



05-4506832



pustaka.upsi.edu.my



Perpustakaan Tuanku Bainun
Kampus Sultan Abdul Jalil Shah



PustakaTBainun



ptbupsi

PERBANDINGAN METODOLOGI PENGELASAN GEODEMOGRAFI DENGAN
MENGUNAKAN ALGORITMA K-MEANS DAN ALGORITMA *SELF-
ORGANIZING MAP*: SATU KAJIAN KES DI NEGERI PERAK

SITI NAIELAH BINTI IBRAHIM @ YUSOF



05-4506832



pustaka.upsi.edu.my



Perpustakaan Tuanku Bainun
Kampus Sultan Abdul Jalil Shah



PustakaTBainun



ptbupsi

TESIS DIKEMUKAKAN BAGI MEMENUHI SYARAT UNTUK MEMPEROLEH
SARJANA SASTERA (MOD PENYELIDIKAN)



05-4506832



pustaka.upsi.edu.my



Perpustakaan Tuanku Bainun
Kampus Sultan Abdul Jalil Shah



PustakaTBainun



ptbupsi



ABSTRAK

Kajian ini dijalankan bertujuan membandingkan dua algoritma utama yang digunakan dalam proses pembentukan pengelasan geodemografi iaitu algoritma k-means dan algoritma *Self-Organizing Map* (SOM). Salah satu daripada algoritma yang dipilih akan digunakan dalam proses pembentukan profil penduduk. Sebanyak 1184 zon banci dan 203 pembolehubah daripada data banci penduduk dan perumahan telah digunakan. Analisis algoritma k-means melibatkan beberapa proses iaitu proses pengurangan pembolehubah, penentuan bilangan kluster optimum, pengubahan dan penyeragaman data. Manakala bagi algoritma SOM sebanyak 2,000 unit jaringan telah dibentuk dan topologi heksagon telah dipilih berserta dengan fungsi kejiranan *gaussian*. Berdasarkan kepada kedua-dua algoritma, k-means telah menghasilkan lima kluster manakala algoritma SOM telah menghasilkan lapan kluster. Hasil bagi kedua-dua algoritma ini dibandingkan dengan menggunakan pengukuran pekali kekuatan. Nilai ralat struktur dan ralat kuadratik bagi algoritma k-means adalah lebih tinggi berbanding algoritma SOM. Manakala nilai ralat pengelasan bagi algoritma k-means adalah lebih rendah berbanding algoritma SOM. Dapatan ini menunjukkan bahawa algoritma k-means memberikan hasil yang lebih baik berbanding algoritma SOM dalam proses pembentukan kluster dan penghasilan pengelasan geodemografi. Pada hierarki pertama, kluster-kluster ini dilabelkan sebagai kluster luar bandar, kluster kelas pertengahan bawahan, kluster perkampungan tradisi, kluster golongan bandar dan bekerja dan kluster pinggir bandar. Kesimpulannya, perbandingan pengelasan geodemografi dengan menggunakan algoritma k-mean dan algoritma SOM telah meningkatkan pencapaian dalam bidang geodemografi dan setanding dengan negara maju yang lain. Hasil daripada kajian ini dapat membantu pihak kerajaan, pihak berkuasa tempatan dan pembuat dasar dalam merumuskan keputusan yang lebih tepat terutamanya dari segi penyediaan kemudahan kesihatan, perkhidmatan awam dan keperluan asas yang lain.





COMPARISON OF GEODEMOGRAPHY CLASSIFICATION METHODOLOGY USING K-MEANS AND SELF-ORGANIZING MAP (SOM) ALGORITHM: A CASE STUDY IN PERAK

ABSTRACT

This study aims to compare two main algorithms used in the process of creating geodemography classification namely; k-means algorithm and Self-Organizing Map (SOM) algorithm. One of the algorithms will be used for the process of creating the population profiles. A total of 1184 census zones and 203 variables from the population and housing census data were used. The k-means algorithm analysis requires several processes; the variable reduction process, determination of optimum cluster number, data transformation and data standardization. For SOM algorithm, 2,000 network units are formed and topologies hexagonal are selected along with the neighborhood Gaussian function. Based on both algorithms, the k-means and SOM algorithms have generated five clusters eight clusters respectively. The results of both algorithms were then compared by using the strength coefficient measurement. The structure error and quadratic error values of *k*-means algorithm are higher than the SOM algorithm, whereas the error classification values for *k*-means algorithm are lower than the SOM algorithm. These findings showed that the k-means algorithm has revealed better results than the SOM algorithm for the cluster creating processes and geodemography classification. In the first hierarchy, these clusters are labeled as rural, lower middle-class, traditional village, city and career and suburban. In conclusion, the method of comparing geodemographic classification by using k-means and SOM algorithms had enhanced the achievement in geodemographic field and the results are comparable with other works in the developed countries. The results from this study would help governments, local authorities and policy makers to formulate more accurate decisions, especially in terms of provision of health facilities, public services and other basic necessities.



