

KLASIFIKASI KATEGORI KEMISKINAN DI MALAYSIA MENGUNAKAN MODEL PENGELOMPOKAN

NUR ADLINA BINTI HAMZAN

DR. NOR SAMSAH SANI

Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia

ABSTRAK

Di Malaysia, pendapatan isi rumah sahaja yang menjadi ukuran kemiskinan sesebuah keluarga. Justeru, pengukuran kemiskinan ini tidak menunjukkan realiti sebenar kekurangan yang dialami oleh sesebuah keluarga. Pengukuran jenis ini iaitu menggunakan kaedah pendapatan dalam mengenalpasti individu atau isi rumah miskin juga dikenali sebagai pengukuran unidimensi. Kemiskinan seharusnya diukur dari pelbagai sudut dimensi yang berbeza, bukan hanya pada pendapatan isi rumah semata-mata. Kajian ini bertujuan untuk mengkaji apakah atribut yang perlu dititikberatkan dalam menentukan kemiskinan pelbagai dimensi bagi memberi keutamaan kepada golongan B40 di Malaysia dengan menggunakan teknik pengelompokan.

B40, M40 dan T20 adalah takrifan yang digunakan bagi mengelaskan pendapatan isi rumah masyarakat di Malaysia. Jabatan Perangkaan Malaysia telah bertanggungjawab untuk menjalankan kajian terhadap julat pendapatan yang dimiliki oleh rakyat Malaysia. Pada tahun 2017, Jabatan tersebut telah mengeluarkan laporan terhadap Penyiasatan Pendapatan Isi Rumah dan Kemudahan Asas (Anon 2019).

Baru-baru ini, bekas Perdana Menteri Tun Dr Mahathir Mohamad melancarkan Wawasan Kemakmuran Bersama 2030 (WKB 2030) yang merupakan hala tuju ekonomi negara bagi tempoh 10 tahun bermula 2021-2030. WKB 2030 telah dirangka oleh Kementerian Hal Ehwal Ekonomi dan Institut Masa Depan Negara (Anon 2019). Salah satu objektif utamanya ialah menangani jurang kekayaan dan pendapatan supaya rakyat yang ketinggalan dapat dibawa bersama dalam arus kemajuan negara (Anon 2019).

Dalam konteks kemiskinan di Malaysia, kemiskinan dijelaskan melalui perspektif kewangan atau pendapatan. Terdapat dua jenis konsep kemiskinan iaitu kemiskinan relatif dan kemiskinan mutlak. Konsep kemiskinan relatif dikaitkan dengan agihan pendapatan. Kemiskinan relatif akan sentiasa ada dan tidak dapat dihapuskan selagi adanya ketidakseimbangan dalam agihan pendapatan. Kemiskinan mutlak menunjukkan keadaan seseorang individu atau isi rumah yang tidak berkemampuan untuk memenuhi keperluan asas untuk menikmati kesejahteraan hidup. Kemiskinan didefinisikan sebagai kegagalan dalam memperoleh pendapatan yang stabil mengikut spesifikasi piawai yang digariskan dalam indeks kemiskinan negara (Wang 2004). Bagi mencapai usaha untuk mengklasifikasikan kemiskinan dalam masyarakat dengan lebih efisien dan berimpak tinggi, suatu pendekatan yang sesuai harus dijalankan supaya permasalahan ini dapat diatasi mengikut keperluan. Salah satu penyelesaian atau usahanya adalah dengan mencipta suatu algoritma yang menggunakan model pengelompokan untuk mengenal pasti atribut yang boleh menjadi penanda aras kepada pengelasan golongan B40.

Atribut yang diguna untuk mengisytiharkan seseorang itu berada dalam kelompok B40 adalah berdasarkan pada jumlah pendapatan isi rumah. Dengan menggunakan hanya satu atribut sahaja untuk mengelaskan seseorang itu dalam golongan B40, ianya tidak mencerminkan realiti dengan tepat akan tahap kemiskinan. Kemiskinan itu sendiri mempunyai pelbagai dimensi seperti aspek kesihatan, pekerjaan, pendidikan dan dan taraf hidup melalui skop bukan berasaskan pendapatan dan penggunaan. Oleh itu, pengukuran yang digunakan masih mempunyai kelemahan dan ketirisan, maka sudah pasti hasil bilangan kemiskinan yang diperoleh tidak menunjukkan angka sebenar kemiskinan. Namun begitu, realitinya pengukuran kemiskinan yang sebegini sudah biasa digunakan dalam menganalisis kemiskinan di negara kurang maju termasuklah Malaysia.

Menurut Tun Dr Mahathir Mohamad (2019), kerajaan Malaysia akan menggunakan Indeks Kemiskinan Pelbagai Dimensi (MPI) sebagai kaedah untuk mengukur kadar kemiskinan di Malaysia (Zanariah Abd Mutalib 2019). Dengan penggunaan MPI ini, penentuan tahap kemiskinan bukan sahaja terhadap pendapatan tetapi turut merangkumi segala layanan, kemudahan dan bantuan yang diberikan kepada rakyat.

Menurut Alkire dan Santos (2009), pendekatan yang menggunakan satu dimensi mempunyai batasanya. Bagi mereka, data yang diperoleh adalah di peringkat isi rumah sahaja dan ianya tidak memberi sebarang maklumat tentang sumber agihan isi rumah. Alkire dan Seth (2009) menemui beberapa kelebihan apabila menggunakan teknik yang dicadangkan oleh Alkire dan Foster (2007) iaitu indeks kemiskinan pelbagai dimensi (MPI) bagi meramal peratusan kemiskinan dari pelbagai dimensi dan mengenalpasti isi rumah yang miskin. Antaranya:

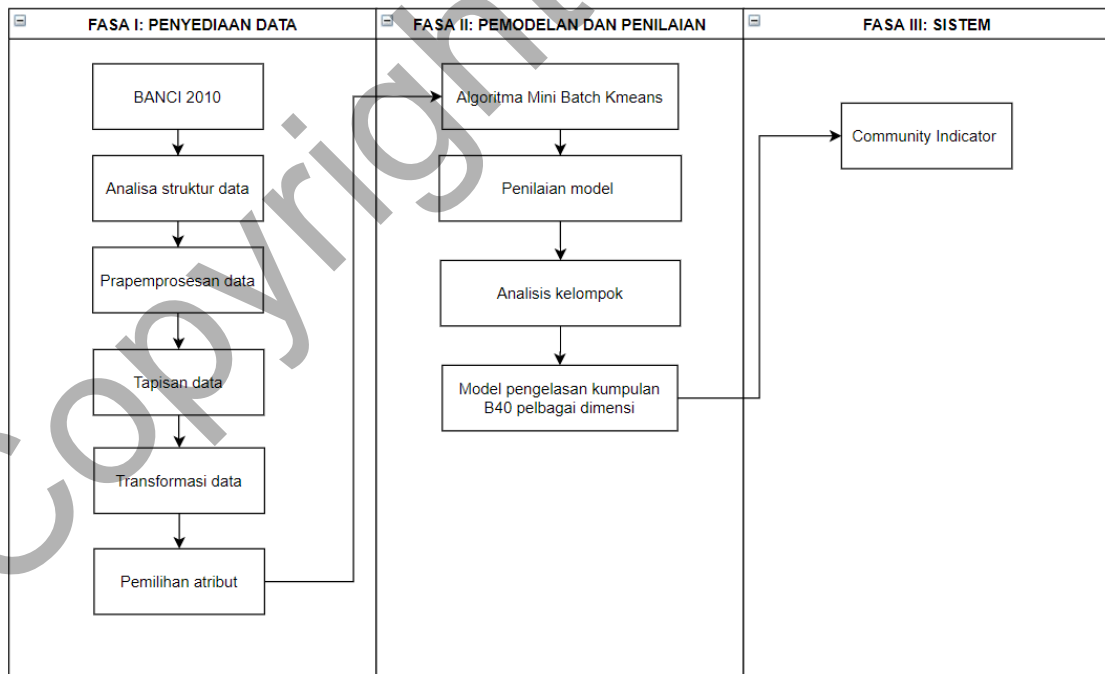
- a) bersifat kukuh dalam mengenalpasti siapa yang termiskin dikalangan yang miskin.
- b) mampu menunjukkan dimensi mana yang menyumbang berlakunya kemiskinan di dalam sesuatu kawasan atau kumpulan tertentu yang dikaji.

Secara spesifiknya, kajian ini dijalankan bagi mencapai objektif - objektif berikut:

1. Membangunkan model pengelompokan bagi mengelask kumpulan B40 berdasarkan indikator yang akan dikenalpasti.
2. Membangunkan sistem yang dapat mengelaskan kumpulan B40 berdasarkan indikator yang akan dikenalpasti.

4 METOD KAJIAN

Secara amnya, projek ini terbahagi kepada tiga fasa iaitu Fasa I: Penyediaan Data, Fasa II: Pemodelan dan Penilaian dan Fasa III: Sistem. Dalam fasa pertama iaitu fasa penyediaan data, data yang diperoleh akan dianalisis dan seterusnya digunakan untuk melalui proses pra-pemrosesan. Bagi fasa kedua iaitu fasa pemodelan dan penilaian, data yang tersedia selepas melalui pra-pemrosesan akan digunakan untuk tujuan proses pengelompokan. Pada fasa ketiga, sistem akan dibangunkan dan indikator kemiskinan pelbagai dimensi dapat dikenalpasti.



Rajah 4.1 Proses pembangunan model

4.1 Fasa Penyediaan Data

Fasa penyediaan data ialah proses untuk mengumpul, membersihkan dan menyatukan data dalam satu fail atau jadual untuk digunakan sewaktu proses analisis data. Berikut adalah langkah-langkah yang diambil sewaktu penyediaan data beserta penerangan.

4.1.1 Pra-pemprosesan Data

Tujuan pra-pemprosesan data adalah untuk mengubah data mentah kepada format atau bentuk yang mudah difahami dan sesuai sebagai persediaan kepada pembangunan model pada fasa seterusnya. Beberapa proses pra-pemprosesan data yang terlibat, antaranya tapisan data, transformasi data dan pemilihan atribut.

i. Tapisan data

Tapisan Warganegara

Fokus utama projek ini adalah tertumpu kepada warganegara Malaysia. Untuk itu, mereka yang bukan warganegara akan ditapis keluar iaitu sebanyak 37,932 individu. Ini menjadikan data projek adalah sebanyak 494,366 individu.

ii. Transformasi data

Bagi atribut yang datanya tidak berbentuk numerik, ia perlu ditransformasi kepada bentuk numerik. Dalam projek ini, transformasi data penting untuk memenuhi keperluan model pengelompokan. Algoritma pengelompokan menggunakan pengesahan bersilang (cross-validation) sebagai teras algoritma dan hanya input berbentuk numerik sahaja dibenarkan. Atribut Kumpulan Umur telah dikenal pasti datanya berbentuk nominal dan ia perlu ditukar menjadi numerik. Jadual 4.1 menunjukkan transformasi data atribut Kumpulan Umur menggunakan operator 'Map' dengan menukar data kepada numerik.

Jadual 4.1 Transformasi data atribut Kumpulan Umur

Kumpulan Umur	Kod Integer	Jumlah Rekod
00-04	1	28,530
05-09	2	33,692
10 - 14	3	34,903
15 - 19	4	29,916
20 - 24	5	18,210

25 - 29	6	17,986
30 - 34	7	17,845
35 - 39	8	18,878
40 - 44	9	18,735
45 - 49	10	16,876
50 - 54	11	13,953
55 - 59	12	11,551
60 - 64	13	9,755
65 - 69	14	6,095
70 - 74	15	4,825
75 ke atas	16	5,348
Jumlah		287,098

iii. Pemilihan atribut

Set data yang telah dianalisis mengandungi banyak atribut dan ini menyebabkan ia tidak relevan untuk diproses kerana mengambil masa yang lama. Maka pemilihan atribut adalah penting sebelum melaksanakan aktiviti pemodelan. Tujuan pemilihan atribut adalah untuk mencari atribut yang minimum dan meningkatkan hasil kualiti pengelompokan yang optimum. Beberapa langkah dalam pemilihan atribut telah dilaksanakan dalam projek ini, antaranya seperti berikut:

- **Membuang atribut yang tidak berguna**

Operator 'Remove Useless Attribute' membuang atribut yang tidak berguna. 4 atribut telah dikenal pasti sebagai atribut tidak berguna. Hasil pemerhatian mendapati atribut Negara Kelahiran, Kewarganegaraan, Negara Kewarganegaraan adalah atribut yang tidak berguna kerana data tersebut hanya mengandungi individu dalam kalangan warganegara Malaysia. Atribut Jenis Kediaman mempunyai nilai yang sama iaitu 1 yang mana ia mewakili kediaman jenis Unit Rumah. Sebanyak 82 atribut yang tinggal selepas proses ini dijalankan.

