

**RAMALAN PERAMBATAN ALIRAN TRAFIK
BERASASKAN KESESAKAN JALAN BERJIRAN
MENGUNAKAN MODEL HIDDEN MARKOV**

BAGUS PRIAMBODO

UNIVERSITI KEBANGSAAN MALAYSIA

RAMALAN PERAMBATAN ALIRAN TRAFIK BERASASKAN KESESAKAN
JALAN BERJIRAN MENGGUNAKAN MODEL HIDDEN MARKOV

BAGUS PRIAMBODO

TESIS YANG DIKEMUKAKAN UNTUK MEMPEROLEHI IJAZAH
DOKTOR FALSAFAH

INSTITUT IR4.0
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI

2023



UNIVERSITI KEBANGSAAN MALAYSIA

The National University of Malaysia

**PERAKUAN TESIS SARJANA / DOKTOR FALSAFAH
(DECLARATION OF MASTER / DOCTOR OF PHILOSOPHY THESIS)**

A. MAKLUMAT PELAJAR/STUDENT DETAILS		
Nama <i>Name</i>	Bagus Priambodo	
No. Pendaftaran <i>Registration No.</i>	P87405	
Fakulti/Institut <i>Faculty/Institute</i>	Institut IR4.0 (IIR4.0)	
Program Pengajian <i>Program of Study</i>	Doktor Falsafah	
Tajuk Tesis <i>Thesis Title</i>	Ramalan Perambatan Aliran Trafik Berasaskan Kesesakan Jalan Berjiran Menggunakan Model Hidden Markov	
Mel-e/Email: bagus.priambodo@mercubuana.ac.id		No. Telefon/Tel.No: 085945239123
B. PERAKUAN / DECLARATION		
Merujuk kepada Dasar Harta Intelek UKM 2018, tesis adalah hak milik UKM seperti keterangan dibawah: 4.2.1 Harta Intelek yang direka cipta oleh pelajar termasuk tesis dan semua hasil penyelidikan ditulis oleh Pelajar UKM dalam tempoh pengajiannya di UKM direka cipta, dibangunkan atau dihasilkan menggunakan kemudahan, bahan, dana, atau sumber lain adalah hak milik UKM melainkan dinyatakan sebaliknya dalam Dasar ini atau dipersetujui sebaliknya oleh UKM dan Pelajar/Penaja. <i>Referring to the UKM Intellectual Property Policy 2018, the thesis is the property of UKM as described below:</i> <i>4.2.1 Intellectual property created by the student, including the thesis and all research output produced by the UKM Student during their studies at UKM, including designed, developed or produced output using facilities, materials, funds, or other resources are the property of UKM unless otherwise stated in this Policy or otherwise agreed by UKM and the Student/Sponsor.</i>		
	RAHSIA CONFIDENTIAL	Mengandungi maklumat rahsia di bawah AKTA RAHSIA RASMI 1972 <i>Consisting of classified information under the OFFICIAL SECRETS ACT 1972</i>
	TERHAD RESTRICTED	Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi / badan di mana penyelidikan dijalankan <i>Consisting of RESTRICTED information which has been determined by the organisation/body where the research was conducted</i>
✓	AKSES TERBUKA/ TIDAK TERHAD OPEN ACCESS/ NON-RESTRICTED	Saya membenarkan tesis ini diterbitkan secara akses terbuka, teks penuh atau dibuat salinan untuk tujuan pengajian, pembelajaran & penyelidikan sahaja <i>I give permission for this thesis to be published through open access, full text or copied only for study, learning, and research purposes.</i>

Bagi kategori Akses Terbuka/Tidak Terhad, saya membenarkan tesis (Sarjana/Doktor Falsafah) ini di simpan di Perpustakaan Universiti Kebangsaan Malaysia (UKM)* dengan syarat-syarat kegunaan seperti berikut:

For the Open Access/Non-Restricted category, I permit this (Master's/Doctoral) Thesis to be kept in the Universiti Kebangsaan Malaysia (UKM) Library with the following conditions:

1. Perpustakaan UKM mempunyai hak untuk membuat salinan untuk tujuan pengajian, pembelajaran, penyelidikan sahaja.

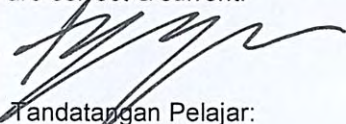
UKM Library has the right to reproduce the thesis only for study, learning and research purposes.

2. Perpustakaan Universiti Kebangsaan Malaysia dibenarkan membuat satu (1) salinan tesis ini untuk tujuan pertukaran antara institusi pengajian tinggi dan mana-mana badan/ agensi kerajaan, tertakluk kepada terma dan syarat.

UKM Library is allowed to make one (1) copy of this thesis for exchange purposes among higher education institutions and any government body/agency, subject to the terms and conditions.

Saya mengakui maklumat & tesis yang diberikan adalah benar & terkini.

I acknowledge the information & the thesis provided are correct & current.



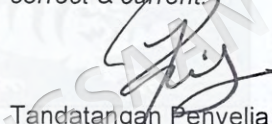
Tandatangan Pelajar:

Student's Signature

Tarikh / Date: 22 March 2023

Saya memperakukan maklumat & tesis yang diberikan adalah benar & terkini.

I verify that the information & the thesis provided are correct & current.



Tandatangan Penyelia / Supervisor's Signature:

Nama / Name:

PROF. MADYA DR. RABIAH ABDUL KADIR

Cop Rasmi / Official Stamp: FELO KANAN

Tarikh / Date: 22/3/2023 Institut IR4.0 (IIR4.0)

Universiti Kebangsaan Malaysia

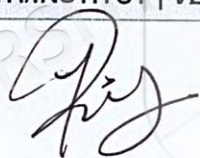
C. PENGESAHAN FAKULTI/INSTITUT | VERIFICATION OF FACULTY/INSTITUTE

Tandatangan/Signature:

Nama/Name:

Tarikh/Date: 27/3/2023

Cop Rasmi/Official Stamp:



PROF. MADYA DR. RABIAH ABDUL KADIR
Pegawai
Institut IR4.0 (IIR4.0)
Universiti Kebangsaan Malaysia

**Kuatkuasa: 15 Julai 2021

PENGAKUAN

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

22 Mac 2023

BAGUS PRIAMBODO
P87405

PENGHARGAAN

Alhamdulillah segala puji dan syukur kami panjatkan kehadiran Allah Subhanahu Wata'ala atas limpahan nikmat, kurnia dan hidayahNya saya dapat menyempurnakan tesis yang mengambil tajuk RAMALAN PERAMBATAN ALIRAN TRAFIK BERASASKAN KESESAKAN JALAN BERJIRAN MENGGUNAKAN MODEL HIDDEN MARKOV, semoga hasil penyelidikan ini dapat memberikan manfaat dalam bidang intelligent transportation system khususnya.

Pertama sekali saya ingin mengucapkan terima kasih kepada Prof Madya Dr Azlina Ahmad dan Prof Madya Dr Rabiah Abdul Kadir yang telah membimbing saya dalam menyempurnakan tesis ini.

Terima kasih khas ditujukan kepada emak saya ibu Hj. Sunarni, atas semua sokongan dan doa beliau yang tiada putusnya di mana akhirnya memudahkan saya menyempurnakan tesis ini.

Ucapan terima kasih tidak terhingga untuk isteri saya yang tercinta Nuri Haqi beserta anak-anak saya Muhammad Rasyid Zufar dan Rumaysha, atas sokongan dan pengorbanannya mengizinkan saya melanjutkan pengajian di negara Malaysia.

ABSTRAK

Kesesakan jalanraya merambat ke seluruh rangkaian disebabkan oleh peningkatan aliran trafik di jalan raya yang berjiran. Ramalan kesan riak (*ripple effect*) ke atas rangkaian ini sangat penting bagi pemandu untuk mengelakkan kesesakan jalan raya. Sebelum ini, kajian perambatan (*propagation*) aliran trafik menggabungkan data trajektori masa nyata. Sistem Kedudukan Global (GPS) dengan maklumat jalan yang berhubung. Masalah timbul apabila ramalan tidak boleh dilakukan dengan kaedah ini jika data trafik sebenar dari sensor seperti Internet of Things (IoT) tidak mempunyai maklumat mengenai jalan raya yang saling berhubung. Oleh itu, objektif utama kajian ini ialah untuk membangunkan satu model Hidden Markov bagi meramal perambatan aliran trafik menggunakan data trafik sebenar tanpa maklumat mengenai jalan raya yang saling berhubung. Untuk mencapai objektif ini, hubungan antara jalan raya berjiran dikenal pasti dan dianalisis menggunakan pola aliran trafik menurut hari dan masa. Faktor spatial dan temporal yang mempengaruhi kesesakan jalan raya di satu kawasan kejiranan juga dikaji. Metodologi kajian adalah berasaskan lima fasa utama dalam kitar hayat pembelajaran mesin iaitu pemerolehan data, pra-pemprosesan data, pemprosesan, latihan dan penilaian. Untuk meningkatkan prestasi ramalan model, satu kaedah penggugusan keadaan trafik dibangunkan menggunakan algoritma pengecaman pola. Hasil penggugusan ini seterusnya divisualisasi dan dianalisis menggunakan data trafik sebenar. Kajian ini mendapati bahawa hubungan antara jalan raya sangat mempengaruhi ketepatan ramalan. Satu model ramalan trafik menggunakan Hidden Markov berasaskan purata jalan berjiran menjadi penanda aras bagi model yang dibangunkan. Model kajian ini menunjukkan prestasi yang tinggi apabila hubungan korelasi antara jalan raya melebihi 0.6. Malah, hasil eksperimen menunjukkan bahawa ketepatan ramalan dengan model Hidden Markov yang dicadangkan mencapai 90% apabila hubungan antara jalan raya melebihi 85%. Terdapat dua sumbangan utama kajian iaitu (i) kaedah penggugusan keadaan trafik untuk mengenalpasti jalan raya yang mempunyai hubungan tinggi dengan jalan sasaran, dan (ii) model Hidden Markov untuk meramal perambatan aliran trafik di jalan raya yang berjiran.

PREDICTING TRAFFIC FLOW PROPAGATION BASED ON CONGESTION AT NEIGHBOURING ROADS USING HIDDEN MARKOV MODEL (HMM)

ABSTRACT

Road congestion propagates throughout a road network due to increase in traffic flow on neighbouring roads. Prediction of this ripple effect is very important for drivers to avoid road congestion. Previously, traffic flow propagation studies incorporated real-time Global Positioning System (GPS) trajectory data which includes information on connected roads. The problem arises when prediction cannot be made using this method when the actual traffic data from sensors such as the Internet of Things (IoT) does not have information on inter-connected roads. Therefore, the main objective of this research is to develop a Hidden Markov model to predict traffic flow propagation using actual traffic data without information on inter-connected roads. To achieve this objective, relationships between neighbouring roads are identified and analysed using traffic flow patterns according to day and time. The spatial and temporal factors that influence traffic flow near a congested road in a neighbouring area were also investigated. The research methodology is based on five main phases of the machine learning life cycle namely data collection, data pre-processing, processing, training and evaluation. To increase the prediction performance of the model, a traffic state clustering method is developed using pattern recognition algorithm. Results of clustering were then visualized and analysed using real traffic data. This study found that relationship between roads greatly affects the accuracy of prediction. A similar Hidden Markov model was used as the benchmark against the developed model. The model of this study shows high performance when the correlation between roads is above 0.6. In fact, the experimental results show that the prediction accuracy with the Hidden Markov model reaches 90% when the relationship between roads is more than 85%. There are two main contributions of this work namely (i) a traffic state clustering method to identify roads that have a high relationship with the target road, and (ii) a Hidden Markov model to predict traffic flow propagation of neighbouring roads.

KANDUNGAN

	Halaman
PENAKUAN	ii
PENGHARGAAN	iii
ABSTRAK	iv
ABSTRACT	v
KANDUNGAN	vi
SENARAI JADUAL	ix
SENARAI SINGKATAN	xv

BAB I	PENDAHULUAN	
1.1	Latar Belakang Penyelidikan	1
1.2	Permasalahan Kajian	2
1.3	Objektif Penyelidikan	5
1.4	Soalan Kajian	6
1.5	Skop Kajian	7
1.6	Sumbangan Kajian	7
1.7	Metodologi Penyelidikan	8
1.8	Struktur Tesis	12
	1.8.1 Bab I	12
	1.8.2 Bab II	12
	1.8.3 Bab III	12
	1.8.4 Bab IV	12
	1.8.5 Bab V	13
1.9	Rumusan	13

BAB II	KAJIAN LITERATUR	
2.1	Pendahuluan	14
2.2	Perambatan Aliran Trafik	14
	2.2.1 Analisis Trafik Masa Nyata	15
	2.2.2 Ramalan Perambatan Trafik	22
2.3	Faktor Ruang Dan Masa Bagi Ramalan Trafik	29
	2.3.1 Ramalan Trafik Berasaskan Jalan Raya Yang Berhubung	30

2.3.2	Ramalan Trafik Jalan Yang Memiliki Hubungan Tinggi Secara Dinamis	30
2.4	Data Raya Dalam Kajian Aliran Trafik	32
2.4.1	Sumber Pengumpulan Data Dalam Ramalan Aliran Trafik	34
2.4.2	Analitik Data Raya Dalam Kajian Trafik	37
2.4.3	Pola Aliran Trafik Yang Sama	41
2.4.4	Ramalan Aliran Trafik	44
2.5	Kenal Pasti Kesesakan Jalan Raya	46
2.6	Penggugusan Jalan Raya	49
2.6.1	Penggugusan Segmen Jalan Menggunakan K-Means	49
2.6.2	Penggugusan Spektrum	54
2.6.3	Matrik Ko-Kejadian Aras Kelabu (<i>Gray Level Co-Occurence Matrix</i>)	59
2.6.4	Menggabungkan Penggugusan Dengan <i>Gray Level Co-Occurence Matrix</i>	60
2.7	Ramalan Keadaan Trafik Menggunakan Hidden Markov Model (Hmm)	62
2.7.1	Rantai Markov (<i>Markov Chain</i>)	62
2.7.2	Model Hidden Markov	66
2.8	Matriks Kekeliruan (<i>Confusion Matrix</i>) Bagi Penilaian Model	72
2.9	Rumusan	75
BAB III	METODOLOGI	
3.1	Pendahuluan	76
3.2	Penentuan Permasalahan	77
3.3	Pengumpulan Data Dan Pra-Pemprosesan	77
3.3.1	Pengumpulan Data	78
3.3.2	Pengiraan Indeks Kesesakan	80
3.3.3	Pengendalian Kehilangan Data	81
3.3.4	Penentuan Kesesakan Trafik	83
3.4	Pembinaan Model	85
3.4.1	Membina Modul Penggugusan	85
3.4.2	Membina Modul Ramalan	97
3.4.3	Modul Ramalan Perambatan Aliran Trafik Jalan Raya Berjiran	101
3.5	Pengujian	103
3.5.1	Matriks Kekeliruan	103

3.5.2	Mengimplementasikan Semula Model Wang Et Al. (2015)	105
3.6	Rumusan	111

BAB IV HASIL DAN PERBINCANGAN

4.1	Pendahuluan	112
4.2	Set Data Eksperimen Kajian	112
4.3	Keputusan Eksperimen Kajian	120
4.3.1	Kaedah Penggugusan Glcm+Spektrum	120
4.3.2	Persilangan Glcm+Spektrum Dan Spektrum Frekuensi	126
4.3.3	Ramalan Perambatan Aliran Trafik Jalan Berjiran	128
4.4	Perbincangan	137
4.5	Rumusan	144