**KANDUNGAN****Muka Surat**

PENGAKUAN	ii
PENGHARGAAN	iii
ABSTRAK	iv
ABSTRACT	v
KANDUNGAN	vi
SENARAI RAJAH	x
SENARAI JADUAL	xii
SENARAI SINGKATAN	xiv

**BAB 1****PENDAHULUAN**

1.1	Pengenalan	1
1.2	Latar Belakang Kajian	1
1.3	Permasalahan Kajian	10
1.4	Persoalan Kajian	12
1.5	Matlamat dan Objektif Kajian	13
1.6	Skop Kajian	13
1.7	Kepentingan Kajian	14
1.8	Organisasi Penulisan Bab	17
1.9	Kerangka Konseptual	19
1.10	Kesimpulan	20



BAB 2**SOROTAN LITERATUR**

2.1. Pengenalan	21
2.2. Definisi	
2.2.1 Pengelasan	22
2.2.2 Pengklusteran	23
2.2.3 Sistem Pengkelasan Geodemografi	25
2.2.4 Aplikasi Pengelasan Geodemografi	28
2.3. Algoritma Kluster	
2.3.1 K-Means	33
2.3.2 <i>Artificial Neural Network</i> (ANN)	34
2.4. Data Banci Penduduk	41
2.5. Kesimpulan	42

BAB 3**KAWASAN DAN METODOLOGI KAJIAN**

3.1 Pengenalan	43
3.2 Kawasan Kajian	44
3.2.1. Kawasan Statistik Terkecil	47
3.3 Metodologi Kajian	49
3.3.1 Data	50
3.3.2 Analisis Data	58
3.3.2.1 Algoritma K-Means	59
3.3.2.2 Self-Organizing Map (SOM)	77
3.4 Perbandingan Kaedah K-means dan SOM	94
3.4.1 Pengukuran Penilaian Kekuatan	94
3.5 Kesimpulan	95

BAB 4**DAPATAN DAN PERBINCANGAN**

4.1	Pengenalan	96
4.2	Objek Pengklusteran	97
4.3	Pembolehubah-Pembolehubah Banci	103
4.4	Algoritma K-Means	
4.4.1	Pemilihan Pembolehubah	104
4.4.1.1	Analisis Korelasi	104
4.4.1.2	Analisis Komponen Utama (PCA)	109
4.4.1.3	Analisis Varian	115
4.4.2	Senarai Pembolehubah Akhir	116
4.4.3	Penghasilan Kluster Geodemografi	119
4.4.4	Pengubahan Data	123
4.4.5	Penyeragaman Data	125
4.5	Kaedah Pengklusteran	130
4.6	Algoritma <i>Self-Organizing Map</i> (SOM)	130
4.7	Perbandingan Antara K-Means dan SOM	133
4.7.1	Pengukuran Penilaian Kekuatan	133
4.8	Kesimpulan	140

BAB 5**PEMBENTUKAN DAN PENGUJIAN SISTEM**

5.1	Pengenalan	142
5.2	Profil Kluster	143
5.3	Nama Kluster	151
5.4	Paparan Kluster Hierarki Pertama	154
5.4.1	Kluster 1: Kluster Luar Bandar	154

05-4506832	pustaka.upsi.edu.my	Perpustakaan Tuanku Bainun	PustakaTBainun	ptbupsi
	5.4.2	Kluster 2: Kluster Kelas Pertengahan Bawah	155	
	5.4.3	Kluster 3: Kluster Perkampungan Tradisi	156	
	5.4.4	Kluster 4: Kluster Golongan Bandar dan Bekerjaya	158	
	5.4.5	Kluster 5: Kluster Pinggir Bandar	159	
	5.5	Profil Kluster Hierarki Kedua	160	
	5.5.1	Pembentukan Kluster Hierarki Kedua	160	
	5.6	Analisis Sensitiviti	179	
	5.7	Perbincangan	182	
	5.8	Kesimpulan	185	
BAB 6	RUMUSAN			
	6.1	Pendahuluan	187	
	6.2	Rumusan Penyelidikan	188	
05-4506832	6.3	Kelemahan Kajian	190	
	6.4	Cadangan Kajian Lanjutan	191	
	6.5	Kesimpulan	193	
	RUJUKAN		194	
	LAMPIRAN			



