

LAPORAN TEKNIK

TAJUK PROJEK

ANALISA CORAK KUNJUNGAN DAN RAMALAN TEMPOH MENUNGGU PESAKIT
LUAR DI KLINIK KERAJAAN TERPILIH NEGERI SELANGOR

KOD PROJEK

PP-FTSM-2020

KETUA PENYELIDIK

PROF. MADYA DR. SUHAILA BINTI ZAINUDIN (K007989)

PENYELIDIK BERSAMA

- 1) DZULHUSNI ANJANG AB. RAHMAN (GP06058)
- 2) DR. AFZAN BINTI ADAM (K016599)

FAKULTI/INSTITUT/PUSAT

FAKULTI TEKNOLOGI & SAINS MAKLUMAT -PUSAT PENYELIDIKAN
TEKNOLOGI KECERDASAN BUATAN (CAIT) UNIVERSITI KEBANGSAAN
MALAYSIA

suhaila.zainudin@ukm.edu.my

TEMPOH PENYELIDIKAN

01/10/2018 HINGGA 12/07/2019 (9 Bulan)

2020

PERAKUAN

Laporan teknikal ini adalah hasil daripada projek yang ditaja oleh

UNIVERSITI KEBANGSAAN MALAYSIA

ABSTRAK

Tumpuan orang ramai ke klinik kesihatan kerajaan (KK) menyebabkan kesesakan seterusnya menjadikan tempoh menunggu yang panjang bagi mendapatkan rawatan. Tempoh menunggu ialah tempoh masa pesakit perlu menunggu bermula dari pendaftaran sehingga konsultasi dengan pegawai perubatan. Menunggu terlalu lama merupakan punca utama pesakit tidak berpuas hati berbanding dengan faktor-faktor lain terutamanya jika tiada kemudahan menunggu yang selesa. Namun, tempoh menunggu yang dimaklumkan kepada pesakit lebih awal dapat mempengaruhi tahap ketidakpuashatian pesakit berkenaan. Objektif kajian ini adalah untuk menganalisa corak kunjungan pesakit luar ke Klinik Kesihatan (KK) dan mencadangkan KK yang patut dikunjungi berdasarkan waktu tertentu serta seterusnya meramalkan tempoh masa menunggu berdasarkan waktu ketibaan. Data yang digunakan dalam kajian ini diperolehi dari lima belas KK di negeri Selangor yang menggunakan sistem rekod perubatan elektronik. Melalui analisa set data yang dibuat, terdapat hubungan korelasi antara tempoh menunggu pesakit dengan waktu, hari dan bulan kunjungan. Faktor umur dan waktu konsultasi juga mempunyai hubungan korelasi dengan tempoh menunggu. Kaedah pembelajaran mesin regresi dan klasifikasi telah digunakan bagi meramalkan tempoh menunggu. Atribut tempoh menunggu (*TimeWaiting*) telah didiskretkan daripada jenis integer dengan nilai berterusan kepada jenis berkategori dengan tiga nilai nominal. Model yang regresi terbaik dihasilkan dengan algoritma Regresi Linear dengan purata ralat (MSE) -0.158882. Model klasifikasi *Linear Discriminant Analysis* (LDA) dan *Support Vector Machine* (SVM) dengan kernel linear masing-masing telah menghasilkan ketepatan ramalan 98.41% dan 95.90%.

PERNYATAAN MASALAH

Jatuh sakit bukan merupakan pilihan tetapi adalah takdir yang perlu dilalui sama ada suka ataupun tidak. Apabila seseorang jatuh sakit, maka klinik atau hospital adalah tempat yang akan ditujui bagi mendapatkan rawatan. Bukan semua penduduk sama ada warganegara atau bukan warganegara Malaysia berkemampuan untuk mendapatkan rawatan di premis-premis swasta yang sudah pasti menelan belanja yang tinggi. Ini menjadikan KK menjadi tumpuan utama bagi mendapatkan rawatan

kesihatan yang berkualiti dengan kos yang amat murah (Nur Syamila Kamarul Arefin 2018). Tumpuan orang ramai ke KK ini menyebabkan kesesakan seterusnya menjadikan tempoh menunggu yang panjang bagi mendapatkan rawatan ("Pesakit dibebani tunggu lama" 2009). Tempoh menunggu didefinisikan sebagai tempoh masa pesakit perlu menunggu bermula dari pendaftaran sehingga konsultasi dengan pegawai perubatan (Belayneh et al. 2017). Menunggu terlalu lama boleh menyebabkan pesakit tidak berpuas-hati terutamanya sekiranya tiada ruang menunggu yang selesa ("Maklumbalas : Kerusi Tak Cukup Cetus Rungutan" 2012).

Dalam kajian yang dibuat oleh Pillay et al. (2011), tempoh menunggu yang lama merupakan punca utama pesakit tidak berpuas hati berbanding dengan faktor-faktor lain seperti kemudahan disediakan dan sikap kakitangan klinik/hospital yang berkenaan. Pesakit yang tidak berpuas hati adalah pelanggan yang tidak akan setia dan berpotensi memberi imej yang buruk kepada klinik/hospital berkenaan. Namun, Bielen & Demoulin (2007) dalam kajiannya menyatakan anggaran tempoh menunggu yang dimaklumkan kepada pesakit lebih awal dapat mempengaruhi tahap ketidakpuashatian pesakit berkenaan. Anggaran waktu menunggu sukar diramal memandangkan tempoh menunggu berkenaan bergantung kepada jenis rawatan kepada pesakit yang datang lebih awal, prosedur-prosedur yang perlu dilalui dan pendekatan yang digunakan oleh doktor bertugas dalam mendiagnosis serta seterusnya merawat pesakit (Chen et al. 2016). Secara umumnya klinik kesihatan kerajaan kelihatan sentiasa penuh dan sesak sepanjang waktu operasinya. Rekod kunjungan pesakit ke KK perlu dianalisa untuk mengenalpasti corak umum dan tersembunyi bagi membolehkan cadangan kepada pesakit mengenai KK mana yang patut dikunjungi pada waktu tertentu bagi meminimumkan tempoh menunggu dan seterusnya meramalkan tempoh menunggu berdasarkan waktu ketibaan pesakit ke KK.