BAB V KESIMPULAN DAN KAJIAN LANJUTAN

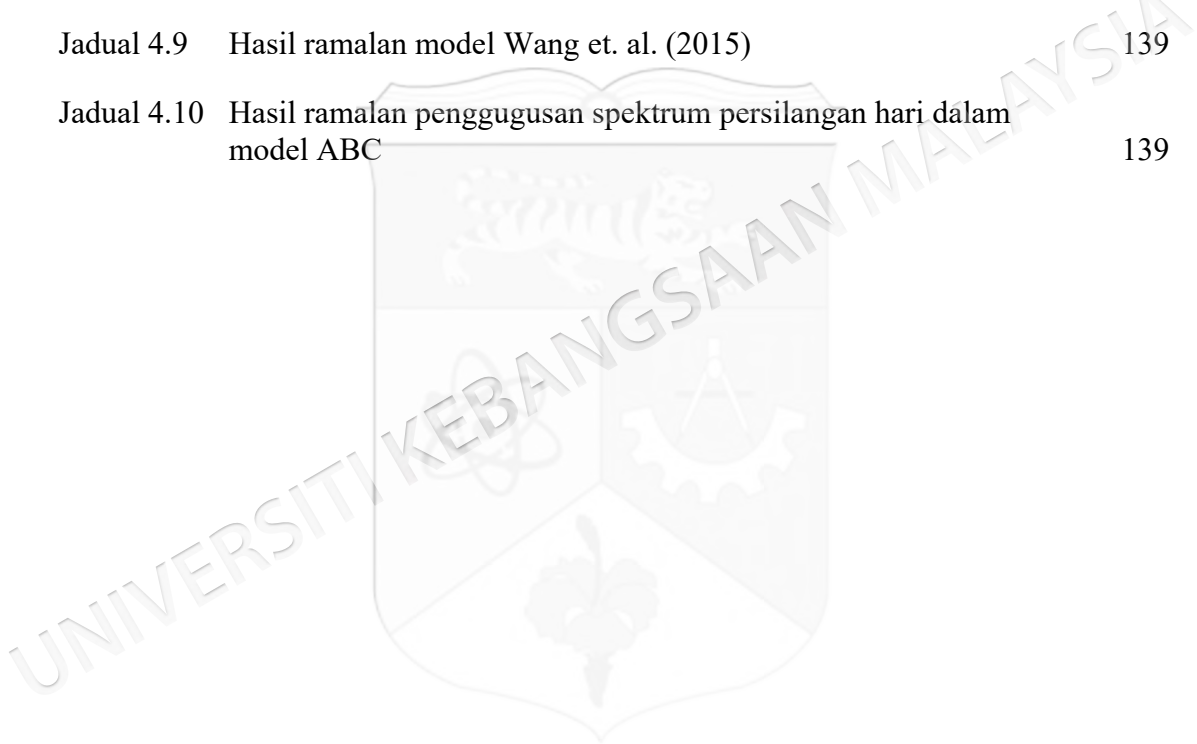
5.1	Pendahuluan	145
5.2	Pencapaian Objektif Kajian	145
5.3	Kekangan	146
5.4	Sumbangan Kajian	147
5.5	Cadangan Kajian Lanjutan	147

RUJUKAN		149
----------------	--	------------

SENARAI JADUAL

No. Jadual		Halaman
Jadual 2.1	Matriks daripada kajian perambatan kesesakan jalan dan kajian ramalan kondisi trafik	25
Jadual 2.2	Contoh nilai daripada pengekstrakan ciri imej	61
Jadual 2.3	Kebarangkalian $P_{qt} + 1_{qt}$ cuaca esok berasaskan cuaca hari ini	65
Jadual 2.4	Matriks keadaan pemerhatian dengan nilai kebarangkalian membawa payung dan tidak membawa payung	66
Jadual 3.1	Contoh data trafik yang diambil daripada pengesan 190100	80
Jadual 3.2	Contoh data trafik yang diambil daripada pengesan 190100	80
Jadual 3.3	Data yang hilang daripada penghitungan indeks kesesakan	82
Jadual 3.4	Pengisian nilai purata kepada data hilang	83
Jadual 3.5	Indeks kesesakan (ci) daripada jalan 158324 dari 05.40 – 09.20 am	84
Jadual 3.6	Pola kesesakan trafik dalam masa dan hari di jalan raya 158536, jalan 173225, dan jalan raya 158536	87
Jadual 3.7	Hasil penggugusan daripada gabungan GLCM+Spektrum dan di jalan 158536	91
Jadual 3.8	Kekerapan kesesakan dalam tempoh 66 hari	92
Jadual 3.9	Hasil penggugusan daripada frekuensi di jalan raya berjiran (158536)	93
Jadual 3.10	Hasil penggugusan bersilang antara frekuensi dan GLCM+Spektrum di jalan 158536	94
Jadual 3.11	Matriks keadaan pemerhatian dengan nilai kebarangkalian emisi pada jalan 158386	98
Jadual 3.12	Contoh nilai keadaan tersembunyi daripada jalan 158386	99
Jadual 3.13	Matriks keadaan pemerhatian dengan nilai kebarangkalian emisi pada jalan 158386	109
Jadual 4.1	Matriks korelasi indeks kesesakan bagi jalan sasaran di kawasan 7	118

Jadual 4.2	Matriks korelasi indeks kesesakan bagi jalan sasaran di kawasan 8	119
Jadual 4.3	Hasil ramalan trafik daripada gugusan GLCM+spektrum	127
Jadual 4.4	Hasil ramalan trafik daripada gugusan spektrum frekuensi.	127
Jadual 4.5	Hasil ramalan trafik daripada gugusan GLCM+spektrum $\cap$ spektrum frekuensi	127
Jadual 4.6	Hasil ramalan bagi jalan sasaran 158565	130
Jadual 4.7	Hasil ramalan bagi jalan sasaran 158565	133
Jadual 4.8	Hasil ramalan model ABC	138
Jadual 4.9	Hasil ramalan model Wang et. al. (2015)	139
Jadual 4.10	Hasil ramalan penggugusan spektrum persilangan hari dalam model ABC	139



SENARAI ILUSTRASI

No. Rajah		Halaman
Rajah 1.1	Keadaan daripada kesesakan jalan	3
Rajah 1.2	Aliran pembinaan model dalam ramalan trafik	8
Rajah 1.3	Langkah-langkah daripada analisis data	10
Rajah 1.4	Garis panduan generik tidak rasmi untuk pemilihan model	11
Rajah 2.1	Pembangunan kesesakan trafik di San Francisco	15
Rajah 2.2	Sorotan maklumat perambatan kesesakan jalan pada jalan berjiran	16
Rajah 2.3	Visual daripada Perambatan trafik	17
Rajah 2.4	Visualisasi sentralitas temporal di Shenzhen	18
Rajah 2.5	Visualisasi analitik perambatan kesesakan trafik	19
Rajah 2.6	Visualisasi laluan perambatan kesesakan di Hangzhou, Zhejiang	19
Rajah 2.7	Visualisasi kesesakan menggunakan peta Google.	20
Rajah 2.8	Visualisasi keadaan trafik di rangkaian jalan Shanghai (Jalan sesak diberikan merah atau merah gelap).	20
Rajah 2.9	Sistem trafik masa nyata daripada ITIS Malaysia.	21
Rajah 2.10	Ramalan perambatan kesesakan jalan	22
Rajah 2.11	Contoh gugusan sub-trajektori untuk dua gugusan ruang yang berbeda.	23
Rajah 2.12	Perambatan kesesakan menggunakan graf	23
Rajah 2.13	Taburan aliran trafik	29
Rajah 2.14	Penggugusan jalan raya dengan K-Means	32
Rajah 2.15	Analisis Data Eksplorasi (Exploratory data analysis EDA) dan kaedah yang berkaitan dalam kajian aliran trafik.	38
Rajah 2.16	Pembelajaran mesin tanpa pengawasan.	41
Rajah 2.17	Pola kesesakan jalan di hari kerja dan hujung minggu	42
Rajah 2.18	Pola kesesakan jalan dalam satu hari.	42

Rajah 2.19	Pola aliran trafik sedang berulang setiap hari	43
Rajah 2.20	Pola aliran trafik di Munich pada waktu pagi dan petang	43
Rajah 2.21	Contoh mencari k terbaik a) sum of squares b) graf <i>silhouette</i>	50
Rajah 2.22	Contoh mencari k terbaik menggunakan lengkung siku	50
Rajah 2.23	Ilustrasi tahap yang berlainan dalam algoritma K-Means	52
Rajah 2.24	(a) Hasil penggugusan k-Means dan (b) Hasil penggugusan Spektrum	54
Rajah 2.25	Contoh pola arah dalam pengenalan imej.	60
Rajah 2.26	Proses penggugusan gambar berasaskan nilai GLCM	61
Rajah 2.27	Contoh nilai daripada pengekstrakan fitur imej	62
Rajah 2.28	Model Markov untuk contoh ramalan cuaca dengan kebarangkalian peralihan keadaan mengikut Jadual 2.3	65
Rajah 2.29	Rantai markov daripada keadaan aliran trafik, menunjukkan keadaan dan peralihan.	68
Rajah 2.30	Kod pseudo daripada algoritma ke hadapan	70
Rajah 2.31	Algoritma viterbi untuk mencari jujukan optimum daripada keadaan tersembunyi	71
Rajah 2.32	Struktur biasa matriks kekeliruan	73
Rajah 2.33	Konsep <i>precision</i> dan <i>recall</i>	75
Rajah 3.1	Metodologi yang dicadangkan.	77
Rajah 3.2	Aliran dalam pengumpulan data dan pra-pemprosesan.	78
Rajah 3.3	Peta daripada pengesan IoT di bandar Aarhus, Denmark	79
Rajah 3.4	Purata kelajuan di jalan raya 158324 setiap lima minit.	84
Rajah 3.5	Aliran dalam pembinaan model.	85
Rajah 3.6	(a) Pola aliran trafik hari kerja , (b) Pola aliran trafik hari hujung minggu	86
Rajah 3.7	Carta Alir Kajian Prestasi GLCM.	88
Rajah 3.8	Penentuan nilai k.	90

Rajah 3.9	Kekerapan kesesakan di kawasan jiran jalan 158536 pada masa yang berbeza.	92
Rajah 3.10	Carta Alir Penggugusan	95
Rajah 3.11	Pseudocode modul penggugusan.	96
Rajah 3.12	Lengkung siku di semua lokasi	97
Rajah 3.13	Penggugusan aras kesesakan dan masa.	98
Rajah 3.14	Pengiraan keadaan trafik menggunakan HMM.	100
Rajah 3.15	Modul HMM(Aras Kesesakan(JB(Masa)+S)).	101
Rajah 3.16	Model Ramalan Perambatan Aliran Trafik Jalan Berjiran	102
Rajah 3.17	Matriks kekeliruan (<i>confusion matrix</i>)	103
Rajah 3.18	Matriks kekeliruan (<i>confusion matrix</i>) Jalan 158536	104
Rajah 3.19	Carta alir penggugusan spektrum persilangan hari.	107
Rajah 3.20	Lengkung siku di semua lokasi	108
Rajah 3.21	Penggugusan aras kesesakan dan masa.	109
Rajah 3.22	Implementasi semula Model Ramalan Trafik Jalan Berjiran Wang et al. (2015)	110
Rajah 4.1	Peta lokasi eksperimen.	113
Rajah 4.2	Sembilan (9) segmentasi kawasan.	114
Rajah 4.3	Nilai korelasi antara jalan pada tujuh (7) Kawasan.	116
Rajah 4.4	Peta empat lokasi eksperimen di bandar Aarhus, Denmark.	120
Rajah 4.5	Carta bara daripada ramalan keadaan trafik jalan raya berjiran 158324.	121
Rajah 4.6	Hasil daripada tiga kaedah penggugusan daripada jalan raya 158324.	122
Rajah 4.7	Carta bara daripada ramalan keadaan trafik jalan raya berjiran 158536.	122
Rajah 4.8	Hasil penggugusan daripada jalan raya 158536.	123
Rajah 4.9	Carta bara daripada ramalan keadaan trafik jalan raya berjiran 158594.	123

Rajah 4.10	Carta bara daripada ramalan keadaan trafik jalan raya berjiran 158954.	124
Rajah 4.11	Carta bara daripada ramalan keadaan trafik jalan raya berjiran 158386.	125
Rajah 4.12	Carta bara daripada ramalan keadaan trafik jalan raya berjiran 158386.	125
Rajah 4.13	Jalan berjiran hasil penggugusan (a), (b) dan (c).	126
Rajah 4.14	Jalan berjiran hasil penggugusan (a), (b) dan (c).	128
Rajah 4.15	Carta garis perbandingan model ramalan ABC dan Wang et. al. (2015).	131
Rajah 4.16	Carta garis perbandingan penggugusan dicadangkan dan penggugusan spektrum persilangan hari, kedua-dua menggunakan HMM dicadangkan.	132
Rajah 4.17	Visualisasi jalan berjiran pada lokasi atau jalan sasaran 158565.	133
Rajah 4.18	Visualisasi jalan berjiran pada lokasi atau jalan sasaran 158565.	137
Rajah 4.19	Carta garis perbandingan ketepatan antara model ramalan.	140
Rajah 4.20	Carta garis perbandingan <i>precision</i> antara model ramalan.	141
Rajah 4.21	Carta garis perbandingan <i>recall</i> antara model ramalan.	142
Rajah 4.22	Carta garis perbandingan skor F1 antara model ramalan.	143
Rajah 4.23	Carta garis perbandingan model ramalan berdasarkan purata pengukuran.	144

SENARAI SINGKATAN

ATIS	<i>Advanced Traveller Information System</i>
CAV	<i>Connected and Autonomous Vehicles</i>
CCTV	<i>Closed Circuit Television</i>
CI	<i>Congestion Index</i>
DB	<i>Davies–Bouldin</i>
DBSCAN	<i>Density-Based Spatial Clustering of Applications with Noise</i>
EDA	<i>Exploratory Data Analysis</i>
FCD	<i>Floating Car Data</i>
FN	<i>False Negative</i>
FP	<i>False Positive</i>
GIGO	<i>Garbage-In-Garbage-Out</i>
GIS	<i>Geographic Information System</i>
GLCM	<i>Gray level of Co-occurrence Matrix</i>
GPS	<i>Global Positioning System</i>
HMM	<i>Hidden Markov Model</i>
ICT	<i>Information and Communication Technologies</i>
IMS	<i>Integrated Information System</i>
IoT	<i>Internet of Things</i>
ITIS	<i>Integrated Transport Information System</i>
ITS	<i>Intelligent Transportation Systems</i>
KNN	<i>K-Nearest Neighbour</i>
LIDAR	<i>Light detection and ranging</i>
ML	<i>Machine Learning</i>
NN	<i>Neural Network</i>
SDN	<i>Software Defined Networking</i>

SSE	<i>Sum of Squared Errors</i>
TN	<i>True Negative</i>
TP	<i>True Positive</i>
VANET	<i>Vehicle Ad Hoc Network</i>
VIDS	<i>Video Image Detection Systems</i>
VMS	<i>Variable Message Signs</i>



BAB I

PENDAHULUAN

1.1 LATAR BELAKANG PENYELIDIKAN

Peningkatan jumlah penduduk menyebabkan peningkatan penggunaan kereta dalam era moden ini, seterusnya menyebabkan kesesakan trafik menjadi lebih teruk. Kerugian daripada kesesakan trafik bukan sahaja menyebabkan kerugian ekonomi (G. Wang & Kim 2016), tetapi juga menyebabkan kerosakan alam sekitar (Jiang et al. 2017), ia membazirkan masa dan tenaga, (Xiaomeng Wang et al. 2015), stres manusia dan pencemaran alam sekitar (Petrovska & Stevanovic 2015) (W. Hu et al. 2014). Bagi mengatasi masalah ini, pemandu memerlukan maklumat trafik yang boleh menjejaskan pemanduan akibat kesesakan trafik, seperti mengubah tabiat pemanduan mereka dan laluan memandu, (G. Zhu et al. 2016) dan ia menyebabkan perubahan dalam rantai trafik di bahagian hulu dan hilir jalan yang saling berhubung dan juga di bahagian lain jalan yang berada dalam rangkaian. Memantau dan memahami kesesakan trafik adalah sangat sukar (Jiang et al. 2017) di mana salah satu kerumitan kesesakan trafik adalah dinamik dan saling berkaitan. Kesesakan trafik merambat dari satu jalan ke jalan berjiran (Z. Wang et al. 2013). Jalan berjiran adalah jalan yang berdekatan, terbagi menjadi jalan yang saling berhubung dan jalan yang tidak berhubung. Jalan berhubung adalah jalan yang berhubungan langsung.