SENARAI RAJAH

No. Rajah	Muka Surat
1.1 Asas dalam pembentukan pengkelasan penduduk	5
1.2 Konsep kotak hitam bagi ANN	7
1.3 Kerangka konseptual perbandingan kaedah algoritma kluster	19
2.1 Perkongsian memori dalam CPU	36
2.3 Perkongsian memori dalam ANN	36
3.1 Peta Lokasi Kajian	46
3.2 Komponen GIS	57
3.3 Proses-Proses Pembentukan SPG	61
3.4 Proses Penyediaan Pembolehubah	61
3.5 Proses Pelaksanaan Algoritma K-Means	68
3.6 Arah Aliran Penyelidikan	76
3.7 Carta alir kaedah penyelidikan menggunakan algoritma SOM	80
3.8 Topologi yang terdapat di dalam perisian SOM iaitu <i>rect</i> (kiri) dan <i>hex</i> (kanan)	85
3.9 Perbezaan bentuk SOM yang boleh diaplikasikan dalam SOM Toolbox dengan menggunakan topologi segiempat (<i>rect</i>): a) sheet sebagai default; b) cylinder; c) toroid	85
3.10 Perbandingan antara bentuk dua dimensi dan bentuk sfera SOM	86
3.11 Fungsi kadar pembelajaran bagi <i>linear</i> , <i>power</i> dan <i>inv.</i>	88
3.12 Pembahagian kumpulan utama paparan SOM	90
3.13 Kategori peta dalam paparan SOM	91
3.14 Kaedah paparan u-matrik	91
3.15 Peta pukulan (hit map) di dalam u-matrik dengan menggunakan perisian SOM	92
4.1 Pecahan Zon Banci Mengikut Daerah Pentadbiran	101
4.2 <i>Screen Plot</i> Bagi Keseluruhan Komponen	111
4.3 Pengiraan Kluster Optimum Dengan Menggunakan Kaedah K-Means	120
4.4 Pengiraan Kluster Optimum Dengan Menggunakan Kaedah Ward	121
4.5 Paparan Peta Dengan Menggunakan Algoritma SOM	131
4.6 Paparan Peta Melalui U-Matrik dan Component Parallel	132



4.7	Jumlah Kluster Dengan Menggunakan Paparan U-Matrik	133
5.1	Carta Radar Kluster 1	144
5.2	Carta Radar Kluster 2	144
5.3	Carta Radar Kluster 3	144
5.4	Carta Radar Kluster 4	144
5.5	Carta Radar Kluster 5	144
5.6	Kluster 1	154
5.7	Kluster 2	155
5.8	Kluster 3	156
5.9	Kluster 4	158
5.10	Kluster 5	159
5.11	Kluster berusia dan kurang berada	162
5.12	Kluster pekerja industri	163
5.13	Kluster pekerja industri kecil dan sederhana	165
5.14	Kluster generasi muda penghuni bandar	166
5.15	Kluster pertengahan bawahan pinggir bandar	167
5.16	Kluster pekerja sektor pelancongan	168
5.17	Kluster berbagai etnik berkeluarga	169
5.18	Kluster pekerja kolar biru berkeluarga	170
5.19	Kluster pekerja kolar biru pinggir bandar	171
5.20	Kluster generasi baru kelas pertengahan bawahan	172
5.21	Kluster muda dan bekerjaya	173
5.22	Kluster berkeluarga dan bekerjaya	174
5.23	Kluster pekerja kolar biru pelbagai etnik	175
5.24	Kluster pekerja perkhidmatan dan separa profesional	177
5.25	Kluster pekerja ladang berkeluarga	178
5.26	Hasil Analisis Sensitiviti 69 Pembolehubah Akhir	180