OBJEKTIF

1. Menganalisa dan visualisasi corak umum dan tersembunyi berkaitan kunjungan pesakit luar ke KK yang terlibat bagi mendapatkan perkhidmatan kesihatan.
2. Mencadangkan waktu kunjungan yang sesuai ke KK tertentu bagi mengelak tempoh menunggu yang panjang.

PERSOALAN KAJIAN

Bagaimana untuk menganalisa dan memvisualisasi corak umum dan tersembunyi berkaitan kunjungan pesakit luar ke klinik kesihatan.

KAJIAN KEPUSTAKAAN

Meramalkan tempoh menunggu pesakit luar di hospital atau klinik adalah aktiviti yang kompleks. Ianya disebabkan proses yang dilalui oleh setiap pesakit adalah berbeza antara satu sama lain. Tempoh menunggu ini juga bergantung kepada proses yang dilalui oleh pesakit yang hadir lebih awal (beratur dihadapan) atau pesakit dengan keutamaan yang dibenarkan untuk memotong barisan menunggu. Terdapat juga faktor luaran yang tidak dapat dikawal seperti pendekatan pegawai perubatan mendiagnos pesakit yang bergantung kepada pegawai perubatan itu sendiri, keupayaan pesakit menerangkan simptom-simptom penyakitnya kepada pegawai perubatan, masalah berkaitan dengan bahasa pertuturan dan kepelbagaian ujian saringan yang perlu dilalui oleh pesakit.

Melalui kajian di klinik Kesihatan Putrajaya yang dilaksanakan oleh Zulkarnain et al. (2010), nombor giliran yang dihasilkan oleh Sistem Pengurusan Giliran (QMS) mempunyai potensi untuk digunakan sebagai salah satu atribut yang penting bagi meramalkan tempoh menunggu. Bagi mencapai tujuan itu nombor giliran QMS perlu ditukarkan kepada atribut yang lebih bermakna dengan menggunakan masa ketibaan

pesakit ke klinik sebagai salah satu komponen bagi menghasilkan atribut yang dinamakan sebagai Ketumpatan Giliran Menunggu (QD).

QD merupakan atribut bagi jumlah bilangan pesakit yang sedang menunggu untuk mendapatkan rawatan pegawai perubatan. Semasa pemprosesan awal data, ianya diasingkan kepada blok waktu setiap tiga puluh (30) minit. Sebagai contoh, QD dikira bagi pesakit yang tiba di klinik pada jam 8.00 hingga 8.30 pagi sebagai satu set QD, pesakit yang hadir pada jam 8.31 hingga 9.00 pagi sebagai set QD yang lain dan begitulah seterusnya sehingga tamat waktu operasi klinik. Seorang pesakit yang tiba di klinik kesihatan dan terdapat sepuluh (10) orang pesakit yang masih menunggu, maka QD bagi pesakit baharu tersebut adalah 10. Set data yang telah ditambah atribut QD ini didedahkan kepada algoritma pembelajaran mesin Pokok Keputusan bagi mendapatkan ramalan tempoh menunggu. Atribut QD berpotensi membantu dalam analisa statistik dan meningkatkan ketepatan peramalan bagi waktu menunggu yang dengan tempoh yang lama. Algoritma dalam Rajah 1 berikut digunakan untuk menjana atribut QD:

Algorithm

For each Current_Pt,

Include Prev_pt_QueueNo in QD_list IF (Current-Pt_Arrival –
Prev_Pt_Arrival_Time <= 30 minutes)

Current_Pt_QD = Current-Pt_QueueNo – minimum (QD_list)

Rajah 1 : Algoritma Penghasilan QD

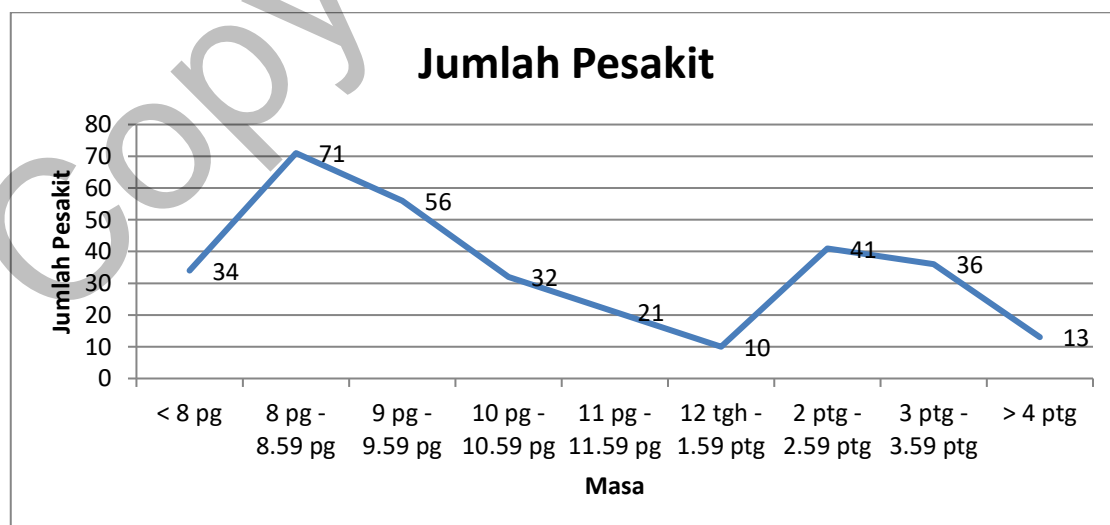
Sumber (Zulkarnain et al. 2010)

Analisa deskriptif yang digunakan adalah berupa statistik ciri-ciri pesakit seperti peratusan mengikut jantina, julat umur, kumpulan pendapatan dan kategori hadir ke klinik sama ada kecemasan atau tidak. Contoh analisa menggunakan jadual adalah seperti Jadual 1 berikut:

Jadual 1 : Analisa statistik pesakit luar Hospital Putrajaya

Ciri-ciri	Frekuensi (%)
Jantina	
Lelaki	21,472 (40.4)
Wanita	31,680 (59.6)
Kumpulan Umur	
0 - 4	9,028 (17.0)
5 - 9	2,937 (5.5)
10 - 12	970 (1.8)
13 - 17	2,243 (4.2)
18 - 29	14,791 (27.8)
30 - 39	9,194 (13.3)
40 - 49	7,026 (13.2)
50 - 59	5,309 (10.0)
Lebih besar dari 60	1,608 (3.0)

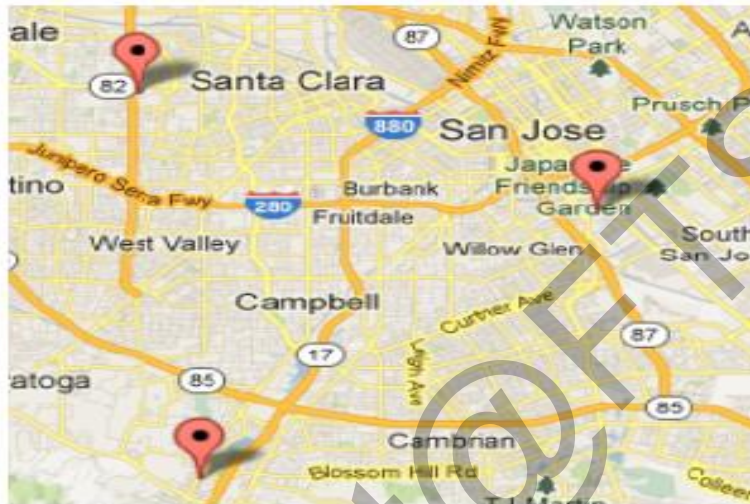
Dalam kajian di Hospital Putrajaya ini juga carta garis digunakan bagi menunjukkan jumlah pesakit yang tiba dihospital mengikut waktu. Paksi x carta garis berkenaan mewakili waktu ketibaan pesakit luar ke Hospital Putrajaya dan paksi y mewakili jumlah pesakit. Penggunaan carta garis berkenaan mewakili data Hospital Putrajaya adalah seperti Rajah 2 berikut. Dengan menggunakan contoh carta garis dibawah jelas digambarkan pesakit luar paling ramai tiba ke Hospital Putrajaya pada jam 8 pagi sehingga sebelum 9 pagi.



Rajah Error! No text of specified style in document. : Carta garis kedatangan pesakit luar berbanding waktu

Ramalan Tempoh Menunggu Pelanggan Di Jabatan Kenderaan Bermotor California, Amerika Syarikat (DMV)

Objektif kajian yang dilakukan oleh Zhang et al.(2013) dengan menggunakan set data DMV adalah untuk meramalkan tempoh menunggu pelanggan yang mendapatkan perkhidmatan pelbagai lokasi pejabat DMV California, Amerika Syarikat.



Rajah 3 : Peta kedudukan pejabat DMV San Jose, Santa Clara dan Los Gatos

◆ Sumber (Zhang et al. 2013)

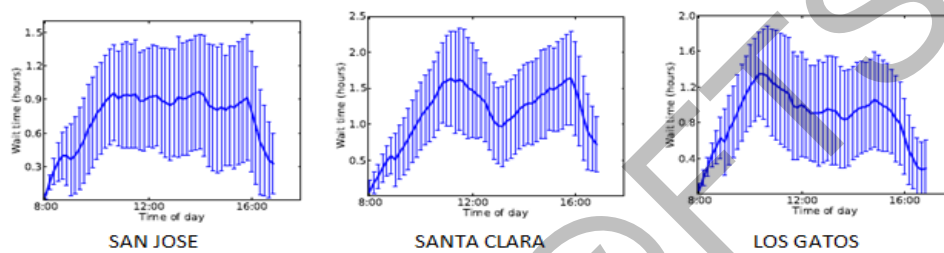
Terdapat sebanyak 176 pejabat DMZ di seluruh California. Ada diantara pejabat-pejabat berkenaan yang berada dalam jarak yang tidak jauh antara satu sama lain. Sebagai contoh pejabat DMV San Jose, Santa Clara dan Los Gatos berada berhampiran antara satu sama lain iaitu dalam lingkungan kurang dari dua puluh (20) kilometer sebagaimana ditunjukkan dalam Rajah di atas (Zhang et al. 2013).

Set data ini telah dibahagikan kepada dua bahagian dengan pembahagian lima puluh peratus (50%) digunakan untuk melatih algoritma dan selebihnya sebagai set data untuk latihan. Seterusnya atribut tempoh menunggu dengan nilai berterusan telah didiskretkan kepada hanya dua puluh (20) nilai. Nilai diskret ini dijana melalui median daripada histogram tempoh menunggu dengan ketinggian yang sama.

Visualisasi Set Data DMV

Kajian ini menggunakan data sejarah tempoh menunggu yang dikumpulkan bagi tempoh tiga (3) bulan iaitu berjumlah tujuh ratus ribu (700,000) rekod. Berdasarkan explorasi dan visualisasi data di Rajah 4 dan Rajah , corak-corak berikut ditemui:

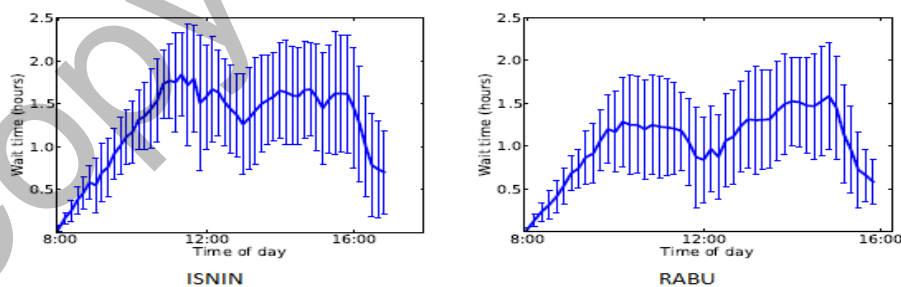
- i) Corak tempoh menunggu adalah berbeza antara pejabat DMZ . Dalam kajian ini, analisa dibuat kepada set data bagi pejabat DMZ San Jose, Santa Clara dan Los Gatos.



