

# KEBOLEHDAPATAN KERJA GRADUAN UKM MENGGUNAKAN PERATURAN SEKUTUAN DALAM PERLOMBONGAN DATA

Chong Wen Yin<sup>1</sup>  
Dr Zalinda Binti Othman<sup>2</sup>

<sup>1</sup>(Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor)

<sup>2</sup>(Pusat Pembangunan Karier (UKM-Karier), Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor)

**Abstract**— Peningkatan jumlah graduan semakin tinggi menyebabkan persaingan untuk mendapat kerja semakin kuat dalam kalangan graduan dan mengakibatkan penambahan pengangguran graduan pada era ini. Dengan lebih memahami keperluan dalam pasaran pekerjaan ini, kajian ini dirancang untuk meramalkan kebolehdapatan kerja graduan UKM. Objektif kajian ini adalah untuk menentukan laluan kerjaya yang sesuai mengikut pencapaian akademik, kokurikulum, psikometrik, disiplin dan sahsiah mengikut pasaran semasa. Oleh itu, graduan dapat latihan berkaitan dan dapat sedia diri sebelum memasuki bidang pekerjaan. Kajian ini berjaya menggunakan jumlah 19792 data graduan yang disediakan oleh Sistem Kebolehpasaran Graduan (SKPG), Kementerian Pendidikan Tinggi dari tahun 2016 hingga tahun 2018. Perlombongan data mengkaji algoritma dan paradigma komputer yang membolehkan komputer mencari corak-corak dan kekerapan dalam pangkalan data, menjalankan anggaran dan ramalan serta secara amnya. Kaedah perlombongan meliputi pemilihan data, membersihkan, pengenalpastian corak dan visualisasi. Algoritma Apriori telah digunakan dalam kajian ini untuk mengenalpastikan hubungan antara faktor-faktor yang akan mempengaruhi kegagalan mendapat kerja dan kebolehdapatan kerja. Atribut yang dipilih dalam kajian ini untuk menentukan peraturan sekutuan ialah e\_muets, e\_keturunan, e\_jantina, e\_pendapatan, e\_bidang, e\_cgpa dan e\_fakulti. Faktor e\_muets = Band 1 hingga Band 3 mempunyai kekerapan yang tertinggi adalah nilai 0.48. Parameter yang baik perlulah mempunyai nilai keyakinan yang tinggi.

**Keywords**—kebolehpasaran; perlombongan data; peraturan sekutuan; algoritma Apriori

## I. PENGENALAN

Pada era globalisasi, masalah pengangguran di kalangan graduan telah menjadi satu isu semasa yang perlu ditangani. Walau bagaimanapun, kadar pengangguran ini masih mengekalkan dalam julat rendah dan menjadi salah satu isu yang serius pada masa ini. Setiap tahun, melahirkan lebih 290,000 graduan dari institusi pengajian tinggi. Malangnya, 1 daripada 5 graduan masih menganggur selepas tamat pengajian tinggi. Hal ini menunjukkan bahawa sijil pemegang ijazah bukan jaminan untuk mendapat pekerjaan. Jumlah graduan akan terus meningkat dan mereka perlulah menghadapi persaingan yang kuat di tengah ekonomi yang tidak stabil ini.

Menurut data yang disediakan oleh Bank Dunia ([https://www.theglobaleconomy.com/Malaysia/youth\\_employment/](https://www.theglobaleconomy.com/Malaysia/youth_employment/)), nilai purata bagi Malaysia dari tahun 1991 hingga 2018 ialah 10.71% dengan minimum 8.45% pada tahun 1997 dan maksimum 12.28% pada tahun 1993. Pada tahun 2017, mempunyai nilai 10.84% manakala mempunyai nilai 11.18% pada tahun 2018, iaitu peningkatan sebanyak 0.34%. Hal ini menerangkan kadar pengangguran graduan di Malaysia masih satu isu yang penting perlu diselesaikan.

Melalui Pengarah Eksekutif MEF, Datuk Shamsuddin Bardan, (MEF: *Skill upgrade is critical, the Stars online* 2019), beliau menerangkan graduan ini menghadapi masalah komunikasi yang ketara, sama ada secara lisan dan bertulis. Pencapaian akademik yang tinggi bukan sebagai satu tiket yang membolehkan graduan masuk dalam bidang pekerjaan sebaliknya graduan harus menguasai beberapa kemahiran yang dapat memenuhi permintaan majikan seperti kemahiran menyelesaikan masalah, kemahiran komunikasi, keupayaan untuk bekerja secara sendiri, kemahiran untuk belajar dan keupayaan untuk bekerja di bawah tekanan.

Projek kajian ini diadakan untuk membangun satu model ramalan kebolehdapatan kerja graduan UKM dengan menggunakan data graduan UKM yang disediakan oleh

Sistem Kebolehpasaran Graduan (SKPG), Kementerian Pendidikan Tinggi dari tahun 2016 hingga tahun 2018. Kaedah yang digunakan dalam kajian ini adalah kaedah perlombongan data dengan menggunakan algoritma Apriori. Kajian ini dijalankan disebabkan pengangguran graduan semakin tambah dan hasil daripada kajian ini akan membantu dalam mengenal pasti ciri-ciri yang diperlukan dalam pasaran pekerjaan pada masa ini. Oleh itu, pelajar boleh sedia diri dan mendapat latihan lebih awal sebelum graduan supaya mudah mendapat kerja.

## II. KAJIAN KESUSASTERAAN

Pada abad ke-21 ini, persekitaran kerja semakin mencabar. Oleh itu, graduan yang berkemahiran dan cemerlang di bidang akademik akan mendapat pekerjaan lebih mudah. Sebaliknya, kekurangan kemahiran yang diperlukan dalam pasaran kerja ini masih menyebabkan banyak graduan tidak dapat pekerjaan di Malaysia.

Menurut Jabatan Statistik Malaysia, Malaysia: Youth Employment (2019), kadar pengangguran di Malaysia masih mengekalkan di 3.30% pada Jun dan Julai 2019. Kadar pengangguran ini berubah setiap bulan, daripada Januari hingga Julai 2019, dengan kadar purata 3.40%. Data ini mencapai tahap tertinggi iaitu 3.60% pada bulan November 2016 dan rekod paling rendah ialah 3.20% pada Januari 2019.