- **Membuang atribut yang mempunyai kolerasi**

Atribut yang mempunyai kolerasi bermaksud atribut tersebut mempunyai kaitan dengan atribut lain. Atribut berkolerasi dibuang kerana mempunyai persamaan tingkah laku yang akan memberi kesan yang sama dalam membuat ramalan. Operator 'Remove Corelated Attribute' akan digunakan untuk menilai nilai kolerasi melebihi 0.95. Sebanyak 8 atribut telah dikenal pasti untuk dibuang dan menjadikan senarai atribut yang tinggal adalah sebanyak 67 atribut. Jadual 4.2 menunjukkan senarai atribut berkolerasi yang dibuang dan pasangan atribut tersebut.

Jadual 4.2 Senarai atribut berkorelasi yang dibuang dan pasangannya

Bil	Atribut berkolerasi	Pasangan Atribut berkolerasi
1	Bayaran Sewa	Sewa Bulanan
2	Bilangan Bilik Tidur	Bilangan Bilik
3	Id Ahli	Id Kediaman
4	Id Isi Rumah	Id Kediaman
5	Industri (1 digit)	Industri (3 digit)
6	Mencari Pekerjaan	Bekerja Kembali
7	Negeri	Kod Negeri/Negara
8	Status Tenaga Buruh	Status Pekerjaan

- **Membuang atribut yang tidak berkepentingan**

Berdasarkan pemerhatian dan analisis yang dilakukan, terdapat beberapa atribut yang perlu dibuang secara manual. Operator 'Select Attribute' akan digunakan bagi proses ini. 9 atribut telah dikenal pasti untuk dibuang dan ini menjadikan 73 atribut yang masih diguna pakai. Jadual 4.3 menunjukkan senarai atribut yang dibuang dan alasan ia dibuang.

Jadual 4.3 Senarai atribut yang dibuang dan alasannya

Bil	Atribut	Alasan Dibuang
1	Umur	Ia mempunyai persamaan dengan atribut Kumpulan Umur.
2	Kod Daerah	Ia mempunyai persamaan dengan atribut Daerah.

3	Kod Negeri	Ia mempunyai persamaan dengan atribut Negeri.
4	Bilangan Lelaki	Kedua-dua bilangan ini adalah pecahan bagi atribut jumlah isi rumah.
5	Bilangan Perempuan	
6	Id Kediaman	Atribut yang mempunyai nilai unik (<i>unique value</i>) perlu dibuang kerana ia akan menghasilkan kelompok kecil.
7	No Individu	
8	No Isi Rumah	
9	No Kediaman	

- **Membuang atribut yang tidak relevan**

Berdasarkan pemerhatian dan analisis yang dilakukan, terdapat beberapa atribut yang tidak relevan perlu dibuang secara manual. Atribut yang dibuang merupakan atribut yang bukan sebahagian daripada rujukan yang digunakan dalam Indeks Kemiskinan Pelbagai Dimensi Malaysia. Atribut tersebut juga tidak memberi gambaran yang jelas dalam menentukan kemiskinan. Jadual 4.4 menunjukkan atribut tidak relevan yang dibuang. Ini menjadikan bilangan atribut yang tinggal adalah sebanyak 16 atribut.

Jadual 4.4 Atribut tidak relevan yang dibuang

Atribut	Atribut	Atribut
Negeri	Bekerja (7 Hari Lalu)	Kamera Digital
Daerah	Bekerja Kembali	Komputer Peribadi
Strata	Jantina	Radio/Hi-Fi
Hubungan	Pekerjaan (3 Digit)	Jenis Isi Rumah
Etnik	Industri (3 Digit)	Bahan Binaan Dinding
Tempat lahir	Agama	Bekalan Air
Negeri kelahiran	1 Kereta	Talian Telefon
Kumpulan umur	1 Motosikal	Kamera Digital
Membaca dan Menulis	2 atau lebih Motosikal	i-pod/PDA
Menggunakan Komputer	Basikal	Tiada Item
Pernah Pergi ke Sekolah	Mesin Basuh	Jenis Rumah
Sijil Tertinggi	Peti Sejuk	Bilangan bilik
Televisyen	Pemain VCD/DVD	Bekalan Elektrik

Kemudahan Tandas	Ketuhar Gelombang Mikro	Jumlah isi rumah
Kemudahan Sampah	Kutipan Tempat kediaman 5 tahun lalu	Alasan Tidak Mencari Kerja

4.2 Fasa Analisa Deskriptif Data

LAMPIRAN (Jadual 1- Jadual 2) merupakan statistik bagi data yang digunakan di dalam kajian ini. LAMPIRAN (Rajah 1) menunjukkan kategori pekerjaan berserta nilai purata gaji tahunan. Individu berkahwin mempunyai bilangan data yang paling tinggi bagi status kahwin manakala inividu bercerai dan berpisah merupakan kumpulan minoriti.

Bagi pendidikan tertinggi pula, 25.51% daripadanya mempunyai pendidikan di peringkat sekolah menengah atas, 20.86% di peringkat sekolah rendah, 16.09% mempunyai lain-lain pendidikan, 13.92% lagi di peringkat sekolah menengah rendah, 10.66% tidak diketahui atau tiada pendidikan dan selebihnya mempunyai pendidikan di peringkat pra-sekolah, pra-universiti, sijil/diploma, ijazah dan ijazah sarjana.

Seterusnya, bilangan isi rumah dalam kediaman untuk Jenis 1 (1 individu) adalah kumpulan majoriti iaitu sebanyak 455,781 dan Jenis 8 (16-21 individu) adalah kumpulan minoriti iaitu sebanyak 18. Sebanyak 281,728 individu yang mempunyai pemilikan kediaman, 104,095 individu tidak mempunyai pemilikan kediaman dan 77,152 individu yang tidak dikeathui pemilikan kediamannya.