Rajah 4 : Purata dan sisihan piawai tempoh menunggu di tiga pejabat DMV

Sumber (Zhang et al. 2013)

- ii) Perbezaan corak tempoh menunggu juga dikesan berdasarkan hari. Dalam kajian ini didapati pada hari Isnin tempoh menunggu lebih panjang berbanding dengan hari Rabu.



Rajah 5 : Tempoh menunggu di Santa Clara pada hari Isnin dan Rabu

Sumber (Zhang et al. 2013)

Teknik Ramalan Tempoh Menunggu Bagi Set Data DMV

Objektif utama bagi kajian yang dilakukan oleh Zhang (2013) ini adalah untuk meramalkan tempoh menunggu akan datang di pejabat DMV. Kajian ini telah menggunakan teknik-teknik seperti ramalan berasaskan masa dan ramalan menggunakan algoritma pembelajaran mesin.

Persediaan Set data

Dalam fasa persediaan data bagi kajian ramalan tempoh menunggu DMV, aktiviti-aktiviti berikut telah dilaksanakan sebelum eksperimen dibuat.:

- a) Set data dipecahkan kepada dua bahagian yang sama sebagai set data latihan dan set data pengujian.
- b) Nilai atribut tempoh menunggu didiskretkan kepada dua puluh sela yang sama rata. Nilai bagi sela masa diambil daripada nilai median setiap sela.

Ramalan Berasaskan Masa

Model pertama yang digunakan dalam kajian ini ialah ramalan berasaskan masa. Pendekatan yang digunakan dalam model ini ialah dengan membuat pengiraan purata waktu bagi setiap jam dalam minggu berkenaan. Sebagai contoh, purata tempoh menunggu pelanggan yang mengunjungi DMV pada hari Isnin jam 8 pagi dikira. Nilai purata ini kemudian dijadikan nilai atribut kelas bagi julat berkenaan. Dengan menggunakan pendekatan ini, purata ralat ramalan yang diperolehi ialah 0.5235 jam.

Ramalan Menggunakan Algoritma Pembelajaran Mesin

Bagi menghasilkan model ramalan tempoh menunggu set data DMV, algoritma pembelajaran mesin yang digunakan adalah algoritma *Hidden Markov* (HMM), algoritma Rangkaian *Dynamic Bayesian*, algoritma *Markov Chain*, algoritma Model Pokok Keputusan, algoritma SVM dan algoritma Regresi Linear. Keputusan yang diperolehi melalui eksperimen dinyatakan dalam Jadual Error! **No text of specified style in document..**

Jadual **Error! No text of specified style in document.** : Purata ralat ramalan berdasarkan algoritma

Bil.	Algoritma	Purata Ralat Ramalan
1.	Rangkaian <i>Dynamic Bayesian</i>	0.7340 jam
2.	<i>Hidden Markov</i>	0.2981 jam
3.	Pokok Keputusan	0.2592 jam
4.	SVM	0.1636 jam
5.	<i>Markov Chain</i>	0.1629 jam
6.	Regresi Linear	0.1582 jam

Model yang dihasilkan menggunakan algoritma regresi linear adalah yang terbaik dengan purata ralat ramalan 0.1582 jam diikuti oleh model *Markov Chain* (ralat ramalan 0.1626 jam) dan seterusnya model SVM (ralat ramalan 0.1636 jam).

Kajian Literatur Analisa Prediktif Tempoh Menunggu Kecemasan Hospital di Negara Portugal

Kajian analisa prediktif tempoh menunggu kecemasan yang dibuat oleh Gonçalves et al. (2019) di salah satu hospital di negara Portugal adalah menggunakan algoritma *Random Forest* (RF). Dalam kajian berkenaan set data yang digunakan mengandungi semua rekod bermula pesakit masuk ke premis hospital, melalui proses *triage*, prosedur, mendapatkan rawatan sehinggalah kepada keluarnya pesakit daripada hospital.

Dalam pemrosesan awal data set data masa ketibaan telah didiskretkan kepada bentuk sela masa dengan menggunakan format masa 24 jam. Nilai sela masa yang digunakan adalah "21–3H", "3–7H", "7–12H", "12–15H", "15–19H" dan "19–21H". Atribut tempoh menunggu juga didiskretkan kepada enam nilai iaitu "really low", "low", "average", "high", "really high". Eksperimen yang menggunakan algoritma RF tersebut hanya menghasilkan ketepatan ramalan sebanyak 50.09%

Parallel Patient Treatment Time Prediction Algorithm For In Queuing Management.

Kajian oleh Sandeep & Harika (2017) yang dibuat di hospital-hospital di India ini mempunyai objektif bagi membangunkan model ramalan tempoh waktu rawatan

pesakit luar dengan menggunakan set data sejarah. Contoh set data yang digunakan adalah seperti dalam Jadual di bawah.

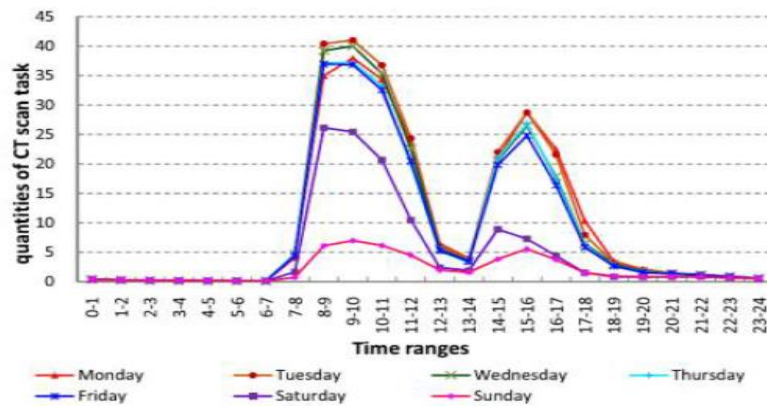
Jadual 3 : Contoh data kajian oleh Sandeep & Harika 2017

