

PENDEKATAN PENGELOMPOKAN DALAM
MENGENAL PASTI INDIKATOR TERLEBIH
BELANJA BERASASKAN KUMPULAN
PENDAPATAN BOLEH GUNA

MOHD SUHAIDI BIN ABDUL RAIS

UNIVERSITI KEBANGSAAN MALAYSIA

PENDEKATAN PENGELOMPOKAN DALAM MENGENAL PASTI INDIKATOR
TERLEBIH BELANJA BERASASKAN KUMPULAN
PENDAPATAN BOLEH GUNA

MOHD SUHAIDI BIN ABDUL RAIS

PROJEK YANG DIKEMUKAKAN UNTUK MEMENUHI SEBAHAGIAN
DARIPADA SYARAT MEMPEROLEH
IJAZAH SARJANA SAINS DATA

FAKULTI TEKNOLOGI DAN SAINS MAKLUMAT
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI

2025

PENGAKUAN

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

09 Februari 2025

MOHD SUHAIDI BIN
ABDUL RAIS
P127270

LIBRARY ETSM

PENGHARGAAN

Dengan nama Allah yang Maha Pengasih lagi Maha Pemurah. Selawat dan salam ke atas junjungan besar Nabi Muhammad SAW, keluarga dan para sahabat Baginda.

Alhamdulillah, setinggi-tinggi kesyukuran ke hadrat illahi atas limpah Rahmat dan kurniaNya juga atas segala kudrat, kekuatan, ketabahan, ilham dan akhirnya atas keizinan-Nya jua untuk saya mampu menyiapkan tesis disertasi bertajuk "**Pendekatan Pengelompokan dalam Mengenal Pasti Indikator Terlebih Belanja Berasaskan Kumpulan Pendapatan Boleh Guna**". Semoga usaha dan penghasilan yang tidak seberapa ini mendapat keberkatan dan keredhaanNya jua, seterusnya dapat memberi manfaat kepada semua. Setinggi-tinggi penghargaan dan terima kasih yang tidak terhingga kepada semua pihak yang telah membantu dan menyokong sepanjang perjalanan saya untuk menyiapkan kajian ini. Pertamanya kepada Pensyarah Penyelia saya, Prof. Azuraliza Abu Bakar yang banyak memberi tunjuk ajar, idea, nasihat serta sokongan kepada saya sepanjang saya melaksanakan kajian ini. Sesungguhnya tanpa segala panduan dan bimbingan beliau, pastinya kajian dan projek ini tidak akan dapat disiapkan dengan jayanya. Seterusnya ucapan terima kasih istimewa kepada kedua ayah bonda saya, Abdul Rais Husin dan Normah Idris yang banyak memberi inspirasi dan mendoakan kejayaan saya. Kepada keluarga kecil saya, Rafidah Mohd Adnan, Muhammad Daniel & Safiya Danisyah yang banyak mendorong dan menyokong *Papa* dalam kajian ini. Terima kasih khas saya kepada Kerajaan Malaysia, khususnya melalui Jabatan Perkhidmatan Awam (JPA) yang menaja pengajian saya dan memungkinan semua ini dapat direalisasikan. Seterusnya kepada Prof. Dekan, para pensyarah dan seterusnya kepada para pegawai dan kakitangan Fakulti Teknologi Sains Maklumat (FTSM) khususnya dan Universiti Kebangsaan Malaysia (UKM) amnya yang tidak lelah mencurahkan ilmu pengetahuan, membimbing dan berkongsi pengalaman seterusnya membantu saya sepanjang tempoh pengajian saya di sini sehinggalah terhasilnya disertasi kajian ini. Kepada Kementarian, jabatan dan agensi yang memberi input dan material berguna kepada kajian ini, terutamanya Kementerian Ekonomi, Jabatan Perangkaan Malaysia (DOSM), Lembaga Penduduk dan Pembangunan Keluarga Negara (LPPKN), UPM, UM, UITM dan begitu juga kepada barisan pakar dan pegawai yang telah memberikan ulasan dan juga input-input bernas terhadap kajian saya ini. Semoga Allah membalas segala jasa kalian semua dengan kebaikan.

ABSTRAK

Pendapatan dan perbelanjaan seseorang individu atau sesebuah isi rumah merupakan antara faktor penentu utama kepada kesejahteraan dan kebebasan kewangan sesebuah institusi keluarga. Ia dapat dilihat secara langsung dari bagaimana sumber pendapatan dan kewangan ini dibelanjakan. Sungguhpun Kerajaan telah memulakan inisiatif untuk beralih kepada pendekatan Kumpulan Pendapatan Boleh Guna Isi Rumah berbanding pendekatan Kumpulan Pendapatan Isi Rumah B40, M40 dan T20 dalam mengklasifikasikan kumpulan pendapatan isi rumah di Malaysia, namun isu terlebih belanja ini masih tetap menjadi isu pokok. Kajian ini adalah bertujuan bagi mengkaji dan mencari pengetahuan (*insight*) dalam menangani isu isi rumah terlebih belanja di Malaysia. Berbanding pendekatan konvensional, pendekatan kajian adalah menggunakan pendekatan pembelajaran mesin dengan teknik pengelompokan atau dikenali sebagai pengklusteran. Set data yang digunakan adalah dari Survei Pendapatan dan Perbelanjaan Isi Rumah tahun 2022 (HIES 2022) yang diperolehi dari sumber rasmi Jabatan Perangkaan Malaysia (DOSM) dengan jumlah 57,145 data isi rumah dengan 33 atribut. Kajian dilaksanakan dengan menggunakan teknik perlombongan CRISP-DM dan seterusnya teknik perlombongan data dengan algoritma pengelompokan pemetaan yang mengaplikasikan kaedah pengukuran jarak Euclidean. Teknik k-M in kemudiannya digunakan dalam algoritma pengelompokan ini bagi mengenal pasti bilangan kluster yang optima menggunakan empat (4) kaedah digunakan iaitu Within-Cluster-Sum of Square (WCSS) atau “Siku” (*Elbow Method*), Indeks Silhouette (*Silhouette Index*), Indeks Davies-Bouldin (*Davies-Bouldin Index*) dan Indeks Calinski-Harabasz (*Calinski-Harabasz Index*). Berdasarkan syor dari teknik-teknik tersebut, nilai optima k yang disyorkan adalah k=3. Berdasarkan nilai sentroid yang diperolehi dari setiap kluster tersebut, atribut tertinggi bagi setiap kluster dikenalpasti sebagai indikator utama bagi kedua-dua kumpulan indikator sosioekonomi dan kumpulan indikator kumpulan perbelanjaan. Kemudian penilaian menggunakan teknik pengelasan Pohon Keputusan J48 dilakukan bagi mendapatkan perbandingan dapatan dan juga peraturan (*rules*) dan pengetahuan (*knowledge*) yang lebih terperinci dan menyeluruh. Maklumat yang dapat diekstrak keluar daripada analisis setiap eksperimen dalam kajian ini dapat memberi pengetahuan dan idea bahawa faktor-faktor yang mempengaruhi isi rumah di Malaysia pada tahun 2022 untuk tergolong dalam tabiat terlebih belanja.

CLUSTERING APPROACH IN IDENTIFYING OVERSPENDING INDICATORS BASED ON DISPOSABLE INCOME GROUPS

ABSTRACT

Income and expenditure of an individual or household are among the key determinants of financial well-being and freedom within a family institution. This can be directly observed through how income sources and financial resources are spent. Although the government has initiated a transition to the Disposable Household Income approach instead of the B40, M40, and T20 Household Income Groups in classifying household income levels in Malaysia, the issue of overspending remains a critical concern. This study aims to examine and gain insights into addressing household overspending issues in Malaysia. Unlike conventional approaches, this research utilizes machine learning techniques, specifically clustering methods. The dataset used is derived from the 2022 Household Income and Expenditure Survey (HIES 2022), obtained from the official source of the Department of Statistics Malaysia (DOSM), comprising 57,145 household records with 33 attributes. The study follows the CRISP-DM data mining framework and employs a clustering algorithm based on mapping techniques that apply the Euclidean distance measurement method. The k-Means technique is then applied to this clustering algorithm to determine the optimal number of clusters using four methods: Within-Cluster Sum of Squares (WCSS) or the Elbow Method, Silhouette Index, Davies-Bouldin Index, and Calinski-Harabasz Index. Based on these techniques, the optimal number of clusters recommended is $k=3$. Based on the centroids obtained from each cluster, the highest-ranking attributes in each cluster were identified as the key indicators for both the socioeconomic and expenditure indicator groups. Subsequently, classification analysis using the J48 Decision Tree technique was conducted to compare findings and derive more detailed and comprehensive rules and insights. The information extracted from the analysis of each experiment in this study provides knowledge and a deeper understanding of the factors influencing households in Malaysia in 2022 to engage in overspending behavior.

KANDUNGAN

		Halaman
PENGHARGAAN		iii
ABSTRAK		iv
ABSTRACT		v
KANDUNGAN		vi
SENARAI JADUAL		ix
SENARAI ILUSTRASI		xi
SENARAI SINGKATAN		xiii
BAB I	PENGENALAN	
1.1	Pendahuluan	1
1.2	Latar Belakang Kajian	1
1.3	Latar Belakang Masalah	6
1.4	Pernyataan Masalah	9
1.5	Persoalan Kajian	10
1.6	Objektif Kajian	10
1.7	Rasional Kajian	11
1.8	Skop Kajian	12
1.9	Kerangka Konsep Kajian	12
1.10	Organisasi Bab Tesis	14
BAB II	KAJIAN LITERATUR	
2.1	Pengenalan	15
2.2	Sosioekonomi Malaysia	15
2.3	Pendapatan Isi Rumah Malaysia	16
2.4	Kesejahteraan Rakyat Malaysia	17
2.5	Pendapatan Isi Rumah Boleh Guna Bersih	19
2.6	Beban Perbelanjaan	21
2.7	Perbelanjaan Melebihi Pendapatan	24
2.8	Teknik Pengelompokan Pembelajaran Mesin	25
	2.8.1 Pengenalan	25
	2.8.2 Algoritma Pengelompokan Pemetaan	28

2.8.3	Algoritma Pengelompokan Hirarki	29
2.8.4	Algoritma Pengelompokan Ketumpatan	30
2.8.5	Algoritma Pengelompokan Grid	31
2.8.6	Algoritma Pengelompokan Model	32
2.8.7	Kelebihan dan Kelemahan Setiap Teknik Pengelompokan	32
2.9	Pelbagai Teknik Perlombongan Data Dalam Kajian Sosioekonomi	33
2.10	Kesimpulan	35
BAB III	METODOLOGI KAJIAN	
3.1	Pengenalan	37
3.2	Perisian Komputer Yang Digunakan	37
3.3	Rangka Kerja Kajian Model Pengelompokan	38
3.4	Fasa Input	39
3.4.1	Pemahaman Perniagaan, Permasalahan, Objektif dan Kajian Literasi	39
3.5	Fasa Pembangunan Model	40
3.5.1	Pemahaman Set Data HIES 2022	40
3.5.2	Analisis Struktur Data	40
3.5.3	Pra-Pemprosesan Data	44
3.5.4	Penjanaan Data/Atribut	47
3.5.5	Tapisan, Pembersihan & Transformasi Data	47
3.5.6	Pengkategorian Kumpulan Atribut	56
3.5.7	Pemilihan Ciri	56
3.5.8	Pembangunan Algoritma Pengelompokan	59
3.5.9	Fasa Penilaian	63
3.6	Fasa Output	64
3.6.1	Analisis Kelompok	64
3.6.2	Analisis Klasifikasi Hasil Kelompok	65
3.7	Kesimpulan	65
BAB IV	DAPATAN DAN PENILAIAN MODEL	
4.1	Pengenalan	66
4.2	Penentuan Eksperimen	66
4.3	Penetapan Model Algoritma	67
4.4	Dapatan Eksperimen	67
4.4.1	Teknik Pengukuran Jarak Euclidean	67
4.4.2	Penentuan Nilai Optima k	68
4.4.3	Pengelompokan k-Min	72

	4.4.4	Pengesahan kualiti Pengelompokan	72
4.5		Analisis Diskriptif	73
	4.5.1	Analisis Statistik Hasil Kelompok	73
	4.5.2	Saiz Kelompok	74
	4.5.3	Penentuan Indikator Penting Setiap Kluster	75
	4.5.4	Visualisasi Nilai Sentroid	77
	4.5.5	Perbincangan dan Ulasan Penemuan	80
4.6		Analisis Klasifikasi Hasil Pengelompokan	84
4.7		Rumusan	87
BAB V		RUMUSAN KAJIAN	
5.1		Pengenalan	88
5.2		Pencapaian Objektif Kajian	88
5.3		Perbandingan kajian Lepas	89
5.4		Sumbangan kajian	90
	5.4.1	Pemahaman Indikator Penyumbang	91
	5.4.2	Pembangunan Model Pengelompokan	91
	5.4.3	Penggunaan Data Perbelanjaan Isi rumah	92
5.5		Batasan Kajian	92
5.6		Penambahbaikan Kajian	93
5.7		Analisis Pengetahuan	94
5.8		Rumusan	95
RUJUKAN			96
LAMPIRAN			
Lampiran A		Emel Kebenaran Penggunaan Data HIES 2022 Oleh DOSM	100
Lampiran B		Ulasan Analisis Pengetahuan Oleh Pakar/Pegawai Terhadap Kajian	101