Kebolehpasaran graduan dari pusat pengajian tinggi dikaitkan dengan kecenderungan bakal majikan mengutamakan graduan memiliki kemahiran insaniah (*soft skills*). Kemahiran insaniah yang tinggi termasuk komunikasi secara efektif, pemikiran kritikal dan penyelesaian masalah, kemahiran kepimpinan dan kreativiti, bertanggungjawab terhadap alam sekitar, beretika dan berprofesional dan kerja sosial dan bertanggungjawab. Kemahiran yang disebutkan telah dilaksanakan oleh banyak universiti awan di Malaysia supaya dapat menanam sifat yang baik kepada mahasiswa Malaysia.

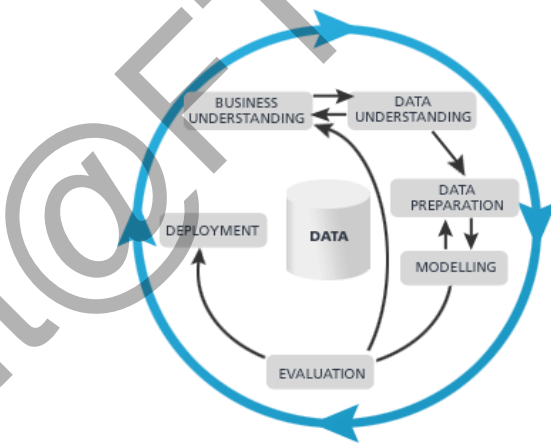
Melalui kajian selidik Abdul Wahab dan Diana (2017), telah menerangkan aspek dan isu akan mempengaruhi kebolehdapatan kerja di Malaysia. Umumnya, kecemerlangan prestasi akademik, kemahiran penguasaan Bahasa Inggeris, kecemerlangan-pendapatan keluarga dan pilihan kursus merupakan faktor utama yang mempengaruhi kebolehdapatan kerja di Malaysia. Rahmah et al (2011) melaporkan bahawa graduan tidak mempunyai kemahiran dan kelayakan yang diperlukan dan kekurangan pengalaman dalam pekerjaan. Selain itu, terdapat aspek kelemahan mengandungi kemahiran pengurusan, penyelesai masalah, komunikasi, kepimpinan, kreativiti, pemikiran kritikal, proaktif, keyakinan diri dan kemahiran interaksi.

Hao-Wen Shih (2018) menggunakan algoritma Apriori dalam perlombongan data untuk menyaring 2,427 rekod akademik supaya mendapat keputusan yang tepat. Berdasarkan tiga-tiga definisi algoritma Apriori, definisi pertama menerangkan minima tahap sokongan  $> 0$ ,

menunjukkan bahawa set item perlu berada dalam set data sekurang-kurangnya 1 kali. Definisi kedua menyokongkan (*confidence*)  $> 0.5$  dan definisi ketiga menyatakan (*lift ratio*)  $> 1$ . Daripada kajian ini, terdapat sejumlah 21 peraturan sekutuan (*association rule*) telah dijumpai.

## III. KAEDAH KAJIAN

Metodologi kajian ini merupakan kaedah yang penting dalam memastikan kajian ini dijalankan dengan lancar. Kaedah yang digunakan dalam kajian ini ialah Proses Lintas-Industri untuk perlombongan data (CRISP-DM). Kaedah ini melibatkan enam fasa dan setiap fasa akan diubah suai bagi memenuhi keperluan kajian. Enam fasa dalam CRISP-DM terlibat fasa pemahaman perniagaan, pemahaman data, penyediaan data, pemodelan, penilaian dan pengaplikasian. Rajah 1.1 menunjukkan fasa-fasa dalam metodologi ini.



Rajah 1.1 Fasa-fasa dalam metodologi CRISP-DM

### A. Fasa Pemahaman Perniagaan

Tugas utama dalam fasa ini ialah memahami pencapaian perniagaan (*Business Goal*) dan soalan perniagaan (*Business Question*). Pencapaian perniagaan untuk kajian ini adalah menyediakan unjuran karier di Malaysia untuk tempoh 5 tahun. Pencapaian kedua mengurangkan jurang pembezaan pekerjaan yang ditawarkan dan bidang di universiti dan diikuti meningkatkan kebolehpasaran kerja graduan. Pencapaian terakhir adalah menentukan laluan kerjaya yang sesuai mengikut pencapaian akademik, kokurikulum, psikometrik, disiplin dan sahsiah mengikut pasaran semasa.

Selain itu, soalan perniagaan juga disediakan sebagai dicatat di bawah.

1. Apakah faktor yang mempengaruhi kebolehpasaran graduan.
2. Apakah hubungan demografi, akademik, dan program pengajian dengan pekerjaan?
3. Mengapa pelajar yang berisiko tidak mendapat pekerjaan selepas bergraduat?

Dalam kajian ini, sebanyak 300 atau lebih atribut akan digunakan untuk meramalkan kebolehpasaran graduan.

## B. Fasa Pemahaman Data

Untuk menentukan status kebolehdapatan kerja terhadap graduan UKM, klasifikasi sifat-sifat data graduan akan dijalankan. Terdapat sebanyak 337 atribut yang berkaitan dengan kebolehpasaran kerja graduan dan data akan dikumpulkan melalui borang yang perlu diisi oleh graduan sebelum mengambil jubah. Atribut ini merangkumi nama, umur, cgpa, tahap MUET dan sebagainya. Daripada data yang dikumpulkan, kita akan mengkategorikan profesion graduan ikuti fakulti untuk membanding mana satu sektor pekerjaan lebih memerlukan pekerja daripada sektor lain. Kajian ini menggunakan data Graduan UKM pada tahun 2016 hingga 2018 yang diperolehi dari SKPG yang dikelolakan daripada Unit Kajian Pengesanan Graduan, Kementerian Pengajian Tinggi Malaysia. Data ini mengandungi 3 tahun data graduan yang bermula dari tahun 2016 hingga tahun 2018. Jumlah data pada tahun 2016 adalah 5889 respons, 7353 respons pada tahun 2017 dan 6550 respons pada tahun 2018. Terdapat sebanyak 19792 data yang merangkumi 353 atribut telah dikumpul.