Sumber (Sandeep & Harika 2017)

Patient No	Gen	Age	Task name	Dept name	Doctor name	Start time	End time
0001	Male	15	Checkup	Surgery	Dr. Chen	2015-10-10 08:30	2015-10-10 08:42
0001	Male	15	Payment	Cashier	Null	2015-10-10 08:50	Null
0001	Male	15	CT scan	CT-5	Dr. Li	2015-10-10 09:20	2015-10-10 09:27
0001	Male	15	MR scan	MR-8	Dr. Pan	2015-10-10 10:05	2015-10-10 10:15
0001	Male	15	Medicine	Pharmacy	Null	2015-10-10 10:42	2015-10-10 10:45

Semasa aktiviti persediaan data semua prosedur yang dibuat telah dikira tempoh masa dengan nilai atribut End time ditolakkan dengan nilai atribut Start time sebagaimana dalam Jadual di atas. Hasil tolakan ini atribut baharu yang dinamakan “tempoh masa prosedur” telah dijana. Set data yang disediakan merangkumi set data lawatan pesakit, set data pesakit, set data rawatan pesakit dan atribut tempoh masa prosedur yang baru dijana. Algoritma *Random Forest* telah digunakan untuk meramalkan tempoh rawatan pesakit.

Analisa deskriptif yang kerap digunakan dalam kajian ini adalah carta garis seperti contoh di Rajah . Dalam carta garis ini paksi x mewakili masa (dalam jam) dan jumlah prosedur imbasan CT pada paksi y. Garisan berwarna pula mewakili hari dalam minggu (Sandeep & Harika 2017). Dengan menggunakan carta garis ini lebih mudah difahami. Dengan menggunakan contoh dibawah, pada jam 9-10 pagi jumlah maksimum prosedur imbasan CT berlaku iaitu dengan jumlah sebanyak 40 prosedur.

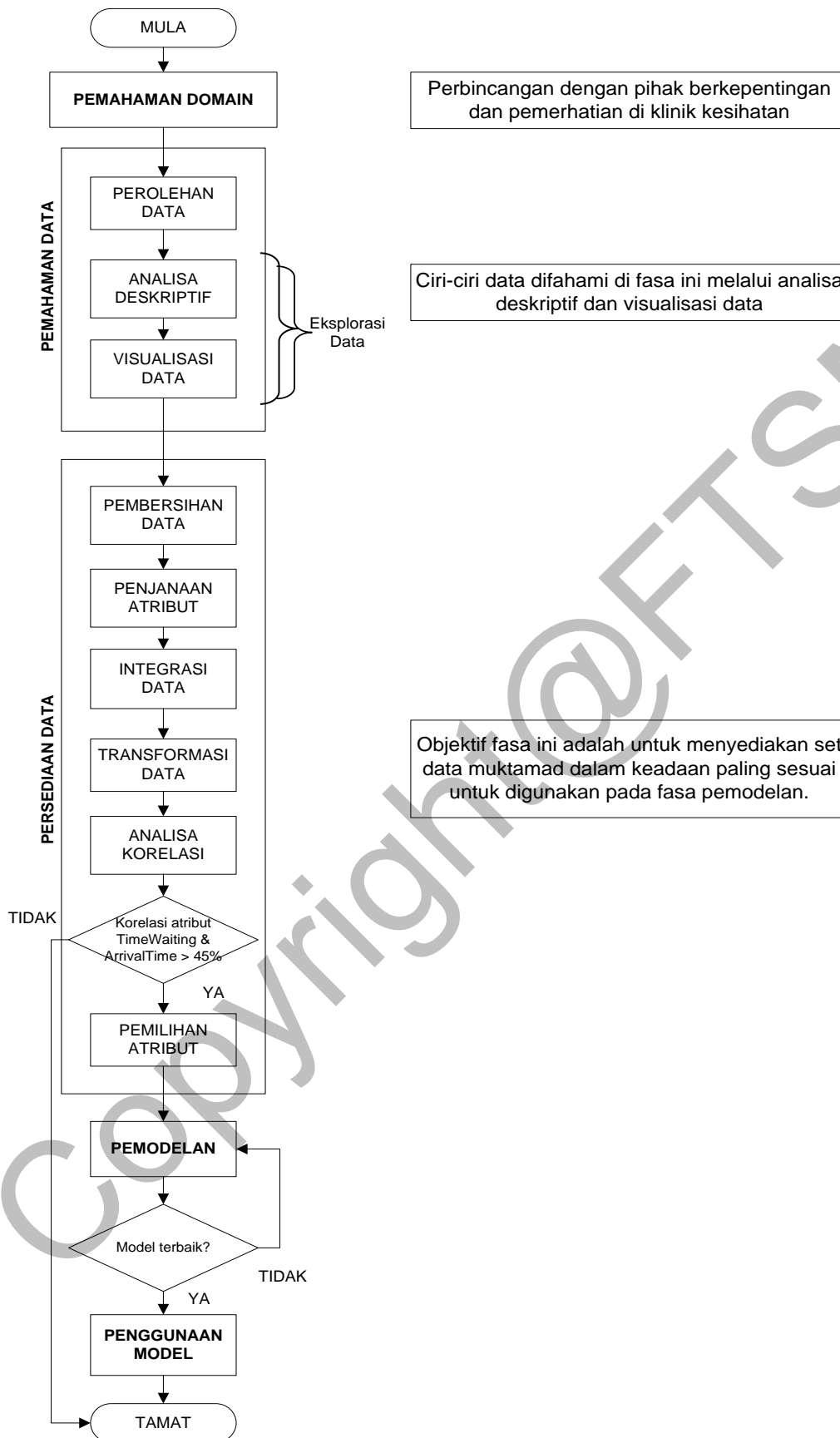


Rajah 6 : Contoh penggunaan carta garis dalam data tempoh menunggu

Sumber (Sandeep & Harika 2017)