SENARAI JADUAL

No. Jadual	Halaman
Jadual 1.1 Sasaran Terpilih RMKe12, 2021-2025	4
Jadual 1.2 Kadar peratusan hutang isi rumah berbanding KDNK	7
Jadual 2.1 Indikator utama sosioekonomi Malaysia bagi tempoh lima tahun terkini	16
Jadual 2.2 Agihan Pendapatan, Pendapatan Isi Rumah Penengah, Purata dan Had Pendapatan mengikut Kumpulan Isi Rumah Desil, Malaysia, 2022	17
Jadual 2.3 Dapatan setiap domain dalam IKK Malaysia, 2019 berbanding 2016	19
Jadual 2.4 Kadar perbelanjaan minima berdasarkan demografi isi rumah dan bandar (RM)	23
Jadual 2.5 Kelebihan dan Kekurangan Teknik Pengelompokan	33
Jadual 3.1 Perincian fungsi prisian dalam pelaksanaan kajian	38
Jadual 3.2 Senarai pemboleh ubah dalam set data	41
Jadual 3.3 Statistik Ringkas Set Data	45
Jadual 3.4 Statistik Ringkas Set Data	48
Jadual 3.5 Statistik Ringkas Set Data	48
Jadual 3.6 Normalisasi kumpulan umur	49
Jadual 3.7 Normalisasi kumpulan Sijil Tertinggi Diperoleh/Pendidikan	49
Jadual 3.8 Normalisasi kumpulan umur	50
Jadual 3.9 Normalisasi kumpulan taraf perkahwinan	50
Jadual 3.10 Normalisasi kumpulan taraf aktiviti	50
Jadual 3.11 Normalisasi kumpulan pekerjaan	51
Jadual 3.12 Normalisasi kumpulan industri	51
Jadual 3.13 Normalisasi kumpulan pendapatan	52
Jadual 3.14 Normalisasi kumpulan pendapatan boleh guna	52

Jadual 3.15	Normalisasi kumpulan perbelanjaan	53
Jadual 3.16	Senarai peratusan perbelanjaan isi rumah mengikut kumpulan belanja	54
Jadual 3.17	Jumlah isi rumah terlebih belanja terlibat mengikut kumpulan perbelanjaan	55
Jadual 3.18	Senarai atribut mengikut kumpulan indikator	56
Jadual 3.19	Senarai akhir atribut dalam set data	57
Jadual 3.20	Perincian Kaedah Penentuan Bilangan Optima Kluster, k	60
Jadual 3.21	Perincian Kaedah Penilaian Pengesahan Hasil Kluster	64
Jadual 4.1	Tetapan Eksperimen Pengelompokan	74
Jadual 4.2	Saiz Kelompok yang terhasil dan nilai AWCD	74
Jadual 4.3	Jadual Sentroid bagi setiap atribut & kluster	75
Jadual 4.4	Senarai ciri utama setiap kluster dan kumpulan indikator mengikut tiga (3) nilai sentroid tertinggi	76
Jadual 4.5	Statistik Deskriptif Antara Kedua-dua Kumpulan Indikator Kluster 1 (Faktor Pekerjaan & Penggajian)	82
Jadual 4.6	Statistik Deskriptif Antara Kedua-dua Kumpulan Indikator Kluster 2 (Faktor Sosio-Demografi)	83
Jadual 4.7	Statistik Deskriptif Antara Kedua-dua Kumpulan Indikator Kluster 3 (Faktor Pendapatan dan Perbelanjaan)	84
Jadual 4.8	Matriks Prestasi Pengkelasan J48 Mengikut Kelas	85
Jadual 4.9	Kepentingan Ciri (feature importance) Setiap Atribut Mengikut Skor	85
Jadual 4.10	Antara Peraturan (<i>Rules</i>) Utama Kepada Isi Rumah Terlebih Belanja Berdasarkan Indikator Utama Yang Dapat Dikenalpasti	86
Jadual 5.1	Perbandingan Indikator Hasil Dapatan Kajian Berkaitan Isi Rumah Terlebih Belanja	90
Jadual 5.2	Analisis Pengetahuan bersama Pakar/Pegawai Dalam Bidang berkaitan	95

SENARAI ILUSTRASI

No. Rajah		Halaman
Rajah 1.1	Piramid Hirarki Maslow	2
Rajah 1.2	Sasaran Baharu Kajian Separuh Penggal RMKe-12	5
Rajah 1.3	Fasa Utama Pelaksanaan Kajian	13
Rajah 1.4	Fasa-fasa Analitik Data menggunakan kaedah CRISP-DM	13
Rajah 2.1	Dapatan IKRM Malaysia Mengikut Sub-Komposit, 2022	19
Rajah 2.2	Hubungan antara Kecerdasan Buatan (AI), Pembelajaran Mesin (ML) dan Pembelajaran Dalam (DL)	25
Rajah 2.3	Taksonomi Teknik Algoritma Pengelompokan (Saxena et al. 2017)	27
Rajah 2.4	Contoh Teknik Algoritma Pengelompokan Yang Lazim Digunakan	27
Rajah 2.5	Perbezaan antara algoritma k-Min dan k-Medoid	29
Rajah 2.6	Idea konsep algoritma DBSCAN (Entezami et al. 2020) (a) Sebuah kluster (b) Titik biru sebagai titik rujukan (c) titik kuning sebagai titik batas (d) Kluster yang ditentukan berdasarkan ketumpatan/kepadatan	31
Rajah 3.1	Proses dan Sub-Proses Mengikut Setiap Fasa Utama Pelaksanaan Kajian	39
Rajah 3.2	Komposisi Perbelanjaan Penggunaan Isi Rumah Bulanan Purata mengikut 13 Kumpulan Utama isi rumah Malaysia, 2019 dan 2022	44
Rajah 3.3	Visualisasi atribut set data	58
Rajah 4.1	Pelaksanaan eksperimen kajian	67
Rajah 4.2	Teknik Siku (<i>Elbow</i>) Untuk Nilai k Optimum	69
Rajah 4.3	Teknik Skor Silhouette Untuk Nilai k Optimum	70
Rajah 4.4	Teknik Indeks Davies-Bouldin Untuk Nilai k Optimum	71
Rajah 4.5	Teknik Indeks Calinski-Harabasz Untuk Nilai k Optimum	72
Rajah 4.7	Penilaian Validasi Kecekapan Kluster, $k=3$	73

Rajah 4.8	Carta Sentroid bagi setiap atribut & kluster	78
Rajah 4.9	Plot taburan mengikut kluster dalam dimensi 2D	79
Rajah 4.10	Plot taburan mengikut kluster dalam dimensi 3D	80
Rajah 4.11	Diagram Pohon Keputusan Model J48 Hasil Pengkelasan Data Kluster k-Min	86

LIBRARY FTSM

SENARAI SINGKATAN

AIR	Ahli Isi Rumah
AWCD	<i>Average Within Centroid Distance</i>
BNM	Bank Negara malaysia
CRISP-DM	The Cross Industry Standard Process for Data Mining
DOSM	Department of Statistics, Malaysia/Jabatan Perangkaan Malaysia
HIES	Survei Pendapatan dan Perbelanjaan Isi Rumah Malaysia
IHP	Indeks Harga Pengguna
IKK	Indeks Kesejahteraan Keluarga Malaysia
IKRM	Indeks Kesejahteraan Rakyat Malaysia
IPT	Institusi Pengajian Tinggi
JPA	Jabatan Perkhidmatan Awam
KDNK	Keluaran Dalam Negeri Kasar
KE	Kementerian Ekonomi
KIR	Ketua Isi Rumah
KSM	Kementerian Sumber Manusia
KSP	Kajian Separuh Penggal
KWSP	Kumpulan Wang Simpanan Pekerja
LPPKN	Lembaga Penduduk dan Pembangunan Keluarga Negara
MOF	Kementerian Kewangan
OECD	<i>The Organisation for Economic Co-operation and Development</i>
PADU	Sistem Pangkalan Data Utama
PBG	Pendapatan Boleh Guna
PNK	Pendapatan Negara Kasar
RMKe-12	Rancangan Malaysia Ke-12

SSB	<i>Sum of Squares Between Cluster</i>
SSE	Ralat Kuasa Dua
SSW	<i>Sum of Squares Within Cluster</i>
SSPA	Sistem Saraan Perkhidmatan Awam
SWRC	Pusat Penyelidikan Kesejahteraan Sosial
UM	Universiti Malaya
UKM	Universiti Kebangsaan Malaysia
UPM	Universiti Putra Malaysia
USD	Dollar Amerika Syarikat

LIBRARY FETSM

BAB I

PENGENALAN

1.1 PENDAHULUAN

Bab ini akan menerangkan secara ringkas mengenai pelaksanaan kajian. Secara umumnya ia merangkumi latar belakang kajian yang berkaitan dengan sosio-ekonomi Malaysia, perilaku terlebih belanja, kumpulan pendapatan isi rumah boleh guna dan pernyataan masalah yang merangkumi permasalahan tentang indikator terlebih belanja di kalangan isi rumah boleh guna di Malaysia pada masa kini. Turut dibincangkan adalah perkembangan teknik dan algoritma perlombongan data dengan kaedah pembelajaran mesin seiring dengan perkembangan teknologi semasa. Selain itu, bab ini juga merangkumi objektif, rasional, skop dan kerangka konsep kajian yang keseluruhannya membincangkan mengenai sasaran, fokus dan kepentingan kajian termasuk ringkasan metodologi kajian yang menghuraikan secara ringkas setiap fasa yang dilaksanakan sepanjang kajian ini. Akhir sekali, bab ini akan menerangkan mengenai organisasi tesis bagi kajian ini.

1.2 LATAR BELAKANG KAJIAN

Dalam era digital dan revolusi perindustrian yang semakin pantas dan pesat membangun, teknik perlombongan data dan pembelajaran mesin telah menjadi alat penting dalam menganalisis data berskala besar dan kompleks. Seiring dengan perkembangan teknologi, kaedah ini bukan sahaja mampu mengenal pasti pola tersembunyi dalam data tetapi juga memberi sokongan kepada pembuatan keputusan yang lebih bijak dan pantas. Dalam konteks sosial dan ekonomi, penggunaan teknik ini semakin relevan untuk memahami isu-isu yang berkaitan dengan kelestarian kewangan, termasuk kajian mengenai isi rumah terlebih belanja.

Dalam dunia yang bergerak pantas pada hari ini, individu atau sesebuah keluarga sudah pastinya akan menghadapi cabaran dalam memenuhi keperluan dan kehendak peribadi serta berusaha menyediakan yang terbaik untuk orang yang disayangi namun pada masa yang sama masih perlu mengimbangi pelbagai tuntutan sosioekonomi. Menelusuri sebuah teori psikologi yang masih lagi relevan hingga kini, iaitu *Maslow's Hierarchy of Needs* atau Hirarki Keperluan Maslow, ia memberikan idea dan gambaran tentang motivasi, keperluan dan kehendak yang mendorong kepada tingkah laku manusia sejak diperkenalkan oleh Abraham Maslow pada 1943. Seperti pada Rajah 1.1, terdapat lima (5) peringkat utama keperluan manusia mengikut hierarki yang dicadangkan oleh teori ini. Ia bermula dari peringkat yang paling asas iaitu keperluan asas atau fisiologi, diikuti dengan keperluan keselamatan dan jaminan diri, seterusnya keperluan untuk berkasih sayang, keperluan untuk penghargaan dan penghormatan diri dan akhirnya keperluan bagi kemajuan dan kesempurnaan sendiri (Maslow 1943). Dapat diperhatikan bahawa melalui teori ini, ia memberikan pandangan mendalam tentang sifat yang kompleks motivasi dan tingkah laku manusia dan lebih memahami faktor-faktor yang menjadi asas kepada pertimbangan tindakan dan keputusan manusia, termasuk dalam perihal berbelanja untuk memenuhi keperluan dan kehendak manusia.



Rajah 1.1 Piramid Hirarki Maslow

Pelbagai faktor dan pertimbangan perlu diambil kira dalam membuat sebarang keputusan sebelum sesuatu keperluan dan kehendak tersebut akan dipenuhi. Antara faktor utama yang sering menjadi pertimbangan adalah seperti pendapatan, corak perbelanjaan, gaya hidup, harga, kos dan lain-lain. Ironinya, masyarakat Malaysia pada

masa kini dilihat kurang memberi perhatian kepada cara perbelanjaan yang baik dan berhemah, walaupun pada hakikatnya ia amatlah penting, terutamanya ketika harga barangan keperluan adalah tidak menentu. Bagi seseorang individu atau isi rumah, corak perbelanjaan pada dasarnya adalah kaedah pengagihan bagaimana sumber pendapatan yang diperolehi dibelanjakan untuk memenuhi sebarang keperluan atau kehendak terhadap barangan dan perkhidmatan yang perlu seperti makanan, pakaian, tempat tinggal, utiliti, dan lain-lain perbelanjaan mengurus. Tindakan bijak dalam pengurusan perbelanjaan dapat memastikan kesejahteraan kewangan yang lebih baik, khususnya dalam menghadapi ketidakpastian ekonomi.