## C. Fasa Penyediaan Data

Selepas data dikumpulkan, proses penyediaan data akan diproses. Dalam fasa ini, data yang akan digunakan akan dibersihkan terlebih dahulu dan diformat. Pembersihan data adalah pra-pemprosesan data untuk mengurangkan atribut dan mengendalikan nilai yang hilang setelah mengkaji kepentingannya akan memberi impak atau kesan yang berguna atau tidak kepada hasil kajian. Pengurangan jumlah data dan pengurangan jumlah atribut termasuk dalam proses pengeluaran data. Pembersihan data juga melibatkan pembetulan data tidak konsisten, mengenal pasti outlier dalam data dan juga menghapus atribut yang sama. Selain itu, disebabkan oleh ketidaksamaan data graduan terhadap ketiga-tiga tahun ini, data akan dijalankan masing-masing daripada secara menggabungkan semua data dalam satu model.

Fasa penyediaan data mempunyai 4 fasa iaitu pembersihan data, pengurangan data, transformasi data dan pelabelan data. Pembersihan data adalah satu proses yang boleh mengurangkan atribut yang sama, mengenalkan outlier, menggantikan kehilangan nilai dan juga pembetulan data yang tidak konsisten. Proses pembersihan di bawah akan menerangkan proses pembersihan untuk ketiga-tiga tahun iaitu tahun 2016, tahun 2017 dan tahun 2018. Pengurangan data adalah satu proses mengurangkan data yang tidak berkaitan dengan kajian ini seperti maklumat alamat, nombor telefon dan sebagainya. Proses transformasi ini mengubah data daripada bererapa kumpulan kepada 2 atau 3 kumpulan untuk memudah proses analisis. Data atribut 'e\_jantina' yang direkodkan 1 dan 2 ditukarkan kepada bentuk nominal iaitu 'lelaki' dan 'perempuan'. Proses akhir adalah proses pelabelan data iaitu menukar label atribut

daripada "e\_40" kepada "status bekerja" untuk senang difahami.

Proses pembersihan data adalah langkah awal dalam pra-pemprosesan. Dalam proses ini, masalah data yang mempunyai ralat, tidak konsisten dan kehilangan akan dapat dibetulkan. Data untuk tahun 2016 mempunyai kehilangan data manakala data untuk tahun 2017 tidak mempunyai kehilangan data. Kehilangan data ini akan digantikan dengan purata atau mod dan menggunakan penapis weka iaitu 'ReplaceMissingValue'. Kehilangan data untuk atribut e\_negeri telah mengurangkan daripada 36 hilang data kepada 0 hilang data. Sejak label Selangor merupakan mod iaitu mempunyai kekerapan tertinggi dalam atribut ini. Oleh itu, kehilangan data akan digantikan dengan label Selangor. Seterusnya, kaedah menyingkirkan outlier adalah untuk menghapuskan data yang tidak sama dengan data lain dan akan membuat data tidak konsisten. Data untuk tahun 2016 tidak mempunyai data tidak konsisten manakala data untuk tahun 2017 dan tahun 2018 tidak mempunyai kehilangan data. Oleh itu, penapis 'InterquartileRange' digunakan untuk mencari outlier dalam set data ini. Selepas mencari outlier, kaedah mengeluarkan outlier telah dijalankan dengan menggunakan penapis 'RemoveWithValue'.

Proses kedua ialah proses pengurangan data adalah langkah kedua dalam proses pra-pemprosesan. Proses ini dijalankan untuk mengurangkan data berlebihan dan tidak berkait dengan kajian ini.

Proses akhir ialah proses transformasi ini digunakan untuk menyatukan beberapa kumpulan dalam satu kumpulan mengikut keperluan data. Selain itu, proses ini juga menukar data dari satu format atau struktur ke format atau struktur lain. Kaedah penyatuan data dapat menukar daripada struktur kumpulan besar kepada struktur kumpulan kecil. Sebagai contoh, atribut e\_umur mempunyai jarak dari umur 19 tahun hingga umur 60 tahun. Selepas menggunakan penapis 'Discretize', 4 kumpulan telah dibahagikan dengan jarak tertentu. Selain itu, kaedah penukaran jenis atribut membantu bertukar jenis atribut daripada 'numeric' kepada 'nominal'. 'Nominal' ialah satu jenis atribut digunakan untuk melabelkan data tanpa memberi sebarang nilai kuantitatif. Penapis 'NumericToNominal' digunakan untuk menukar jenis atribut dalam kajian ini. Proses akhir adalah proses pelabelan data iaitu menukar label atribut daripada "e\_40" kepada "status bekerja" untuk senang difahami. Selain menukar label atribut, label kumpulan juga boleh ditukarkan dengan menggunakan excel.

## D. Pemodelan

Fasa ini bertujuan untuk membangunkan model ramalan. Model yang dibangunkan dalam kajian ini akan menggunakan algoritma Apriori. Algoritma Apriori adalah salah satu kaedah dalam peraturan sekutuan. Fasa pemodelan ini boleh dilakukan dengan ujian dan percubaan model yang berbeza ciri. Ujian ini boleh diulang

dengan berbeza perubahan parameter, dapat memantau keputusan dan membuat kesimpulan awal pada model yang digunakan. Sebagai contoh, KNN, Naive Bayes, Pohon Keputusan, Rangkaian Neural, Regresi Logistik dan SVM dan sebagainya digunakan sebagai pembelajaran mesin untuk proses pemodelan dan eksperimen.

Pembelajaran Peraturan Sekutuan adalah satu kaedah yang menonjol untuk menentukan hubungan antara penbolehubah dalam pangkalan data yang besar. Sebagai contoh,  $I = \{i_1, i_2, i_3, \dots, i_n\}$  menjadi satu set  $n$  atribut yang dipanggil item dan  $D = \{t_1, t_2, \dots, t_n\}$  sebagai satu set transaksi. Set  $i$  dan set  $D$  menjadi satu set data.  $T_i$  dalam set  $D$  mempunyai urusan unik dalam setiap transaksi dan set  $D$  ini terdiri daripada subset in  $I$ . Kaedah ini boleh didenigiskan sebagai satu implikasi iaitu  $X \rightarrow Y$  mana  $X$  dan  $Y$  adalah subset  $I$  ( $X, Y \subseteq I$ ) dan mereka tidak mempunyai elemen yang sama. Peraturan Sekutuan mempunyai ukuran penting seperti *support*, *confidence*, *lift* dan *conviction*. (Association rule learning)

#### 1. Support

Support dari set item  $X$ ,  $\text{supp}(X)$  adalah perkadaran transaksi dalam pangkalan data di mana item  $X$  muncul.