METODOLOGI PENYELIDIKAN

Metodologi kajian ini adalah berasaskan kepada metodologi CRISP DM dengan beberapa penyesuaian dibuat bagi menyesuaikan dengan keperluan kajian. Metodologi yang digunakan dalam kajian ini dapat dirumuskan seperti di Rajah berikut:



Rajah 7 : Aliran aktiviti kajian

Pemahaman Bisnes (Kajian domain)

Objektif utama aktiviti pemahaman bisnes adalah untuk memahami persekitaran bisnes semasa. Aktiviti yang dilaksanakan bagi memahami bisnes adalah perbincangan dengan pihak berkepentingan iaitu pegawai-pegawai di KKM bagi mendapatkan maklumat dan membuat pemerhatian secara tanpa mengganggu di klinik kesihatan yang berkenaan. Berdasarkan aktiviti-aktiviti di atas potensi kajian dapat difahami dan diukur kepentingannya. Elemen penting yang perlu difahami dan diuruskan ialah teori dan konsep umum berkaitan dengan menunggu giliran serta proses dan elemen-elemen yang berkaitan dengan menunggu giliran bagi mendapatkan perkhidmatan di klinik kesihatan kerajaan.

Pemahaman Data

Aktiviti-aktiviti yang terlibat dalam fasa pemahaman data adalah perolehan data, analisa deskriptif dan visualisasi data serta mengenalpasti isu dan masalah data. Pada peringkat awal pemahaman data perisian Microsoft Excel digunakan memandangkan data yang diterima adalah dalam bentuk Microsoft Excel. Pemprosesan lanjutan data seterusnya dilaksanakan menggunakan bahasa pengaturcaraan Python versi 3.6.4 dengan menggunakan Microsoft Visual Studio Code versi 1.33.1.

Perolehan Data

Aktiviti pertama dalam fasa pemahaman data ialah membuat perolehan data itu sendiri. Pada peringkat awal, data bagi kajian ini diperolehi daripada portal data terbuka kerajaan Malaysia iaitu <http://www.data.gov.my>. Walau bagaimanapun data berkenaan tidak mencukupi dari sudut bilangan atribut dan jumlah rekod data, tambahan pula data berkenaan juga telah dibersih dan diringkaskan. Data yang lengkap bagi kajian ini diperolehi daripada Bahagian Pembangunan Kesihatan Keluarga, Kementerian Kesihatan Malaysia (KKM). Data yang diperolehi ini merupakan rekod kehadiran pesakit luar ke klinik-klinik kerajaan di negeri Selangor bagi bulan Januari hingga Disember 2018 daripada Sistem TPC.

Penerangan Ciri-ciri Set Data

Set data rekod kehadiran pesakit luar ke klinik-klinik kerajaan di negeri Selangor diterima dalam bentuk Microsoft Excel (*.xlsx). Data berkenaan disusun dalam dua belas (12) lembaran kerja mengikut bulan iaitu Januari 2018 sehingga bulan Disember 2018. Data bagi setiap bulan adalah merangkumi keseluruhan data bagi kesemua klinik kerajaan bagi negeri Selangor yang menggunakan Sistem TPC. Jumlah keseluruhan rekod data tersebut adalah berjumlah 1,320,459 rekod data. Pecahan rekod data mengikut bulan ditunjukkan dalam Jadual dibawah.

Jadual 4: Jumlah rekod mengikut bulan

Bulan	Jumlah Rekod
Januari	115,879
Februari	102,267
Mac	109,740
April	108,648
Mei	97,214
Jun	87,446
Julai	127,304
Ogos	115,860
September	99,733
Oktober	124,801
November	117,480
Disember	114,087
JUMLAH BESAR	1,320,459

Set data yang diterima dari KKM ini mempunyai tiga belas (13) atribut yang merupakan rekod lawatan pesakit luar ke klinik kerajaan. Atribut-atribut bagi set data ini terdiri dari pelbagai. Selain daripada menggunakan fungsi-fungsi yang terdapat dalam Microsoft Excel, keterangan Set Data juga diperolehi dengan menggunakan perisian Python menggunakan fungsi `shape()` dan `describe()`. Fungsi Python `dtypes()` juga digunakan untuk melihat jenis atribut-atribut yang terdapat dalam set data tersebut.

Eksplorasi Data

Di peringkat awal aktiviti eksplorasi data, fungsi-fungsi Microsoft Excel digunakan bagi memahami data. Keseluruhan rekod data tidak dapat dikumpulkan ke dalam satu lembaran kerja memandangkan jumlah baris maksimum Microsoft Excel

hanyalah sebanyak 1,048,576 baris sedangkan keseluruhan data berjumlah 1,320,459 iaitu melebihi kapasiti sebanyak 271,883. Data-data yang sebelum ini dikumpulkan mengikut bulan telah diasingkan berdasarkan daerah masing-masing dengan menggunakan fungsi *Advanced Filter* yang sedia terdapat dalam Microsoft Excel. Dengan menggunakan fungsi ini, set data telah dipecahkan mengikut daerah dan klinik. Set data ini kemudian disimpan dalam (5) lima fail dalam bentuk Microsoft Excel mengikut nama daerah masing-masing. Eksplorasi data juga dibuat secara berasingan mengikut daerah dan klinik kesihatan. Aktiviti eksplorasi data yang dibuat bagi set data ini adalah seperti berikut:

Analisa Deskriptif

Analisa deskriptif dilaksanakan sebanyak dua kali iaitu pada set data asal dan pada set data selepas pembersihan dan penjana atribut baharu. Objektif analisa deskriptif adalah untuk memahami dan menilai kualiti set data serta mendapatkan *insight* diperingkat awal kajian. Analisa deskriptif yang dibuat ialah mendapatkan purata, nilai minimum dan maksimum data, sisihan piawai serta pecahan pada kedudukan 25%, 50% dan 75%. Fungsi Python `describe()` digunakan dalam analisa ini. Objektif utama kajian ini adalah bagi meramalkan tempoh menunggu pesakit. Atribut kelas atau label bagi kajian ini ialah `TimeWaiting`.