Tidak dinafikan bahawa di Malaysia, pendapatan dan perbelanjaan isi rumah merupakan subjek yang sentiasa menjadi perhatian, sama ada dari pihak Kerajaan, pemain industri, para penyelidik mahupun kepada rakyat keseluruhannya. Oleh kerana pendapatan menjadi sesuatu yang dekat di hati rakyat di mana ia menjadi penunjuk utama kepada tahap kesejahteraan dan sosioekonomi rakyat selain dari menjadi pengukur kepada Pariti Kuasa Beli (*Purchasing Power Parity*) dalam aspek kewangan mahupun perbelanjaan, ia juga menjadi indikator utama kepada pengukuran kadar insiden kemiskinan negara dan juga sejauh mana dasar-dasar kerajaan berjaya meningkatkan taraf kehidupan dan sosioekonomi rakyat.

Dalam dekad terakhir ini, dapat dilihat pelbagai usaha dan komitmen berterusan Kerajaan dalam terus meningkatkan taraf hidup rakyat, menghapuskan kemiskinan, merapatkan jurang ekonomi dan seterusnya memastikan inisiatif bersepadu untuk meningkatkan kesejahteraan rakyat serta pendapatan dan kuasa beli isi rumah melalui pelbagai dasar dan pendekatan menyeluruh. Menyedari bahawa peri pentingnya faktor kedudukan kewangan isi rumah yang kukuh serta peningkatan kuasa beli oleh rakyat dalam mendorong pertumbuhan ekonomi dan kesejahteraan sosial, pemegang taruh dan pembuat dasar negara terus merangka dan melaksanakan pelbagai inisiatif untuk meningkatkan pendapatan rakyat, mempromosikan dasar inklusiviti kewangan, dan serta terus memastikan corak konsumerisma yang mampan.

Kerajaan melalui Rancangan Malaysia ke-12, bagi tempoh 2021-2025 (RMKe-12) yang dibentangkan oleh Yang Amat Berhormat Perdana Menteri pada September

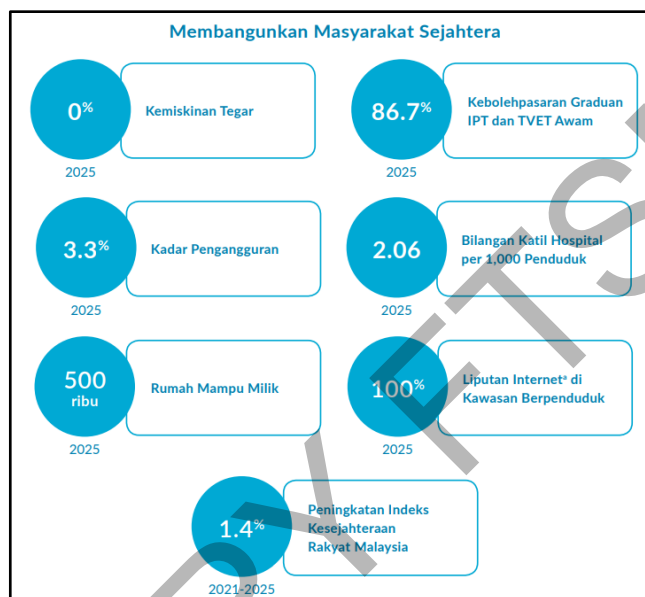
2021 telah menetapkan pendekatan dan sasaran utama yang bertujuan untuk memulihkan dan merangsang semula ekonomi negara pasca tempoh pandemik Covid-19. Antara fokus utama dalam rancangan ini adalah agenda Menangani Kemiskinan dan Membangunkan Masyarakat Inklusif yang turut terkandung dalam senarai utama Matlamat Pelbagai Dimensi. Melalui agenda ini, Kerajaan mensasarkan untuk terus komited dalam menangani kemiskinan, mengurangkan ketidaksamarataan, memperkasa golongan 40% Terendah (B40), memberikan sokongan kepada golongan 40% Penengah bagi mewujudkan masyarakat yang saksama sehinggalah kepada memperkasa kumpulan sasar khusus (KE_RMKe-12 2021). Disamping itu, kerajaan juga berusaha melaksanakan penetapan gaji minimum RM1,500 untuk seluruh majikan dalam semua sektor pekerjaan di Malaysia (KSM KSM 2022). Kesemua inisiatif ini adalah berpaksikan kepada sasaran bagi menjana ekonomi negara, merencanakan ekonomi domestik, meningkatkan pendapatan isi rumah, meningkatkan kuasa beli pengguna dan seterusnya dapat meningkatkan kesejahteraan rakyat Malaysia.

Jadual 1.1 Sasaran Terpilih RMKe12, 2021-2025

Indikator	Pencapaian RMKe-11	Sasaran RMKe-12
Pertumbuhan KDNK, setahun.	2.7 %	4.5 % -5.5 %
PNK Per Kapita, akhir tempoh	RM42,503	RM57,882
Pertumbuhan Produktiviti Buruh, setahun.	1.1 %	3.6 %
Pampasan Pekerja, akhir tempoh (% daripada KDNK).	37.2 %	40.0 %
Purata Pendapatan Bulanan Isi Rumah, akhir tempoh.	RM7,160	RM10,065
Pertumbuhan Indeks Kesejahteraan Rakyat Malaysia, setahun.	0.5 %	1.2 %

Seusai dengan perubahan tampuk kepimpinan negara pada 2022, Kerajaan menunjukkan komitmen yang teguh dalam membela nasib rakyat secara keseluruhan. Pelancaran Ekonomi MADANI pada Julai 2023 dan pembentangan laporan Kajian Separuh Penggal (KSP) RMKe-12 pada September 2023 yang disesuaikan dengan aspirasi dan visi Ekonomi MADANI, yang bertujuan untuk merencanakan ekonomi negara bagi tempoh 10 tahun serta memacu pembangunan negara dan memperkasa kualiti hidup rakyat. Bermatlamatkan "Malaysia Madani: Mampan, Sejahtera dan Berpendapatan Tinggi", pendekatan dan kaedah baru dalam KSP ini dilaksanakan sebagai respons terhadap ketidakstabilan politik dan pasca wabak Covid-19 yang telah

menjejaskan kehidupan rakyat seperti kehilangan mata pencarian, menyusut purata pertumbuhan ekonomi serta merosotnya pelaburan asing (Kem. Ekonomi, 2023). Ini turut ditambah dengan komitmen kerajaan dalam memartabatkan dan memperkasakan perkhidmatan awam melalui Kajian Sistem Saran Perkhidmatan Awam (SSPA) (JPA_SSPA JPA 2023).



Rajah 1.2 Sasaran Baharu Kajian Separuh Penggal RMKe-12

Sumber : Kementerian Ekonomi, 2023

Dalam pembentangan Belanjawan 2024 oleh YAB Perdana Menteri di Parlimen pada 13 Oktober 2023, belanjawan ini dianggap sebagai langkah penting dalam menyokong perbelanjaan pengguna serta kumpulan isi rumah rentan (*vulnerable*) dengan peruntukan besar ditumpukan kepada perbelanjaan pembangunan serta projek infrastruktur rakyat yang sedang berjalan (MOF_Bajet2024 2023). Ia selaras dengan hasrat kerajaan yang akan memfokuskan kepada kesejahteraan rakyat, khususnya dalam meningkatkan taraf hidup, menambah pendapatan dan juga memperkenalkan strategi subsidi bersasar. Ia juga adalah konsisten dengan strategi Kerajaan yang bercadang untuk menghapuskan pengelasan pendapatan berdasarkan kategori 40% terbawah (B40), 40% Pertengahan (M40) dan 20% Tertinggi (T20) dan beralih kepada pengelasan mengikut pendapatan bersih boleh guna mulai 2024. Ia bertujuan untuk memastikan agar isi rumah yang benar-benar layak dan memerlukan dapat dikenalpasti seterusnya penyampaian bantuan dapat disalurkan dengan lebih efektif dan saksama.

Kesemua inisiatif ini dirancang untuk mencapai sasaran kesejahteraan rakyat dengan memfokuskan kepada peningkatan pendapatan isi rumah dan seterusnya dapat menguruskan corak perbelanjaan secara keseluruhan.

1.3 LATAR BELAKANG MASALAH

Permasalahan isi rumah terlebih belanja Malaysia semakin menjadi satu cabaran dan isu yang serius, terutamanya dalam konteks ekonomi Malaysia. Keadaan ini diburukkan oleh peningkatan kos sara hidup, kadar inflasi yang tidak menentu yang ditambah dengan ketidaktentuan ekonomi global, isi rumah Malaysia terutamanya golongan rentan kini rata-ratanya semakin menghadapi tekanan kewangan yang kian serius. Fenomena ini tidak hanya memberi kesan negatif terhadap kesejahteraan individu dan keluarga, namun ia juga boleh menjejaskan kestabilan ekonomi negara secara kolektif. Walaupun pelbagai usaha dan inisiatif dilaksanakan oleh Kerajaan untuk mengurangkan impak isu ini, data dan laporan kajian konvensional menunjukkan bahawa corak perbelanjaan isi rumah yang masih kekal sebagai cabaran utama yang perlu diatasi.

Kesan langsung yang jelas daripada fenomena isi rumah terlebih belanja adalah peningkatan jumlah hutang isi rumah. Malaysia mencatatkan kadar hutang isi rumah yang tertinggi berbanding negara-negara lain di rantau Asia Tenggara. Bank Negara Malaysia (BNM) melaporkan bahawa kadar nisbah hutang isi rumah Malaysia berbanding Keluaran Dalam Negeri Kasar (KDNK) bagi tahun 2023 adalah 84.2% dengan jumlah hutang berjumlah RM1.53 trilion. Walaupun terdapat sedikit penurunan berbanding tahun 2022 dan 2021, namun jumlah hutang ini masih lebih tinggi daripada hutang Kerajaan Persekutuan, yang berjumlah RM1.147 trilion pada tahun 2023 (BNM_HHDebt 2023).

Jadual 1.2 Kadar peratusan hutang isi rumah berbanding KDNK

Tahun	Peratusan hutang isi rumah kepada KDNK (%)
2006	68.8
2010	76.0
2016	84.4
2019	82.7
2020	93.2
2021	89.0
2022	84.5
2023	84.2

Sumber : Bank Negara Malaysia, 2023

Sebagai respon proaktif, Kerajaan melalui Bank Negara Malaysia mengambil pendekatan lebih berhati-hati dalam mengurus risiko keberhutangan isi rumah bagi memastikan perkara itu tidak menimbulkan kesan kepada kestabilan sistem kewangan negara memandangkan kadar hutang isi rumah ini masih mencatatkan angka melebihi 80 peratus yang masih boleh dianggap membimbangkan. Pelbagai inisiatif kerajaan dirangka untuk dilaksanakan seperti reformasi dan perubahan struktur terhadap ekonomi domestik berteraskan Kerangka Ekonomi MADANI supaya daya saing dan prestasi ekonomi negara dapat dipertingkatkan seterusnya secara tidak langsung dapat mempertingkatkan pendapatan serta kesejahteraan rakyat yang diharap dapat membantu mengurangkan peratusan hutang isi rumah.

Senario keberhutangan yang membimbangkan ini mencerminkan kekurangan dalam perancangan dan kecelikan kewangan serta ditambah oleh tingkahlaku perbelanjaan yang boleh menyebabkan kepada keberhutangan yang tidak terkawal. Tindakan seperti belanja berlebihan melebihi pendapatan, pembelian yang tidak dirancang, ketidakmampuan membayar semula hutang pinjaman, serta kecenderungan mengutamakan kehendak berbanding keperluan, semuanya menjadi faktor penyumbang kepada fenomena ini. Akibatnya, kes-kes kemuflihan dan kebangkrapan semakin meningkat di Malaysia direkodkan atas pelbagai faktor (BNM_HHDebt 2023).

Selain itu, pada 2023 Kerajaan telah bercadang untuk memansuhkan pengkelasan kumpulan pendapatan isi rumah B40, M40 dan T20 sepertimana yang digunakan sebelum ini. Sebaliknya, Kerajaan akan mula menggunakan pendekatan

pendapatan boleh belanja bersih isi rumah (yang juga dikenali sebagai pendapatan boleh guna) dalam mengenal pasti isi rumah yang benar-benar layak dalam pemberian subsidi dan bantuan Kerajaan. Pendapatan boleh guna merujuk kepada jumlah pendapatan selepas ditolak bayaran pindahan semasa termasuklah bayaran wajib yang dikenakan ke atas isi rumah seperti cukai, zakat dan caruman kepada skim keselamatan sosial serta pindahan semasa yang diterima secara mata benda. Pendapatan boleh guna ini juga merupakan pendapatan isi rumah yang masih ada untuk dibelanjakan ke atas perbelanjaan penggunaan dan untuk simpanan sepanjang tempoh rujukan.

Sungguhpun begitu, tidak seperti pengkelasan kumpulan pendapatan isi rumah B40, M40 dan T20 sebelum ini, permasalahan utama dalam menjalankan kajian ini adalah penentuan nilai ambang (*threshold*) pengkelasan kumpulan Pendapatan Boleh Guna isi rumah Malaysia yang masih belum dimuktamadkan. Ketiadaan nilai ambang menjadikan set data yang akan digunakan dalam kajian ini tiada label kelas (*no class label*). Sehingga kajian ini ditulis, perkara ini adalah masih dalam kajian pihak Kerajaan seiring dengan pengenalan Sistem Pangkalan Data Utama (PADU) yang baru diluncurkan pada awal 2023. Walaubagaimanapun, pengkelasan sedia ada yang sedang digunakan oleh agensi-agensi Kerajaan dan pihak berkuasa dalam laporan-laporan rasmi akan dirujuk dan digunakan sebaiknya dalam kajian ini.