Formula:

$$\text{supp}(X) = \frac{\text{Bilangan transaksi yang berlaku}}{\text{Jumlah transaksi}}$$

#### 2. Confidence

Ukuran ini menandakan kemungkinan item  $Y$  dibeli apabila item  $X$  dibeli.

Formula :

$$\text{conf}(X \rightarrow Y) = \frac{\text{supp}(XUY)}{\text{supp}(X)}$$

#### 3. Lift

Ukuran ini menandakan kemungkinan set item  $Y$  yang dibeli apabila item  $X$  dibeli kira mengambil kira populariti  $Y$ .

Formula :

$$\text{lift}(X \rightarrow Y) = \frac{\text{supp}(XUY)}{\text{supp}(X) * \text{supp}(Y)}$$

#### 4. Conviction

Ukuran ini mengukur kekuatan implikasi dari kebebasan statistik.

Formula :

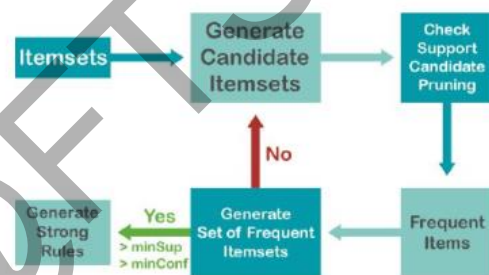
$$\text{conv}(X \rightarrow Y) = \frac{1 - \text{supp}(Y)}{1 - \text{conf}(X \rightarrow Y)}$$

Algoritma Apriori merupakan model klasifikasi dalam kajian ini. Algoritma ini merupakan salah satu algoritma dalam pembelajaran Association Rule dan dicipta oleh R.Agrawal dan R.Srikant pada tahun 1994 untuk

mencari kekerapan set item dalam set data. Pengguna menggunakan pendekatan berulang di mana kekerapan  $k$  set item digunakan untuk mencari  $k + 1$  set item. (Association rule learning)

Mekanisme untuk algoritma Apriori:

1. Mencipta satu jadual kekerapan untuk semua item yang muncul dalam semua transaksi.
2. Menemui semua kekerapan item yang mempunyai kemunculan yang berkali lebih tinggi atau sama dengan  $\text{minSup}$
3. Menghasil calon daripada item yang mempunyai kekerapan tinggi.
4. Memeriksa dan mencocok hasil dengan item yang mempunyai kekerapan tinggi.
5. Akhirnya, menjana peraturan yang sesuai



Rajah 3.1 Mekanisme algoritma Apriori

## E. Penilaian

Pengelasan dan penilaian model anggaran adalah salah satu perkara yang penting dalam sebarang proses perlombongan data. Kriteria penilaian utama dalam perspektif klasifikasi adalah kriteria ketepatan seluruh diperolehi oleh kaedah pengesahan model. Salah satu kaedah yang terkenal adalah kaedah memvisualisasikan prestasi ramalan metrik yang boleh didapati dengan membina matrik kekeliruan.

Peraturan akan dilahirkan melalui weka dan melahirkan hubungan di antara set data untuk mengetahui bagaimana ciri-ciri atribut mempengaruhi status bekerja atau tidak bekerja. Sokongan ialah kekerapan munculan dalam pangkalan data. Sebagai contoh, sekiranya set sokongan  $> 0.1$ , kekerapan kurang daripada  $0.1$  tidak akan dikirakan oleh weka. Keyakinan menunjukkan kemungkinan munculan item  $Y$  sekiranya sama ada kemunculan item  $X$ .

## F. Pengaplikasian

Dalam fasa pengaplikasian, keputusan didapati terhadap model klasifikasi yang digunakan memberi data yang berguna dan berkesan dalam menjalani kajian ini. Model klasifikasi yang berbeza akan memberi ketepatan yang berbeza. Model klasifikasi menyediakan ketepatan yang tertinggi atas parameter yang berbeza untuk menganalisis atribut yang memberi kesan kepada

kebolehpasaran kerja graduan. Selepas kajian ini disiapkan, hasil dan keputusan akan disampaikan kepada penyelia untuk mengubahsuai. Penyelidikan ini akan diterbitkan sekiranya dilakukan dengan sempurna dan tidak bersalah supaya dapat memberi pengajaran dan sebagai rujukan untuk graduan pada masa akan datang.

#### IV. KEPUTUSAN DAN PERBINCANGAN

Data ketiga-tiga tahun ini akan digabungkan dalam satu data berdasarkan atribut yang dimiliki oleh data ketiga-tiga tahun ini. Ukuran parameter untuk data ketiga-tiga tahun adalah mengubah nilai keyakinan (*confidence*) dan nilai sokongan (*support*). Model ini dijalankan dengan beberapa parameter yang berbeza. Perbezaan pemilihan parameter ini akan menghasilkan keputusan model yang berbeza malah peraturan tidak akan dilahirkan kerana nilai keyakinan dan nilai sokongan tidak memenuhi syarat algoritma Apriori.

Untuk data ketiga-tiga tahun ini, nilai sokongan mempunyai julat daripada nilai 0.1 hingga nilai 0.4 manakala nilai keyakinan mempunyai julat daripada nilai 0.4 hingga nilai 0.6. Perubahan parameter akan melahirkan keputusan yang berbeza daripada segi bilangan sampel, bilangan kitaran yang dilaku dan saiz kumpulan item. Oleh itu, jadual 4.1 akan menunjukkan menerangkan perbezaan antara perubahan parameter-parameter tersebut. Selain parameter yang dicatatkan, parameter-parameter lain tidak akan melahirkan peraturan sekutuan ataupun mempunyai peraturan sekutuan yang sama dengan parameter yang dicatatkan.

Jadual 4.1 Perbezaan antara perubahan parameter

Pengubahan parameter				
<i>Sup port</i>	<i>Confi dence</i>	Bilangan sampel	Bilangan kitaran	Jumlah kumpulan item
0.1	0.4	934	18	4
0.1	0.5	934	18	4
0.1	0.6	934	18	4
0.2	0.4	1868	16	3
0.2	0.5	1868	16	3
0.3	0.4	2802	14	2
0.3	0.5	2802	14	2
0.4	0.4	3736	12	2
0.4	0.5	3736	12	2

##### a) Pengetahuan Analisis

Menurut Hainnur Aqma Rahim(2014), selain pencapaian akademik cemerlang, pelajar perlu mempunyai kemahiran insaniah yang tinggi termasuk komunikasi secara berkesan, kemahiran kepimpinan, kerja berpasukan, pemikiran kritis,

dan keusahawanan. Oleh itu, cgpa bukan satu-satunya piawaian dalam kerjaya malah *soft skills* dan faktor peribadi juga akan dipertimbangkan.