SENARAI JADUAL

No. Jadual	Muka Surat
1.1 Pembahagian Model ANN Mengikut Pembelajaran	9
2.1 Kajian Kegunaan SPG Dalam Bidang Pencegahan Jenayah	32
2.2 Beberapa Model ANN Yang Paling Menonjol Dalam Sejarah Pembangunannya	38-40
3.1 Taburan Jantina Penduduk Mengikut Daerah Di Perak Pada Tahun 2000	44
3.2 Pelan Pembangunan dan Data Banci Yang Digunakan	52
3.3 Petunjuk Korelasi Guilford	63
4.1 Kekerapan Zon Banci Di Perak	98-100
4.2 Pertumbuhan Penduduk Perak Tahun 1991 Hingga Tahun 2000	102
4.3 Pembolehubah Yang Berkorelasi	105-106
4.4 Jadual Matrik Korelasi Bagi Pembolehubah Umur	108
4.5 Penentuan Saiz Komponen Dengan Menggunakan <i>Screen Plot</i>	110
4.6 Saiz Komponen Dengan Menggunakan Kaedah Kriteria Kaiser	113
4.7 Senarai Pembolehubah Dengan Nilai Varian Yang Kecil	116
4.8 Senarai Pembolehubah Akhir Untuk Analisis Kluster	117-119
4.9 Taburan Kluster Data Yang Belum Diseragamkan	122
4.10 Nilai Statistik Kepencongan Asal, Kuasa Dua dan Logaritma	123-125
4.11 Taburan Kluster Yang Diseragamkan Menggunakan Penyeragaman Berkadar	126
4.12 Taburan Kluster Yang Diseragamkan Menggunakan Z-Skor	127
4.13 Taburan Kluster Yang Diseragamkan Menggunakan Penyeragaman Persentil	128
4.14 Perbezaan Nilai SSE Bagi Setiap Kaedah Penyeragaman	129
4.15 Jadual Ralat Kuasa Dua Bagi Pengiraan Ralat Kuadratik	134-138
4.16 (a) Jadual Pengelasan Pusat bagi Algoritma SOM	139
4.16(b) Jadual Pengelasan Pusat bagi Algoritma K-means	139
4.17 Jadual Pengukuran Penilaian Kekuatan Bagi Algoritma SOM Dan K-Mean	140
5.1 Jadual Anova 62 Pembolehubah Akhir	145-151
5.2 Label Kluster Hierarki Pertama dan Hierarki Kedua	152-153



5.3	Pengiraan Bilangan Kluster Optimum	161
5.4	Pembolehubah yang Menyebabkan Pengurangan Daripada Pusat Kluster	181
5.5	Pembolehubah Yang Mempunyai Kesan Kecil Terhadap Perubahan Jarak Purata	182



SENARAI SINGKATAN

ACORN	A classification of Residential Neighbourhood
ANN	<i>Artificial Neural Network</i>
BP	Blok Perhitungan
CPU	Unit Pemprosesan Berpusat
DEO	Jabatan Alam Sekitar
GIS	<i>Geographical Information Systems</i>
JPM	Jabatan Perangkaan Malaysia
NKRA	National Key result Areas
OA	<i>Output Area</i>
OAC	<i>Output Area Classification</i>
ONS	<i>Office for National Statistics</i>
PCA	<i>Principal Component Analysis</i>
SOM	<i>Self Organizing Map</i>
SPG	Sistem Pengelasan Geografi
SPSS	<i>Statistical Package for The Social Sciences</i>
TK	Tempat Kediaman





SENARAI SINGKATAN

ACORN	A classification of Residential Neighbourhood
ANN	<i>Artificial Neural Network</i>
BP	Blok Perhitungan
CPU	Unit Pemprosesan Berpusat
DEO	Jabatan Alam Sekitar
GIS	<i>Geographical Information Systems</i>
JPM	Jabatan Perangkaan Malaysia
NKRA	National Key result Areas
OA	<i>Output Area</i>
OAC	<i>Output Area Classification</i>
ONS	<i>Office for National Statistics</i>
PCA	<i>Principal Component Analysis</i>
SOM	<i>Self Organizing Map</i>
SPG	Sistem Pengelasan Geografi
SPSS	<i>Statistical Package for The Social Sciences</i>
TK	Tempat Kediaman





BAB 1

PENDAHULUAN

1.1 Pengenalan



Bab ini membincangkan tentang latar belakang kajian, permasalahan kajian, matlamat dan objektif kajian, persoalan kajian, organisasi penulisan bab serta kerangka konseptual. Selain itu, skop kajian juga akan dibincangkan bagi memberikan gambaran sebenar halangan serta batasan kajian ini. Kepentingan kajian juga penting bagi memastikan kajian ini dapat memberikan manfaat kepada pihak berwajib.