Justeru, melihat kepada tren dan perkembangan semasa ini, di sebalik penggunaan pendekatan analisis statistik yang kebiasannya digunakan, kajian ini mengambil langkah yang lebih inovatif dengan menggunakan pendekatan pembelajaran mesin dan analitik data raya. Pendekatan ini bertujuan untuk membantu menganalisis dan memahami konteks belanja berlebihan atau terlebih belanja di kalangan kumpulan pendapatan isi rumah di Malaysia. Dengan menggunakan teknik ini, kajian ini berusaha untuk mengenal pasti indikator-indikator perbelanjaan yang mungkin mempengaruhi senario dan situasi sosioekonomi serta kewangan isi rumah pada masa kini. Justeru penggunaan teknik dan analisa kajian yang lebih maju seperti algoritma perlombongan data dan pembelajaran mesin adalah sangat sesuai. Pendekatan ini bukan sahaja mampu melihat dan mengesan corak perbelanjaan yang tersembunyi dalam data berskala besar tetapi juga dapat membantu mengenal pasti indikator-indikator utama yang menyumbang kepada masalah isi rumah terlebih belanja. Justeru, melalui analisis ini,

penyelesaian yang lebih tepat, strategik, dan bersandarkan bukti dapat dirangka untuk mengurangkan beban kewangan isi rumah, sekali gus meningkatkan kualiti hidup masyarakat Malaysia.

1.4 PERNYATAAN MASALAH

Pengaplikasian analitik data raya dapat membantu pengkaji dalam mentafsir, memberikan gambaran, trend dan corak tersembunyi data yang biasanya kurang jelas kelihatan dengan menggunakan analitik data biasa. Banyak kajian lepas yang mengaplikasikan algoritma pembelajaran mesin dalam mengkaji domain sosioekonomi isi rumah seperti pendapatan, perbelanjaan, ekonomi kewangan, kesejahteraan, kemakmuran, kemiskinan dan lain-lain. Sungguhpun begitu, pendekatan algoritma pembelajaran mesin menggunakan teknik pengelompokan bagi domain sosioekonomi adalah masih kurang berbanding teknik pengkelasan. Teknik pengkelasan, sungguhpun mudah dan ringkas namun memerlukan set data berlabel kelas (*class label*). Oleh kerana kebanyakan set data sosioekonomi adalah data tanpa label kelas, kebanyakan pengkaji akan melabelkan data dengan label kelas secara sendiri. Ia memberi sedikit risiko kepada pelabelan kelas yang kurang tepat dan boleh memberi kesan kepada hasil dapatan akhir (Abdul Rahman et al. 2021). Maka adalah dicadangkan agar algoritma pengelompokan yang merupakan kaedah pembelajaran mesin tanpa selia dan sesuai dengan set data tanpa label kelas bersama-sama dengan tambahan pengujian melalui teknik pengkelasan J48 akan digunakan sebagai alternatif kepada permasalahan ini. Teknik pengelompokan ini akan bekerja dengan mengkategorikan Kumpulan isi rumah terlebih belanja ini kepada beberapa kelompok berdasarkan ciri-ciri persamaan dan kemiripan yang dimiliki oleh setiap isi rumah dan pengujian melalui teknik pengkelasan akan memberikan pengetahuan baharu terhadap perkara ini.

Menyentuh mengenai bidang kajian, Kerajaan pada 2023 telah mula mengusulkan cadangan untuk beralih daripada pengkelasan kumpulan pendapatan isi rumah B40, M40 dan T20 kepada kaedah pengkelasan berasaskan nilai pendapatan boleh guna bersih (*net household disposable income*). Konsep pendapatan boleh guna bersih ini diperkenalkan bertujuan untuk mengukur kesejahteraan rakyat dan dijadikan asas baharu dalam mengenalpasti kumpulan sasaran yang benar-benar layak serta bagi mencerminkan situasi sosio-ekonomi isi rumah Malaysia dengan lebih menyeluruh.

Justeru, dengan menggabungkan analisis statistik tradisional bersama pendekatan pembelajaran mesin yang lebih moden, kajian ini berupaya untuk menganalisis data berasaskan sosio-ekonomi berskala besar dengan lebih mendalam untuk memperoleh pengetahuan bagi memahami perbelanjaan berisiko isi rumah serta ciri-ciri sosio demografi yang mempengaruhinya. Pendekatan ini juga membolehkan pengenalpastian ciri demografi dan sosioekonomi yang signifikan, menjadikannya landasan penting dalam memahami dan menangani isu isi rumah terlebih belanja dan permasalahan pendapatan boleh guna bersih secara menyeluruh.

1.5 PERSOALAN KAJIAN

Berdasarkan set data HIES 2022 serta beberapa permasalahan yang dikenalpasti dalam situasi dan senario semasa tersebut, berikut adalah persoalan-persoalan kajian bagi kajian ini:

1. Sejauh mana model k-Min adalah teknik terbaik dalam pembangunan model pengelompokan kumpulan isi rumah terlebih belanja berasaskan kumpulan pendapatan boleh guna?
2. Sejauh mana teknik pembelajaran mesin menggunakan algoritma pengelompokan dapat menentukan indikator penting dalam kelompok kumpulan isi rumah terlebih belanja berasaskan kumpulan pendapatan boleh guna?

1.6 OBJEKTIF KAJIAN

Secara khususnya berdasarkan set data HIES 2022 tersebut juga, kajian ini dilaksanakan bagi mencapai objektif-objektif seperti berikut:

1. Membangunkan model pengelompokan kumpulan isi rumah terlebih belanja berasaskan kumpulan pendapatan boleh guna menggunakan teknik k-Min.
2. Mengenal pasti indikator yang signifikan bagi setiap kategori atribut dan setiap kluster yang diperoleh.

1.7 RASIONAL KAJIAN

Kajian ini dilaksanakan dengan tujuan utama untuk mengkaji dan memperoleh pengetahuan (*insight*) yang lebih mendalam dalam menangani isu isi rumah terlebih belanja di Malaysia, dengan pendekatan yang lebih elegan, inovatif dan teknologi terkini berbanding kajian konvensional lain. Teknik perlombongan data bersama pembelajaran mesin berkeupayaan dalam menganalisis data yang kompleks dan berskala besar, terutama data berkaitan corak perbelanjaan dan faktor sosioekonomi isi rumah. Berbanding pendekatan tradisional, teknik ini membolehkan pengesanan corak tersembunyi, ramalan yang lebih tepat, dan pengenalpastian faktor risiko utama yang menyumbang kepada tingkah laku isi rumah terlebih belanja. Kajian ini bukan sahaja relevan dalam konteks memahami isu terlebih belanja, tetapi juga memberikan nilai tambah kepada penggubalan dasar kewangan yang lebih berfokus. Dengan menggunakan teknik seperti pengelompokan (*clustering*) untuk mengenal pasti kumpulan isi rumah dengan ciri perbelanjaan yang serupa serta algoritma ramalan (*predictive algorithms*) untuk mengenal pasti risiko terlebih belanja, dapatan kajian ini mampu memberikan gambaran yang lebih jelas kepada pembuat dasar.

Secara keseluruhannya, inflasi global pada 2022 mencatatkan kadar 8.8%. Ia banyak dipengaruhi oleh harga makanan dan bahan api global yang lebih tinggi. Sebagai rekod, kadar inflasi di kalangan negara-negara maju merekodkan 7.3% manakala 9.9% dalam kalangan negara membangun. Indeks Harga Pengguna (IHP) Malaysia pada 2022 merekodkan kadar 3.3%. Sepertimana faktor global, ia banyak dipengaruhi oleh tekanan ke atas harga makanan segar tertentu yang berpunca daripada kenaikan harga komoditi bukan tenaga serta peningkatan permintaan pascapandemik. Secara langsung menjadikan inflasi makanan dan minuman bukan alkohol kekal tinggi sepanjang tahun dengan merekodkan 5.8% pada 2022 berbanding 1.7% pada 2021. Walau bagaimanapun, peningkatan ini telah dibendung oleh penguatkuasaan kawalan harga dan peruntukan subsidi kerajaan bagi barangan keperluan terpilih.

Selain itu, Malaysia juga merekodkan tahap Indeks Kesejahteraan Rakyat Malaysia (IKRM) yang meningkat kepada 2.5 peratus pada tahun 2022. Menurut Jabatan Perangkaan Malaysia (DOSM) dalam Laporan Indeks Kesejahteraan Rakyat Malaysia (IKRM) 2022, ia banyak disokong oleh pertumbuhan positif kesejahteraan

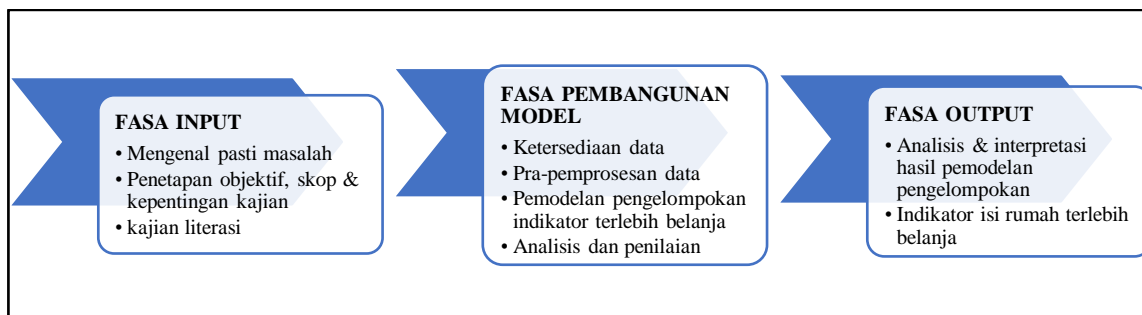
ekonomi dan sosial (UtusanMalaysia-IKRM 2023). Kesejahteraan ekonomi terdiri daripada lima komponen iaitu pengangkutan, komunikasi, pendidikan, pendapatan dan pengagihan dan persekitaran kerja. Kesejahteraan sosial pula meliputi sembilan komponen iaitu perumahan, hiburan dan rekreasi, keselamatan awam, penyertaan sosial, tadbir urus, kebudayaan, kesihatan, alam sekitar dan keluarga (BeritaHarian_IKRM 2023).

1.8 SKOP KAJIAN

Kajian ini memfokuskan kepada isi rumah warganegara Malaysia dengan menggunakan data dari survei Pendapatan dan Perbelanjaan Isi Rumah Malaysia tahun 2022 (HIES 2022) yang dilaksanakan oleh pihak Jabatan Perangkaan Malaysia (DOSM). Daripada data tersebut, kumpulan pendapatan isi rumah yang terlibat pula adalah kumpulan isi rumah terlebih belanja iaitu isi rumah yang mencatatkan jumlah perbelanjaan melebihi jumlah pendapatan berdasarkan pengelasan Kumpulan isi rumah boleh guna. Penyediaan dan analisis data serta pembangunan model pengelompokan dan seterusnya pengklasifikasi indikator isi rumah terlebih belanja, termasuk eksperimen, analisis terperinci, pemodelan, visualisasi dan pelaporan akan menggunakan beberapa perisian iaitu *Anaconda Python*, *Rapid Miner*, *Microsoft Excel* dan *Word 365* dan juga *Power BI*.

1.9 KERANGKA KONSEP KAJIAN

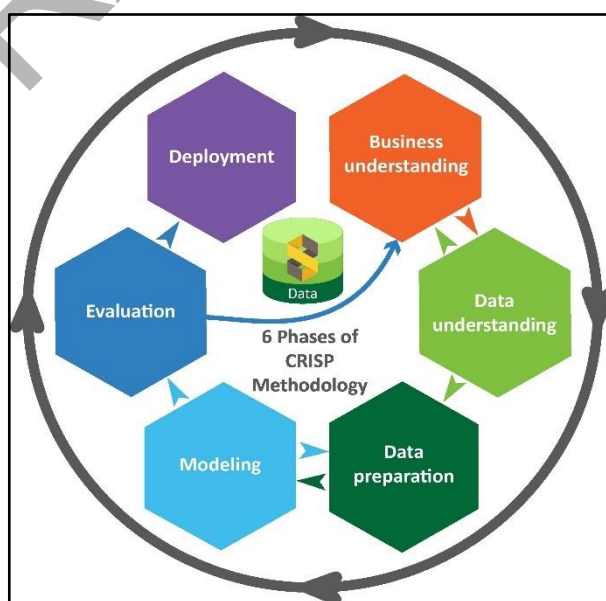
Kerangka Konsep Kajian secara umumnya diwujudkan sebagai panduan dan hala tuju sepanjang pelaksanaan kajian ini bagi memenuhi dan mencapai objektif kajian yang telah ditetapkan. Bagi kajian ini, kerangka konsep adalah terdiri daripada tiga (3) fasa utama iaitu, Fasa Input, Fasa Pembangunan Model; dan Fasa Output.