Peraturan sekutuan yang dikumpulkan daripada model ini boleh menunjukkan kebolehpasaran graduan yang diperlukan pada masa kini. Berdasarkan data ketiga-tiga tahun, e\_muets, e\_keturunan, e\_jantina, e\_bidang, e\_pendapatan, e\_cgpa dan e\_fakulti merupakan faktor paling penting dalam menentukan status bekerja dan status tidak bekerja. Keputusan yang paling tepat adalah parameter mempunyai nilai sokongan 0.1 dan nilai keyakinan 0.6. Hal ini berpunca daripada parameter ini mempunyai jumlah kumpulan set yang tertinggi dan mempunyai nilai keyakinan paling tinggi. Oleh itu, keputusan faktor dan kekerapan faktor dalam sebanyak 88 peraturan sekutuan akan dibincangkan untuk status bekerja manakala mempunyai sebanyak 77 peraturan sekutuan untuk status tidak bekerja.

Kekerapan faktor yang semakin tinggi bermakna faktor tersebut mempunyai jumlah kemunculan yang tinggi dalam jumlah keseluruhan peraturan kesatuan dilahirkan sebaliknya kekerapan faktor yang semakin rendah bermakna faktor tertentu mempunyai kemunculan yang rendah. Jadual 4.2 menunjukkan faktor-faktor kebolehpasaran untuk status bekerja manakala jadual 4.3 menunjukkan faktor-faktor kebolehpasaran untuk status tidak bekerja.

Jadual 4.2 Faktor-faktor kebolehpasaran untuk status bekerja

FAKTOR	KEKERAPAN
e_muets = Band 1 hingga Band 3	0.48
e_keturunan = Melayu	0.43
e_jantina = Perempuan	0.34
e_bidang = Sastera & Sains Sosial	0.27
e_pendapatan = <RM2000	0.26
e_cgpa = 3.10-3.50	0.17
e_pendapatan = RM2000-RM5000	0.09
e_jantina = Lelaki	0.05
e_fakulti = Fakulti Ekonomi dan Pengurusan	0.05
e_fakulti = Fakulti Sains Sosial dan Kemanusiaan	0.01
e_bidang = Sains	0.01

Berdasarkan faktor-faktor kebolehpasaran untuk status bekerja, faktor e\_muets = Band 1 hingga Band 3 mempunyai kekerapan yang tertinggi adalah nilai 0.48. Hal ini demikian kerana majikan tidak mementingkan keputusan MUET graduan asalkan dapat berkomunikasi dalam bahasa Inggeris. Faktor e\_keturunan = Melayu mempunyai nilai 0.43 kerana golongan keturunan Melayu mempunyai nisbah yang tertinggi daripada keturunan lain. Faktor e\_jantina untuk perempuan dan lelaki mempunyai nilai kekerapan 0.34 dan 0.05 masing-masing. Selain itu, graduan daripada bidang sastera & sains sosial dan bidang sains memiliki nilai

kekerapan 0.27 dan 0.01 masing-masing. Hal ini berpendapat demikian kerana kedua-dua bidang ini lebih mempunyai peluang pekerjaan daripada bidang lain. Seterusnya,  $e_{cgpa} = 3.10 - 3.50$  menunjukkan nilai kekerapan 0.17 dan ini menerangkan majikan tidak semestinya memerlukan graduan mendapat keputusan yang cemerlang. Di samping itu, pendapatan keluarga bawah RM2000 mempunyai nilai kekerapan 0.26 dan pendapatan keluarga daripada RM2000 hingga RM5000 mempunyai nilai kekerapan 0.09. Keseluruhan pendapatan keluarga bawah RM5000 dipercayai membawa impak positif kepada graduan dalam pencarian kerja dengan bersungguh-sungguh. Akhir, graduan daripada fakulti ekonomi dan pengurusan, dan fakulti sains sosial dan kemanusiaan adalah nilai kekerapan 0.05 dan 0.01.

Jadual 4.3 Faktor-faktor kebolehpasaran untuk status tidak bekerja

FAKTOR	KEKERAPAN
$e_{muet} = \text{Band 1 hingga Band 3}$	0.45
$e_{keturunan} = \text{Melayu}$	0.44
$e_{jantina} = \text{Perempuan}$	0.40
$e_{pendapatan} = < \text{RM2000}$	0.23
$e_{bidang} = \text{Sastera \& Sains Sosial}$	0.19
$e_{cgpa} = 3.10-3.50$	0.14
$e_{bidang} = \text{Sains}$	0.14
$e_{fakulti} = \text{Fakulti Sains dan Teknologi}$	0.10
$e_{pendapatan} = \text{RM2000-RM5000}$	0.06
$e_{jantina} = \text{Lelaki}$	0.01
$e_{fakulti} = \text{Fakulti Sains Sosial dan Kemanusiaan}$	0.01

Daripada jadual yang dipaparkan, sebanyak nilai kekerapan 0.45 menunjukkan graduan mendapat band 1 hingga band 3 dalam pemeriksa MUET mereka. Hal ini demikian kerana majoriti graduan mendapat keputusan ini adalah sebanyak 8016 daripada 9339 graduan. Selain itu, graduan daripada fakulti sains dan teknologi akan lebih sukar dalam pencarian kerja daripada fakulti lain. Perempuan mempunyai nilai kekerapan 0.40 lebih tinggi daripada nilai kekerapan lelaki 0.01 kerana nisbah perempuan lebih daripada lelaki.

Untuk menganalisis keputusan dalam cara yang lebih jelas, keempat-empat jadual di bawah memaparkan peraturan sekutuan dengan parameter yang berlainan iaitu nilai keyakinan 0.4 dan 0.5 serta nilai kesokongan 0.1 dan 0.2.