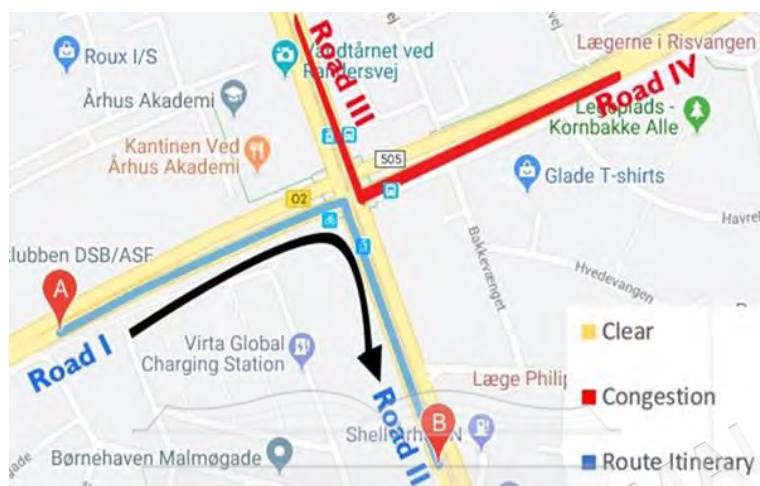
Kebanyakan kajian dalam aliran trafik melibatkan pengvisualan. Pengvisualan digunakan untuk menganalisis kesesakan trafik dan perambatan aliran trafik. Sebagai contoh, beberapa kajian telah dijalankan untuk menganalisis ruang (spatial) dan temporal, seperti kajian terhadap trajektori GPS menggunakan pengvisualan data trajektori masa nyata (Tominski et al. 2012) (Jianqin et al. 2015) (Bast et al. 2014) (Z. Wang et al. 2013) (Huang et al. 2015). Kebanyakan kajian pengvisualan trafik masa

nyata tidak meramalkan kesan keadaan trafik di jalan berjiran. Walau bagaimanapun ramalan perambatan trafik sangatlah diperlukan, hasil ramalan perambatan boleh digunakan untuk mengurus, menganalisis trafik dan untuk mengelakkan kesesakan trafik (meminimumkan terjejas teruk). Terdapat beberapa kajian yang terlibat dalam meramal perambatan kesesakan trafik. Kajian oleh (Xiong & Vahedian 2018) (Z. Chen et al. 2018) (Nguyen et al. 2016a), menggunakan pendekatan graf untuk menganggarkan perambatan kesesakan. Satu lagi kajian oleh (Keler et al. 2017) garis poli berkelompok (*over polyline*) digunakan untuk mengesan kesesakan. Semua kajian ini menggunakan sub-graf yang kerap untuk mencari punca perambatan yang kerap dan semuanya menggunakan data trajektori GPS untuk mengenal pasti laluan perjalanan dan jalan yang berhubung. Kajian lain cuba meramal perambatan dengan berkelompok secara dinamik (gugusan semula) berdasarkan masa. Termasuk penyelidikan menggunakan pendekatan graf oleh (Besenczi et al. 2020) (Z. Chen et al. 2018) (Nguyen et al. 2016a), dan trajektori perjalanan oleh (Ye et al. 2016).

Kebanyakan kajian perambatan kesesakan jalan raya memerlukan set data yang mempunyai maklumat mengenai jalan yang berhubung dan arah trafik. Penyelidikan ini berfokus kepada meramal perambatan kesesakan trafik jalan raya berjiran yang terjejas oleh kesesakan jalan raya menggunakan set data yang tidak memiliki maklumat mengenai jalan yang berhubung dan arah trafik.

1.2 PERMASALAHAN KAJIAN

Pengurusan trafik yang proaktif memerlukan keupayaan untuk meramal aliran trafik. Salah satu tanggungjawab utama Sistem Pengurusan Trafik Termaju (ATMS - *Advanced Traffic Management System*) adalah meramal keadaan trafik dalam jarak pendek hingga jarak pertengahan (Smith et al. 1994). Ramalan ini boleh digunakan untuk menguruskan trafik dan untuk mengelakkan kesesakan trafik dan meminimumkan kesannya. Sebagai contoh, perjalanan dari titik A ke titik B menggunakan jalan I dan jalan II boleh dipengaruhi oleh keadaan trafik di jalan III dan jalan IV. Keadaan ini ditunjukkan dalam Rajah 1.1.



Rajah 1.1 Keadaan daripada kesesakan jalan

Berdasarkan Rajah 1.1, pemandu menjangkakan aliran trafik di jalan III atau jalan IV juga akan memberi kesan kepada aliran trafik jalan berjiran seperti jalan I dan jalan II. Menurut Wang et al. (2016), dalam 20 minit kesesakan boleh merambat sekitar 5 km, kerana itu kajian ini menjalankan penyelidikan untuk meramal keadaan kesesakan trafik jalan berjiran yang terjejas lima (5) km akibat kesesakan daripada jalan sasaran.

Kajian yang berkaitan dengan penyelidikan ini adalah kajian perambatan kesesakan aliran trafik yang menggunakan pendekatan graf oleh (Besenczi et al. 2020)(Z. Chen et al. 2018)(Nguyen et al. 2016a),(Ye et al. 2016). Pendekatan graf perlu membina graf yang dihubungkan dan sub-graf yang saling berhubung secara rekursif semasa proses ramalan kesesakan jalan raya. Walau bagaimanapun pendekatan graf memerlukan set data yang mempunyai maklumat mengenai jalan yang saling berhubung dan arah trafik.

Selain daripada kaedah graf, kaedah lain yang boleh digunakan adalah membuat persamaan matematik dan kaedah kebarangkalian. Masalah yang dihadapi dengan pendekatan persamaan matematik adalah kesesakan aliran trafik tidak dapat diramalkan. Perkara ini menyebabkan pendekatan persamaan matematik untuk

meramal aliran trafik adalah sangat sukar. Oleh itu, pendekatan yang sesuai untuk meramal kesesakan trafik adalah menggunakan kaedah statistik atau kebarangkalian (Rapant et al. 2016). Kaedah statistik digunakan untuk membezakan keadaan trafik antara jalan. Pelbagai penyelidikan lepas banyak menggunakan kaedah statistik model Markov tersembunyi (HMM) untuk meramalkan aliran trafik, diantaranya kajian oleh (Qi & Ishak 2014), (Xiaomeng Wang et al. 2015), (G. Zhu et al. 2016), (Rapant et al. 2016), (Zaki et al. 2019). Kajian terdahulu yang menggunakan HMM untuk meramal keadaan kesesakan trafik mempertimbangkan meramal kesesakan trafik berdasarkan aras kesesakan (Qi & Ishak 2014), sementara itu kajian oleh (Xiaomeng Wang et al. 2015) mempertimbangkan aras kesesakan masa jalan berjiran sebagai emisi (*emission*), sedangkan penyelidikan ini meramal kesesakan mempertimbangkan aras kesesakan masa jalan berjiran dan jalan sasaran.. Kaedah statistik kontras digunakan oleh (Qi & Ishak 2014) (Zaki et al. 2019) untuk mengenalpasti keadaan trafik. Manakala kaedah statistik kontras digunakan dalam kajian ini untuk mengelompokkan jalan berjiran berdasarkan hari dan masa.

Penggugusan (*clustering*) kebiasaanya mengelompokkan jalan ke dalam satu gugusan berdasarkan data kelajuan atau ketumpatan kenderaan. Terdapat beberapa kajian terdahulu mengguguskan jalan berjiran menggunakan K-Means, iaitu kajian penggugusan jalan raya berasaskan ketumpatan kenderaan dan kelajuan oleh (Mondal & Rehena 2019), dan (Necula 2015). Seterusnya kajian oleh (Xiaomeng Wang et al. 2015) menggunakan kaedah spektrum yang berdasarkan data kelajuan. Didapati kajian sebelum ini hanya menggunakan satu parameter pada satu masa, iaitu apabila pengelompokkan jalan raya menggunakan data kelajuan maka jalan raya akan dikelompokkan berdasarkan persamaan data kelajuan dan apabila menggunakan data ketumpatan kenderaan, maka jalan raya akan dikelompokkan berdasarkan persamaan data ketumpatan kenderaan. Ini menyebabkan ketepatan ramalan aliran trafik rendah kerana menurut (El Sherief et al. 2016) aliran trafik sebenarnya dipengaruhi oleh data kelajuan dan ketumpatan kenderaan, malah menurut (W. X. Wang et al. 2018) aliran trafik juga dipengaruhi oleh nilai normal waktu pemanduan.

Penyelidik menggunakan visualisasi untuk menunjukkan kesesakan aliran trafik pada satu jalan tertentu (El Sherief et al. 2016) (W. X. Wang et al. 2018). Walau

bagaimanapun visualisasi tersebut tidak menunjukkan jalan raya yang berjiran ataupun saling berhubung. Seterusnya kajian ini menggunakan set data yang tiada maklumat mengenai jalan yang bersambung dan arah kenderaan untuk memplotkan visualisasi pada peta diperlukan bagi menggambarkan jalan yang saling berhubung atau tidak berhubung.

Dalam kajian ini, pendekatan yang berbeza digunakan iaitu pengelompokan tidak berdasarkan data kelajuan atau ketumpatan kenderaan, malah jalan yang memiliki pola aliran trafik yang sama dikelompokkan menggunakan kaedah penggugusan berdasarkan keadaan aliran trafik. Nilai statistik kontras keadaan trafik jalan raya dalam hari dan waktu yang sama di ekstrak untuk mencari pola trafik pada hari dan masa. Penggugusan berdasarkan nilai statistik kontras ini terbukti meningkatkan ketepatan dalam pengelompokan imej seperti kajian oleh (Oleiwi, 2017), (M.-W. Lin et al. 2007) (X. Zhang et al. 2008)(Yun & Shu 2011). Dalam penyelidikan ini model Markov tersembunyi (HMM) digunakan untuk meramal aliran trafik jalan berjiran yang terjejas akibat kesesakan jalan sasaran pada set data tiada maklumat jalan berhubung dan arah trafik. Dalam kajian ini aras kesesakan, masa dan kesesakan jalan sasaran digunakan sebagai pemboleh ubah untuk meramal perambatan trafik dengan set data tanpa maklumat tentang jalan yang berhubung dan arah trafik.

1.3 OBJEKTIF PENYELIDIKAN

Objektif utama kajian ini adalah untuk meramal perambatan aliran trafik jalan berjiran akibat daripada kesesakan jalan sasaran dengan menggunakan set data yang tidak mempunyai maklumat tentang jalan berhubung. Berikut adalah objektif kajian yang terperinci iaitu:

1. Mengenal pasti jalan berjiran dengan menganalisis kesesakan trafik aliran trafik jalan berjiran tanpa maklumat tentang jalan berhubung menggunakan kaedah penggugusan persilangan GLCM+Spektrum dan Spektrum Frekuensi;
2. Memvisualisasikan hubungan kesesakan trafik antara jalan berjiran pada peta lokasi;

3. Mereka bentuk model ramalan perambatan kesesakan trafik jalan berjiran berdasarkan analisis ruang (*spatial*) dan temporal dari jalan raya sasaran dengan menggunakan model Markov tersembunyi (HMM);

1.4 SOALAN KAJIAN

Objektif utama dalam penyelidikan ini boleh dicapai dengan merujuk kepada beberapa soalan kajian terhadap perincian objektif seperti berikut:

1. Mengenal pasti jalan berjiran dengan menganalisis kesesakan trafik aliran trafik jalan berjiran tanpa maklumat tentang jalan berhubung menggunakan kaedah penggugusan persilangan GLCM+Spektrum dan Spektrum Frekuensi. Objektif ini dicapai dengan menangani soalan-soalan berikut:
 - a. Bagaimana pola arus trafik dalam hari dan masa ?
 - b. Bagaimana kesamaan pola aliran trafik antara jalan raya berjiran?
 - c. Bagaimana menentukan kaedah penggugusan (*clustering*) yang sesuai untuk menganalisis hubungan antara jalan raya?
2. Memvisualisasikan hubungan kesesakan trafik antara jalan berjiran pada peta lokasi. Objektif ini dicapai dengan menangani soalan berikut:
 - a. Bagaimana memvisualisasikan hubungan antara jalan raya pada peta?
 - b. Bagaimana memvisualisasikan ramalan perambatan aliran trafik yang terkesan daripada jalan yang sesak ?
3. Mereka bentuk model ramalan perambatan kesesakan trafik jalan berjiran berdasarkan analisis ruang (*spatial*) dan temporal dari jalan raya sasaran dengan menggunakan model Markov tersembunyi (HMM). Objektif ini dicapai dengan menangani soalan-soalan berikut:
 - a. Bagaimana menentukan kesesakan trafik?
 - b. Bagaimana menentukan aliran pemerhatian (*observation state*) untuk meramal aliran trafik menggunakan Model Hidden Markov?

1.5 SKOP KAJIAN

Menurut kajian oleh (Keler et al. 2017) dan (Akoglu 2018) bahawa jalan raya memiliki hubungan yang kuat apabila memiliki nilai korelasi pearson lebih dari 0.6. Kajian ini memberi tumpuan kepada penyelidikan untuk meramal kesesakan jalan berjiran yang disebabkan oleh kesesakan pada jalan sasaran. Sebab itu jalan raya harus memiliki hubungan kuat korelasi antara satu sama lain, atau suatu jalan memberi kesan kepada jalan lainnya. Oleh itu dalam penyelidikan ini ketepatan model ramalan yang dicadangkan dalam kajian bergantung kepada nilai purata korelasi aliran trafik pada jalan sasaran dan jalan berjiran iaitu melebihi 0.6.

Selain itu kajian lain menerangkan bahawa kesesakan berlaku dalam masa 0.5 jam hingga 2 jam (Susilo & Imanuel 2018)(Christidis & Ibanes Rivas 2012), dengan demikian median masa kesesakan adalah 80 minit. Oleh itu, menentukan minimum frekuensi kesesakan 80 minit perhari juga diperlukan untuk meningkatkan ketepatan ramalan perambatan aliran trafik jalan berjiran. Kerana dalam kajian ini indeks kesesakan dikira setiap 20 minit, maka untuk 80 minit tempo kesesakan akan berlaku 4 kesesakan. Sehingga untuk set data dengan tempoh masa 3 bulan ditentukan jumlah frekuensi kesesakan ialah 192. Selain itu kajian ini menetapkan batas keadaan sesak adalah apabila indeks kesesakan jalan raya melebihi 3.03 seperti diterangkan pada Bab 3.3.4 Penentuan kesesakan trafik.

1.6 SUMBANGAN KAJIAN

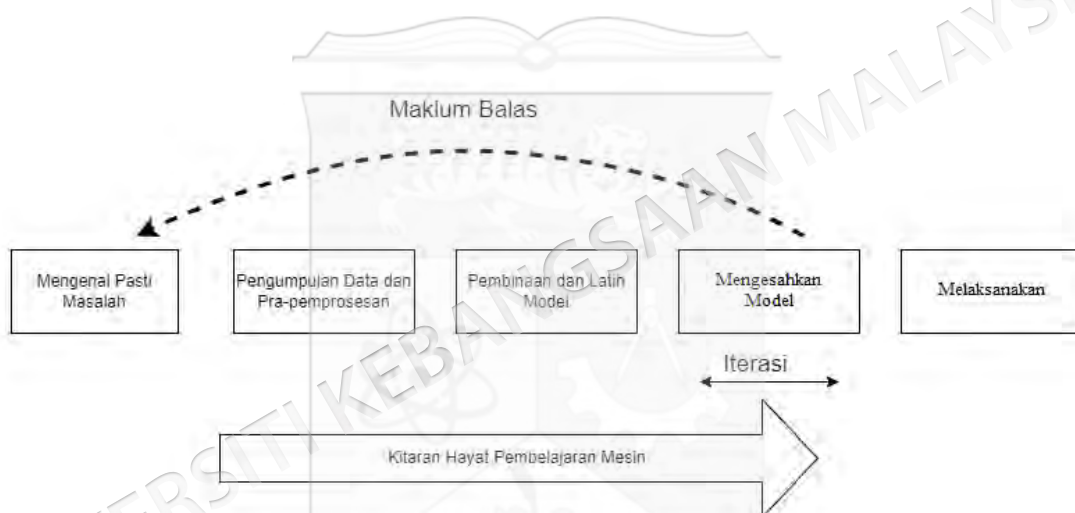
Kajian ini memperkenalkan modul penggugusan untuk mengelompokkan jalan dengan berasaskan pola kesesakan trafik pada hari dan masa. Di mana kebanyakan penyelidik sebelum ini melaksanakan penggugusan jalan berdasarkan data kelajuan dan ketumpatan kenderaan atau bilangan kenderaan di jalan raya.

Model yang dicadangkan dapat dilaksanakan dalam aplikasi mudah alih untuk memberikan pemandu maklumat trafik yang boleh menjejaskan pemanduan dari kesesakan trafik

Model yang dicadangkan dapat digunakan untuk untuk set data yang tidak mempunyai maklumat arah dan hubungan antara pembolehubah. Sebagai contoh untuk ramalan cuaca berasaskan pola cuaca pada hari, masa dan lokasi yang berbeza.