Status pemilikan kediaman persendirian mempunyai data yang paling tinggi iaitu sebanyak 376,965 manakala swasta adalah yang paling rendah iaitu sebanyak 9,811. Majoriti atau 73.00% individu mempunyai telefon bimbit. Peratus bagi individu yang tidak mempunyai saluran TV berbayar, penghawa dingin, komputer riba, 2 kereta, langganan internet, penapis air serta 3 atau lebih kereta adalah melebihi 60% bagi setiap atribut.

4.3 Fasa Reka Bentuk

4.3.1 Reka Bentuk Antara Muka

Reka bentuk muka adalah suatu proses dimana penghasilan antara muka yang membenarkan pengguna untuk berinteraksi dengan perisian. Lakaran awal reka bentuk antara muka untuk pengguna telah dibuat dengan menggunakan *Sublime*. Lakaran ini dibuat untuk tujuan menampakkan gambaran idea utama bagaimana ia akan digunakan. Rajah 4.2 menunjukkan aliran bagi pengguna. Mengikut Rajah 4.2 paparan utama skrin ketika sistem dimulakan. Pengguna perlu mengisi data yang diminta berdasarkan paparan. Selepas pengguna mengisi data yang diperlukan, maka pengelasan kumpulan B40 akan dipaparkan.

Community Indicator

Person Information

Marital Status: Select

Highest level of Education: Select

Occupation: Select

Have 2 cars?: Select

Have 3 cars or more?: Select

Have air-conditioner?: Select

Have laptop?: Select

Have handphone?: Select

Have paid TV channel?: Select

Have Internet subscription?: Select

Have Water filter?: Select

Resident Ownership: Select

Ownership Status: Select

Total person in Living Quarters: Select

Income Information

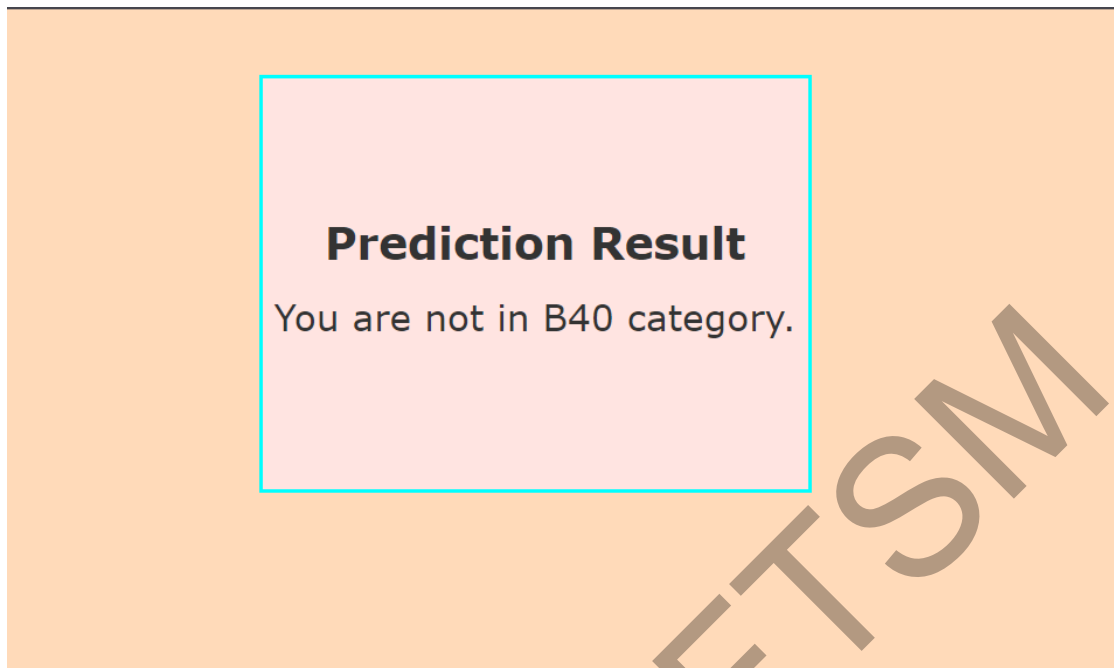
Total Monthly Household Income(RM):

Income(RM):

Predict B40 category

Pengguna perlu memasukkan data





Papar keputusan selepas kemasukan data

Rajah 4.2 Aliran untuk pengguna

4.3.2 Reka Bentuk Algoritma

Algoritma adalah satu prosedur langkah demi langkah, yang mentakrifkan satu set arahan yang akan dilaksanakan dalam aturan tertentu bagi mendapatkan hasil yang diinginkan. Dalam projek ini, algoritma yang digunakan adalah *Mini Batch K-Means*.

i. *Mini Batch K-Means*

Pembelajaran mesin tanpa selia merupakan cabang pembelajaran mesin yang mana modelnya berfungsi dengan sendirinya untuk mengetahui maklumat yang diperolehi. Ia biasanya digunakan terhadap data yang tidak berlabel. Tujuan utama merangka algoritma adalah untuk memudahkan proses pembangunan sistem. Fungsi yang terlibat dalam proses membangunkan sistem Community Indicator ini dapat diperlihatkan secara jelas dengan penggunaan algoritma yang pilihan. Algoritma telah dipilih untuk mereka bentuk sistem Community Indicator adalah *Mini Batch K-Means*.

Pada dasarnya, algoritma *Mini Batch K-Means* berfungsi sama seperti *K-Means*. Perbezaan antara *Mini Batch K-Means* dan *K-Means* adalah *Mini Batch K-Means* menggunakan kumpulan data rawak yang kecil dengan saiz yang tetap supaya ia boleh disimpan didalam memori. Idea utama algoritma *Mini Batch K-Means* adalah

menggunakan kumpulan data rawak yang kecil dengan satu ukuran tetap sehingga dapat disimpan dalam memori. Sampel rawak yang baru dari dataset diperoleh bagi setiap iterasi dan digunakan untuk mengemaskini kelompok dan ia berulang sehingga berlaku penumpuan.