Rajah 1.3 Fasa Utama Pelaksanaan Kajian

Sepertimana kajian berkaitan perlombongan data dan pembelajaran mesin yang lain, metodologi yang digunakan bagi kajian ini adalah dengan menggunakan metodologi *Cross-Industry Process for Data Mining* atau CRISP-DM. CRISP-DM ialah kaedah universal terbaik dan efektif dalam industri perlombongan data. Metodologi ini terbahagi kepada enam bahagian utama iaitu:

1. Pemahaman perniagaan;
2. Pemahaman data;
3. Penyediaan data;
4. Pembinaan model;
5. Analisis & Penilaian; dan
6. Pelaksanaan dan laporan.



Rajah 1.4 Fasa-fasa Analitik Data menggunakan kaedah CRISP-DM

1.10 ORGANISASI BAB TESIS

Keseluruhannya, terdapat lima (5) bab secara keseluruhan di dalam tesis ini yang merangkumi:

1. Bab I: Bab ini membincangkan secara keseluruhan mengenai pelaksanaan kajian termasuk latar belakang kajian, pernyataan masalah, objektif kajian, rasional kajian, skop kajian serta kerangka konsep kajian.
2. Bab II: Bab ini akan membincangkan lebih terperinci mengenai sorotan kajian-kajian kepustakaan lalu dan tinjauan lepas yang khusus berkaitan dengan status sosioekonomi Malaysia, tingkahlaku terlebih belanja, kumpulan pendapatan isi rumah Malaysia serta teknik pembelajaran mesin dengan algoritma pengelompokan.
3. Bab III: Bab ini adalah merupakan perbincangan berkaitan kaedah dan metodologi kajian secara terperinci dalam pelaksanaan kajian ini. Ia juga membincangkan mengenai fasa data secara terperinci termasuk memahami dan mengenal pasti data, analisis struktur data serta proses-proses terlibat di dalam fasa pra-pemrosesan data dan seterusnya eksperimen kajian.
4. Bab IV: Bab ini pula akan menerangkan mengenai dapatan hasil kajian proses pembangunan model, iaitu model pengelompokan dan model pengklasifikasian terutamanya bagi meneroka pengetahuan dengan menjalankan analisis terhadap hasil pengelompokan yang diperoleh bagi mengenal pasti indikator dan ciri utama bagi isi rumah terlebih belanja. Model-model ini dinilai dan dianalisis kualiti dan keberkesanan pelaksanaan pembelajaran mesin melalui matrik-matrik penilaian dan seterusnya mendapatkan hasil, dapatan dan perbandingan indikator terbaik.
5. Bab V: Bab terakhir ini akan merumuskan secara keseluruhan tentang kajian yang telah dilaksanakan termasuk perbandingan kajian lepas, sumbangan kajian, cadangan penambahbaikan dan analisis pengetahuan.

BAB II

KAJIAN LITERATUR

2.1 PENGENALAN

Kajian literatur atau kepustakaan merujuk kepada perbincangan mengenai penyelidikan dan analisis kajian-kajian yang telaah dilaksanakan oleh pengkaji-pengkaji sebelum ini. Bab ini membincangkan mengenai kajian-kajian lepas yang mengkaji tentang sosioekonomi isi rumah serta tingkahlaku terlebih belanja mengikut kumpulan pendapatan isi rumah di Malaysia dan juga kajian-kajian berkaitan di luar negara. Selain itu, pengaplikasian teknik pengelompokan dalam pelbagai domain dan bidang kajian yang telah dijalankan oleh pengkaji terdahulu juga akan turut dibincangkan di dalam bab ini. Bab ini amat penting dalam menentukan hala tuju kajian khususnya kerana kajian kepustakaan lepas boleh membantu sebagai panduan dalam mengaplikasikan dan membangunkan model berdasarkan pelbagai pandangan dan rujukan daripada kajian terdahulu. Oleh itu, kajian kepustakaan ini diharapkan akan menjadi panduan terbaik dalam menghasilkan kajian dan model yang lebih baik.

2.2 SOSIOEKONOMI MALAYSIA

Sejak lebih dua dekad, Malaysia dikategorikan sebagai sebuah negara berpendapatan menengah atas atau *upper-middle-income nation*. Ia adalah klasifikasi yang diberikan oleh Bank Dunia kepada sesebuah negara berdasarkan kepada beberapa indikator utama, antaranya adalah nilai Pendapatan Negara Kasar (PNK) perkapita sesebuah negara dan diselaraskan dalam nilai matawang Dollar Amerika bagi tujuan perbandingan di peringkat antarabangsa. Bagi tahun 2022, Bank Dunia telah menetapkan nilai ambang kadar PNK per kapita bagi sesebuah negara untuk dikategorikan sebagai negara berpendapatan tinggi adalah USD12,536 dan PNK per kapita Malaysia untuk tahun tersebut adalah USD11,780. Sungguhpun ia adalah peningkatan agak signifikan 9.99 peratus berbanding USD10,710 pada 2021, namun

masih dibawah sasaran yang ditetapkan oleh indikator daripada Bank Dunia tersebut. Sungguhpun begitu, Malaysia tetap komited dalam memenuhi indikator ditetapkan dan seterusnya bagi mencapai taraf negara berpendapatan perkapita tinggi seperti yang telah disasarkan dalam RMKe-12 termasuk juga beberapa indikator sosioekonomi lain agar setaraf dengan negara-negara maju diperingkat antarabangsa (KE_RMKe-12 2021).

Jadual 2.1 Indikator utama sosioekonomi Malaysia bagi tempoh lima tahun terkini

Indikator	2018	2019	2020	2021	2022
Penduduk (juta)	32.4	32.5	32.4	32.6	32.7
Tenaga Buruh (juta)	15.3	15.6	15.7	15.8	15.9
Pengangguran (%)	3.3	3.3	4.5	4.6	4.9
KDNK Nominal (RM Bilion)	1,447.8	1,512.7	1,418.0	1,545.4	1,3208.5
GNI Perkapita (USD)	10,735	10,935	10,145	11,115	11,746

Sumber: Kementerian Ekonomi, 2022

2.3 PENDAPATAN ISI RUMAH MALAYSIA

Sepertimana umum sedia maklum, lanskap sosioekonomi isi rumah Malaysia adalah dinamik dan fleksibel kerana ia banyak dipengaruhi oleh pelbagai elemen seperti pendapatan dan perbelanjaan isi rumah, kecelikan atau literasi kewangan, dan juga tahap kesejahteraan. Secara umumnya, pengkategorian kumpulan utama pendapatan isi rumah di Malaysia adalah berdasarkan kepada peratusan atau persentil (percentile) yang disusun dari tingkat pendapatan terendah sehinggalah ke tertinggi. Ia dikelaskan hasil daripada dapatan dan penemuan survei Pendapatan dan Perbelanjaan Isi Rumah (HIES) yang dilaksanakan oleh Jabatan Perangkaan Malaysia (DOSM) sekitar dua atau tiga tahun sekali. Sungguhpun ia adalah kaedah persampelan statistik atas faktor kekangan kos dan masa, ia tetap meliputi rangka (frame) isi rumah seluruh negara dan merupakan antara kutipan data primer terbaik sekurang-kurangnya buat masa ini. Sebagai rekod, jumlah pengemaskinian penyenaaran rangka (frame) isi rumah Malaysia untuk diliputi dalam survei HIES tahun 2022 adalah sebanyak 7.9 juta isi rumah dan data dikutip dari setiap isi rumah yang telah terpilih secara rawak (DOSM-HIES 2023).

Sebagai natijah daripada survei tersebut, terdapat tiga (3) kategori kumpulan pendapatan utama bagi penduduk Malaysia iaitu kumpulan isi rumah berpendapatan 20 peratus tertinggi (T20), 40 peratus pertengahan (M40), dan 40 peratus terendah (B40). Berdasarkan kepada laporan HIES 2022, had pendapatan bagi kumpulan isi rumah T20

adalah isi rumah yang mempunyai pendapatan bulanan berjumlah RM11,820 atau lebih. Manakala had pendapatan bagi kumpulan isi rumah M40 pula adalah merujuk kepada isi rumah yang berpendapatan di antara RM5,250 sehingga RM11,819 dan kumpulan isi rumah B40 pula merujuk kepada isi rumah yang mempunyai had pendapatan bulanan RM5,250 dan kurang. Laporan HIES 2022 juga memperincikan mengenai kumpulan pendapatan mengikut desil (*decile*) yang mana ketiga-tiga kumpulan pendapatan itu tadi dipecahkan lagi kepada 10 sub-kumpulan kecil bagi melihat had agihan pendapatan secara lebih jelas dan terperinci.

Jadual 2.2 Agihan Pendapatan, Pendapatan Isi Rumah Penengah, Purata dan Had Pendapatan mengikut Kumpulan Isi Rumah Desil, Malaysia, 2022

Kumpulan Desil	Agihan Pendapatan	Pendapatan Isi Rumah Penengah	Pendapatan Isi Rumah Purata	Had Pendapatan (RM)	
T20	T2	30.3	20,776	25,719	15,870 dan lebih
	T2	30.3	20,776	25,719	15,870 dan lebih
	T1	16.0	13,475	13,585	11,820 - 15,869
M40	M4	12.5	10,552	10,577	9,450 - 11,819
	M3	10.1	8,523	8,536	7,690 - 9,449
	M2	8.2	6,979	6,989	6,340 - 7,689
	M2	8.2	6,979	6,989	6,340 - 7,689
	M1	6.8	5,770	5,782	5,250 - 6,339
B40	B4	5.6	4,764	4,771	4,310 - 5,249
	B3	4.6	3,875	3,874	3,440 - 4,309
	B2	3.6	3,034	3,018	2,560 - 3,439
	B1	2.3	2,012	1,941	Kurang daripada 2,560

Sumber: HIES DOSM, 2022

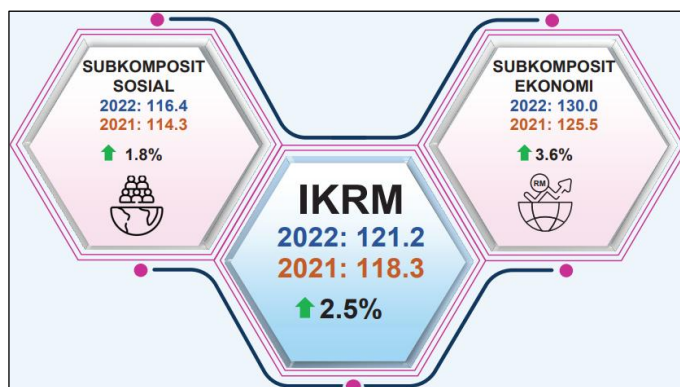
2.4 KESEJAHTERAAN RAKYAT MALAYSIA

Berasaskan dari Teori Hirarki Keperluan Maslow, sarjana kontemporari telah banyak memberikan pelbagai definisi terhadap pengertian dan maksud kesejahteraan. Ia mencerminkan kepelbagaian perspektif dan dimensi yang berbeza, namun masih dalam lingkungan yang hampir sama. Walaubagaimanapun terdapat beberapa definisi yang signifikan dalam memberikan maksud yang sesuai dalam konteks kajian ini. Antaranya adalah kesejahteraan didefinisikan sebagai satu penilaian terhadap kesejahteraan fizikal, material, sosial, dan emosi (Felce & Perry 1995). Selain itu, kesejahteraan diperihalkan sebagai situasi perasaan atau emosi yang positif iaitu kegembiraan dan kepuasan serta kehadiran elemen fungsi positif iaitu penglibatan dan pengiktirafan diri.

(Ryan & Deci 2001). Dalam kajian mengenai kesejahteraan yang lebih terkini, terdapat pelbagai pendekatan yang digunakan terhadap definisi dan pengukuran kesejahteraan yang jauh melangkaui pengukuran ekonomi lama. Ini termasuk obligasi terhadap jenis barangan atau perkhidmatan kepenggunaan seperti tempat tinggal atau perlindungan, perubatan, dan pendidikan sebagai keperluan asas (Kasri & Ahmed 2019).

Di Malaysia, umumnya kesejahteraan hidup rakyat diukur berasaskan kepada beberapa indikator dan pengukur yang sesuai. Ia dilaksanakan oleh beberapa agensi Kerajaan berkaitan dan laporan-laporan mengenainya banyak diterbitkan dari masa ke masa. Tahap kesejahteraan rakyat di Malaysia kebiasannya diukur melalui dapatan indeks. Antara indeks-indeks utama berkaitan pengukuran kesejahteraan rakyat yang terkait dengan sosioekonomi rakyat adalah Laporan Indeks Kesejahteraan Rakyat Malaysia (IKRM) yang dikeluarkan oleh DOSM dan juga Laporan Indeks Kesejahteraan Keluarga (IKK) Malaysia yang dikeluarkan oleh Lembaga Penduduk dan Pembangunan Keluarga Negara (LPPKN). Sehingga kajian ini ditulis, laporan IKRM terkini tersedia bagi tahun rujukan 2022, manakala laporan IKK hanya tersedia bagi tahun rujukan 2019.