1.7 METODOLOGI PENYELIDIKAN

Secara umumnya, penyelidikan yang berkaitan dengan saintis data akan melalui beberapa langkah seperti yang ditunjukkan pada Rajah 1.2 (Rao 2019). Dengan itu, kajian ini juga melalui langkah tersebut dalam meramal perambatan aliran trafik berjiran seperti beberapa kajian terdahulu yang dijalankan oleh (Jain et al. 2017),(Lopes et al. 2010), (Liang & Wakahara 2014).



Rajah 1.2 Aliran pembinaan model dalam ramalan trafik

Setiap langkah dari aliran pada Rajah 1.2 dijelaskan seperti berikut:

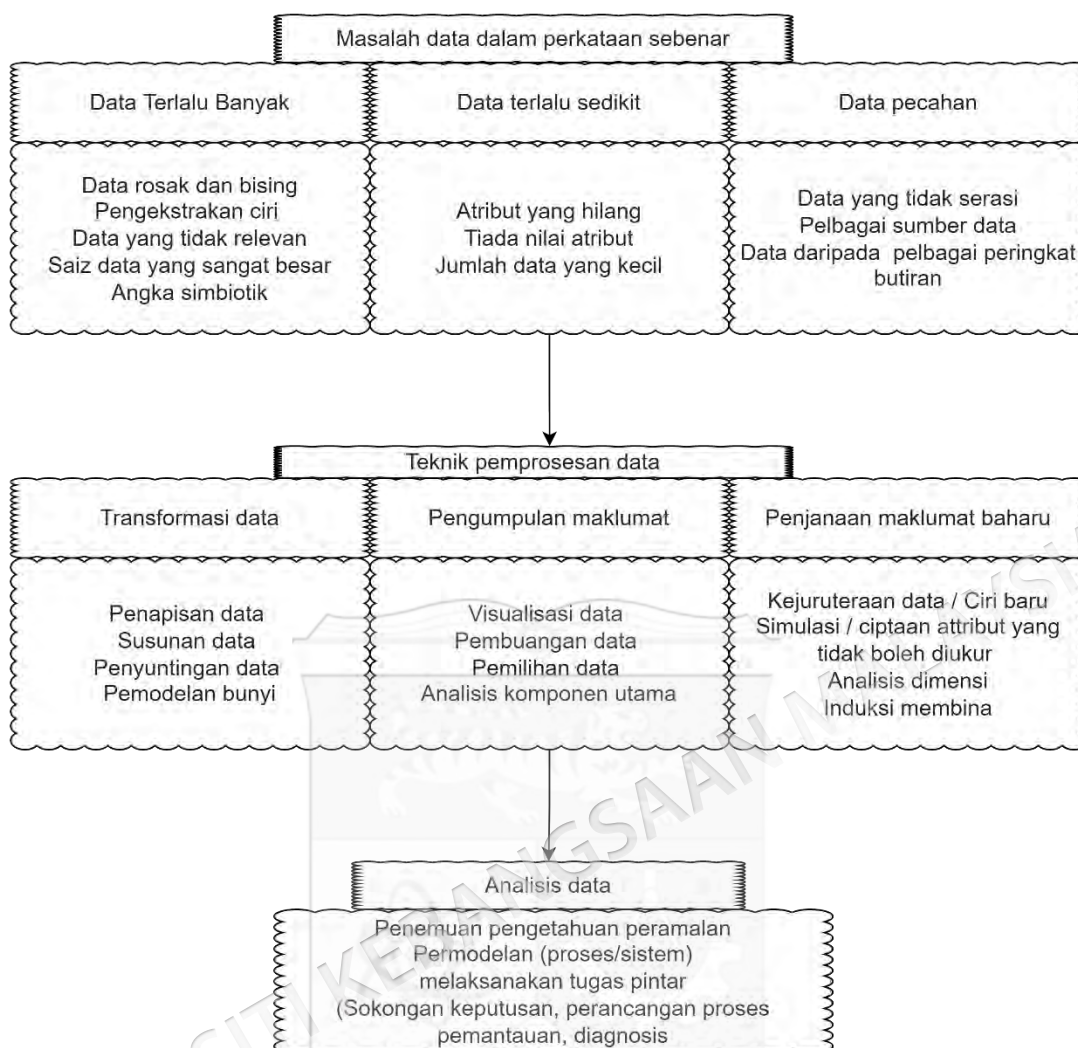
Mengenal pasti masalah

Langkah ini merupakan langkah pertama dalam memulakan penyelidikan untuk memantapkan Kebenaran Asas (*Ground Truth*) bagi masalah yang telah dikenal pasti sebelum beralih ke langkah seterusnya iaitu mengumpul dan memproses data. Langkah ini akan mentakrifkan masalah yang perlu diselesaikan dengan jelas. Dalam kajian sistem pengangkutan pintar, khususnya meramal aliran trafik, penentuan aras kesesakan jalan raya adalah penting (Afrin & Yodo 2021). Pelbagai definisi kesesakan trafik telah ditentukan dalam kajian terdahulu. Rothenberg menyatakan

kesesakan trafik adalah keadaan di mana bilangan kenderaan di jalan raya melebihi kapasiti maksimum perkhidmatan piawai sesuatu jalan (Yingya Zhang et al. 2016). Dalam kajian ini kesesakan trafik dikenal pasti dengan mendapatkan indeks kesesakan iaitu mengambil kira tahap kepadatan kenderaan, kelajuan dan gabungan keduanya (W. X. Wang et al. 2018).

Pengumpulan dan pra-pemprosesan data

Langkah ini mengumpul, membersihkan dan menyediakan data yang dikenali sebagai peringkat pra-pemprosesan data. Dalam kajian ini, sumber data dikumpul daripada alat pengesan IoT, GPS, video dan VANET. Walau bagaimana pun, ada ketika di mana alat pengesan IoT atau pautan komunikasi tidak berfungsi dengan baik atau gagal menghasilkan nilai yang tepat disebabkan oleh kegagalan kuasa, gangguan, servis peranti, kerja di permukaan jalan. Ini mengakibatkan data rosak atau hilang (Bellemans et al. 2000). Masalah data ialah situasi yang menghalang penggunaan mana-mana alat analisis data secara cekap atau yang boleh mengakibatkan hasil yang tidak boleh diterima (Famili et al. 1997). Pra-pemprosesan data ialah langkah yang sangat penting untuk mendapatkan set data yang bersih dan boleh digunakan untuk latihan model ramalan yang dicadangkan. Pelbagai teknik pra-pemprosesan data dibangunkan untuk menyelesaikan masalah data. Rajah 1.3 menunjukkan set data yang telah dikumpul yang mempunyai pelbagai masalah data dan bagaimana pra-pemprosesan dilaksanakan sebelum analisis data.

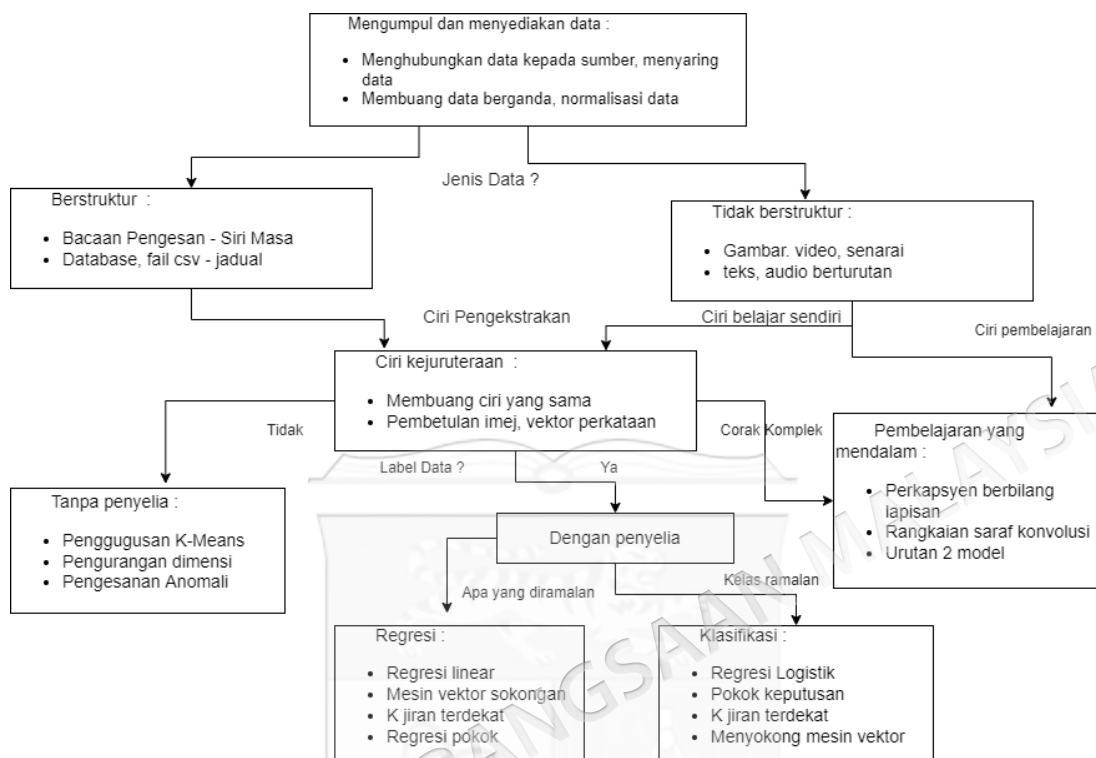


Rajah 1.3 Langkah-langkah daripada analisis data (Rao 2019).

Membina dan latihan model

Dalam perlaksanaan langkah ini, pemahaman tentang representasi model diperlukan iaitu model diterangkan sebagai hubungan antara output atau pemboleh ubah tindak balas dan input yang sepadan atau pemboleh ubah bebas dalam set data. Dalam perlaksanaan langkah ini, pemahaman tentang representasi model diperlukan iaitu model diterangkan sebagai hubungan antara output atau pemboleh ubah tindak balas dan input yang sepadan atau pemboleh ubah bebas dalam set data (Rao 2019) diaplikasikan dalam latihan model. Langkah seterusnya ialah membangunkan model dan latihan. Rajah 1.4 menunjukkan strategi peringkat tinggi yang terbaik diamalkan untuk memilih model yang sesuai. Dalam kajian ini, jenis data dikenal pasti dan

seterusnya pemilihan jalan berjiran dijalankan menggunakan daripada kaedah penggugusan.



Rajah 1.4 – Garis panduan generik tidak rasmi untuk pemilihan model (Rao 2019).

Pengesahan model

Pengesahan model yang digunakan dalam model ramalan yang dijalankan adalah merupakan langkah penting daripada keseluruhan pembinaan model Pembelajaran Mesin (*Machine Learning*). Matrik penentuan model sangat bergantung kepada jenis model yang dipilih, sebagai contoh, matrik untuk model regresi akan berbeza daripada model klasifikasi atau model penggugusan (Rao 2019). Dalam kajian ini, matrix kekeliruan (*confusion matrix*) digunakan untuk menilai prestasi model ramalan yang dicadangkan.

1.8 STRUKTUR TESIS

Tesis ini mengandungi lima (5) bab. Berikut adalah ringkasan penerangan bagi setiap bab:

1.8.1 Bab I

Bab ini menyediakan ulasan latar belakang penyelidikan, pernyataan masalah, objektif penyelidikan dan persoalan kajian. Selain daripada itu bab ini juga menerangkan skop kajian dan sumbangan yang dihasilkan daripada penyelidikan.

1.8.2 Bab II

Bab ini menyediakan ulasan literatur yang lengkap mengenai kerja penyelidikan yang berkaitan dengan kajian ini. Topik yang dikaji termasuk analisis trafik masa nyata, perambatan aliran trafik, faktor ruang dan masa dalam ramalan, hubungan spatial antara jalan raya, data raya dalam peramalan trafik, pembelajaran mesin, mengenalpasti kesesakan jalan, pola aliran trafik, penggugusan jalan menggunakan k-means, penggugusan jalan menggunakan kaedah spektrum, kaedah GLCM, penggabungan kaedah GLCM dan kaedah penggugusan, dan diakhiri dengan asas model HMM.

1.8.3 Bab III

Bab ini menerangkan aliran kajian yang dijalankan berdasarkan metodologi penyelidikan. Bab ini juga memberi penerangan yang terperinci tentang pengumpulan dan ciri-ciri set data, pra-pemprosesan data, pembentukan modul penggugusan dan modul ramalan. Penerangan tentang pembinaan model ramalan perambatan aliran trafik jalan berjiran dan penilaian model mengakhiri bab III.

1.8.4 Bab IV

Bab ini membincangkan dan menganalisis hasil eksperimen kajian pada setiap peringkat implementasi pembinaan model ramalan perambatan aliran trafik jalan berjiran. Peningkatan ketepatan ramalan daripada model ramalan yang dicadangkan juga diterangkan dan dibincangkan dengan lebih terperinci.

1.8.5 Bab V

Akhir sekali, Bab V akan merumuskan penemuan dan sumbangan utama kajian ini. Bab ini juga membincangkan peluang penyelidikan masa hadapan yang merupakan lanjutan daripada dapatan dan kekangan kajian ini.

1.8.6 RUMUSAN

Bab awal ini telah memberi penjelasan mengenai perkara-perkara yang melatari kajian penyelidikan. Seterusnya penjelasan berkaitan dengan permasalahan kajian, objektif kajian, soalan kajian, dan sumbangan kajian juga dijelaskan secara terperinci. Bab awal ini juga memberi penjelasan mengenai metodologi penyelidikan yang merangkumi proses daripada menentukan masalah, membina melatih model, dan akhir sekali mengesahkan model.

BAB II

KAJIAN LITERATUR

2.1 PENDAHULUAN

Tujuan utama kajian adalah melaksanakan penyelidikan yang berkaitan dengan perambatan aliran trafik kepada jalan berjiran yang dipengaruhi oleh kesesakan jalan raya tanpa mempunyai maklumat tentang jalan berhubung dan arah trafik. Bab ini akan mengupas penyelidikan, terminologi, kaedah dan pendekatan yang berkaitan dengan kajian ini. Dalam kajian ini, aliran trafik adalah pergerakan para pemandu individu dan kenderaan yang saling berinteraksi antara satu sama lain (Daganzo 1997). Manakala, perambatan aliran trafik adalah proses beransur-ansur merebaknya kesesakan dalam rangkaian jalan. Rangkaian penyebaran kesesakan menyatakan kesesakan menyerebak dari satu (jalan) segmen ke jalan yang berjiran (Nguyen et al. 2016b). Kajian terdahulu tentang perambatan aliran trafik akan dibincangkan dalam seksyen berikutnya, ini diikuti dengan penerangan tentang ramalan aliran trafik yang dipengaruhi oleh ruang dan masa iaitu dua pemboleh ubah yang digunakan dalam kajian ini. Kajian aliran trafik melibatkan penganalisan data raya (*big data*) dari pelbagai sumber data seperti GPS, pengesan IoT dan video. Bab ini juga menjelaskan bagaimana menentukan kesesakan jalan raya. Asas modul penggugusan dan modul ramalan yang digunakan dalam kajian ini juga dijelaskan.