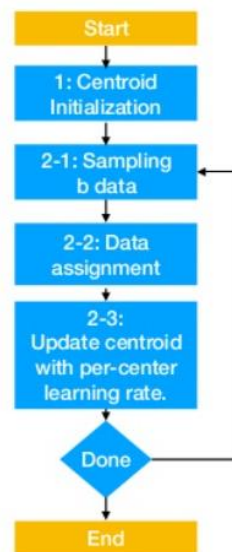
Setiap kumpulan mini akan mengemaskini kelompok menggunakan kombinasi cembung antara nilai prototaip dan data dengan menerapkan kadar pembelajaran yang menurun dengan jumlah iterasi. Kadar pembelajaran ini adalah songsangan dari jumlah data yang diberikan kepada kelompok semasa proses. Apabila bilangan iterasi meningkat, pengaruh data baru berkurang, sehingga penumpuan dapat dikesan ketika tidak ada perubahan dalam kelompok yang terjadi dalam beberapa iterasi berturut-turut.

Algoritma ini mengambil secara rawak kumpulan kecil data yang dipilih untuk setiap iterasi. Setiap data dalam kumpulan ditugaskan ke kelompok, bergantung pada lokasi sebelumnya dari *cluster centroids*. Ia kemudian mengemas kini lokasi pusat kluster berdasarkan titik baru dari kumpulan.

Sekiranya diberi dataset $T = \{x_1, x_2, \dots, x_p\}$, $x_i \in \mathbb{R}^{m \times n}$ dimana x_i mewakili satu rekod rangkaian yang merupakan vektor sebenar n -dimensi. Nilai m menunjukkan bilangan rekod yang terdapat dalam dataset T . Objek masalah pengelompokan adalah untuk mencari set C yang merupakan pusat kelompok $c \in \mathbb{R}^{m \times n}$ yang meminimumkan dataset T bagi rekod $c \in \mathbb{R}^{m \times n}$ fungsi berikut.

$$\min \sum_{x \in T} \|f(C, x) - x\|^2$$

Dimana, $f(C, x)$ merujuk pusat kelompok $c \in C$ yang terhampit kepada rekod x . $|C| = K$ dan K adalah bilangan kelompok yang dicari. Bagi menentukan pusat kelompok, nilai K yang dipilih secara rawak akan direkodkan menggunakan *KMeans++* (Bahmani 2012) dan pusat kelompok C ditetapkan nilainya sama seperti nilai K yang dipilih secara rawak. Berdasarkan Rajah 4.3 menunjukkan bagaimana aliran menggunakan *Mini Batch K-Means*.

Rajah 4.3 Aliran *Mini Batch K-Means*

4.3.3 Reka Bentuk Pangkalan Data (Dataset)

Reka bentuk pangkalan data merupakan suatu kumpulan maklumat yang dianjurkan agar dapat diurus, dikemas kini serta diakses dengan mudah. Ia berperanan penting dalam menguruskan maklumat yang dikelaskan secara tertentu. Data yang disimpan dan bakal disimpan perlu difahami terlebih dahulu agar proses penyusunan data dapat dikelaskan dan disimpan dengan teratur. Pangkalan data diperoleh hasil pra-pemprosesan data. Pangkalan data ini disimpan dalam satu fail *Excel*, dilabelkan sebagai *data.csv*.

4.4 Fasa Sistem

Dalam pembangunan sistem, teknologi kecerdasan buatan (pembelajaran mesin) akan disistemkan dengan menggunakan perisian Sublime Text dan Anaconda Powershell dalam bahasa pengaturcaraan *Sublime* dan *Python*. Terdapat empat fail komponen penting sewaktu membangunkan perisian iaitu *model.py*, *app.py*, *request.py* dan *HTML*.

Fail *model.py* mengandungi model pembelajaran mesin untuk meramal klasifikasi kategori kemiskinan. Fail *app.py* mengandungi *Flask API* yang menerima perincian data melalui panggilan *GUI* atau *API*, menghitung nilai yang diramalkan berdasarkan model dan mengembalikannya. File *request.py* pula digunakan terhadap model untuk memanggil *API* yang ditentukan dalam *app.py* dan memaparkan nilai yang dikembalikan. File *HTML* mengandungi templat *HTML* yang membolehkan pengguna memasukkan perincian yang diminta dalam sistem dan memaparkan ramalan klasifikasi kategori kemiskinan.

4.4.1 Pengeksportan Fail Model

Setelah algoritma *Mini Batch K-Means* dimodelkan dan model tersebut mendapat nilai ketetapan yang tinggi, maka model tersebut boleh dieksport. Model *Mini Batch K-Means* akan dieksport menggunakan fungsi *Pickle* dalam perisian *Jupyter Notebook*. Fungsi *Pickle* biasanya digunakan untuk menterjemahkan dan menyusun struktur objek *Python*. Dengan kata mudah, ia adalah cara untuk menulis objek python pada cakera yang boleh dipindahkan kemana sahaja dan kemudian dibaca kembali oleh skrip python. Selepas fungsi *Pickle* diimplementasi, satu file format .pkl akan dihasilkan. Fail yang dihasilkan dinamakan *model.pkl*.

4.4.2 Pengekodan Sistem

Bahagian seterusnya adalah membuat *API* yang menerima perincian indikator melalui *GUI* dan meramalkan klasifikasi kategori kemiskinan berdasarkan model *Mini Batch K-Means*. Untuk itu, perterjemahan model *Pickle* akan dibuat dalam bentuk objek *Python*. Halaman utama dtetapkan menggunakan *Interface.html*. Dengan menghantar

nilai menggunakan fungsi *POST*, keputusan klasifikasi kategori kemiskinan dapat diramalkan.

Keputusan dapat dipaparkan dengan menggunakan fungsi *POST* yang lain kepada direktori/*result*. Ia menerima input *JSON*, menggunakan model terlatih untuk membuat ramalan dan mengembalikan ramalan itu dalam format *JSON* yang dapat diakses melalui titik akhir *API*.

4.4.3 Pengeksportan Fail Perisian

Apabila laman web telah berjaya dibangunkan menggunakan *local server*, maka proses mengeksport fail perisian boleh dijalankan. Dalam fasa ini laman web atau sistem yang telah dibangunkan akan dieksport ke dalam *server* agar laman web ini mudah diakses oleh pengguna. Dalam projek ini, platform *Heroku* dan perisian *Git Bash* telah digunakan bagi menghasilkan fasa ini.