Dapatan daripada laporan IKRM bagi tahun 2022, indeks kesejahteraan rakyat secara keseluruhannya meningkat sebanyak 2.5 peratus kepada 121.2 mata indeks berbanding 118.3 mata indeks pada 2021. Ia adalah indikator yang baik memandangkan negara masih dalam fasa peralihan dari fasa pandemik Covid-19 kepada fasa endemik yang mana sektor-sektor ekonomi telah mula dibuka sepenuhnya bermula April 2022. Kedua-dua subkomposit yang terlibat dalam pengukuran indeks ini telah menunjukkan peningkatan. Subkomposit Sosial melonjak naik dari 1.8 peratus kepada 116.4 mata indeks pada tahun 2022 berbanding 114.3 mata indeks pada tahun 2021. Sementara itu, Subkomposit Ekonomi turut mencatatkan peningkatan kepada 130.0 mata indeks berbanding 125.5 mata indeks pada 2021 iaitu peningkatan 3.6 peratus.



Rajah 2.1 Dapatan IKRM Malaysia Mengikut Sub-Komposit, 2022

Sumber: Indeks Kesejahteraan Rakyat Malaysia 2022, DOSM

Begitu juga dengan indeks IKK Malaysia yang dikeluarkan oleh Lembaga Penduduk dan Pembangunan Keluarga Negara (LPPKN). Bagi laporan IKK tahun rujukan 2019, indeks menunjukkan peningkatan kepada 7.72 berbanding 7.33 mata indeks bagi laporan tahun rujukan 2016 (LPPKN 2021). Justeru dapat diperhatikan bahawa dari tahun ke tahun, hampir kesemua indikator sosioekonomi terutamanya subkomposit dan domain yang menjadi pengukuran kepada IKRM dan IKK Malaysia menunjukkan peningkatan yang stabil.

Jadual 2.3 Dapatan setiap domain dalam IKK Malaysia, 2019 berbanding 2016

Domain	2019	2016
Hubungan Kekeluargaan	8.35	7.90
Ekonomi Keluarga	7.67	7.05
Kesihatan Keluarga	7.44	7.18
Keselamatan Keluarga	7.86	7.96
Keluarga dan Keterlibatan Komuniti	8.00	7.84
Keluarga, Peranan Agama dan Amalan Kerohanian	8.25	8.04
Perumahan dan Persekitaran	7.35	6.24
Keluarga dan Teknologi Komunikasi	6.82	6.38

2.5 PENDAPATAN ISI RUMAH BOLEH GUNA BERSIH

Seperti dibincang pada bab sebelum ini, pada 2023 Kerajaan telah mengusulkan cadangan bagi peralihan klasifikasi kumpulan pendapatan isi rumah B40, M40 dan T20 kepada kaedah berdasarkan nilai pendapatan boleh guna atau juga disebut pendapatan

boleh belanja isi rumah bersih (*net household disposable income*). Konsep pendapatan boleh guna bersih ini diperkenalkan bertujuan untuk mengukur kesejahteraan rakyat dan dijadikan asas serta pertimbangan utama dalam usaha mengurangkan risiko keciciran dan juga mengurangkan risiko tersalah kumpulan sasar khususnya dalam pemberian dan penyaluran subsidi dan bantuan agar ianya bersifat lebih adil dan saksama (Parlimen Parlimen 2024).

Pendapatan Boleh Guna (PBG) adalah merujuk kepada jumlah pendapatan selepas ditolak bayaran pindahan semasa termasuklah bayaran wajib yang dikenakan oleh kerajaan samada persekutuan, negeri atau tempatan ke atas isi rumah seperti cukai, zakat dan caruman kepada skim keselamatan sosial serta sebarang pindahan semasa yang diterima secara mata benda (*in-kind*). Pendapatan boleh guna ini juga merupakan pendapatan isi rumah yang masih ada atau berbaki untuk dibelanjakan ke atas perbelanjaan penggunaan dan juga untuk tujuan simpanan (DOSM-HIES 2023).

$$\text{PBG} = \text{Pendapatan} - (\text{cukai} + \text{zakat} + \text{caruman keselamatan} + \text{nilai mata benda})$$

Konsep pendapatan boleh guna ini merupakan konsep yang paling hampir dengan konsep pendapatan isi rumah sepertimana dalam teori ekonomi. Ia telah diiktiraf sebagai pengukuran utama di kebanyakan negara, memandangkan ia merupakan salah satu elemen penting dalam manual Sistem Akaun Negara (SNA) 2008 yang menjadi rujukan utama negara-negara dalam penyusunan Keluaran Dalam Negara Kasar (KDNK). Dalam manual tersebut, konsep ini di rujuk sebagai Pendapatan Boleh Guna Isi Rumah Disesuaikan (OECD 2024).

Dalam kajian mengenai kesan perbelanjaan isi rumah oleh faktor-faktor seperti harga petrol kenderaan, pendapatan boleh guna isi rumah dan kadar naik-turun tukaran matawang di Afrika Selatan, didapati bahawa adanya hubungan jangka panjang antara faktor-faktor tersebut terutamanya pengaruh oleh kumpulan pendapatan dan kadar tukaran matawang, selain dari kepentingan pendapatan boleh guna isi rumah dalam mempengaruhi pola perbelanjaan isi rumah. (Habanabakize 2021). Dalam kajian lain,

sistem cukai dan pindahan tunai di Denmark telah berjaya mengurangkan ketidaksamaan pendapatan isi rumah dengan mengurangkan kesan perubahan pendapatan bukan sahaja dalam jangka pendek malah untuk jangka sederhana hingga panjang. Reformasi cukai yang menurunkan kadar cukai pendapatan margin secara signifikan telah menyumbang kepada pertumbuhan pendapatan boleh guna isi rumah, bahkan melebihi pertumbuhan pendapatan bagi kebanyakan individu dan isi rumah. (Leth-Petersen & Sæverud 2022).

Sehubungan itu, cadangan Kerajaan untuk merangka halatuju baharu dalam peralihan klasifikasi kumpulan pendapatan isi rumah B40, M40 dan T20 kepada kaedah berdasarkan nilai pendapatan boleh guna adalah suatu langkah yang sangat tepat dan relevan dengan keadaan semasa. Sebagai sebuah negara yang progresif dan responsif terhadap cabaran persekitaran ekonomi global yang semakin kompleks dan tidak menentu, Malaysia perlu memiliki keberanian dan wawasan yang tinggi dalam mengambil pendekatan sebegini, walaupun berdepan dengan risiko-risiko tidak dapat diramalkan.

2.6 BEBAN PERBELANJAAN

Fenomena beban perbelanjaan atau lebih mudah dikenali sebagai beban kewangan adalah permasalahan yang sentiasa dialami oleh individu atau keluarga dalam memenuhi tuntutan dan keperluan kehidupan seharian. Dalam konteks isi rumah, ia adalah sangat berkait rapat dengan keadaan pendapatan, taraf sosioekonomi, gaya hidup, pola perbelanjaan, harga barangan & perkhidmatan, lokasi tempat tinggal dan lain-lain faktor. Ia boleh terjadi bukan hanya kepada golongan berpendapatan rendah, namun juga boleh berlaku kepada golongan berpendapatan tinggi. Keadaan akan lebih buruk sekiranya seseorang individu itu kurang kesedaran dan kecelikan kewangan dalam mengurus sumber pendapatan dan perbelanjaan selain elemen kehendak melebihi keperluan akibat perubahan gaya hidup yang mendadak. Justeru, keperluan agar dapat diwujudkan sebuah dasar dan polisi yang berupaya untuk membantu meningkatkan kesedaran dan tahap kecelikan kewangan semua golongan rakyat adalah sebuah agenda penting dan harus diusahakan demi kepentingan sesebuah negara dan masyarakat untuk jangka panjang (Levantesi & Zacchia 2021).

Dalam inisiatif panduan kewangan “Belanjawanku” tahun 2022/2023 yang dibangunkan oleh Kumpulan Wang Simpanan Pekerja (KWSP) dengan kerjasama Pusat Penyelidikan Kesejahteraan Sosial (SWRC) Universiti Malaya (UM), jumlah perbelanjaan bulanan yang minimum bagi seseorang individu bujang yang tinggal di Lembah Klang dan hanya menggunakan pengangkutan awam adalah RM1,930. Manakala di Alor Setar, mencatatkan kadar perbelanjaan terendah bagi golongan yang sama dengan RM1,530. Panduan Belanjawanku ini sebenarnya memfokuskan kepada perincian perbelanjaan minimum mengikut kategori isi rumah (bujang, keluarga atau warga emas) cadangan mengenai peruntukan perbelanjaan, perbelanjaan pilihan dan juga simpanan persendirian. Ia bukan hanya untuk bandar-bandar di Lembah Klang, malah juga untuk bandar-bandar utama seluruh negara.

Jadual 2.4 Kadar perbelanjaan minima berdasarkan demografi isi rumah dan bandar (RM)

Bandar/ Kategori Isi Rumah	Bujang Tanpa Kereta	Bujang Dengan Kereta	Berkahwin Tanpa Anak	Berkahwin (1 Anak)	Berkahwin (2 Anak)	Warga Emas (Bujang)	Warga Emas (Berkahwin)	Ibu/Bapa Tunggal (1 Anak)	Ibu/Bapa Tunggal (2 Anak)
Lembah Klang	1,930	2,600	4,630	5,980	6,890	2,520	3,210	4,740	5,650
Alor Setar	1,530	2,060	3,680	4,760	5,430	2,020	2,630	3,660	4,340
K.Kinabalu	1,710	2,230	4,000	5,130	5,840	2,290	2,930	4,000	4,710
Johor Bahru	1,760	2,290	4,110	5,360	6,100	2,330	3,020	4,200	4,940
K.Terengganu	1,630	2,160	3,820	4,920	5,610	2,090	2,730	3,820	4,510
Kuching	1,680	2,210	3,920	5,020	5,720	2,160	2,790	3,890	4,590
Kuantan	1,710	2,230	3,910	5,030	5,740	2,130	2,780	3,900	4,610
Kota Bharu	1,540	2,110	3,750	4,830	5,520	2,050	2,690	3,720	4,400
Georgetown	1,830	2,430	4,360	5,640	6,370	2,450	3,140	4,460	5,190
Ipoh	1,680	2,270	3,970	5,140	5,850	2,190	2,840	4,020	4,730
Seremban	1,720	2,300	4,170	5,400	6,130	2,250	2,910	4,250	4,980
Bandar Melaka	1,730	2,270	4,010	5,140	5,850	2,180	2,830	4,020	4,720

Sumber: Belanjawanku 2022/2023, KWSP & UM

2.7 PERBELANJAAN MELEBIHI PENDAPATAN

Perbelanjaan melebihi pendapatan atau terlebih belanja adalah istilah yang merujuk kepada tingkahlaku individu yang berbelanja melebihi daripada sumber wang atau peruntukan yang ada atau yang telah diperuntukkan bagi sesuatu tempoh atau pembelian tertentu. Terlebih belanja boleh mengakibatkan kepada pertambahan hutang, menambah tekanan kewangan yang tidak diperlukan dan secara tidak langsung meningkatkan cabaran untuk memenuhi keperluan perbelanjaan. Ini sekaligus menjadikannya sebagai tingkah laku yang boleh menyumbang kepada ketidakstabilan kewangan, hutang tidak terkawal dan pelbagai kesan negatif terhadap kesejahteraan kewangan dan keseluruhan kualiti hidup seseorang individu. Terma "hutang" digunakan apabila seseorang mempunyai komitmen kewangan yang belum diselesaikan, iaitu bayaran yang sepatutnya dibayar tetapi belum dilunaskan. Namun, istilah "terlebih berhutang" merujuk kepada hutang yang tidak boleh dilunaskan atau mungkin hanya boleh diselesaikan dengan kesukaran yang sangat besar (Achtziger et al. 2015).

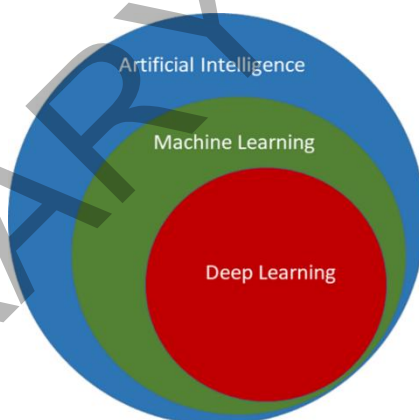
Belanja berlebihan mungkin berpunca daripada keperluan asas keluarga. Namun, ia juga disebabkan oleh gaya hidup individu yang tidak mengukur kemampuan diri sendiri. Walaubagaimanapun, tidak dinafikan bahawa terlebih belanja boleh berlaku disebabkan oleh kecemasan atau peristiwa yang tidak dijangka yang mungkin berlaku sekali-sekala (Othman et al. 2020). Di sesetengah negara, terdapat peratusan individu cenderung untuk berbelanja secara berlebihan yang akhirnya mendapat hutang besar. (Lea 2021). Hutang peribadi merupakan antara produk hutang yang popular dan lebih diterima di beberapa negara, namun akhirnya dapat dilihat bahawa ramai peminjam menyesal mengambil hutang pengguna. (Ong et al. 2019). Sebagai contoh, lebih daripada 40% daripada isi rumah di Amerika Syarikat sentiasa merasa bahawa mereka berbelanja melebihi had, dan menyesal akan pilihan kewangan mereka namun mereka tidak mengetahui apakah jenis fenomena terlebih belanja yang sedang dialami (Sui et al. 2021). Pengguna di China juga didapati semakin terjejas dengan fenomena belanja berlebihan dan juga terhadap hutang pengguna yang hampir tidak dapat dibayar semula (Song et al. 2020). Sebuah kajian menyimpulkan bahawa, dua faktor utama iaitu penawaran produk atau pilihan hutang dan kredit yang semakin banyak serta meningkat dan juga ditambah dengan strategi pemasaran hutang pengguna yang lebih meluas dan

berkesan telah dikenalpasti antara punca utama kenaikan fenomena terlebih belanja dan peningkatan hutang pengguna (Hamid & Loke 2021).