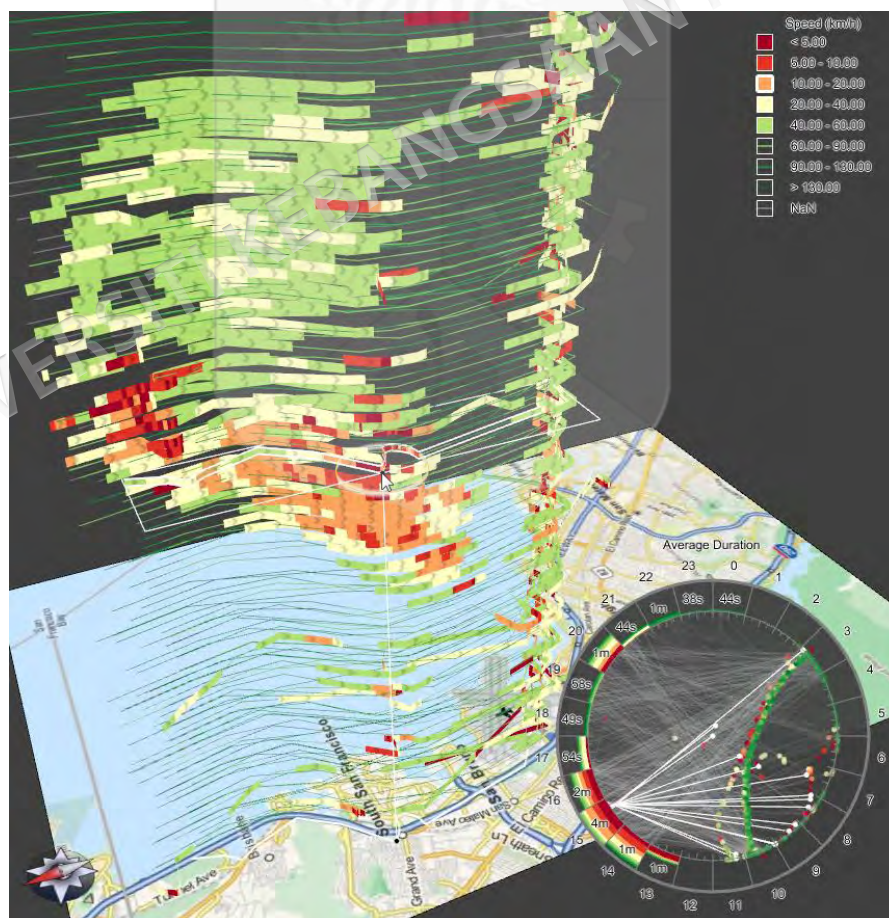
2.2 PERAMBATAN ALIRAN TRAFIK

Bahagian ini membincangkan kajian lepas dalam perambatan aliran trafik. Berasaskan pemerhatian dari kajian literatur, penganalisis perambatan aliran trafik boleh dibahagikan kepada dua pendekatan. Pendekatan pertama menggunakan penggambaran untuk menganalisis aliran trafik (peta, histogram, graf), dan

pendekatan lain melakukan ramalan dengan mempelajari pola perambatan aliran trafik berasaskan data sebelumnya.

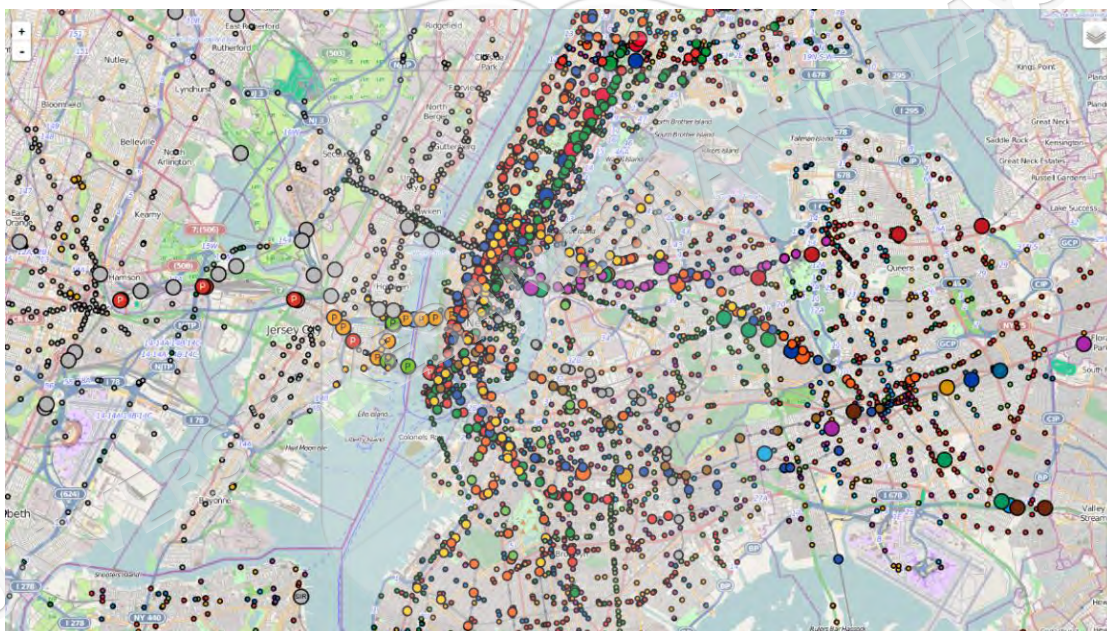
2.3 ANALISIS TRAFIK MASA NYATA

Visualisasi adalah kaedah yang sering digunakan untuk menganalisis trafik masa nyata, kerana mata manusia adalah penganalisis terbaik. Terdapat banyak aktiviti dalam menganalisis aliran trafik yang melibatkan analisis trafik, kesesakan trafik, dan kesan perambatan trafik. Sebelum membincangkan kajian perambatan trafik, terdapat banyak teknik untuk memvisualkan trajektori GPS, kerana maklumat temporal. Kajian oleh (Tominski et al. 2012) menggunakan kiub masa ruang(space-time cube) untuk GPS. Visualisasi trajektori dengan z axis diterokai untuk perwakilan masa supaya dinding trajektori GPS diperolehi, seperti digambarkan dalam Rajah 2.1.



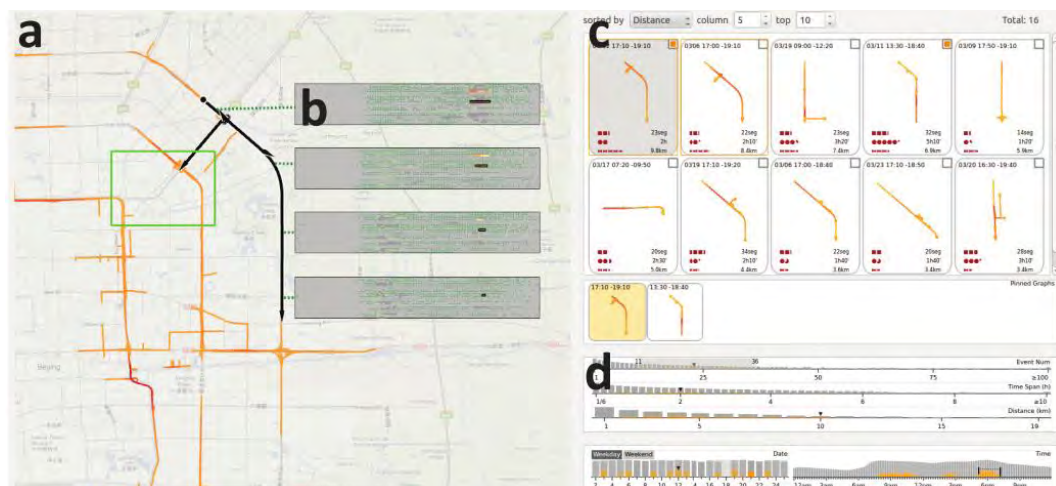
Rajah 2.1 Pembangunan kesesakan trafik di San Francisco (Tominski et al. 2012)

Kajian lain menggunakan peralihan animasi untuk mewakili maklumat masa ruang masa nyata. Kajian oleh (Bast et al. 2014) menganalisis trafik masa nyata dari General Transit Feed Specification (GTFS). GTFS adalah spesifikasi data yang membolehkan agensi transit awam menerbitkan data transit mereka dalam format yang boleh dimakan oleh pelbagai aplikasi perisian. Agensi transit boleh menghasilkan suapan GTFS untuk berkongsi maklumat transit awam mereka dengan pemaju, yang menulis alat yang menggunakan suapan GTFS untuk memasukkan maklumat transit awam ke dalam aplikasi mereka. Hari ini, format data GTFS digunakan oleh beribu-ribu penyedia pengangkutan awam (Bast et al. 2014). Contoh analisis trafik masa nyata dari GTFS digambarkan dalam Rajah 2.2.



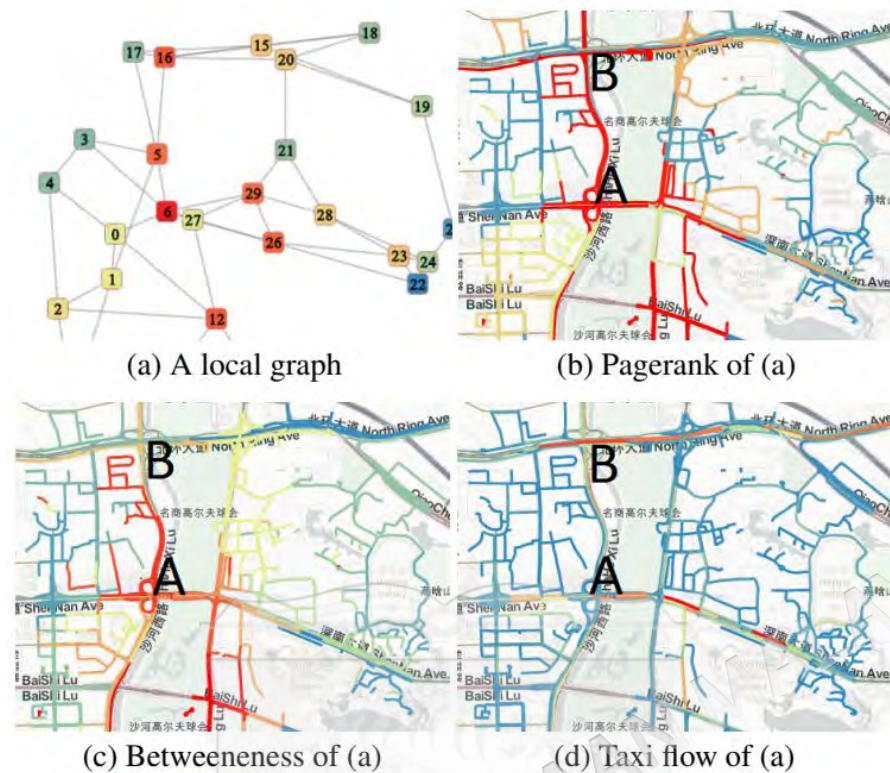
Rajah 2.2 Sorotan maklumat perambatan kesesakan jalan pada jalan berjiran (Bast et al. 2014).

Visualisasi membantu analisis perambatan aliran trafik dengan pemahaman yang lebih baik. Terdapat pelbagai kajian lepas yang menganalisis perambatan aliran trafik menggunakan kaedah visual, sebagai contoh kajian oleh (Z. Wang et al. 2014) meneroka data trafik menggunakan visualisasi seperti yang ditunjukkan dalam Rajah 2.3. Pelbagai teknik visual seperti graf dinamik, histogram dan animasi tempatan digunakan untuk membentangkan kekuatan, kelajuan dan trafik yang tidak normal. Pengguna boleh memilih lokasi sel untuk mengendali korelasi kesesakan membandingkan antara sel pusat dengan laluan



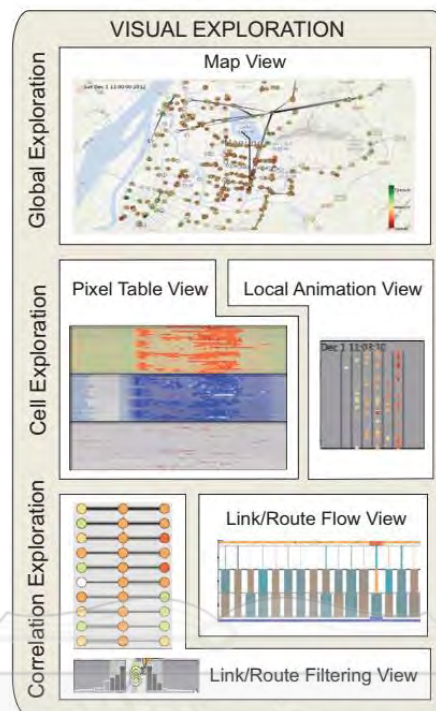
Rajah 2.3 Visual daripada Perambatan trafik (Z. Wang et al. 2014).

Kajian lain yang dilakukan oleh (Huang et al. 2015) menggunakan graf, dengan menukarkan trajektori GPS ke dalam graf yang berkaitan. Kajian mereka menjalankan *PageRank* dan *betweenness* untuk menganalisis banyak trajektori yang berhubungan antara satu sama lain. *PageRank* pada mulanya digunakan oleh Google untuk menentukan ranking atau kedudukan sebuah halaman web berdasarkan kualiti dan relevansi kandungan di dalamnya serta populariti laman web yang merujuk kepada halaman tersebut. Merujuk hal tersebut, maka pentingnya segmen jalan dinilai berdasarkan konsep bahawa pautan ke segmen jalan yang mempunyai skor lebih tinggi akan meningkatkan skor lebih banyak berbanding dengan pautan ke segmen jalan yang mempunyai skor lebih rendah. Jalan-jalan dengan skor Pagerank yang tinggi adalah jalan utama yang diutamakan (Huang et al. 2015). Keutamaan pusat antara (*Betweenness centrality*) memaklumkan bahawa satu titik (vertex) adalah penting jika ia terletak pada banyak laluan terpendek antara dua titik. Ia boleh mengukur sama ada satu jalan/rantau adalah tulang belakang dalam rangkaian bandar (Huang et al. 2015). Contoh *PageRank* dan *Betweenness* ditunjukkan pada Rajah 2.4

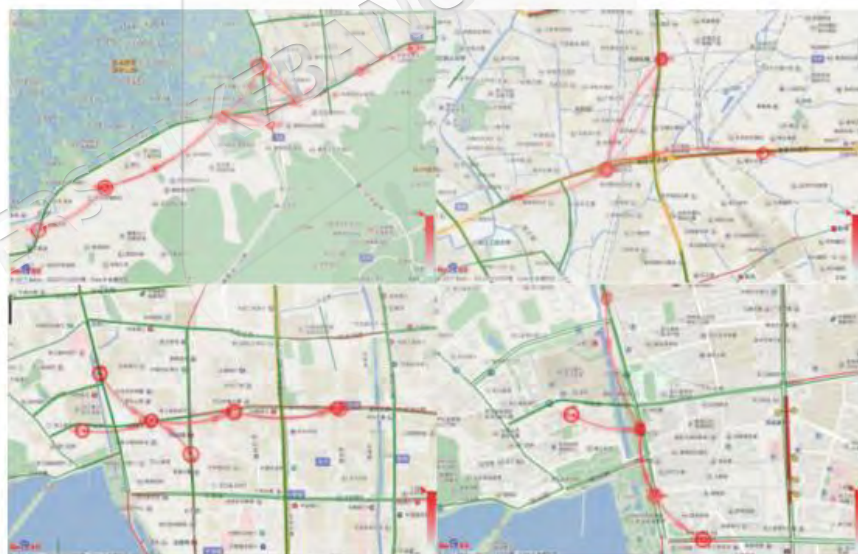


Rajah 2.4 Visualisasi sentralitas temporal di Shenzhen (Huang et al. 2015)

Satu kajian dalam visualisasi aliran trafik yang spesifik dalam analitik perambatan kesesakan trafik adalah kajian oleh (Z. Wang et al. 2013). Mereka membentangkan visualisasi yang lebih baik dalam analisis kesesakan trafik dan perambatan kesesakan trafik. Kajian mereka menjelaskan ciri untuk mengemukakan analisis kesesakan trafik dengan histogram, contoh visualisasi kajian oleh (Z. Wang et al. 2013) dipaparkan dalam Rajah 2.5. Kajian oleh (Z. Shan et al. 2018) memvisualisasikan perambatan kesesakan dengan kaedah graf ditunjukkan dalam Rajah 2.6.

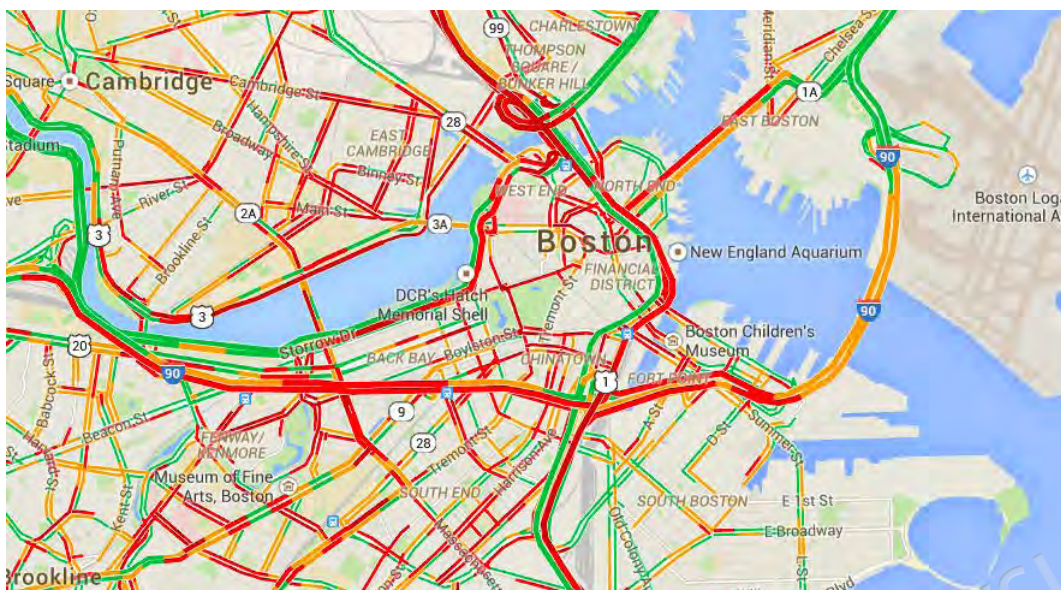


Rajah 2.5 Visualisasi analitik perambatan kesesakan trafik (Z. Wang et al. 2013).



