangan ini disimpan sebagai dua entiti yang berbeza dalam pangkalan data. Martin dan Longley (1995) menyatakan bahawa data digital yang menghubungkan atribut ruangan dan bukan ruangan ini juga dirujuk sebagai model data vektor.

Kesedaran umum tentang kepentingan untuk memanfaatkan sepenuhnya data dalam jumlah yang besar bukan perkara baru, ia telah diperkatakan sejak dari tahun 1980-an lagi. Antara pernyataan popular yang sering dirujuk oleh para pengkaji ialah "*we are drowning in information but starved for knowledge*" yang disebut dalam kajian Naisbitt pada tahun 1982 (Larose, 2005). Keadaan ini menunjukkan bahawa masyarakat pada hari ini bukan lagi berhadapan dengan masalah kekurangan data, tetapi sebaliknya menghadapi kesukaran untuk memanfaatkan data-data tersebut. Han dan Kamber (2006) pula menggambarkan keadaan masyarakat pada hari ini sebagai "*data rich but information poor*" yang merujuk kepada kerja-kerja pengutipan data secara berterusan sehingga menyebabkan pembangunan pangkalan simpanan data secara besar-besaran yang melampaui keupayaan semulajadi manusia untuk menganalisis data-data tersebut.

Vickers (2007) menyatakan bahawa masalah utama untuk menganalisis data banci penduduk dan perumahan ialah disebabkan oleh kerencaman data banci itu sendiri. Jumlah data banci yang besar dan mengandungi berbagai maklumat adalah sukar untuk difahami sekiranya maklumat-maklumat yang terdapat dalam set data tersebut tidak diringkaskan. Salah satu daripada kaedah yang boleh digunakan untuk mengubah data yang kompleks ini kepada bentuk yang lebih mudah difahami, ialah melalui penghasilan sistem geodemografi (SG).

Vickers dan Rees (2007) dalam kajian penggunaan data banci tahun 2001 di UK menyatakan bahawa jumlah kawasan hasilan (*output area*) yang lebih daripada 230 ribu unit itu terlalu kompleks untuk diproses oleh otak manusia pada sesuatu masa. Ia perlu diproses terlebih dahulu dengan menggunakan analisis kluster yang membahagikan kawasan-kawasan hasilan kepada kumpulan-kumpulan yang mempunyai persamaan maksimum sesama mereka dan mempunyai perbezaan maksimum dengan kumpulan yang lain.

Proses pembentukan kumpulan ini juga melibatkan pengurangan data yang membolehkan pemerhatian terhadap perbezaan corak taburan dilakukan dan seterusnya meningkatkan pemahaman pengkaji terhadap ciri-ciri di sesebuah kawasan. Ini bersesuaian dengan pandangan Lobo et al. (2004) yang menyatakan bahawa walaupun proses meringkaskan data banci yang besar ini mempunyai pelbagai masalah, sistem pengkelasan ini sebenarnya telah berjaya menjadi alat bantu yang berkesan dalam pelbagai bidang. Kelebihan ini menyebabkan kajian pembangunan sistem geodemografi perlu dijalankan supaya kualiti dan kebolegunaan sistem ini mampu ditingkatkan.

Kesukaran untuk membangunkan pangkalan data individu ini membawa kepada pengenalan sistem geodemografi yang mengelaskan individu mengikut tempat mereka tinggal yang dirujuk sebagai lokasi geografi. Pada setiap peringkat kelas lokasi geografi ini, maklumat secara agregat daripada data banci seperti umur pertengahan, pendapatan isi rumah, pemilikan dan kepadatan penduduk boleh diketahui (Harris et al., 2005).

Abdul Rahman dan Norfariza (2007) menjelaskan bahawa beberapa kriteria perlu diikuti bagi menentukan bidang yang bersesuaian untuk dijadikan sebagai soalan banci. Soalan yang dikemukakan perlu mengambilkira kesesuaian keperluan pengguna yang pelbagai, mempunyai perbandingan secara maksimum dengan negara lain, maklumat yang boleh dimanfaatkan oleh masyarakat umum dan kemampuan kewangan untuk menjalankan banci. Perbandingan pemilihan soalan yang dimasukkan dalam borang banci pada tahun 1970, 1980, 1991 dan 2000 adalah seperti yang ditunjukkan dalam jadual 2.2.

Jadual 2.2: Perbandingan soalan banci tahun 1970, 1980, 1991 dan 2000

SOALAN BANCİ	1970	1980	1991	2000
A) POPULASI				
CIRI-CIRI GEOGRAFI				
Tempat individu ditemui pada hari banci	P	P	P	O
Tempat tinggal biasa pada hari banci	O	O	P	P
CIRI-CIRI DEMOGRAFI DAN SOSIAL				
Jantina	P	P	P	P
Umur	P	P	P	P
Tarikh lahir	P	P	P	P
Status perkahwinan	P	P	P	P
Bangsa	P	P	P	P
Agama	P	P	P	P
Status warganegara	P	P	P	P
Warna kad pengenalan	P	P	O	O
Dialek	P	P	O	O
Kecacatan	P	P	O	P
KADAR KELAHIRAN DAN KEMATIAN				
Bilangan kelahiran hidup kanak-kanak	P	P	O	P
Bilangan kanak-kanak hidup	P	P	O	P
Umur pada perkahwinan pertama	O	P	O	O
Bilangan perkahwinan	P	P	O	O
Tahun perkahwinan	P	O	O	O
CIRI-CIRI MIGRASI				
Tempat lahir	P	P	P	P
Tempoh masa tinggal di Malaysia	P	P	O	O
Tempoh masa tinggal di lokasi semasa	P	P	O	O
Tempat tinggal terdahulu	O	P	O	O
Alasan bermigrasi	O	P	O	O
Tempat tinggal lima tahun lepas	O	O	P	P
Tahun sampai ke Malaysia	P	O	O	P