Heroku adalah platform *cloud* berasaskan Perkhidmatan sebagai Perkhidmatan (PaaS). Pembangun menggunakan Heroku untuk membangun dan mengurus aplikasi moden. Platform ini fleksibel dan mudah digunakan serta menawarkan kepada pembangun cara termudah untuk memasarkan aplikasinya.

Permulaan bagi melaksanakan fasa ini adalah pembangun perlu menghasilkan dua buah fail iaitu *requirements.txt* dan *Procfile*. Kedua-dua fail ini mempunyai spesifikasi yang hendaklah dipasang terlebih dahulu agar sistem boleh dijalankan dalam platform Heroku.

4.4.4 Implementasi

Pada fasa ini sistem akan dibangunkan menggunakan *Git Bash* dan *Heroku*. Setelah sistem ini telah siap dibangunkan, maka sistem *Community Indicator* boleh digunakan. Pautan yang diperoleh hasil pembangunan sistem ini adalah <https://communityindicator.herokuapp.com/>. Sistem ini bermula dengan paparan kemasukan data oleh pengguna. Pengguna boleh mengisi data yang diminta dan menekan butang '*Predict B40 Category*' untuk menghantar data yang dimasukkan.

5 HASIL PENGUJIAN

Sebelum membangunkan sistem, pengujian terhadap model perlu dilaksanakan agar model yang dipilih berkemampuan untuk menghasilkan nilai ketepatan yang tinggi. Jadual 5.1 menunjukkan hasil perbandingan antara model yang dipilih bagi mengenal pasti model yang terbaik. Apabila telah memilih model pilihan, maka sistem boleh dibangunkan. Setelah laman web dibangunkan, maka pengujian perlu dilakukan terhadap laman web Community Indicator bagi memastikan hasil ramalan dapat diperoleh setelah pengguna memasukkan input. Jadual 5.2 dan Jadual 5.3 masing-masing menunjukkan contoh input yang dimasukkan bagi kumpulan B40 dan kumpulan bukan B40. Bagi Rajah 5.1 dan Rajah 5.2, masing-masing menunjukkan hasil ramalan yang diperoleh bagi kumpulan B40 dan bukan kumpulan B40.

5.1 Hasil Perbandingan Antara Model

Jadual 5.1 Hasil perbandingan antara model

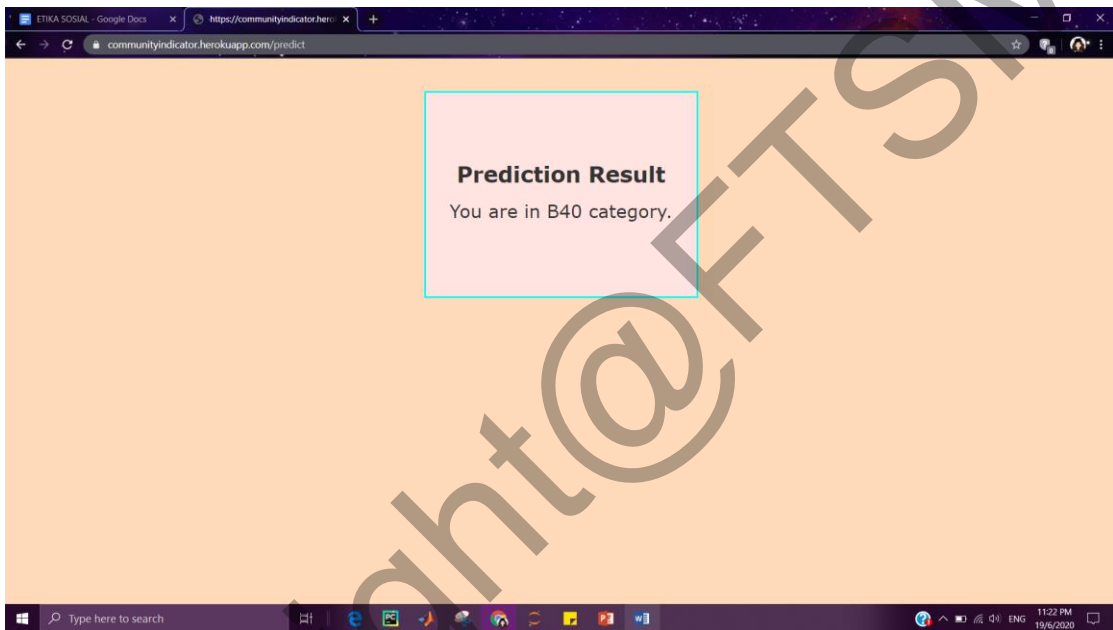
Model	<i>Silhouette Coefficient</i>
<i>Mini Batch K-Means</i>	0.859
<i>DBSCAN</i>	0.743
<i>Mean Shift Clustering</i>	0.666

5.2 Hasil Ujian Input Data Kumpulan B40

Jadual 5.2 Contoh input kumpulan B40

Atribut	Input
Status Kahwin	Berkahwin
Pendidikan Tertinggi	Ijazah sarjanamuda
Pekerjaan	Perkhidmatan dan penjualan
Mempunyai 2 kereta	Tidak
Mempunyai 3 kereta atau lebih	Tidak
Mempunyai penghawa dingin	Tidak
Mempunyai komputer riba	Ya
Mempunyai saluran tv berbayar	Tidak
Mempunyai langganan internet	Ya

Mempunyai penapis air	Tidak
Pemilikan kediaman	Tiada
Status pemilikan	Lain-lain
Jumlah bilangan isi rumah	5 orang
Jumlah pendapatan isi rumah	RM 5,876
Gaji	RM 3,560