Rajah 2.6 Visualisasi laluan perambatan kesesakan di Hangzhou, Zhejiang Province (Z. Shan et al. 2018).

Pada masa ini pengguna lebih banyak menggunakan peta google (Rajah 2.7) dan peta baidu yang dibuat oleh china (Z. Chen et al. 2018) (Rajah 2.8) untuk melihat tahap kesesakan berasaskan data trafik semasa.



Rajah 2.7 Visualisasi kesesakan menggunakan peta Google.

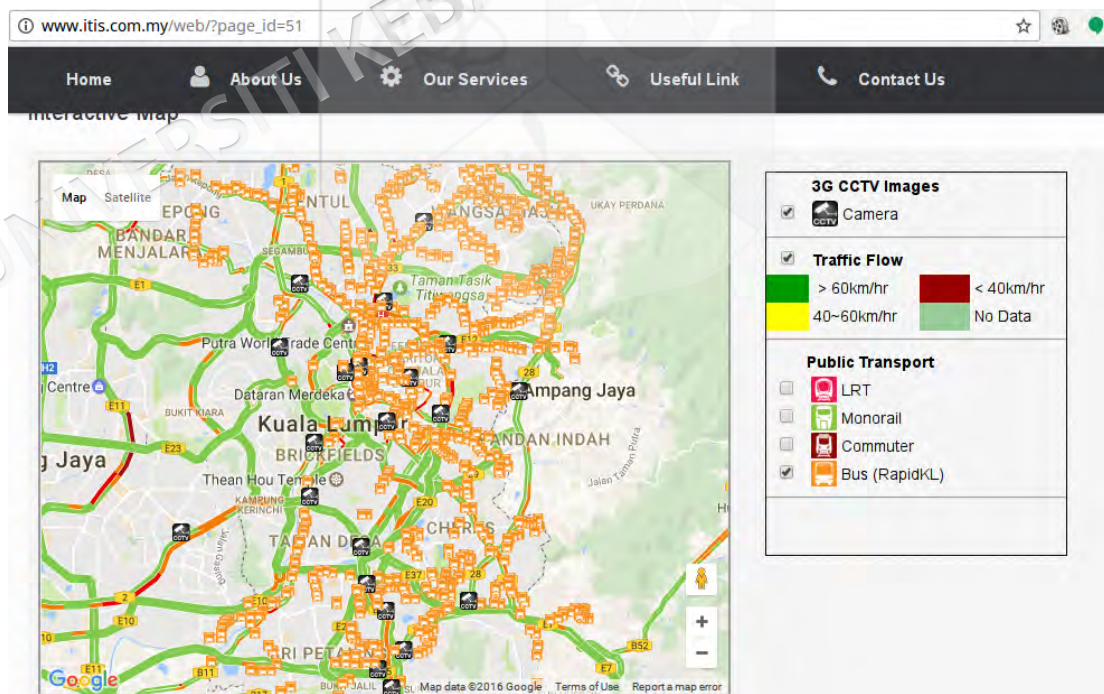


Rajah 2.8 Visualisasi keadaan trafik di rangkaian jalan Shanghai (Jalan sesak diberikan merah atau merah gelap).

Di Malaysia, kerajaan telah membangunkan Sistem Maklumat Pengangkutan Bersepadu (*Integrated Transport Information System (ITIS)*). Data trafik diambil dari CCTV yang dipasang di seluruh bahagian yang berlainan di bandar untuk pengawasan trafik dan pengumpulan data trafik. Dewan Bandaraya Kuala Lumpur telah memilih ITIS sebagai pendekatan utama untuk menambah baik dari menangani isu-isu trafik di Kuala Lumpur. ITIS mempunyai dua sistem pengendalian teras, iaitu IMS (Sistem Maklumat Bersepadu) dan ATIS (Sistem Maklumat Pengembara *Advanced*). Dalam

modul IMS, ia akan memaparkan peta GIS lanjutan yang memberikan maklumat mengenai pengurusan insiden. Di samping itu, ia akan menguruskan CCTV untuk pengawasan trafik dan jenayah bagi tujuan untuk memberi maklumat mengenai isu-isu berkaitan trafik/jenayah. Modul IMS dapat mengemas kini keadaan jalan raya masa nyata, peristiwa, beberapa insiden dan memaparkan maklumat mengenai VMS (*variable message sign*). Dengan mengakses ke peta GIS, ia akan memaparkan lokasi CCTV dan VMS di mana orang ramai boleh melihat keadaan trafik dan jalan lokasi tertentu, dan juga dapat melihat mesej di papan VMS. Dalam modul IMS, ia mempunyai peta kesesakan yang memaparkan keadaan status jalan raya yang lancar, perlahan dan sesak.

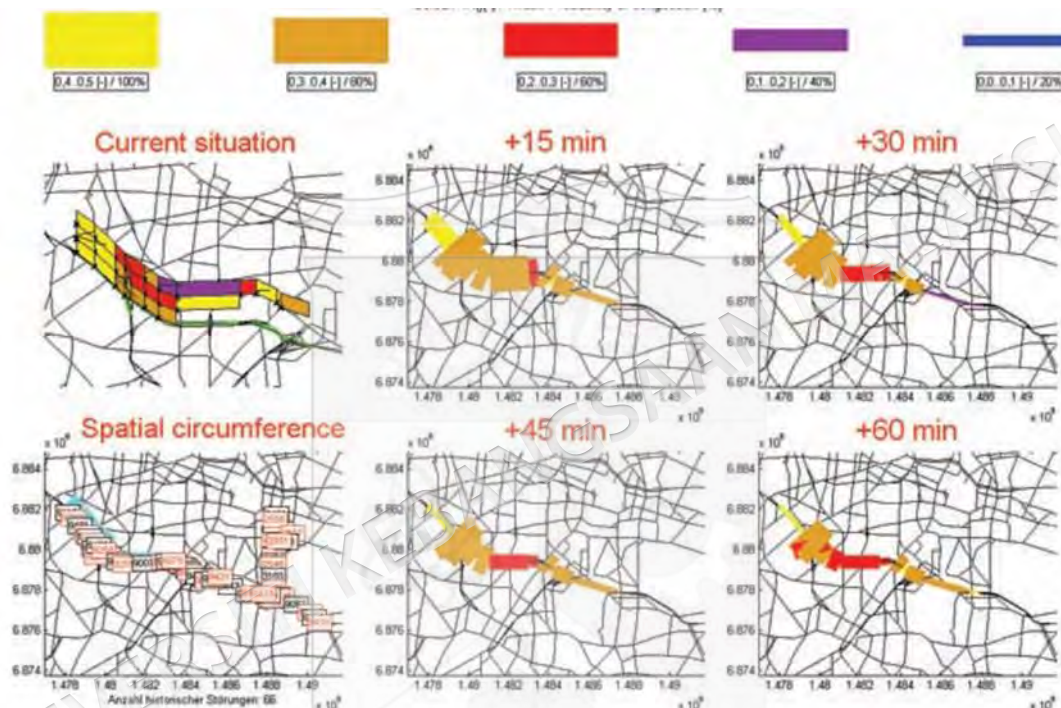
ATIS modul akan menyampaikan maklumat trafik melalui kamera dengan kualiti 3G ke pengguna jalan raya dalam bentuk peta kesesakan, peta kelajuan perjalanan dan peta kejadian melalui internet. Ia juga mempunyai ciri-ciri perancang perjalanan dan imej CCTV yang boleh disegarkan setiap 60 minit. Rajah 2.9 menunjukkan contoh output dari ITIS Malaysia.



Rajah 2.9 Sistem trafik masa nyata daripada ITIS Malaysia.

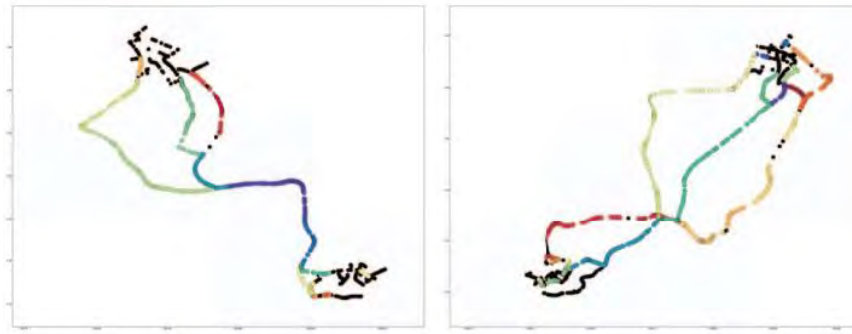
2.3.1 Ramalan perambatan trafik

Dalam bahagian ini kajian lepas dalam ramalan perambatan aliran trafik dikaji semula. Pada tahun 2008, kajian oleh (Maier et al. 2008) mencuba untuk membuat visual daripada perambatan aliran trafik dengan membandingkan keadaan semasa dengan beberapa aliran trafik sebelumnya yang serupa dalam situasi semasa, contoh visualisasi dipaparkan dalam Rajah 2.10.



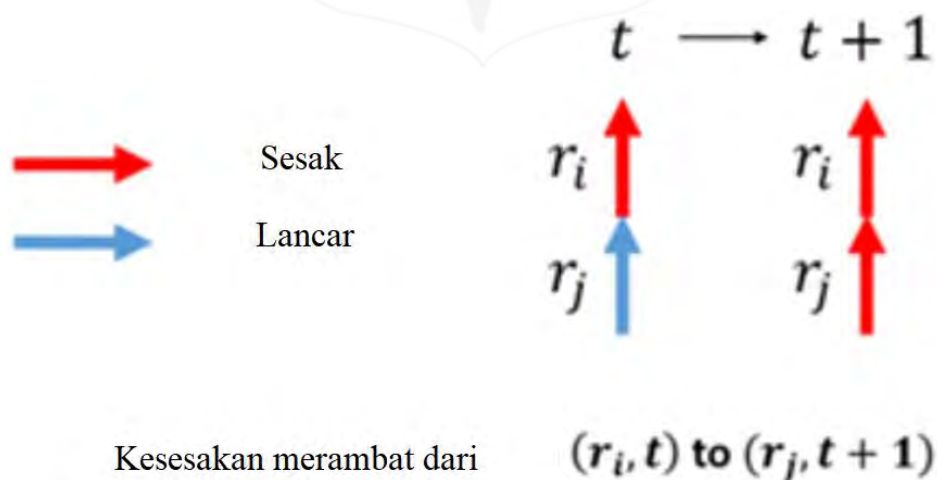
Rajah 2.10 Ramalan perambatan kesesakan jalan (Maier et al. 2008).

Beberapa kajian lain oleh (Hamner 2010), (Barbosa et al. 2013), (Deep Singh et al. 2015), menggunakan set data jalan yang saling berhubung dan berarah, dibangunkan daripada laluan perjalanan berasaskan trajektori GPS untuk meramal masa perjalanan. Contoh visualisasi dari lintasan dan sub lintasan dipaparkan dalam Rajah 2.11 (Deep Singh et al. 2015).



Rajah 2.11 Contoh gugusan sub-trajektori untuk dua gugusan ruang yang berbeda.

Kajian lain oleh (Z. Chen et al. 2018) (Nguyen et al. 2016a) menggunakan jalan berasaskan arah trajektori jejak masa sebelumnya, mereka menggunakan pendekatan grafik dalam kajian mereka, terdapat juga kajian oleh (Ye et al. 2016), (Xiong & Vahedian 2018) yang menggunakan garis poli (*polyline*) untuk menggambarkan trajektori perjalanan, Mereka menggunakan garis poli yang dihubungkan untuk ramalan perambatan aliran trafik. Graf yang saling berkaitan dikembangkan berasaskan sub graf yang kerap diperolehi dengan mencari punca perambatan yang sering berlaku atau menggunakan garis poli *overlay* (garis poli yang saling bertumpukan). Contoh perambatan kesesakan menggunakan graf ditunjukkan dalam Rajah 2.12, Sekiranya r_i sesak di t , maka segmen hulu yang menghubungkan ke r_i mungkin menjadi sesak di $t + 1$.



Rajah 2.12 Perambatan kesesakan menggunakan graf (Xiong et al. 2018)

Selain menggunakan pendekatan graf, kajian lain mempelajari perambatan trafik menggunakan kaedah penggugusan, seperti kajian oleh (Saeedmanesh & Geroliminis 2017), (Ambühl et al. 2023) dan (Yang et al. 2017) iaitu analisis perambatan trafik dilakukan dengan cara mengguguskan jalan raya berjiran berdasarkan masa kesesakan dan masa lancar. Kajian oleh (Yang et al. 2017) menganalisis pola perambatan trafik menggunakan penggugusan spektrum berdasarkan masa, dimana hasil penggugusan jalan raya dilakukan berulang kali pada masa lancar dan pada masa sesak.

Selain meramal perambatan trafik, terdapat juga kajian meramal keadaan trafik seperti kajian oleh (Xiaomeng Wang et al. 2015) menggunakan HMM untuk meramal keadaan aliran trafik yang tidak diketahui (hilang) berdasarkan data dari data kenderaan terapung (*floating car data*). Terdapat juga kajian oleh (G. Zhu et al. 2016) menggunakan HMM untuk memodelkan peralihan keadaan aliran trafik untuk rangkaian jalan di bandar berdasarkan data VMS (*variable message sign*).

Daripada kajian literatur, terdapat dua teknik yang boleh digunakan untuk menganalisis perambatan aliran trafik. Teknik pertama merupakan teknik yang menggunakan data trafik berdasarkan trajektori GPS dimana arah pergerakan kenderaan diketahui. Untuk set data yang berdasarkan trajektori GPS, lintasan ditukar kepada graf, dan kemudian diplotkan pada peta. Seterusnya pergerakan graf dianalisis oleh penglihatan manusia (Z. Wang et al. 2013) (Huang 2015). Untuk teknik yang melibatkan grafik, pembinaan graf dan sub graf yang saling berkaitan dibina secara rekursif. Teknik dilaksanakan dengan menyusuri jalan-jalan yang berhubung sesuai arah pergerakan kenderaan. Teknik kedua merupakan teknik yang menggunakan data trafik dari pengesanan seperti (FCD, VMS, IoT). Secara garis besar teknik ini menggunakan pendekatan kebarangkalian dan penggugusan. Seperti kajian oleh (X. Wang et al. 2015) (Zhu et al. 2016) menggunakan model Markov tersembunyi untuk meramal peralihan aliran trafik sedangkan kajian oleh (Saeedmanesh & Geroliminis 2017), (Ambühl et al. 2023) menggunakan penggugusan untuk menganalisis perambatan trafik. Perbandingan kajian perambatan trafik lebih detil diterangkan dalam Jadual 2.1.

Jadual 2.1 Matriks daripada kajian perambatan kesesakan jalan dan kajian ramalan kondisi trafik

No	Kajian	Sumber Data	Perambatan trafik	Visualisasi	Ramalan Trafik	Maklumat berhubung	Teknik
1	Pembentangan berasaskan penumpukan bagi data atribut lintasan	GPS	×	✓	×	✓	Graf
2	Kaedah analisis visualisasi ruang masa bagi operasi teksi di Beijing	GPS	×	✓	×	✓	Graf
3	TrajGraph : Pendekatan analitik visual berasaskan graf untuk kajian lintasan	GPS	×	✓	×	✓	Graf
4	Analisis kesesakan lalu lintas visual berdasarkan data lintasan	GPS	✓	✓	×	✓	Graf
5	Memprediksi Corak Penyebaran Kesyakan Lalu Lintas: Pendekatan Graf Penyebaran	GPS	✓	✓	✓	✓	Graf
6	Mengesan Corak Penyebaran Kesyakan Lalu Lintas Bandar dengan Data Lintasan Teks	GPS	✓	×	×	✓	Graf
7	Mengesan Corak Penyebaran Kesyakan dalam Data Lalu Lintas Spasio-Temporal	GPS	✓	✓	✓	✓	Graf
8	Pemahaman penyebaran kesesakan dengan menggabungkan teori perkolasi dengan diagram fundamental makroskopik	GPS	✓	×	×	✓	Penggugusan
9	Penggugusan dan penyebaran dinamik kesesakan dalam rangkaian lalu lintas bandar yang tercongested secara heterogen	GPS	✓	×	×	✓	Penggugusan

bersambung...