1.2 Latar Belakang Kajian

Proses pembandaran berlaku seiring dengan perkembangan pembangunan negara sekaligus menyebabkan berlaku pertumbuhan penduduk yang pesat dalam kawasan



bandar khususnya selepas kemerdekaan dan pelaksanaan Dasar Ekonomi Baru (DEB).

Keadaan ini bukan sahaja menyebabkan berlakunya penghijaran dari kawasan luar bandar ke kawasan bandar, tetapi turut menyebabkan berlakunya rombakan gaya hidup, struktur pekerjaan, dan pendapatan. Perubahan yang berlaku ini turut memberikan perubahan kepada statifikasi sosial dengan kemunculan kelas sosial yang baharu (Abdul Rahman Embong, 2011). Menyedari masalah yang dihadapi oleh pihak kerajaan untuk menyalurkan peruntukan dan menyediakan kemudahan seperti kemudahan infrastuktur dan perumahan kepada kumpulan sasaran, kajian pembangunan sistem pengkelasan geodemografi yang menggunakan sempadan kawasan terkecil ini perlu dijalankan. Sumber data utama yang digunakan ialah blok penghitungan yang diperolehi daripada Jabatan Perangkaan Malaysia (JPM) dan maklumat-maklumat lain seperti pendidikan, perumahan, pekerjaan dan sosio-ekonomi yang diperolehi daripada data Banci Penduduk dan Perumahan Malaysia

Istilah geodemografi mula wujud pada awal tahun 1970an apabila wujudnya sistem *A classification of Residential Neighbourhood* (ACORN) yang dibangunkan oleh CACI di Great Britain dan sistem PRIZM oleh Claritas di Amerika Syarikat (Thomas & Stephen, 2004). Sejak sistem itu diperkenalkan alatan tersebut telah menjadi satu piawaian kerana ianya boleh digunakan untuk mengenalpasti ciri penduduk secara khusus. Oleh kerana berlaku ledakan di dalam *Geographical Information System* (GIS), satu istilah baru telah digunakan iaitu geodemografi yang mana ianya merujuk kepada penggunaan sistem pengkelasan kejiranan untuk aktiviti khidmat pelanggan dan aktiviti pemasaran (Bolton Council, 2008).



Demografi dirujuk sebagai pembahagian kepada penduduk mengikut data-

data penduduk dan tingkah laku (Schwabe & O'Donovan, 1999). Geodemografi adalah kajian ke atas penduduk mengikut tempat tinggal mereka (Sleight, 2014). Menurut O'Malley et al. (1995) istilah geodemografi diperolehi daripada gabungan dua istilah berbeza, iaitu geografi dan demografi. Demografi ditakrifkan oleh Siegel (2002) sebagai kajian saintifik tentang penduduk yang meliputi pada kajian terhadap ciri-ciri seperti saiz, taburan, struktur umur, jantina, aspek sosio-ekonomi, kadar kelahiran, kematian dan migrasi.

Pendefinisian geodemografi menurut Tedd et al. (1997) adalah pengkelasan kejiranan mengikut maklumat demografi dan data gaya hidup. Geodemografi mampu memberikan maklumat berkaitan ciri penduduk mengikut kawasan kejiranan mereka. Kajian di dalam bidang geodemografi mempunyai pelbagai kegunaan dalam sektor awam dan sektor swasta. Secara umumnya, ianya digunakan untuk mengenal pasti kumpulan penduduk di dalam sesebuah populasi yang mampu untuk menggambarkan tingkah laku penduduknya mengikut kawasan (Bolton Council, 2008).

Pengelasan secara umumnya merujuk kepada proses menyusun dan mengumpul sesuatu benda mengikut kumpulan, berasaskan kepada kesamaan dan ketidaksamaan. Hasil daripada proses ini, ianya dikenal pasti sebagai sistem pengkelasan (Batley, 2005). Pelopor kepada kajian taksonomi ialah Aristotle, iaitu seorang ahli falsafah Greek yang terkenal. Aristotle telah memperkenalkan dua konsep pengkelasan yang utama iaitu pengkelasan organisma mengikut jenis dan pengkelasan mengikut definisi binomial (Tilton, 2009). Kajian berkaitan dengan pengkelasan berasal daripada taksonomi yang melibatkan kajian tentang sains





pengkelasan dengan tumpuan khusus kepada pengkelasan organisma hidup dan hubungan semulajadi yang wujud (Tilton, 2009).