Visualisasi Data

Matlamat aktiviti visualisasi data dibuat adalah untuk mendapatkan gambaran, corak dan ciri-ciri set data dalam bentuk yang lebih mudah difahami. Kualiti data, isu dan masalah set data juga dapat dikenalpasti dalam aktiviti visualisasi data ini. Aktiviti visualisasi ini akan menjadi input bagi fasa persediaan data.

Persediaan Data

Fasa persediaan data merupakan fasa yang terpenting dalam kajian sains data kerana aktiviti ini berperanan bagi menghasilkan set data muktamad dan boleh dipercayai sebelum didedahkan kepada algoritma pembelajaran mesin.

Aktiviti-aktiviti di bawah fasa persediaan data ini adalah pembersihan data, penjanaan atribut baharu daripada atribut sedia ada, memuatnaik data ke dalam perisian Python, integrasi set data dengan set data pihak ketiga, transformasi data kepada bentuk yang boleh digunakan oleh algoritma pembelajaran mesin dan pemilihan atribut untuk kajian. Fasa ini akan dilakukan secara berulang-ulang sehingga mencapai matlamatnya iaitu data set terakhir yang boleh dipercayai dan berpotensi untuk digunakan dalam kajian ini. Set data yang sebelum ini disusun dalam bentuk lima (5) fail mengikut daerah sebagaimana diterangkan dalam dibawah aktiviti eksplorasi data juga telah dipecahkan kepada lima belas (15) fail mengikut klinik kesihatan memandangkan kajian akan dibuat secara berasingan berdasarkan klinik kesihatan.

Pembersihan Data

Set data yang diterima telah diteliti dengan menyeluruh dalam fasa ini sebagai salah satu aktiviti dibawah persediaan data. Penelitian awal terhadap set data diperingkat ini masih lagi menggunakan Microsoft Excel.

Penjanaan Atribut Baharu

Terdapat atribut sedia ada dalam set data ini boleh dimanipulasi bagi menghasilkan atribut baharu yang berpotensi dan bermakna dalam kajian ini. Muatnaik Data Ke Dalam Python

Set data kehadiran pesakit ke klinik kerajaan bagi tempoh dua belas bulan iaitu dari 1 Januari 2018 hingga 31 Disember 2018 dimuatnaik ke dalam perisian Python 3.6.4 (Microsoft Visual Studio Code) dengan menggunakan fungsi `Panda.ExcelFile()` dan `Parse()`. Set data yang akan digunakan adalah mengikut daerah yang merangkumi semua klinik kesihatan dalam daerah berkenaan dan set data berdasarkan lima belas klinik secara berasingan.

Integrasi Set Data

Set data yang telah dibersihkan dan dimuat naik dalam persekitaran Python tidak mempunyai atribut hari dalam minggu (jumlah hari dalam seminggu), hari dalam bulan (hari kalender) dan bulan. Untuk menghasilkan laporan dan kajian yang lebih bermakna atribut berkenaan diperlukan.

Bagi memenuhi keperluan berkenaan, set data ini telah diintegrasikan dengan set data yang mengandungi atribut hari dalam minggu, hari dalam bulan dan bulan. Fungsi Python merge() digunakan bagi tujuan ini dengan menggunakan ReportDate sebagai kunci.

Transformasi Data

Transformasi data dibuat ke atas set data ini bagi memenuhi keperluan algoritma pembelajaran mesin yang akan digunakan.

Penilaian Korelasi Set data

Set data yang digunakan dalam kajian ini agak besar yang terdiri daripada set data bagi lima belas klinik kesihatan kerajaan dari enam daerah. Terdapat corak berbeza bagi set data setiap klinik. Memandangkan kajian ini akan dibuat secara berasingan mengikut klinik kesihatan, penilaian korelasi perlu dibuat bagi memastikan hanya set data dengan atribut yang mempunyai hubungan korelasi yang kuat sahaja dipilih untuk kajian seterusnya.

Persediaan Data Semasa Eksperimen

Semasa fasa pemodelan, antara eksperimen yang dibuat adalah dengan memadamkan rekod data atribut TimeWaiting yang lebih besar dari 160 minit. Perkara ini selaras dengan visualisasi data dimana nilai rekod yang melebihi 160 minit merupakan data hingar. Sekiranya tidak dipadam, data hingar ini dikhuatiri akan menghasilkan keputusan yang tidak relevan.

Seterusnya atribut TimeWaiting dibahagikan kepada tiga bahagian secara statistik dengan menggunakan fungsi Python `Numpy.linspace()`

Pemilihan Atribut Untuk Kajian Ramalan Tempoh Menunggu

Atribut-atribut yang akan digunakan dalam kajian ini dipilih dengan menggunakan kaedah korelasi atribut menggunakan fungsi Python *Feature Importance* (FI) dan *Recursive Feature Elimination* (RFE). Pemilihan atribut sangat penting bagi memastikan hanya atribut-atribut yang relevan kepada output kajian sahaja dipilih supaya ketepatan optimum algoritma pembelajaran mesin dapat diperolehi. Bilangan atribut yang lebih kecil juga akan memendekkan tempoh masa pembelajaran mesin serta mengurangkan kebergantungan kepada komputer yang berkapasiti tinggi.

PEMODELAN

Fasa pemodelan ini dilaksanakan secara berulang-ulang sehingga keputusan dengan ketepatan terbaik diperolehi. Dalam fasa pemodelan ini, eksperimen dibuat dengan mendedahkan set data mentah dan set data dimanipulasi kepada dua belas algoritma pembelajaran mesin bagi kaedah regresi dan klasifikasi.

Set data mentah didefinisikan sebagai set data yang telah melalui fasa persediaan data tetapi belum dilaksanakan pemilihan fitur/atribut dan pendiskretan. Atribut kelas yang dikaji adalah atribut TimeWaiting dari jenis integer dengan nilai

Set data tersebut dibahagikan kepada dua bahagian iaitu set data latihan yang merupakan tujuh puluh peratus (70%) dari keseluruhan data dan baki tiga puluh peratus (30%) set data dijadikan set data pengujian.