Jadual 4.4 10 peraturan kesatuan yang pertama dengan nilai sokongan 0.1 dan nilai keyakinan 0.4

<b>Support : 0.1</b> <b>Bekerja</b>	<b>Confidence: 0.4</b> <b>Tidak Bekerja</b>
Rule:	Rule:
1. $e_{jantina} = \text{Perempuan}$ $e_{keturunan} = \text{Melayu}$ $e_{muet} = \text{Band 1 hingga Band 3}$ $e_{pendapatan} = < \text{RM2000}$ $3034 \implies e_{status} = \text{Bekerja}$ 1487 conf:(0.49)	1. $e_{muet} = \text{Band 1 hingga Band 3}$ $e_{cgpa} = 3.10-3.50$ $e_{pendapatan} = < \text{RM2000}$ $1022 \implies e_{status} = \text{Tidak Bekerja}$ 1022 conf:(0.49)
2. $e_{fakulti} = \text{FAKULTI SAINS SOSIAL DAN KEMANUSIAAN}$ 1918 $\implies e_{status} = \text{Bekerja}$ 940 conf:(0.49)	2. $e_{muet} = \text{Band 1 hingga Band 3}$ $e_{bidang} = \text{Sastera \& Sains Sosial}$ $e_{pendapatan} = < \text{RM2000}$ 2548 $\implies e_{status} = \text{Tidak Bekerja}$ 1188 conf:(0.47)
3. $e_{jantina} = \text{Perempuan}$ $e_{keturunan} = \text{Melayu}$ $e_{pendapatan} = < \text{RM2000}$ 3228 $\implies e_{status} = \text{Bekerja}$ 1581 conf:(0.49)	3. $e_{keturunan} = \text{Melayu}$ $e_{muet} = \text{Band 1 hingga Band 3}$ $e_{cgpa} = 3.10-3.50$ 2985 $\implies e_{status} = \text{Tidak Bekerja}$ 1380 conf:(0.46)
4. $e_{jantina} = \text{Perempuan}$ $e_{muet} = \text{Band 1 hingga Band 3}$ $e_{pendapatan} = < \text{RM2000}$ 3462 $\implies e_{status} = \text{Bekerja}$ 1690 conf:(0.49)	4. $e_{keturunan} = \text{Melayu}$ $e_{muet} = \text{Band 1 hingga Band 3}$ $e_{bidang} = \text{Sastera \& Sains Sosial}$ $e_{pendapatan} = < \text{RM2000}$ 2237 $\implies e_{status} = \text{Tidak Bekerja}$ 1033 conf:(0.46)
5. $e_{bidang} = \text{Sains}$ 2665 $\implies e_{status} = \text{Bekerja}$ 1134 conf:(0.43)	5. $e_{jantina} = \text{Perempuan}$ $e_{keturunan} = \text{Melayu}$ $e_{muet} = \text{Band 1 hingga Band 3}$ $e_{bidang} = \text{Sastera \& Sains Sosial}$ 2815 $\implies e_{status} = \text{Tidak Bekerja}$ 1277 conf:(0.45)

Daripada jadual di atas, biasanya sesiapa yang berjantina perempuan, berbangsa Melayu, mendapat Band 1 hingga Band 3 dalam pemeriksaan MUET dan pendapatan keluarga bawah RM2000 lebih mudah mencari kerja selepas graduan. Hal ini demikian kerana peraturan ini mendapat nilai keyakinan sebanyak 0.49 dan diikuti oleh pelajar dari fakulti sains sosial dan kemanusiaan. Untuk status tidak bekerja, graduan dengan mendapat Band 1 hingga Band 3 dalam pemeriksaan MUET, pendapatan keluarga bawah RM2000 dan cgpa antara 3.10-3.50 lebih sukar dalam pencarian kerja dan mendapat 0.49 dalam nilai keyakinan manakala graduan dengan mendapat Band 1 hingga Band 3 dalam pemeriksaan MUET, pendapatan keluarga bawah RM2000 dan belajar dalam bidang sastera & sains sosial mempunyai nilai keyakinan 0.47.

Jadual 4.5 10 peraturan kesatuan yang pertama dengan nilai sokongan 0.2 dan nilai keyakinan 0.4

<b>Support : 0.2</b>	<b>Confidence: 0.4</b>
<b>Bekerja</b>	<b>Tidak Bekerja</b>
Rule:	Rule:
1. e_jantina=Perempuan e_pendapatan=<RM2000 3850 ==> e_status=Bekerja 1876 conf:(0.49)	1. e_keturunan=Melayu e_pendapatan=<RM2000 4297 ==> e_status=Tidak Bekerja 2124 conf:(0.49)
	2. e_jantina=Perempuan 6752 ==> e_status=Tidak Bekerja 3333 conf:(0.49)
	3. e_keturunan=Melayu e_muets=Band 1 hingga Band 3 e_pendapatan=<RM2000 4039 ==> e_status=Tidak Bekerja 1992 conf:(0.49)
	4. e_jantina=Perempuan e_keturunan=Melayu 5626 ==> e_status=Tidak Bekerja 2768 conf:(0.49)
	5. e_jantina=Perempuan e_muets=Band 1 hingga Band 3 5812 ==> e_status=Tidak Bekerja 2843 conf:(0.49)
	6. e_jantina=Perempuan e_keturunan=Melayu e_muets=Band 1 hingga Band 3 5103 ==> e_status=Tidak Bekerja 2489 conf:(0.49)
	7. e_keturunan=Melayu 7738 ==> e_status=Tidak Bekerja 3644 conf:(0.47)
	8. e_muets=Band 1 hingga Band 3 8016 ==> e_status=Tidak Bekerja 3757 conf:(0.47)
	9. e_keturunan=Melayu e_muets=Band 1 hingga Band 3 7017 ==> e_status=Tidak Bekerja 3273 conf:(0.47)

Hanya satu peraturan untuk status bekerja dengan nilai keyakinan 0.49 iaitu berjantina perempuan dan pendapatan keluarga bawah RM2000. Faktor e\_jantina = perempuan, faktor e\_muets = Band 1 hingga Band 3 dan faktor e\_keturunan = Melayu merupakan faktor penting dalam menentukan status tidak bekerja dalam ukuran parameter ini. Nilai keyakinan yang paling tinggi iaitu 0.49 manakala nilai keyakinan paling rendah iaitu 0.47.