...sambungan

10	Analisis corak variasi keadaan lalu lintas untuk rangkaian jalan bandar berasaskan pengklusteran spektral	GPS	x	✓	✓	✓	Penggugusan
11	Ramalan lintasan kenderaan berdasarkan Model Markov Tersembunyi	GPS	x	✓	✓	✓	HMM
12	Model Markov Tersembunyi untuk anggaran trafik skala bandar menggunakan data kereta terapung	Pengesan	x	x	✓	✓	HMM
13	Model transisi keadaan aliran trafik untuk rangkaian jalan bandar berasaskan Model Markov Tersembunyi	Pengesan	x	x	✓	x	HMM
14	Model Markov Tersembunyi untuk ramalan jangka pendek keadaan lalu lintas di lebuh raya	Pengesan	x	x	✓	x	HMM
15	Ramalan Kelajuan Lalu Lintas Menggunakan Model Markov Tersembunyi untuk Lebuhraya Republik Czech	Pengesan	x	x	✓	x	HMM
16	Ramalan kesesakan lalu lintas berdasarkan Model Markov Tersembunyi dan ukuran kontras	Pengesan	x	x	✓	x	HMM
17	Model Markov Tersembunyi Dinamik untuk Ramalan Aliran Trafik Metropolitan	Pengesan	x	x	✓	✓	HMM

Seperti telah diterangkan diatas, untuk teknik yang melibatkan grafik. Teknik hanya perlu menyusuri jalan-jalan yang berhubung sesuai arah pergerakan kenderaan.

Perambatan kesesakan dari satu jalan ke jalan yang lain disusun dengan menyusuri jalan yang berhubung dengan jalan sasaran yang mengalami kesesakan pada saat yang bersamaan dengan jalan sasaran secara rekursif. Akan tetapi membina graf bersambung dan graf subsambung secara rekursif semasa kesesakan jalan raya memakan masa. Selain itu memerlukan set data yang mempunyai maklumat tentang jalan berhubung dan arah daripada aliran trafik. Untuk data trafik yang tidak memiliki maklumat mengenai jalan berhubung dan arah, aliran trafik pendekatan graf secara rekursif ini tidak dapat digunakan kerana menyusuri jalan-jalan yang berhubung tidak boleh dilakukan. Untuk situasi seperti ini perlu dipertimbangkan kaedah lain seperti pendekatan menggunakan pendekatan kebarangkalian dan penggugusan.

Seterusnya, kaedah penyelidikan dalam ramalan trafik, pada dasarnya boleh dibahagikan kepada dua kategori: pendekatan parametrik dan non-parametrik (Sun et al. 2019). Pendekatan parametrik adalah berdasarkan data empirikal dan andaian teoritis untuk menentukan parameter model, seperti model penapis Kalman dan model berasaskan Autoregressive Moving Average (ARIMA), yang merupakan model siri masa (*time series*) klasik (Jin et al. 2013), Model Autoregressive juga digunakan oleh (Kerner & Herrtwich 2006) dan (Davoodi et al. 2016) untuk meramal aliran trafik. Kelebihan kaedah parametrik termasuk (1) kerumitan model yang rendah dan (2) kemudahan pelaksanaan. Walau bagaimanapun, rangkaian trafik adalah rumit dan menunjukkan perubahan tak linear dan stokastik, kerana itu pendekatan parametrik tidak dapat mengendalikan ketidakpastian dengan baik kerana prestasi ramalan kaedah parametrik akan dipengaruhi oleh banyak faktor luar dari aliran trafik.

Untuk pendekatan nonparametrik, data latihan dengan jumlah besar diperlukan untuk membina struktur model. Model nonparametrik yang biasa digunakan adalah rangkaian saraf buatan (ANN), jiran K-terdekat (KNN), dan model rangkaian Bayesian (Sun et al. 2019). ANN dapat memodelkan hubungan kompleks dalam aliran trafik, kelemahannya yang tidak dapat dielakkan adalah jumlah masa yang panjang yang diperlukan untuk melatih rangkaian (Kindzerske & Ni 2007), kerana struktur rangkaian saraf mempunyai terlalu banyak parameter dan proses latihan rumit, yang menyebabkan kecacatan kos pengiraan yang terlalu tinggi (Zhao et al. 2019). Manakala KNN bergantung daripada data trafik lepas yang ideal yang mana ia

memilih jiran terdekat daripada data trafik lepas (Kindzerske & Ni 2007). Rangkaian bayes dinamik paling mudah adalah model markov tersembunyi (HMM) (Yun Zhang et al. 2022). Kajian terdahulu oleh (Zhou et al. 2020), menjumpai bahawa pola trafik berulang setiap hari, seperti harga saham, dimana kajian ramalan harga saham berdasarkan siri masa menggunakan HMM mencapai hasil yang sangat baik, kerana itu HMM sangat baik digunakan untuk ramalan aliran trafik yang berulang setiap hari (Zhao et al. 2019). Manakala, untuk meramal seberapa tepat mungkin aliran trafik, maka maklumat dari jalan bersebelahan (terutamanya, hulu) dan jalan semasa harus dipertimbangkan bersama. Sebab itu HMM dapat mengambil kira hubungan kausal sepenuhnya antara pemboleh ubah rawak secara statistik, dan dengan itu boleh digunakan untuk memodelkan dan menganalisis aliran trafik jalan berjiran (Kerner & Herrtwich 2006).

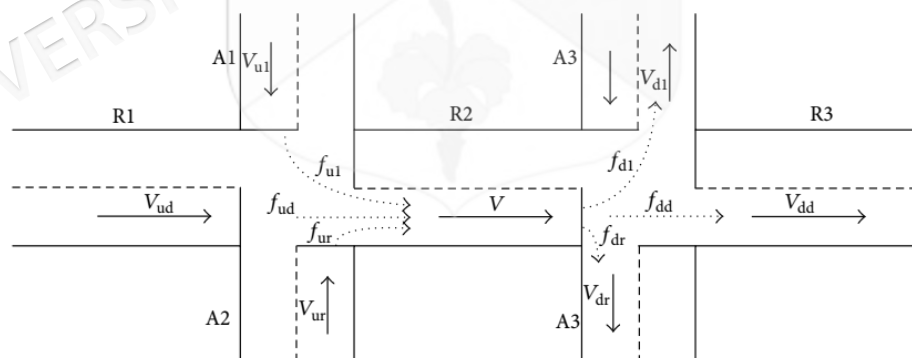
Seperti yang telah diberikan dalam Jadual 2.1, beberapa kajian menggunakan pendekatan kebarangkalian model Markov tersembunyi untuk meramal keadaan trafik seperti penelitian yang dilakukan oleh (Xiaomeng Wang et al. 2015), (Qi & Ishak 2014), (G. Zhu et al. 2016), (Rapant et al. 2016), (Zaki et al. 2019). Pelbagai kajian lepas meramal keadaan trafik. Walau bagaimanapun pelbagai penelitian tersebut tidak secara spesifik meramal perambatan atau tidak meramal jalan yang mengalami impak daripada jalan yang mengalami kesesakan.

Untuk itu kajian ini menggabungkan antara analisis perambatan trafik menggunakan penggugusan seperti dilakukan oleh (Saeedmanesh & Geroliminis 2017), (Ambühl et al. 2023) dan (Yang et al. 2017), tetapi alih-alih melakukan penggugusan berulang kali pada masa sesak dan masa lancar, kaedah GLCM digabungkan dengan penggugusan untuk menyelidiki pola kesesakan daripada jalan yang berdekatan. Seterusnya jalan raya berjiran dikelompokkan berasaskan pola trafik. Hasil daripada penggugusan jalan yang memiliki kesamaan pola yang kuat dihasilkan dengan jalan sasaran atau dengan kata lain memiliki hubungan yang kuat. Seterusnya pendekatan statistik kebarangkalian digunakan untuk meramal kesan kesesakan jalan raya terhadap keadaan trafik di kawasan jiran.

2.4 FAKTOR RUANG DAN MASA BAGI RAMALAN TRAFIK

Faktor ruang dan masa yang mempengaruhi aliran trafik telah lama diselidiki sebagai satu daripada faktor yang mempengaruhi aliran trafik. Seperti yang telah diketahui secara umumnya banyak faktor boleh menjejaskan keadaan trafik di jalan raya, termasuk pelbagai data maklumat lampau iaitu: data trajektori GPS, kelajuan kenderaan, bilangan kenderaan, cuaca dan kemalangan, serta peristiwa khas. (Kwoczek et al. 2014). Bahagian ini, menjelaskan faktor yang paling banyak memberi kesan yang boleh menjejaskan aliran trafik di jalan raya. Faktor yang paling banyak memberi kesan itu ialah keadaan trafik di kawasan berjiran (satu kawasan yang berdekatan).

Dua atau lebih jalan di kawasan berdekatan mungkin mengalami korelasi kesesakan jika semua jalan sesak pada masa yang sama (Y. Wang et al. 2016). Dengan kata lain, dua atau lebih jalan boleh menyebabkan kesesakan jalan raya di jalan hilir. Sebagai contoh, diketahui empat arah jalan iaitu V , V_{dd} , V_{dr} , V_{dt} . Maka seperti ditunjukkan pada rajah 2.13 bahawa V boleh menjejaskan kesesakan di dua arah jalan V_{dd} , V_{dr} dan V boleh menjejaskan kesesakan di tiga jalan raya V_{dt} , V_{dd} , V_{dr} .



Rajah 2.13 Taburan aliran trafik (Deng, & Qu 2016).

Terdapat empat arah jalan iaitu V , V_{dd} , V_{dr} , V_{dt}

V boleh menjejaskan kesesakan di dua arah jalan V_{dd} , V_{dr} .

$V \rightarrow (V_{dd}, V_{dr})$

V boleh menjejaskan kesesakan di tiga jalan raya V_{dt} , V_{dd} , V_{dr} .

$$V \rightarrow (V_{dt}, V_{dd}, V_{dr})$$

2.4.1 Ramalan trafik berasaskan jalan raya yang berhubung

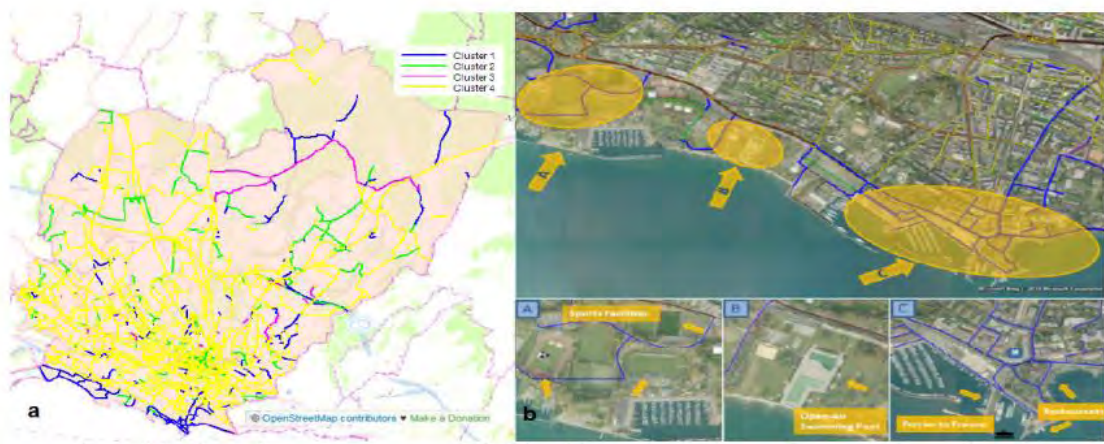
Pemboleh ubah ruang dan masa ialah satu daripada pemboleh ubah untuk meramal aliran trafik. Bahagian ini membincangkan jalan yang berhubung secara statik yang digunakan sebagai pemboleh ubah ruang dalam ramalan aliran trafik. Jalan sekitar, terutamanya jalan-jalan yang berhubung dihilir, dan jalan yang berhubung secara langsung dianggap sebagai faktor spatial yang mempengaruhi aliran trafik. Kajian yang dilakukan oleh (Cheng & Lu 2018) menggunakan jalan berhubung yang mengelilingi jalan sasaran sebagai faktor untuk meramal aliran trafik. Kajian yang dilakukan oleh (C. Hu et al. 2008) menunjukkan bahawa ramalan aliran trafik menggunakan jalan yang berhubung di sekitar jalan sasaran menghasilkan hasil yang lebih baik daripada ramalan trafik yang tidak melibatkan jalan berhubung di sekitarnya atau hanya menggunakan data jalan sasaran sahaja. Bahkan kajian yang dilakukan (Yu et al. 2016) menyatakan bahawa meramal keadaan trafik menggunakan KNN dengan jalan bersambung di hulu dan di hilir, menghasilkan ketepatan yang lebih baik berbanding dengan ramalan keadaan trafik yang menggunakan hanya satu jalan berhubung. Sementara itu, satu kajian lanjutan menggunakan jalan berhubung dengan jalan berhubung lainnya (*sub level*). Kajian yang dilakukan oleh (Hong et al. 2015) melanjutkan satu aras jalan berhubung di hulu dan hilir sebagai pemboleh ubah ruang. Dalam kajian yang dilakukan oleh (Min & Wynter 2011) menyatakan jalan berhubung lanjutan (lebih dari satu aras) meningkatkan ketepatan ramalan keadaan trafik berbanding dengan ramalan yang hanya menggunakan jalan berhubung sahaja (satu aras). Hal ini lebih membuktikan bahawa melibatkan jalan-jalan yang berhubung dengan jalan sasaran akan meningkatkan ketepatan ramalan arus trafik.

2.4.2 Ramalan trafik jalan yang memiliki hubungan tinggi secara dinamis

Sebaliknya, selain kaedah yang menggunakan jalan berhubung secara langsung, terdapat kajian lain yang menggunakan jalan yang memiliki hubungan secara dinamik untuk meramal aliran trafik. Jalan yang memiliki kaitan secara dinamik bermaksud

jalan raya berjiran yang mempunyai persamaan keadaan trafik (C. Hu et al. 2008). Kajian oleh (Y. Wang et al. 2016) menentukan perkaitan antara segmen jalan dengan memerhati keadaan trafik antara dua jalan pada masa yang sama dalam jarak tertentu. Menurut Wang et al. (2016) terdapat korelasi dari segmen jalan A ke B pada jarak tertentu d , apabila terdapat kesesakan di bahagian jalan A pada masa t dan pada masa $t + T$. Kajian lain oleh (Rempe et al. 2016) juga menggunakan kaedah yang sama, bezanya ialah segmen jalan dimasukkan dalam kluster kesesakan yang sama jika terdapat laluan antara segmen jalan dan kedua-duanya dalam keadaan kesesakan yang sama pada masa T . Selain itu, kajian lain cuba menggunakan pendekatan statistik dan pembelajaran mesin iaitu kajian oleh (Liu et al. 2016). Mereka menggunakan regresi linear berganda untuk mencari jalan yang memiliki keterkaitan yang tinggi satu sama lain. Kajian lain oleh (J. Y. Ahn et al. 2015) menggunakan rantai Markov untuk mencari pola persamaan antara jalan dalam ruang dan masa. Terdapat juga kajian lain oleh (Han et al. 2014; Han & Moutarde 2012; Xing et al. 2015) menggunakan tensor PCA untuk mendapati kaitan antara jalan raya dengan cara meminimumkan ciri daripada data jalan raya berjiran.