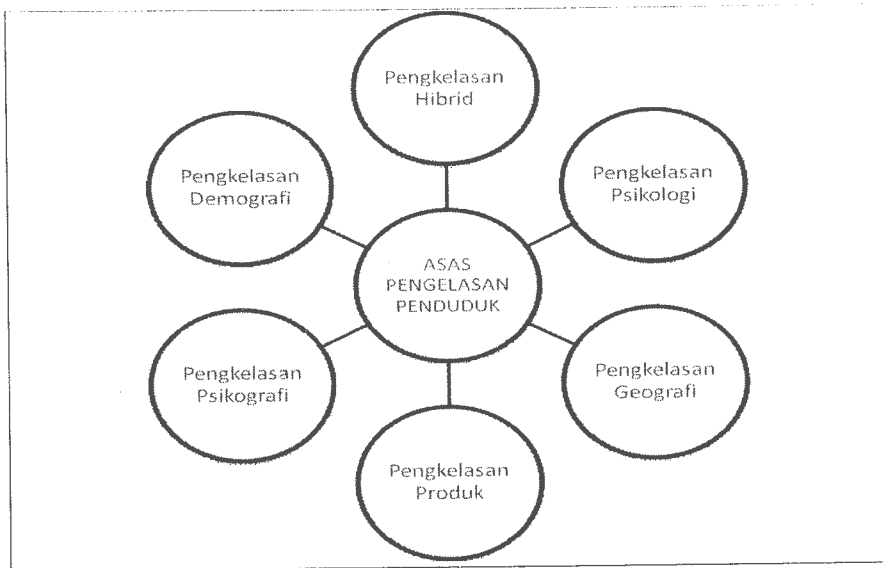
Gordon (1999) melihat sistem pengkelasan sebagai suatu alat yang boleh digunakan untuk menyiasat perhubungan yang wujud dalam satu kumpulan objek, sama ada objek tersebut diringkaskan atau tidak dalam kumpulan yang lebih kecil. Kumpulan kecil dibentuk dalam mengelaskan objek yang serupa dalam kelas yang sama. Malahan, terdapat juga teori-teori pembangunan pengkelasan yang dibincangkan oleh beberapa orang pengkaji terdahulu yang melibatkan sistem pengkelasan dalam bidang kajian yang berbeza (Gordon, 1999). Walau bagaimanapun, matlamat utama penggunaan sistem pengkelasan adalah untuk meringkaskan data dan membuat tanggapan terhadap corak yang wujud di dalamnya.



Rajah 1.1 menunjukkan enam asas utama yang boleh digunakan untuk pengelasan penduduk, antaranya adalah pengelasan hibrid, pengelasan psikologi, pengelasan geografi, pengelasan produk, pengelasan psikografi dan pengelasan demografi. Pengelasan hibrid merupakan pengelasan campuran seperti campuran pengelasan demografi dan pengelasan geografi dan telah membentuk pengelasan geodemografi manakala pengelasan psikologi mengelaskan beberapa jenis masalah mental. Pengelasan produk mengelaskan pengguna berdasarkan persamaan yang terdapat dalam pemilihan jenama, kecenderungan penggunaan dan penilaian pulangan kepuasan daripada produk tersebut manakala pengelasan psikografi mengelompokkan penduduk mengikut ciri psikologi yang sepadan, nilai sosio budaya dan cara hidup. Pengelasan geografi mengelaskan penduduk kepada lokasi mereka berada, manakala



pengelasan demografi membahagikan penduduk kepada ciri yang sama seperti jantina, umur, pendapatan, pekerjaan, pendidikan dan saiz isi rumah.



Rajah 1.1 Asas dalam pembentukan pengelasan penduduk. Dirujuk dari

Bagi membentuk sistem geodemografi penduduk di sesebuah kawasan, adalah penting dalam memastikan agar:

- Maklumat yang dipaparkan adalah mengikut kawasan kajian.
- Sistem itu merangkumi keseluruhan kawasan dan unit ruangan yang paling kecil.
- Definisi bagi kategori geodemografi adalah saintifik bagi mengelakkan sifat bias akan mempengaruhi pembahagian atau penerangan dalam sesebuah kumpulan.
- Keseluruhan maklumat asas demografi digunakan sepenuhnya bagi memastikan segala maklumat berkaitan populasi dimasukkan.