Jadual 4.6 10 peraturan kesatuan yang pertama dengan nilai sokongan 0.1 dan nilai keyakinan 0.5

<b>Support : 0.1</b>	<b>Confidence: 0.5</b>
<b>Bekerja</b>	<b>Tidak Bekerja</b>
Rule:	Rule:
1. e_jantina=Perempuan e_muets=Band 1 hingga Band 3 e_bidang=Sastera & Sains Sosial 3183 ==> e_status=Bekerja 1741 conf:(0.55)	1. e_jantina=Perempuan e_pendapatan=<RM2000 3850 ==> e_status=Tidak Bekerja 1974 conf:(0.51)
2. e_bidang=Sastera & Sains Sosial e_cgpa=3.10-3.50 1854 ==> e_status=Bekerja 1013 conf:(0.55)	2. e_jantina=Perempuan e_muets=Band 1 hingga Band 3 e_pendapatan=<RM2000 3462 ==> e_status=Tidak Bekerja 1772 conf:(0.51)
3. e_jantina=Perempuan e_bidang=Sastera & Sains Sosial 3475 ==> e_status=Bekerja 1897 conf:(0.55)	3. e_jantina=Perempuan e_pendapatan=<RM2000 3228 ==> e_status=Tidak Bekerja 1647 conf:(0.51)
4. e_muets=Band 1 hingga Band 3 e_bidang=Sastera & Sains Sosial e_cgpa=3.10-3.50 1722 ==> e_status=Bekerja 938 conf:(0.54)	4. e_fakulti=FAKULTI SAINS SOSIAL DAN KEMANUSIAAN 1918 ==> e_status=Tidak Bekerja 978 conf:(0.51)
5. e_keturunan=Melayu e_muets=Band 1 hingga Band 3 e_bidang=Sastera & Sains Sosial e_pendapatan=<RM2000 2237 ==> e_status=Bekerja 1204 conf:(0.54)	5. e_jantina=Perempuan e_keturunan=Melayu e_muets=Band 1 hingga Band 3 e_pendapatan=<RM2000 3034 ==> e_status=Tidak Bekerja 1547 conf:(0.51)

Terdapat tiga peraturan mempunyai nilai keyakinan yang tertinggi 0.55 untuk status bekerja iaitu e\_jantina=Perempuan e\_muets=Band 1 hingga Band 3 e\_bidang=Sastera & Sains Sosial, e\_bidang=Sastera & Sains Sosial e\_cgpa=3.10-3.50 dan e\_jantina=Perempuan e\_bidang=Sastera & Sains Sosial. Terdapat bilangan sebanyak 1918 orang pelajar daripada fakulti sains sosial dan kemanusiaan tidak dapat bekerja selepas graduan dan nilai keyakinan ialah 0.51.

Jadual 4.7 10 peraturan kesatuan yang pertama dengan nilai sokongan 0.2 dan nilai keyakinan 0.5

<b>Support : 0.2</b>	<b>Confidence: 0.5</b>
<b>Bekerja</b>	<b>Tidak Bekerja</b>
Rule:	Rule:
1. e_keturunan=Melayu e_muets=Band 1 hingga Band 3 e_bidang=Sastera & Sains Sosial 3842 ==> e_status=Bekerja 2212 conf:(0.58)	1. e_jantina=Perempuan e_pendapatan=<RM2000 3850 ==> e_status=Tidak Bekerja 1974 conf:(0.51)
2. e_keturunan=Melayu e_bidang=Sastera & Sains Sosial 4086 ==> e_status=Bekerja 2351 conf:(0.58)	
3. e_muets=Band 1 hingga Band 3 e_bidang=Sastera & Sains Sosial 4320 ==> e_status=Bekerja 2482 conf:(0.57)	
4. e_jantina=Perempuan e_bidang=Sastera & Sains Sosial 3475 ==>	



	e_status=Bekerja	1897	conf:(0.55)
5.	e_jantina=Perempuan e_keturunan=Melayu e_muets=Band 1 hingga Band 3 5103 ==> e_status=Bekerja 2614		conf:(0.51)
6.	e_jantina=Perempuan e_muets=Band 1 hingga Band 3 5812 ==> e_status=Bekerja 2969		conf:(0.51)
7.	e_keturunan=Melayu e_muets=Band 1 hingga Band 3 e_pendapatan=<RM2000 4039 ==> e_status=Bekerja 2047		conf:(0.51)
8.	e_keturunan=Melayu e_pendapatan=<RM2000 4297 ==> e_status=Bekerja 2173		conf:(0.51)
9.	e_muets=Band 1 hingga Band 3 ==> e_status=Bekerja 2342		conf:(0.5)

Peraturan e\_keturunan=Melayu e\_muets=Band 1 hingga Band 3 e\_bidang=Sastera & Sains Sosial dan e\_keturunan=Melayu e\_bidang=Sastera & Sains Sosial memiliki nilai keyakinan tertinggi iaitu 0.58 dalam status bekerja. Nilai ini dapat mengesahkan bidang sastera & sains sosial lebih diperlukan dalam pemasaran kerja daripada bidang lain. Sebaliknya, e\_jantina=Perempuan e\_pendapatan=<RM2000 mendapat nilai keyakinan 0.51 dalam status tidak bekerja. Hal ini berpendapat majoriti graduan daripada golongan perempuan dan golongan keluarga miskin.

## b) Perbincangan Faktor-faktor yang Mempengaruhi Keputusan

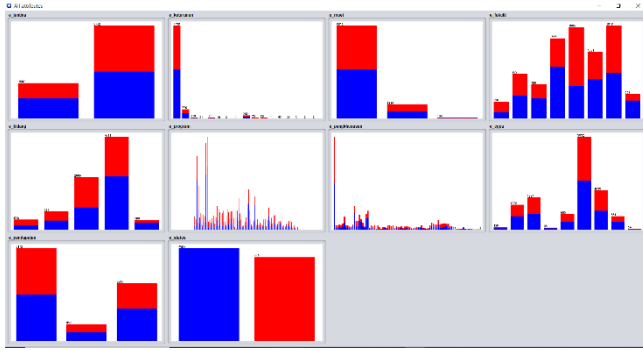
Bahagian ini membincangkan faktor yang mempengaruhi keputusan kajian ini. Kualiti data dalam kajian ini merupakan faktor yang penting supaya dapat meningkat ketepatan keputusan kajian. Kualiti data yang baik boleh diperolehi melalui fasa pra-pemprosesan. Sebelum proses ini, nilai data asal telah terlalu bising, tidak konsisten dan mempunyai nilai data yang hilang. Selepas pra-pemprosesan, masalah kerumitan data asal telah dapat diselesaikan. Fasa pra-pemprosesan melibatkan empat proses utama iaitu pembersihan data, pengurangan data, transformasi data dan melabel data. Proses pembersihan data untuk mengendalikan kehilangan data, kebisingan data dan data yang tidak konsisten. Selain itu, proses pengurangan data mengurangkan dimensi dan atribut data yang tidak berguna atau lebih manakala proses transformasi data transform jenis data daripada 'numeric' kepada 'nominal'. Akhirnya, proses melabel data adalah melabel data dalam bentuk yang lebih mudah difahami. Jadual 4.8 menunjukkan laporan kualiti data untuk atribut yang digunakan.