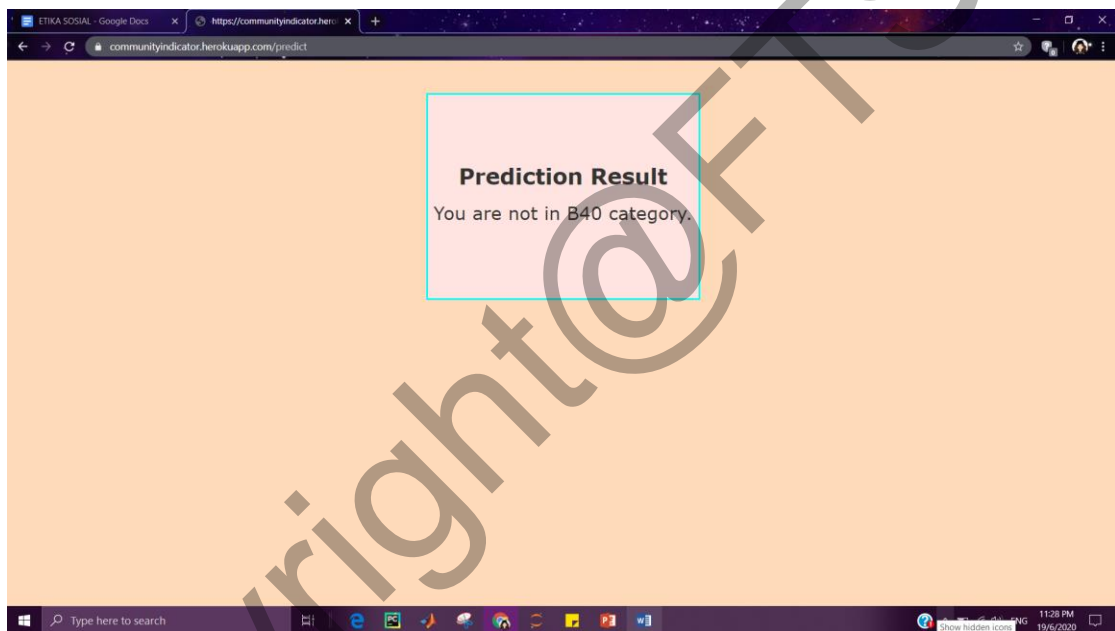
Rajah 5.1 Hasil keputusan kemasukan input kumpulan B40

5.3 Hasil Ujian Input Data Bukan Kumpulan B40

Jadual 5.3 Contoh input kumpulan bukan B40

Atribut	Input
Status Kahwin	Berkahwin
Pendidikan Tertinggi	Pra-universiti
Pekerjaan	Profesional
Mempunyai 2 kereta	Tidak
Mempunyai 3 kereta atau lebih	Ya
Mempunyai penghawa dingin	Ya
Mempunyai komputer riba	Ya

Mempunyai saluran tv berbayar	Ya
Mempunyai langganan internet	Ya
Mempunyai penapis air	Ya
Pemilikan kediaman	Ya
Status pemilikan	Persendirian
Jumlah bilangan isi rumah	6 -10 orang
Jumlah pendapatan isi rumah	RM 12,690
Gaji	RM 4,702



Rajah 5.2 Hasil keputusan kemasukan input kumpulan bukan B40

6 KESIMPULAN

Secara kesimpulannya, sistem Community Indicator ini telah dibangun dan telah memenuhi objektif yang telah ditetapkan pada awal tahap pembangunan yang telah dibincangkan. Pembangunan sistem ini merupakan sistem yang meramal klasifikasi kategori kemiskinan berdasarkan hasil input yang dimasukkan oleh pengguna. Walaupun sistem ini mempunyai beberapa kelebihan, namun aplikasi ini juga terdapat beberapa kelemahan yang perlu diperbaiki. Oleh itu, kajian yang mendalam terhadap sistem ini harus diteruskan untuk meningkatkan kualiti sistem ini.

Copyright@FTSM

7 RUJUKAN

Abhinav Sagar. 2019. How to Easily Deploy Machine Learning Models Using Flask. <https://towardsdatascience.com/how-to-easily-deploy-machine-learning-models-using-flask-b95af8fe34d4> [13 Jun 2020]

Alkire, S. dan Foster, J. 2007. *Counting and Multidimensional Poverty Measures*, Oxford Poverty & Human Development initiative OPHI Working paper 7.

Alkire, S. 2007. *Choosing Dimension: The Capability Approach and Multidimensional Poverty*, CPRC Working Paper 88, Chronic Poverty Research Center.

Alkire, S. dan Foster, J. 2009. *Counting and Multidimensional Poverty Measurement*, OPHI Working paper 32. Oxford University: Oxford Poverty & Human Development initiative.

Alkire, S., dan Foster, J. 2009. *Counting and Multidimensional Poverty*, In Von Braun J. (Ed.) *The Poorest and Hungry: Assessment, Analysis and Actions*. Washington D.C.: International Food Policy Research Institute.

Alkire, S and Foster, J. 2011. *Understanding and Misunderstanding of Multidimensional Poverty Measurement*. OPHI Working Paper No. 43. University of Oxford.

Amunategui, M. and Roopaei, M., 2018. *Monetizing Machine Learning: Quickly Turn Python ML Ideas into Web Applications on the Serverless Cloud*. Apress.

Anon. 2019. Apa Itu B40 M40 Dan T20 Maksud Golongan Isi Rumah Ikut Pendapatan. <https://www.mysumber.com/b40-m40-t20.html> [23 September 2019]

Anon. 2019. Tentang Wawasan Kemakmuran Bersama 2030. <https://www.malaysiakini.com/news/494547> [5 Oktober 2019]

Anon. 2019. Wawasan Kemakmuran Bersama 2030. <https://www.pmo.gov.my/ms/2019/10/wawasan-kemakmuran-bersama-2030/> [5 Oktober 2019]

Anon. 2019. Golongan Isi Rumah B40, M40 dan T20 di Malaysia Berdasarkan Pendapatan. <https://www.comparehero.my/budgets-tax/articles/klasifikasi-pendapatan-di-malaysia-t20-m40-dan-b40>[23 September 2019]

Bahmani, B., Moseley, B., Vattani, A., Kumar, R. and Vassilvitskii, S., 2012. Scalable k-means++. *arXiv preprint arXiv:1203.6402*.

Dr. Michael J. Garbade. 2018. Understanding K-means Clustering in Machine Learning.

Halko, N., Martinsson, P. G., & Tropp, J. A. 2009. Finding structure with randomness: Stochastic algorithms for constructing approximate matrix decompositions.