2.8 TEKNIK PENGELOMPOKAN PEMBELAJARAN MESIN

2.8.1 Pengenalan

Kecerdasan Buatan (*Artificial Intelligent*) adalah suatu program kecerdasan pengkomputeran yang disematkan di dalam komputer atau mesin khusus bagi memberinya satu kemampuan dan kecerdasan untuk cuba berfikir dan meniru mengikut cara dan kaedah sebagaimana kecerdasan semulajadi manusia. Objektif utamanya adalah untuk memberikannya kemampuan untuk membuat keputusan sendiri tanpa intervensi manusia. Pada dasarnya, Kecerdasan Buatan mengandungi kaedah pembelajaran yang dibahagikan kepada dua pecahan utama, iaitu Pembelajaran Mesin (*Machine Learning*) dan Pembelajaran Mendalam (*Deep Learning*).



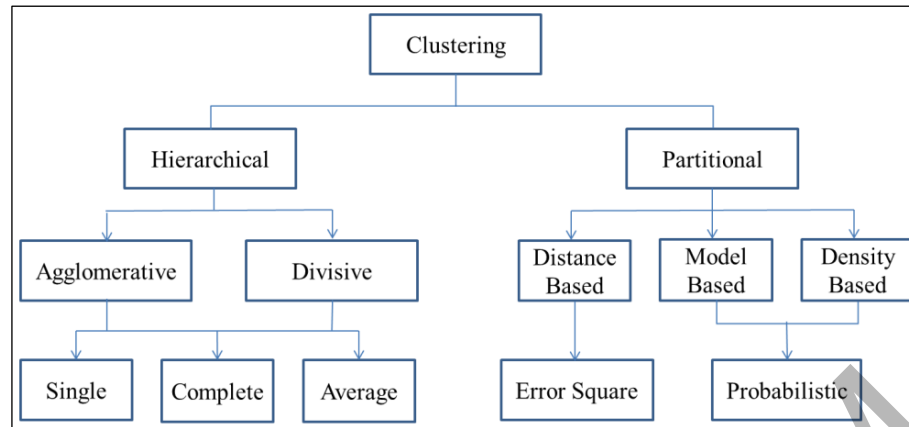
Rajah 2.2 Hubungan antara Kecerdasan Buatan (AI), Pembelajaran Mesin (ML) dan Pembelajaran Dalam (DL)

Menggunakan algoritma pembelajaran mesin dalam analitik data adalah tidak asing lagi dalam dunia penyelidikan pada hari ini. Ia adalah program atau pengaturcaraan yang dapat belajar dan dilatih dari pengalaman data tanpa perlunya campurtangan manusia dalam membuat penilaian dan keputusan. Konsep belajar ini merangkumi pelbagai skop dan tugas, seperti mempelajari struktur, corak dan tren tersembunyi dalam data berlabel atau tidak berlabel, pemetaan berasaskan input ke output,

menghasilkan label atau ramalan dan seterusnya menghasilkan model baharu atau keputusan akhir berasaskan semua yang telah dipelajari dari data.

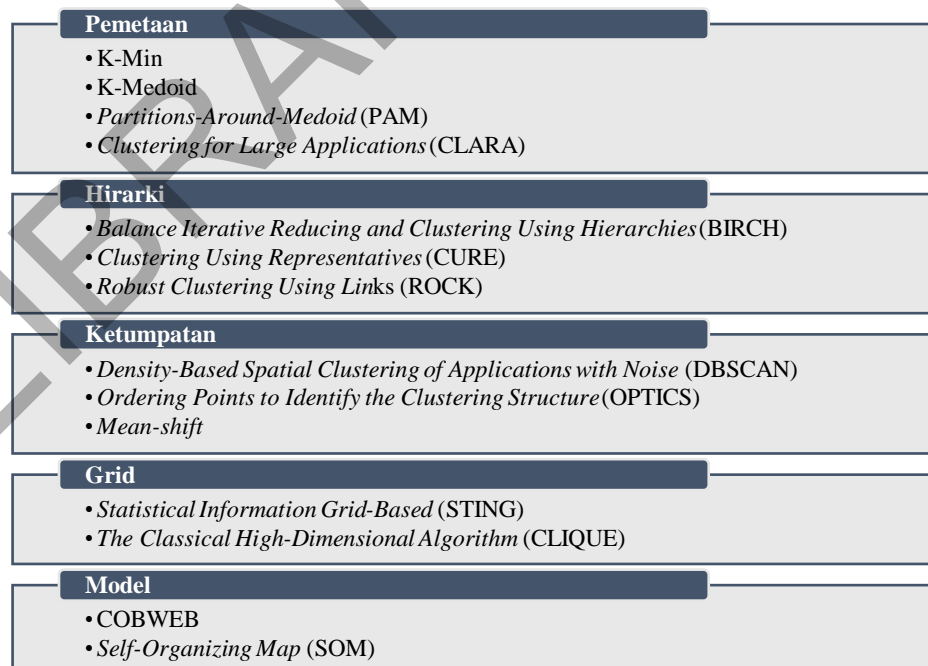
Pendekatan yang digunakan ini telah banyak membantu para pengkaji khususnya dalam analisis data untuk meneroka dan menemukan maklumat dan pengetahuan yang tepat, signifikan dan penuh makna dalam pelbagai domain bidang kajian menggunakan set data atau data raya yang besar. Ia menawarkan pelbagai tugas analitik seperti pengkelasan, pengelompokan, peramalan, pengesanan sisihan dan diagnostik. (Sevakula et al. 2020). Umumnya, terdapat empat jenis algoritma pembelajaran mesin yang utama, bergantung kepada jenis input atau data dan juga jenis output yang dikehendaki. Ia adalah pembelajaran dengan penyeliaan (*supervised learning*), pembelajaran tanpa penyeliaan (*unsupervised learning*), pembelajaran semi-penyeliaan (*semi-supervised learning*) dan pembelajaran pengukuhan (*reinforcement learning*).

Selain dari pengkelasan, antara teknik perlombongan data dengan pembelajaran mesin yang popular digunakan adalah teknik pengelompokan (*clustering*). Pengelompokan adalah satu teknik pembelajaran tanpa penyeliaan (*unsupervised learning*) yang mana membolehkan pengelompokan dibuat tanpa label data (*data label*). (Ahmed et al. 2020). Objektif utama teknik pengelompokan adalah bagi mengenal pasti dan mengumpulkan objek dan data yang mempunyai ciri-ciri persamaan ke dalam satu kelompok khusus dan setiap kelompok-kelompok tersebut mempunyai perbezaan antara satu dengan yang lain. Pengelompokan memungkinkan pengkaji untuk mengenal pasti sesuatu corak atau susunan objek dan elemen yang saling berkaitan. Justeru dapatlah diketahui sebarang persamaan atau perbezaan yang wujud dalam sesebuah set data (Bahari & Dollah@Md. Zain 2005). Teknik pengelompokan sangat bergantung kepada cara bagaimana untuk memisahkan objek yang sama atau mempunyai kemiripan dengan yang berbeza atau tiada kemiripan antara satu sama lain. (Entezami et al. 2020).



Rajah 2.3 Taksonomi Teknik Algoritma Pengelompokan (Saxena et al. 2017)

Teknik pengelompokan biasanya dibahagikan kepada dua jenis yang utama iaitu pembahagian (partitional) dan hirarki (hierarki). Namun, terdapat lima teknik utama bagi algoritma pengelompokan yang umum dan ianya boleh dibahagikan kepada lima teknik pendekatan utama iaitu pendekatan pengelompokan secara pemetaan, hirarki, ketumpatan, penggridan (grid) dan pemodelan (modelling). (Abu Bakar et al. 2018). Jadual berikut memperihalkan contoh-contoh algoritma lazim digunakan bagi kelima-lima kaedah pendekatan pengelompokan tersebut.

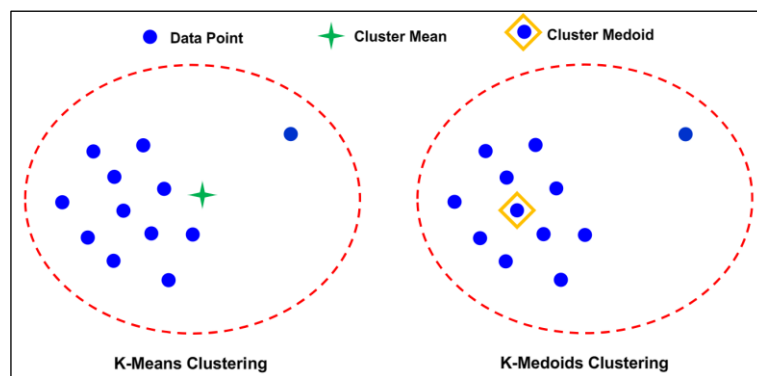


Rajah 2.4 Contoh Teknik Algoritma Pengelompokan Yang Lazim Digunakan

2.8.2 Algoritma Pengelompokan Pemetaan

Antara algoritma pengelompokan yang popular dan sentiasa menjadi pilihan pengkaji dalam pembelajaran mesin adalah pengelompokan menggunakan algoritma k-Min (*k-Mean*). Ia adalah sebahagian daripada algoritma pemetaan dan menjadi pilihan kerana ianya agak mudah (*simple*) di samping menjimatkan masa. Algoritma k-Min adalah sebuah teknik pembelajaran tanpa penyeliaan di mana ianya berupaya menganalisis set data yang tidak mempunyai data label. Objektif utama algoritma ini adalah bagi mencari kelompok atau kumpulan yang mempunyai kemiripan yang hampir sama dari sesebuah set data di mana pembolehubah k mewakili jumlah kumpulan atau kelompok yang diinginkan. Pada asasnya, idea algoritma k-Min adalah algoritma pengelompokan yang berasaskan sentroid (*centroid*) atau pusat bentuk yang bertujuan untuk mengelompokkan titik data atau objek ke dalam k kluster, di mana k adalah parameter yang telah ditentukan lebih awal. Algoritma ini secara berulang akan menentukan setiap objek ke sentroid kluster terdekat dan kemudiannya akan terus memperbaharui titik sentroid berdasarkan purata jarak dari objek yang ditugaskan ke setiap kluster. Proses ini akan berterusan sehinggalah tahap titik tumpuan (*convergence*) dicapai, biasanya ketika sentroid telah stabil atau ketika jumlah ulangan yang ditentukan telah tercapai. Analisis pengelompokan k-min adalah kaedah pengelompokan terbaik untuk mengurangkan darjah ralat kesilapan kerana ia mengklasifikasikan objek data ke dalam kelompok terhampir berdasarkan kiraan jarak paling minimum yang biasanya menggunakan jarak *Euclidean* (Yee et al. 2023).

Selain k-Min, algoritma k-Medoids juga digunakan dalam kajian pengelompokan pembelajaran mesin. Algoritma k-Medoid secara teknikalnya adalah merupakan varian dan versi penambahbaikan kepada algoritma k-Min itu sendiri. Sedikit berbeza dengan k-Min yang menggunakan pusat bentuk atau sentroid, k-Medoid memilih titik data sebenar yang disebut medoid pada sesebuah kumpulan objek untuk mewakili sesuatu kelompok. Hasil dari pengiraan jarak di antara medoid dengan objek non-medoid ini tadi maka akan terhasillah kluster. Antara lain, algoritma-algoritma berdasarkan pemetaan lain yang biasa digunakan ialah seperti *Partitions-Around-Medoid* (PAM) dan *Clustering for Large Applications* (CLARA).



Rajah 2.5 Perbezaan antara algoritma k-Min dan k-Medoid

Sumber: (Entezami et al. 2020)

2.8.3 Algoritma Pengelompokan Hirarki

Pengelompokan hirarki adalah sebuah kaedah yang boleh dikatakan agak klasik, namun masih relevan dalam bidang data analitik dan algoritma pengelompokan. Bermula dengan idea kaedah dendogram pada 1930an dan seterusnya kepada kaedah Jiran-Terdekat (*nearest-neighbour*) pada tahun-tahun 50an dan 60an, ia terus berkembang pada dekad-dekad seterusnya dimana diperluaskan kepada disiplin ilmu seperti seperti statistik, sains komputer, biologi dan perubatan. Algoritma pengelompokan hirarki adalah merupakan struktur yang menyamai sistem pelbagai-tingkat (*multi-level*), piramid atau pohon. Idea kepada algoritma ini adalah untuk mengumpulkan atau mengelompokkan kumpulan objek atau data yang mempunyai persamaan dan kemiripan yang paling hampir. Ia kemudiaannya diikuti oleh objek yang memiliki persamaan kemiripan yang kedua terhampir dan sehinggalah proses seterusnya hingga dapat membentuk kelompok seperti bentuk piramid atau pepohon. Piramid ini menunjukkan susunan hirarki yang jelas di antara objek dan data daripada yang mempunyai kesamaan atau kemiripan tertinggi kepada yang terendah.