Faktor yang kedua adalah nilai data yang tidak seimbang. Sebagai contoh, nilai perempuan mempunyai 6752 data manakala nilai lelaki mempunyai 2587 data. Hal ini menunjukkan keputusan kajian tidak boleh mencapai keadilan kerana data ini majoriti daripada golongan perempuan sahaja. Daripada histogram atribut, faktor e\_jantina, e\_keturunan, e\_muets dan e\_bidang mempunyai jurang yang besar daripada nilai data yang lain. Rajah 4.1 memaparkan histogram atribut untuk menjelaskan ketidakseimbangan data.

Jadual 4.8 Laporan kualiti data

Ciri-ciri	Kiraan	Kardinaliti	Mod	Kekerapan		Kekerapan		
				Mod	Mod %	Mod Kedua	Mod Kedua %	
e_jantina	9339	2	Perempuan	6752	72.30	Lelaki	2587	27.70
e_keturunan	9339	17	Melayu	7738	82.86	Cina	775	8.30
e_muets	9339	3	Band 1 hingga Band 3	8015	85.82	Band 4 hingga Band 6	1214	13.00
e_fakulti	9339	8	Fakulti Sains Sosial Dan Kemanusiaan Sastera & Sains	1918	20.54	Teknologi	1885	20.18
e_bidang	9339	5	Sosial	4711	50.44	Sains	2665	28.54
e_program	9339	158	Pentadbiran Perniagaan	459	4.91	Islam	430	4.60
e_pengkhususan	9339	88	Tidak Berkaitan	1840	19.70	Syariah	468	5.01
e_cgpa	9339	9	3.10-3.50	3832	41.03	3.50-4.00	1638	17.54
e_pendapatan	9339	3	<RM2000	5172	55.38	RM2000- RM5000	3215	34.43
e_status	9339	2	Bekerja	4905	52.52	Tidak Bekerja	4434	47.48





Rajah 4.1 Histogram atribut

## V. KESIMPULAN

Secara keseluruhannya, kajian ini dijalankan untuk mengenal pasti faktor-faktor kebolehpasaran dengan paraturan sekutuan yang dilahirkan daripada dataset graduan dari tahun 2016 hingga tahun 2018. Kajian ini dilaksanakan dengan menggunakan teknik perlombongan data dan algoritma Apriori.

Kajian graduan ini dijumpai melalui SKPG daripada Kementerian Pendidikan Tinggi. Dengan menggunakan teknik peraturan sekutuan dalam kajian ini, objektif kajian iaitu mengenal pasti faktor-faktor kebolehpasaran graduan, membangun satu model peraturan sekutuan bagi kebolehpasaran graduan dan mengenal pasti hubungan antara kebolehdapatan kerja dan kebolehpasaran graduan akan dicapai.

Selain itu, kajian kesusasteraan terhadap aplikasi teknik algoritma Apriori dan teknik peraturan sekutuan telah dibaca untuk memahami konsep tentang kedua-dua teknik tersebut. Kaedah yang digunakan dalam kajian ini ialah Proses Lintas-Industri untuk perlombongan data (CRISP-DM). Kaedah ini melibatkan enam fasa dan setiap fasa akan diubah suai bagi memenuhi keperluan kajian. Enam fasa dalam CRISP-DM terlibat fasa pemahaman perniagaan, pemahaman data, penyediaan data, pemodelan, penilaian dan pengaplikasian.

Pra-pemprosesan data adalah penting dalam fasa penyediaan data kerana pembersihan data dan transformasi data boleh mengurangkan kelebihan data dan menyeragamkan data supaya memudahkan proses pemodelan. Proses ini telah dijalankan dengan menggunakan perisian Weka.

Atribut yang dipilih dalam kajian ini untuk menentukan paraturan sekutuan ialah e\_muet, e\_keturunan, e\_jantina, e\_pendapatan, e\_bidang, e\_cgpa dan e\_fakulti. Faktor e\_muet = Band 1 hingga Band 3 mempunyai kekerapan yang tertinggi adalah nilai 0.48. Parameter yang baik perlulah mempunyai nilai keyakinan yang tinggi

## VI. RUJUKAN

Othman, Z., Shan, S. W., Yusoff, I., & Chang, P. K. (2018). Classification techniques for predicting graduate employability. *International Journal on*

*Advanced Science, Engineering and Information Technology*, 8(4-2), 1712-1720.

Abdul Rahman, Nor Azziaty and Tan, Kian Lam and Lim, Chen Kim (2017) *Supervised and unsupervised learning in data mining for employment prediction of fresh graduate students*. Journal of Telecommunication, Electronic and Computer Engineering, 9 (2). pp. 155-161. ISSN 2180-1843

Sapaat, M. A., Mustapha, A., Ahmad, J., & Chamili, K. (2011). A Data Mining Approach to Construct Graduates Employability Model in Malaysia, 1(4), 1086–1098.

Ismail, Rahmah & Wei Sieng, Lai. (2011). Employers' Perceptions on Graduates in Malaysian Services Sector. *International Business Management*. 5. 184-193. 10.3923/ibm.2011.184.193.

Zhang, Leibao. (2019). Association Rule Mining for Career Choices Among Fresh Graduates. *Applied and Computational Mathematics*. 8. 37. 10.11648/j.acm.20190802.13.

Abdul Wahab, Diana Binti (2017) *Graduate Labour Market Analysis in Malaysia*. PhD thesis, University of Leeds.

Hao-Wen Shih. (2018). Analysis of the Association Rules of Graduates' Employment Types and Achievements in Core Academics using Data Mining at School. *The International Journal Management Studies*, 13(2), 8-18

Malaysia: Youth Employment (2019) [https://www.theglobaleconomy.com/Malaysia/youth\\_unemployment/](https://www.theglobaleconomy.com/Malaysia/youth_unemployment/)

Association rule learning [https://en.wikipedia.org/wiki/Association\\_rule\\_learning](https://en.wikipedia.org/wiki/Association_rule_learning)

Hainnur Aqma Rahim (2014). Definisi kebolehpasaran. 29 Apr