Pengelasan boleh dilakukan dengan menggunakan *Articial Neural Network*

(ANN). Definisi formal ANN yang dikemukakan oleh Haykin (2008) adalah ANN merupakan satu prosesor pembahagian selari besar-besaran yang mempunyai kecenderungan semulajadi untuk menyimpan pengetahuan melalui pengalaman dan menjadikannya sedia untuk digunakan. Ketika ini ANN sudah mampu untuk menjalankan beberapa tugas seperti pengelasan, pengecaman, penganggaran, ramalan, pengklusteran, simulasi memori dan beberapa tugas lain lagi. ANN menyerupai fungsi otak manusia melalui 2 perkara iaitu:

- 1 pengetahuan yang diperolehi daripada rangkaian adalah melalui proses pembelajaran
- 2 kekuatan sambungan antara neuron dikenali sebagai pemberat sinaptik yang digunakan untuk menyimpan pengetahuan.



Kebanyakan model ANN lebih mirip kepada beberapa teknik analisis statistik yang popular seperti *generalized linear model*, *polynomial regression*, *nonparametric regression*, *discriminant analysis*, *projection pursuit regression*, *principal components*, *cluster analysis* dan beberapa teknik statistik yang lain (Sarle, 1994). Salah satu daripada perbezaan ANN dengan kaedah analisis statistik ialah ia boleh mengesan corak perhubungan di dalam satu set data yang kecil dan boleh bekerja dalam keadaan data input yang tidak lengkap. Hal ini kerana ANN bersifat boleh untuk mengadaptasi perubahan yang berlaku di sekelilingnya melalui proses pembelajaran (Dawson & Wilby, 1998). Keupayaan yang ada pada ANN membolehkannya mengadaptasi dengan perubahan yang berlaku di persekitarannya ketika proses pembelajaran berlaku (Dawson & Wilby, 1998) serta mampu untuk



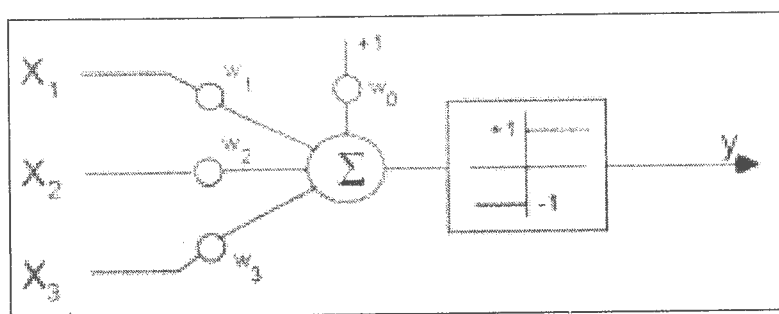


bertoleransi dengan sebarang kesilapan. Struktur data input yang dikendalikan secara bebas dan tidak mempunyai had maksimum membolehkan ANN memperbaiki kesilapannya apabila ia dilatih secara berulang (Zhang & Stanley, 1997).

Selain tu, sifat ANN yang boleh belajar daripada sampel dan membentuk perhubungan kotak hitam untuk melihat sesuatu perhubungan di antara data-data juga menyebabkannya tidak memerlukan algoritma matematik untuk membuat permodelan (Zhang & Stanley, 1997). Rajah 1.2 menunjukkan konsep hitam bagi ANN iaitu ianya menerima beberapa input yang berbeza dan akan melakukan beberapa proses. Seterusnya ia akan mengeluarkan satu output yang signifikan dan digunakan dalam kajian. Namun, keadaan ini akan menyebabkan kesukaran apabila pengkaji mahu menerangkan berkaitan perhubungan matematik dengan menggunakan model ANN yang mereka gunakan kerana ianya tidak dapat difahami oleh pengkaji. Sarle (1994)



menyatakan, pengguna ANN terdiri daripada jurutera, ahli fizik, ahli neurologi, psikologi dan ahli sains komputer yang mana kebanyakan mereka mempunyai pengetahuan yang sedikit berkaitan dengan aplikasi statistik dan sering gagal untuk memahami bagaimana kaedah matematik digunakan.



Rajah 1.2 Konsep kotak hitam bagi ANN





ANN mewakili disiplin kejuruteraan berkaitan dengan sistem penyesuaian pemprosesan maklumat bukan program yang mana ianya dibangunkan untuk menyatukan (menukar atau pemetaan) antara objek sebagai tindak balas kepada persekitaran mereka. ANN adalah sejenis senibina perkomputeran selari yang besar-besaran berdasarkan kepada pengkodan maklumat seperti otak dan pemprosesan model menunjukkan kelakuan seperti otak seperti pembelajaran, hubungan, pengkategorian, kelaziman, fungsi pengeluaran dan pengoptimuman. ANN dibangunkan untuk cuba mengatasi dan meniru fungsi kebijaksanaan biologikal dengan ciri-ciri yang boleh dipelajari sama ada melalui contoh atau pengalaman. ANN juga berupaya untuk mengira dengan cepat apabila berurusan dengan jumlah set data yang besar (Craig, 2001).