Hans-Hermann, B. O. C. K. 2008. Origins and extensions of the k-means algorithm in cluster analysis. *Journal Electronique d'Histoire des Probabilités et de la Statistique Electronic Journal for History of Probability and Statistics*, 4(2).

Jamil, N. & Che Mat, S. H. 2014. Realiti Kemiskinan: Satu Kajian Teoritik. *Jurnal Ekonomi Malaysia* 48(1): 167–177.

Mandy Sidana. 2017. Types of classification algorithms in Machine Learning.

<https://medium.com/@Mandysidana/machine-learning-types-of-classification-9497bd4f2e14>[15 Oktober 2019]

Mariah Abdul Rahman. 2019. Klasifikasi Kemiskinan Pelbagai Dimensi Bagi Kumpulan B40 di Malaysia Menggunakan Teknik Pengelompokan.

Meng, S., Fu, Y., Liu, T. and Li, Y., 2015, December. Principal component analysis for clustering temporomandibular joint data. In *2015 8th International Symposium on Computational Intelligence and Design (ISCID)* (Vol. 1, pp. 422-425). IEEE.

Nabila Yasmin Razib. 2019. Ke mana hala tuju AI Malaysia

<https://www.utusan.com.my/bisnes/ekonomi/ke-mana-hala-tuju-ai-malaysia-1.770394>[15 November 2019]

Omar, Mohamed Zaini. 2010. Pembasmian kemiskinan di Malaysia: Pengalaman Amanah Ikhtiar Malaysia (Penerbit USM). Penerbit USM.

Peng, K., Leung, V.C. and Huang, Q., 2018. Clustering approach based on mini batch kmeans for intrusion detection system over big data. *IEEE Access*, 6, pp.11897-11906

Prof Nazari Ismail. 2019. Masalah kemiskinan di Malaysia: Punca dan Penyelesaian. <http://www.astroawani.com/berita-malaysia/masalah-kemiskinan-di-malaysia-punca-dan-penyelesaian-216007> [27 November 2019].

Sidath Asiri. 2018. Machine Learning Classifier. <https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623> [23 November 2019]

Tentang I-BR1M. 2014. <https://www.takaful-ikhlas.com.my/i-BR1M/about.html> [23 September 2019]

V. Ganti and A. D. Sarma. 2013. *Data Cleaning: A Practical Perspective, ser. Synthesis Lectures on Data Management*. Morgan & Claypool Publishers.

Wang, Y. P. 2004. *Urban poverty, housing and social change in China*. New York: Routledge.

Wirth, R. and Hipp, J. 2000. CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (pp. 29-39). London, UK: Springer-Verlag.

Zanariah Abd Mutalib. 2019. BH Online. Kerajaan akan guna MPI ukur kadar kemiskinan. <https://www.bharian.com.my/berita/nasional/2019/10/614860/kerajaan-akan-guna-mpi-ukur-kadar-kemiskinan> [29 November 2019]

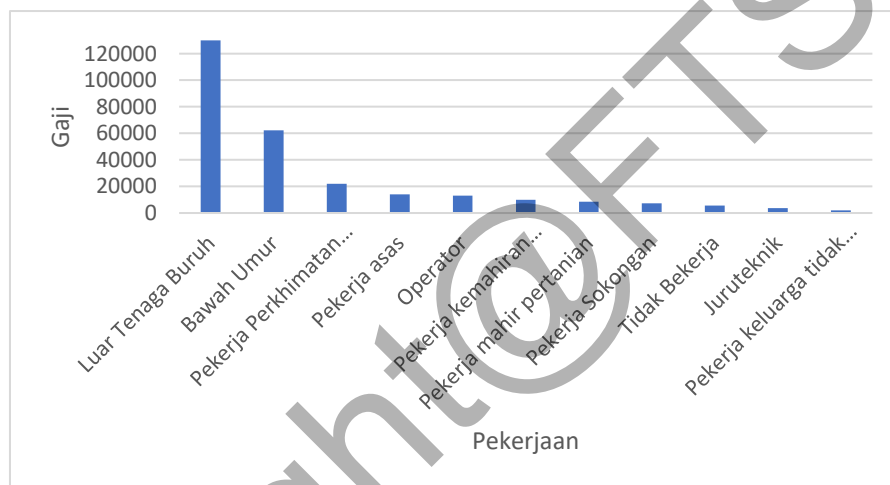
LAMPIRAN

Jadual 1 Peratus bagi data berdasarkan atribut pilihan

Data	Peratus (%)
Status kahwin:	
• Berkahwin	41.63
• Tidak pernah berkahwin	33.96
• Bawah Umur	20.22
• Balu, Bercerai, Berpisah	4.19
Pendidikan tertinggi:	
• Sekolah Menengah Atas	25.51
• Sekolah Rendah	20.86
• Lain-lain	26.66
• Sekolah Menengah Rendah	13.92
• Pra-universiti	4.25
• Pra Sekolah	3.35
• Sijil/Diploma	3.06
• Ijazah/Ijazah sarjana/Diploma Lanjutan	2.06
• Pendidikan selepas Menengah	0.23
Bilangan isi rumah dalam kediaman:	
• 1 individu	98.45
• 2 individu	1.03
• 3 individu	0.33
• 4 individu	0.09
• 5 individu	0.05
• 6-10 individu	0.006
• 11-15 individu	0.005
• 16-21 individu	0.004
• 21 individu dan keatas	0.009
Pemilikan kediaman:	
• Ya	60.85
• Tidak	22.48
• Tidak diketahui	16.66
Status pemilikan kediaman:	
• Persendirian	81.42
• Lain-lain	12.14
• Kerajaan	4.31
• Swasta	2.12

Jadual 2 Peratus bagi data output berbentuk binari berdasarkan atribut pilihan

Atribut	Ya (%)	Tidak (%)
Telefon Bimbit	73.00	26.99
Saluran TV Berbayar	38.89	61.10
Penghawa Dingin	21.46	78.54
Komputer Riba	19.46	80.54
2 Kereta	16.36	83.64
Langganan Internet	14.55	85.44
Penapis Air	20.00	80.00
3 atau lebih Kereta	4.93	95.07



Rajah 4.4 Kategori pekerjaan berserta nilai purata gaji