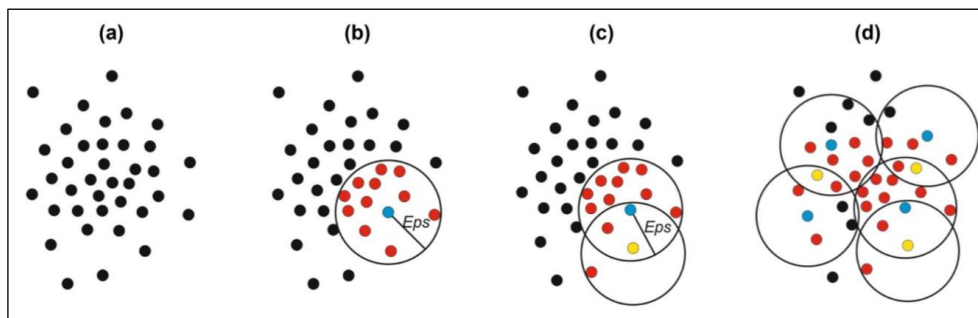
Secara umumnya, pengklusteran hierarki boleh dibahagikan kepada dua jenis utama iaitu pemusatan (*agglomerative*) dan pemecahan (*divisive*). Pengklusteran pemusatan (*agglomerative clustering*) atau biasanya disebut sebagai *agglomerative nesting* (AGNES) menggunakan pendekatan pengelompokan secara bawah-atas (*bottom-up*). Manakala pengklusteran pemecahan (*divisive clustering*) yang biasa

disebut sebagai divisive analysis (**DIANA**) melaksanakan tugas pengelompokan data secara atas-bawah (*top-down*). Kedua-dua kaedah ini dibezakan berdasarkan pendekatan dalam melakukan pengelompokan, samada kaedah secara bawah-atas (*bottom-up*) atau atas-bawah (*top-down*) (Dewi 2012).

2.8.4 Algoritma Pengelompokan Ketumpatan

Algoritma pengelompokan berasaskan ketumpatan atau kepadatan adalah sebahagian daripada pengelompokan pembahagian (*partitional clustering*). Objektif kaedah pengelompokan ini adalah memisahkan objek data yang “berketumpatan tinggi” (*high density*) dengan objek data yang “berketumpatan rendah” (*low density*). Algoritma ini juga dapat menentukan jumlah kluster yang akan dihasilkan dan ianya mampu untuk mengesan sebarang *noise* atau *outlier*. Teknik yang popular bagi algoritma berasaskan ketumpatan ini adalah *Density-Based Spatial Clustering of Applications with Noise* atau lebih dikenali sebagai DBSCAN.

DBSCAN diperkenalkan oleh Martin Ester, Hans-Peter Kriegel, Jörg Sander, dan Xiaowei Xu dalam pembentangan kertas seminar berjudul "*A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*" yang diterbitkan pada tahun 1996 (Ester et al. 1996). Oleh kerana ia direkabentuk khusus bagi menyesuaikan perlombongan data bagi data set berbentuk spatial, algoritma ini bekerja dengan mengelompokkan objek-objek bagi data spatial berdasarkan ketumpatannya, tanpa memerlukan pengguna untuk menentukan jumlah kluster sebelumnya. Justeru, ini akan mendefinisikan kluster sebagai kawasan padat titik data yang dipisahkan oleh kawasan ketumpatan yang lebih rendah.



Rajah 2.6 Idea konsep algoritma DBSCAN (Entezami et al. 2020) (a) Sebuah kluster (b) Titik biru sebagai titik rujukan (c) titik kuning sebagai titik batas (d) Kluster yang ditentukan berdasarkan ketumpatan/kepadatan

2.8.5 Algoritma Pengelompokan Grid

Seterusnya, kaedah pengelompokan berasaskan grid memperkenalkan sedikit perbezaan dalam pendekatannya berbanding dengan algoritma pengelompokan konvensional. Algoritma ini adalah teknik yang digunakan untuk membahagikan kumpulan data objek menjadi kelompok-kelompok yang berasaskan kepada struktur grid nilai ruang (*value space*) yang ditetapkan kepada objek data. Secara asasnya, ia mengubah ruang data asal kepada struktur grid dengan saiz tertentu untuk tujuan pengelompokan. Kelebihan utama pengelompokan berasaskan grid ini adalah ia agak mudah dikendalikan dan mengurangkan tahap kompleksiti pengkomputeran terutamanya apabila melibatkan analitik data bagi pemprosesan data bersaiz besar.

Antara algoritma pengelompokan berasaskan grid yang biasa digunakan adalah *Statistical Information Grid* atau STING. Ia diperkenalkan oleh Wang, Yang, dan Muntz pada 1997. STING dirancang untuk mengatasi cabaran data spatial yang bersaiz besar secara berkesan dengan membahagikan ruang data objek kepada ruang bentuk struktur grid agar kerja-kerja pengelompokan dilakukan di dalam dan berasaskan kepada setiap ruang grid tersebut. (Wang et al. 1997). Selain itu, *The Classical High-Dimensional Algorithm* (CLIQUE) juga adalah alternatif kepada teknik algoritma yang digunakan dalam pengelompokan berasaskan grid ini. Secara keseluruhan, algoritma pengelompokan berasaskan grid ini menawarkan pendekatan yang sederhana namun berkesan untuk membahagikan data spatial bersaiz besar menjadi kelompok-kelompok bermakna berdasarkan struktur grid.

2.8.6 Algoritma Pengelompokan Model

Pengelompokan berasaskan model adalah sebuah pendekatan statistik bagi kajian pengelompokan dalam perlombongan data. Data yang dianalisa (multivariat) diandaikan adalah terhasil daripada campuran model komponen yang terbatas. Setiap model komponen adalah taburan kebarangkalian, yang biasanya adalah parameter taburan multivariat (Banerjee & Shan 2010) Sebagai contoh, dalam model taburan campuran multivariat Gauss, setiap komponen adalah taburan multivariat Gauss. Algoritma pengelompokan berasaskan model secara umumnya adalah bertujuan untuk memilih model tertentu serta menentukan model yang paling sesuai untuk sesebuah kelompok bagi menjelaskan data objek yang menjadi asas dalam mengenal pasti kelompok berdasarkan parameter model.

Selain kaedah pembelajaran berdasarkan statistik seperti *Categorical Object Web* (COBWEB) yang bekerja dengan melaksanakan pengkelasan berdasarkan kriteria heuristik tertentu dengan anggapan bahawa taburan kebarangkalian bagi setiap atribut tidak berkait, terdapat juga kaedah pembelajaran berdasarkan rangkaian neural. Kaedah pembelajaran rangkaian neural yang umum digunakan adalah *Self-Organizing Map* (SOM) yang pada asasnya melakukan proses pengelompokan dengan membentuk jaringan kohonen di mana banyak digunakan untuk mengurangkan dimensi, visualisasi data dan seterusnya tugas pengelompokan yang mana pengelompokan tersebut dilaksanakan berdasarkan ciri-ciri datanya.

2.8.7 Kelebihan dan Kelemahan Setiap Teknik Pengelompokan

Tidak dinafikan, dalam setiap teknik dan algoritma yang diaplikasikan bagi kajian berasaskan pembelajaran mesin, sudah pastinya terdapat kelebihan dan kelemahan bagi setiap algoritma tersebut. Begitu juga dengan teknik dan algoritma pengelompokan. Sebilangan pengkaji telah menjalankan kajian tentang kelebihan dan kelemahan algoritma ini. (Xu & Tian 2015) dalam kajian komprehensif yang dilaksanakan terhadap beberapa teknik pengelompokan mendapati bahawa setiap pendekatan bagi kaedah dan teknik pengelompokan mempunyai beberapa kelebihan dan kelemahan teknik itu sendiri. Jadual berikut memperincikan secara ringkas mengenai setiap kelebihan dan kekurangan bagi teknik-teknik algoritma pengelompokan yang dibincangkan.

Jadual 2.5 Kelebihan dan Kekurangan Teknik Pengelompokan

Teknik Algoritma Pengelompokan	Kelebihan	Kelemahan
Pemetaan	<ol style="list-style-type: none"> Keputusan pengelompokan dengan kecekapan tinggi Proses agak cepat dan pantas kerana kompleksiti masa yang rendah 	<ol style="list-style-type: none"> Perlu menetapkan bilangan kelompok awal Sangat sensitif kepada parameter dan outliers Hasil pengelompokan yang sensitif terhadap jumlah kelompok disasarkan Hasil kelompok bersifat local optima
Hirarki	<ol style="list-style-type: none"> Tugas berskala besar Sesuai dengan set data berbentuk arbitrari Dapat mengesan hubungan antara kelompok 	<ol style="list-style-type: none"> Bilangan kelompok yang perlu ditetapkan awal Tugas agak kompleks Proses agak mengambil masa
Ketumpatan	<ol style="list-style-type: none"> Sesuai untuk set data arbitrari Kecekapan tinggi dalam tugas pengelompokan 	<ol style="list-style-type: none"> Ruang memori yang besar diperlukan bagi data bersaiz besar Pengelompokan menjadi kurang berkualiti jika kepadatan ruang objek data tidak seragam/sekata Agak sensitif kepada pembolehubah semasa tugas pengelompokan
Grid	<ol style="list-style-type: none"> Amat sesuai untuk pemprosesan secara serentak/selari atau pengemaskinian secara berperingkat Proses agak pantas dan kompleksiti masa rendah Berskala besar 	<ol style="list-style-type: none"> Kecekapan hasil pengelompokan yang tinggi namun kualiti kelompok yang rendah Sensitiviti hasil pengelompokan terhadap data mikro/granularity yang akan mengurangkan ketepatan pengelompokan
Model	<ol style="list-style-type: none"> Setiap model mempunyai ciri-ciri istimewa yang boleh memberi maklumat tambahan dalam set data bidang tertentu Model yang dibina adalah baik dan bervariasi 	<ol style="list-style-type: none"> Hasil pengelompokan adalah agak sensitif kepada pembolehubah yang mana model yang telah dipilih Proses agak mengambil masa bergantung kepada jenis dan saiz data

2.9 PELBAGAI TEKNIK PERLOMBONGAN DATA DALAM KAJIAN SOSIOEKONOMI

Teknik perlombongan data dengan menggunakan algoritma pengelompokan bukanlah asing lagi dalam kajian-kajian ilmiah pada era ini. Berdasarkan kajian literasi yang dijalankan sehingga kini, banyak kajian didapati selesai dan bersesuaian mengaplikasikan teknik pengelompokan dalam pelbagai bidang dan domain kajian, termasuklah dalam domain kajian berkaitan sosio-ekonomi isi rumah. Kajian (Yee et

al. 2023) berkenaan analisis pengelompokan k-Min dan model *Multi Linear Regression* (MLR) terhadap pendapatan isi rumah di Malaysia mendapati kombinasi analisis pengelompokan k-Min dengan model MLR mampu untuk menghasilkan pendekatan yang lebih efektif dalam mengurangkan kadar ralat jika dibandingkan dengan hanya menggunakan MLR semata-mata. Justeru, menurut mereka dengan mengelompokkan data sebelum melakukan analisis regresi, kajian ini memperoleh nilai *Mean Square Error* (MSE) yang lebih kecil, sekaligus menunjukkan anggaran pendapatan isi rumah yang lebih tepat.

Kajian pengelompokan k-Min dari penggunaan elektrik menggunakan data masa dari meter pintar mencadangkan bahawa pentingnya untuk menggunakan data meter pintar dan teknik pengelompokan untuk memahami dan mengkategorikan kepenggunaan elektrik berdasarkan tingkahlaku penggunaan oleh pengguna. (Okereke et al. 2023). Dalam kajian berkenaan prestasi akademik mahasiswa di Institut Pengajian Tinggi (IPT), teknik pengelompokan telah digunakan sebagai algoritma pembelajaran mesin dan pengkaji telah memperoleh gambaran tentang pola prestasi, tingkahlaku dan pencapaian mahasiswa, yang sekaligus dapat memberikan idea untuk sebarang intervensi dan strategi dengan sasaran bagi membantu kejayaan mahasiswa di IPT (Nafuri et al. 2022).

Dalam kajian berkenaan klasifikasi kemiskinan pelbagai dimensi bagi kumpulan B40 di Malaysia dengan menggunakan teknik pengelompokan, dari sejumlah 532,000 data isi rumah dari Banci Penduduk & Perumahan 2010 yang digunakan dalam telah berjaya menghasilkan tiga (3) model pengelompokan bagi mengelompok kumpulan pendapatan B40 menggunakan algoritma k-Min. Hasil dr pengelompokan tersebut, kajian ini telah berjaya menghasilkan tiga (3) set indikator kemiskinan pelbagai dimensi bagi kumpulan B40 tersebut berdasarkan jenis ketersisihan yang dialami. Sebagai penemuan utama kajian, terdapat dimensi baru yang dikenal pasti iaitu dimensi teknologi selain dimensi sedia ada iaitu pendidikan, pendapatan dan taraf hidup yang dilihat penting dalam mengelompok kumpulan B40 di Malaysia (Abdul Rahman et al. 2021).

Banyak kajian lain turut memperihalkan tentang penggunaan dan kelebihan mengaplikasikan teknik-teknik perlombongan data dalam pelbagai domain kajian, khususnya dalam kajian-kajian berkaitan sosio-ekonomi terutamanya pendapatan dan