Berikut adalah pembahagian model ANN mengikut pembelajaran iaitu



pembelajaran dipantau dan pembelajaran tanpa pemantauan sebagaimana yang ditunjukkan dalam jadual 1.1.



Jadual 1.1

Pembahagian model ANN mengikut pembelajaran

	Panggil Semula ke Belakang	Panggil Semula ke Hadapan
Pembelajaran dipantau	Brain-state-in box	Perceptron
	Fuzzy Cognitive Map	Multi-Layer Perceptron
		ADALINE, MADALINE
		Boltzmann Machine
		CMAC
		Radial Basis Function
		Network
		Probabilistic Neural
		Network
		General Regression Neural
Pembelajaran tanpa pantauan		Network
		Modified Probabilistic
		Neural Network
	Adaptive Resonance Theory	Kohonen
		Counterpropagation
	Hopfield Neural Network	Fuzzy Associative Memory
	Bidirectional Associative	
	Memory	

Dirujuk dari Zaknich, 2003



1.3 Permasalahan Kajian

Dewasa ini, para pengguna dan penyelidik tidak lagi menghadapi 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 (Kamarul, 2013). Pernyataan ini jelas membuktikan masalah sebenar yang sedang dihadapi oleh pengguna bukan ketiadaan data tetapi jumlah data yang semakin bertambah sehingga menimbulkan kesukaran untuk memanfaatkan data-data tersebut (Larose, 2005). Malahan, lambakan data ini juga menyukarkan pengguna dalam menganalisis data agar hasil yang diperolehi itu cepat dan efisien.

Bagi sesetengah negara maju seperti Jepun, Australia dan Kanada, proses pengutipan data banci dilaksanakan dalam tempoh masa lima tahun, manakala di negara-negara sedang membangun kerja-kerja merekodkan maklumat mengenai penduduk ini dijalankan 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). Data banci Penduduk dan Perumahan Negara merupakan sumber data utama dalam proses pembentukan sesuatu sistem geodemografi bersifat rencam.

Proses pengutipan data secara berterusan akan menyukarkan para pengguna kerana bilangan data mentah semakin bertambah setiap kali banci dilakukan. Hal ini





kerana, data tersebut akan disesuaikan dengan maklumat penduduk bagi memastikan ketepatan maklumat penduduk di suatu kawasan. Bagi mengatasi masalah ini, kaedah analisis yang boleh meringkaskan maklumat banci tersebut perlu dilakukan kepada bentuk yang lebih ringkas agar boleh memudahkan capaian kepada pengguna dan mudah untuk difahami. Oleh itu, pengkelasan sistem geodemografi dengan menggunakan algoritma pengklusteran dalam penyelidikan ini merupakan kaedah yang berkesan untuk mengurangkan jumlah data dan menyusun data tersebut dalam kumpulan kecil yang lebih mudah difahami dan diuruskan. Di dalam konteks kajian ini, sejumlah 203 pembolehubah berpotensi untuk proses analisis pengelasan geodemografi dan 1,884 unit zon banci di negeri Perak digunakan sebagai input dalam pembentukan sistem ini.



Penyelidikan ini dijalankan disebabkan oleh kesukaran mendapatkan maklumat individu dalam sesebuah kawasan. Maklumat individu seperti umur, tempat tinggal, pekerjaan, taraf pendidikan penting terutamanya bagi kajian demografi. Proses pembentukan pangkalan data yang menyimpan maklumat individu hampir mustahil untuk direalisasikan di dalam dunia nyata. Hal ini kerana proses untuk menghitung segala maklumat yang berkaitan sangat sukar dijalankan kerana ianya bukan sahaja melibatkan kos yang besar, malah data yang dikutip ini perlulah disimpan secara sistematik.

Melalui kajian yang dilakukan oleh Vickers (2006), beliau mencadangkan agar kaedah algoritma konvensional iaitu kaedah algoritma k-means digunakan dalam menganalisis data banci. Begitu juga dengan kajian yang dilakukan oleh Kamarul (2013) yang mana menggunakan kaedah algoritma k-means. Namun, bagi pengkaji