Set data latihan didedahkan kepada algoritma-algoritma pembelajaran mesin regresi dan klasifikasi manakala set data pengujian akan didedahkan kepada model terbaik yang terhasil daripada eksperimen. Dalam melaksanakan eksperimen menggunakan algoritma regresi, satu teknik tambahan diuji iaitu penskalaan set data bagi memberi julat nilai setiap atribut hanya dari 0 hingga 1 sahaja. Fungsi Python `StandardScaler()` digunakan bagi tujuan ini.

Sebelum eksperimen dijalankan, set data telah dibahagikan kepada dua bahagian iaitu set data latihan yang merupakan tujuh puluh peratus dari keseluruhan data dan baki tiga puluh peratus set data dijadikan set data pengujian. Teknik *10 fold cross validation* digunakan bagi menilai ketepatan model yang dihasilkan dengan menggunakan set data latihan.

Eksperimen dalam fasa pemodelan ini hanya dibuat menggunakan set data tiga klinik kesihatan yang mempunyai nilai peratus korelasi atribut TimeWaiting tertinggi iaitu dengan peratus korelasi melebihi 45%.

Penalaan Dan Kernel Algoritma

Eksperimen kaedah klasifikasi yang menggunakan algoritma SVM dibuat secara berulang-ulang dengan menggunakan empat jenis kernel iaitu kernel *linear*, kernel polinomial, kernel *radial basis* dan kernel *sigmoid*. Tiada penalaan parameter dibuat pada algoritma LDA kerana tiada parameter yang boleh dibuat penalaan (datascience.stackexchange.com 2017).

PENGGUNAAN MODEL PEMBELAJARAN MESIN

Model terbaik yang dihasilkan akan diuji dengan set data pengujian yang merupakan tiga puluh peratus daripada jumlah keseluruhan set data.

PENEMUAN PENYELIDIKAN

Maklumbalas yang diterima daripada KKM digunakan sebagai asas bagi melaksanakan aktiviti-aktiviti dalam fasa pemahaman dan persediaan data seperti pemadaman dan manipulasi rekod data. Set data yang diterima ini agak besar tetapi hanya bagi tempoh satu tahun iaitu pada tahun 2018. Oleh itu kajian ini tidak akan dibuat sebagai kajian analisa data dengan siri masa. Secara keseluruhan, corak kunjungan pesakit luar di keseluruhan klinik kesihatan adalah berbeza. Dapat dirumuskan ciri-ciri data bagi kesemua KK adalah berbeza. Ini menyebabkan eksperimen bagi set data ini perlu dilaksanakan secara berasingan mengikut set data klinik yang berkenaan.

Kedua-dua teknik pemilihan fitur telah mencadangkan atribut ArrivalTime, ConsultationTime dan ConsultationEndTime sebagai penting. Ketiga-tiga atribut ini beserta tiga atribut tambahan AgeFinal, Month dan DayOfWeek akan digunakan dalam eksperimen memandangkan salah satu daripada teknik pemilihan fitur turut mencadangkannya sebagai penting.

OUTPUT PENYELIDIKAN

MODAL INSAN

1. Bilangan Pelajar Sarjana : 1

RUJUKAN

- Belayneh, M., Woldie, M., Berhanu, N. & Tamiru, M. 2017. GLOBAL JOURNAL OF MEDICINE AND PUBLIC HEALTH 1 www The determinants of patient waiting time in the general outpatient department of Debre Markos and Felege Hiwot hospitals in Amhara regional state, North West, Ethiopia. *Gjmedph.Com* 6(5). Retrieved from www.gjmedph.com
- Bielen, F. & Demoulin, N. 2007. Waiting time influence on the satisfaction-loyalty relationship in services. *Managing Service Quality* 17(2): 174–193. doi:10.1108/096045207110735182
- Chen, J., Li, K., Tang, Z., Bilal, K. & Li, K. 2016. A parallel patient treatment time prediction algorithm and its applications in hospital queuing-recommendation in a big data environment. *IEEE Access*. doi:10.1109/ACCESS.2016.2558199
- Gonçalves, F., Pereira, R., Ferreira, J., Vasconcelos, J. B., Melo, F. & Velez, I. 2019. Predictive Analysis in Healthcare: Emergency Wait Time Prediction. *Springer Nature Switzerland AG 2019* 806: 138–145. doi:10.1007/978-3-030-01746-0
- Maklumbalas : Kerusi Tak Cukup Cetus Rungutan. 2012. *Kementerian Kesihatan Malaysia*. <http://www.moh.gov.my/index.php/pages/view/541> [7 July 2019].

- Nur Syamila Kamarul Arefin. 2018. Rakyat pilih klinik kerajaan walau terpaksa tunggu 3 jam. *FMT News*.
<https://www.freemalaysiatoday.com/category/bahasa/2018/09/01/rakyat-pilih-klinik-kerajaan-walau-terpaksa-tunggu-3-jam/> [21 February 2019].
- Pillay, M. S., Mohd Ghazali, R. J., Abd Manaf, N. H., Abdullah, A. H. A., Abu Bakar, A., Salikin, F., Umaphy, M., et al. 2011. Hospital waiting time: the forgotten premise of healthcare service delivery? *International Journal for Healthcare Quality Assurance* 24(7): 506–522. doi:DOI 10.1108/09526861111160553
- Sandeep, T. S. & Harika, K. 2017. Parallel Patient Treatment Time Prediction Algorithm for in Queuing Management by Big Data. *International Journal of Innovative Research in Science, Engineering and Technology* Vol. 6(Issue 6): 12272–12284. doi:10.15680/IJRSET.2017.0606298
- Zhang, Ye., Nguyen, L. T., & Zhang, J., 2013. Wait time prediction: how to avoid waiting in lines? In Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication (UbiComp '13 Adjunct). Association for Computing Machinery, New York, NY, USA, 481–490. DOI:<https://doi.org/10.1145/2494091.2495980>
- Zulkarnain, A.K., Sararaks, S., Azman, A.B., & Nora, M.S., 'Queue Density as a Predictors for Outpatient Waiting Time,' *Journal of Health Management*, 9/2 (2010): 10.