Pendekatan penggugusan (*clustering*) didapati sebagai kaedah alternatif untuk mencari keadaan trafik yang sama antara jalan raya. Kaedah penggugusan mengelompokkan jalan raya asas data terdahulu. Pendekatan ini mengumpulkan jalan raya dengan keadaan trafik yang sama dalam gugusan yang sama. Beberapa kajian terdahulu mengguguskan jalan raya menggunakan K-Means berasaskan ketumpatan dan kelajuan (Mondal & Rehena 2019), kajian lain mengguguskan berasaskan lintasan GPS (Necula 2015). Contoh pemvisualan penggugusan jalan raya dengan K-Means ditunjukkan pada Rajah 2.14. Selain itu terdapat kajian lain yang dilakukan oleh (Xiaomeng Wang et al. 2015) menggunakan penggugusan spektrum untuk mengguguskan jalan berasaskan purata kelajuan di kawasan jiran.



Rajah 2.14 Penggugusan jalan raya dengan K-Means (Necula 2015).

2.5 DATA RAYA DALAM KAJIAN ALIRAN TRAFIK

Data raya menjanjikan banyak harapan untuk menambah baik perancangan dan pengurusan aktiviti trafik dengan meningkatkan jumlah atau ketersediaan berkaitan mobiliti secara radikal, bahkan mendekati masa nyata. Begitu juga, akses kepada data yang lebih terperinci dan boleh diambil tindakan berkenaan pengendalian kenderaan dan persekitaran di mana ia beroperasi memegang banyak harapan untuk meningkatkan keselamatan pengangkutan. Ketiga-tiga bidang ini iaitu operasi, perancangan dan keselamatan, adalah bidang dimana pihak berkuasa mesti menilai secara kritis penilaian dimana dan bagaimana data baharu atau yang baharu tersedia serta pemerhatian berkaitan data, boleh menambah baik pengangkutan (*transport policy*) (Crist 2015).

Bilangan kenderaan, semakin bertambah menyebabkan trafik dan kesesakan yang kerap berlaku di jalan raya setiap hari, memerlukan penyelesaian inovatif untuk pengurusan infrastruktur dan trafik. Memandangkan komponen sistem trafik menjadi lebih autonomi dan lebih pintar (contohnya, keupayaan komunikasi baharu kenderaan dan infrastruktur), terdapat peningkatan keperluan untuk kerjasama di kalangan sistem transportasi cerdas (ITS) untuk pengurusan pengangkutan dan pemantauan alam sekitar untuk menambah baik strategi pengurusan trafik. Dalam sistem pengurusan perniagaan generasi baharu ini, pengurusan rangkaian pengangkutan disepadukan

dengan strategi perniagaan dan model operasi syarikat pengangkutan dan pelanggan individu, memberikan impak yang besar kepada syarikat dari segi perancangan perniagaan, kualiti perkhidmatan dan penyesuaian kepada keperluan pelanggan. serta untuk pengguna individu dari segi penjimatan masa dan wang, perancangan perjalanan yang dapat disesuaikan dan sokongan tingkah laku sosial yang saling menguntungkan. Sumber data raya dalam transportasi, meliputi data terstruktur dan semi-terstruktur, dikumpulkan oleh sistem transportasi cerdas (ITS)(Iliashenko et al. 2021), ITS bertindak sebagai penjana data dan sumber yang membawa kepada sejumlah besar data yang tersedia dengan kadar kemas kini yang singkat. Pertumbuhan jumlah data ini didorong oleh individu dan peningkatan penggunaan mereka dalam media sosial (*social network*), pengesanan keluaran terbaharu dan keupayaan komunikasi dalam kenderaan dan infrastruktur trafik; aplikasi teknologi maklumat dan komunikasi (ICT) terkini (*Cloud computing, Internet of Things (IoT)*) dengan percambahan peranti dan sistem yang disambungkan ke internet. Saiz set data semakin besar dan semakin meningkat dikumpul oleh peranti *ubiquitous information sensing mobile devices, aerial sensory technologies* (pengesanan jauh), log perisian, kamera, mikrofon, *radio-frequency identification readers*, dan *wireless sensor networks*. Terdapat juga pecutan dalam bahagian data yang diproduksi oleh mesin dan tidak berstruktur (foto, video, media sosial feeds, dll.) (Jelena et al. 2013). Sistem pengangkutan pintar (ITS) telah dibangunkan sejak awal tahun 1970-an. ITS adalah arah masa depan sistem pengangkutan. ITS menggabungkan teknologi canggih yang termasuk teknologi pengesanan elektronik, teknologi penghantaran data dan teknologi kawalan pintar ke dalam sistem pengangkutan. Tujuan ITS adalah untuk menyediakan perkhidmatan yang lebih baik untuk pemandu dan pengendali kenderaan dalam sistem pengangkutan (L. Zhu et al. 2019). Pemahaman awal tentang data raya mentakrifkan data raya sebagai 3V, yang bermaksud *Volume*, *Velocity*, dan *Variety*. *Volume* merujuk kepada saiz data yang sangat besar manakala *Variety* merujuk kepada pelbagai jenis data dan dikumpulkan daripada pelbagai sumber dan *Velocity* merujuk kepada kelajuan penciptaan data (Adrian et al. 2016). Data yang dikumpul oleh sistem pengangkutan cerdas (ITS) semakin kompleks dan sangat besar. *Volume* merujuk kepada kuantiti data yang dihasilkan oleh pelbagai sumber dan masih berkembang. Dengan pertumbuhan jumlah trafik dan pengesanan, ketumpatan data dalam

pengangkutan telah meningkat dengan ketara. Selain itu, pengembara, barangan dan kenderaan menjana lebih banyak data apabila transponder penjejakan digunakan. Data yang dijana daripada infrastruktur, pemantauan alam sekitar dan meteorologi juga meningkat sebagai bahagian kritikal data pengangkutan. *Variety* tertumpu terutamanya pada semua jenis data yang dihasilkan oleh pengesanan(*sensor*), dan juga media sosial. *Variety* data berkaitan pengangkutan telah meningkat dengan ketara. Sebagai contoh, kenderaan terkemas kini boleh melaporkan telemetri sistem dalaman dalam masa nyata dan maklumat semua anak kapal dan penumpang. *Velocity* data dalam pengangkutan telah meningkat disebabkan oleh peningkatan teknologi komunikasi, peningkatan kuasa pemprosesan dan kelajuan pemantauan dan pemprosesan. Perkembangan teknologi dalam ITS telah membawa kepada peningkatan dalam kerumitan, kepelbagaian dan jumlah data yang dijana dan dikumpulkan daripada kenderaan, dan pergerakan manusia. Kajian oleh (L. Zhu et al. 2019) mengenal pasti enam (6) kategori sumber data yang berbeza dalam ITS. Namun bahagian ini hanya menjelaskan empat (4) sumber utama pengumpulan data dalam ramalan aliran trafik.

2.5.1 Sumber pengumpulan data dalam ramalan aliran trafik

a. Data raya daripada GPS

Global Positioning System (GPS) ialah alat yang paling popular untuk penjejakan lokasi. Data trafik boleh dikumpul dengan lebih cekap dan selamat dengan pengesanan lokasi melalui GPS. Menggabungkan sistem maklumat geografi (GIS) atau teknologi paparan peta lain, GPS merupakan alat yang berpotensi untuk pengumpulan data, atau yang dikumpul boleh digunakan untuk menangani banyak isu trafik, seperti cadangan laluan (J. Dai et al. 2015)(Cui et al. 2018), dan analisis aliran trafik menggunakan pengvisualan (Xin Wang, Zhao, et al. 2017) (Z. Wang et al. 2014) (Z. Wang et al. 2013) (Z. Wang et al. 2013) (Tominski et al. 2012)(Jianqin et al. 2015) (Bast et al. 2014) (Huang et al. 2015), kajian perambatan aliran trafik (Besenczi et al. 2020) (Z. Chen et al. 2018) (Nguyen et al. 2016a), dan (Ye et al. 2016).

b. Data raya daripada video

Kamera video digunakan secara meluas dalam ITS. Seperti yang ditunjukkan dalam sistem pengurusan trafik lanjutan (ATMS), sistem pengesanan imej video (VIDS) ialah alternatif yang baik berbanding dengan sensor konvensional untuk tugas seperti identifikasi kenderaan dan pengesanan aliran trafik. Satu kelebihan VIDS ialah kos yang rendah (Xin Wang, Zhao, et al. 2017). Sensor pengimejan lebar raya yang menggunakan data video yang besar telah berjaya digunakan untuk mengesan dan mengklasifikasikan kenderaan (Mallikarjuna et al. 2009). Selain daripada pengurusan trafik (Xin Wang, Zhao, et al. 2017), jurutera dan perancang pengangkutan yang mengumpul data video kenderaan yang lebih tepat boleh menambah baik sistem pemprosesan imej supaya menjadi lebih baik dalam membuat permintaan pengangkutan umum berkenaan model pelepasan kenderaan.

c. Data raya daripada pengesan

Peralatan pengesan yang dipasang dalam ITS digunakan untuk mengumpul data seperti kelajuan kenderaan, ketumpat kenderaan, aliran trafik dan masa perjalanan. Sensor jalan raya tradisional, contohnya pengesan infrared dan gelombang mikro, telah berkembang untuk memperolehi, mengira dan memindahkan data trafik (Lopes et al. 2010). Seperti kajian oleh (Lopes et al. 2010), pengumpulan data daripada pengesan boleh dibahagikan kepada tiga sumber: data tepi jalan, data kereta terapung (*floating car data*) dan data kawasan luas (Antoniou et al. 2011). Data tepi jalan terutamanya merujuk kepada data yang dikumpul oleh pengesan yang terletak di sepanjang tepi jalan. Contoh daripada pengesan tepi jalan terdahulu iaitu *inductive magnetic loops*, *pneumatic road tubes*, *piezoelectric loops arrays* dan *microwave radars*, benda ini telah digunakan selama bertahun-tahun. Pengesan tepi jalan generasi baharu seperti sistem pengesan ultrasonik dan akustik, pengesan kenderaan magnetometer, sistem *infrared*, pengesanan dan julat cahaya (*light detection and ranging LIDAR*), dan sistem pemprosesan dan pengesanan imej video secara beransur-ansur muncul dengan perkembangan teknologi terkini. Data kereta terapung terutamanya merujuk kepada data mobiliti kenderaan di lokasi berbeza dalam ITS, di mana pengesan tersuai dibenamkan dalam kenderaan (Xiaomeng Wang et al. 2015).

Sesetengah pengesanan *onboard* menyediakan maklumat yang yakin dan cekap untuk pemilihan dan anggaran laluan perjalanan. Dengan perkembangan teknik pengesanan kenderaan, teknik pengesanan FCD yang popular termasuk identifikasi kenderaan automatik (AVI), pengecaman plat lesen (LPR), dan transponder seperti kenderaan siasatan dan tanda tol elektronik. Data kawasan luas merujuk kepada data aliran trafik kawasan luas yang dikumpul oleh pelbagai teknik penjejakan sensor seperti pemprosesan fotogrametri, rakaman bunyi, pemprosesan video dan radar berasaskan ruang.

d. Data raya CAV dan VANET

Kenderaan bersambung dan autonomi (CAV) ialah teknologi baharu di kawasan ITS yang menggabungkan perubahan radikal reka bentuk kenderaan dan interaksinya dengan infrastruktur jalan raya. Kenderaan bersambung dan autonomi menggabungkan pelbagai teknologi yang berbeza, memudahkan pergerakan orang dan barangan yang selamat dan cekap. Sistem trafik yang menggunakan CAV telah menunjukkan potensi besar untuk mengurangkan kesesakan, mengurangkan kelewatan perjalanan dan meningkatkan prestasi keselamatan (Uhlemann 2015), (C. Chen et al. 2017). CAV juga menjana sejumlah besar data pengangkutan masa nyata yang berkaitan dengan alam sekitar, seperti koordinat, kelajuan, pecutan, data keselamatan (Uhlemann 2015). Menggunakan teknologi rangkaian terkini seperti *Software Defined Networking* (SDN). SDN ialah seni bina rangkaian yang membolehkan rangkaian dikawal secara pintar dan berpusat. Ini membantu pengendali mengurus keseluruhan rangkaian secara konsisten dan holistik, tanpa mengira teknologi rangkaian asas, data boleh diambil dengan cekap (Xin Wang, Wang, et al. 2017). Data ini boleh digunakan untuk mencipta maklumat yang boleh diambil tindakan untuk menyokong dan memudahkan pilihan pengangkutan hijau dan digunakan pada kawalan isyarat adaptif masa nyata (Xin Wang, Wang, et al. 2017) (King 2015). Rangkaian *Ad Hoc* Kenderaan (VANET) ialah sejenis rangkaian *ad hoc* mudah alih yang menggunakan kenderaan dan elemen infrastruktur sebagai nod untuk meningkatkan kawasan liputan dan keupayaan komunikasi. Sebagai bahagian penting ITS, VANET menjana sejumlah besar data (Jindal 2014). Penyediaan data dan keputusan masa nyata adalah tugas yang mencabar untuk analisis berskala besar.

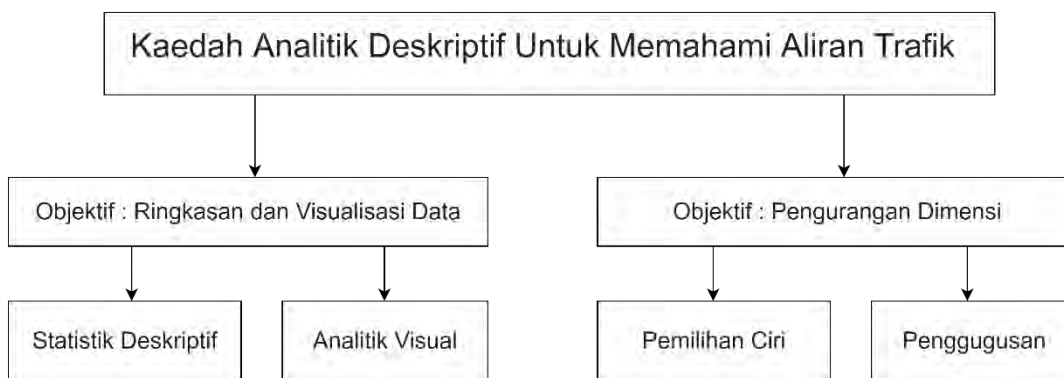
Menggunakan analitik data raya, cabaran VANET (Contreras-Castillo et al. 2016) yang berkaitan boleh ditangani, seperti ramalan aliran trafik (Jindal 2014).

2.5.2 Analitik data raya dalam kajian trafik

Selepas menjelaskan jumlah besar sumber data trafik, langkah seterusnya ialah analitik. Secara umum, terdapat dua jenis analitik yang biasa digunakan dalam kajian aliran trafik seperti yang diterangkan di bawah:

a. Analitik deskriptif

Analitik deskriptif adalah analisis yang paling mudah dan juga sangat penting. Kaedah ini menerangkan data dengan jelas. Analitik ini berkaitan persoalan "apa yang berlaku". Sebagai contoh dalam situasi kemalangan, analitik deskriptif melihat ranap yang berlaku, mempersoal soalan asas seperti apa yang berlaku; bilangan kematian, bilangan pejalan kaki yang terlibat; dan bilangan kecederaan serius. Outputnya merupakan ringkasan statistik. Contoh ringkasan statistik yang digunakan ialah nilai minimum, median, maksimum atau sisihan piawai. Tambahan pula, bantuan visual seperti carta dan histogram boleh membantu manusia memahami cerapan data. Berkaitan dengan analisis aliran trafik, analisis data penerokaan (*exploratory data analysis* EDA) biasa digunakan dalam sains data. Aktiviti EDA menyiasat data untuk menemui pola, atau untuk mengesan anomali, atau untuk menguji hipotesis dan untuk menyemak andaian dengan bantuan statistik ringkasan dan perwakilan grafik. Analisis data penerokaan dalam aliran trafik diterangkan dalam Rajah 2.15.



Rajah 2.15 Analisis Data Eksplorasi (Exploratory data analysis EDA) dan kaedah yang berkaitan dalam kajian aliran trafik.

i. Ringkasan data dan visualisasi

Ringkasan data termasuk kedua-dua *univariate* (cth. kecenderungan memusat, serakan) dan *multivariate* (cth., korelasi), adalah pelengkap kepada ringkasan data, yang merupakan pendekatan ringkas untuk memahami arah aliran, pola dan anomali dalam data. Pengvisualan membantu menganalisis aliran trafik dengan pemahaman yang lebih baik. Terdapat kajian lepas bahawa aliran trafik menggunakan visualisasi dalam menganalisis. Kajian yang dilakukan oleh (Z. Wang et al. 2014) mengkaji data trafik menggunakan graf dinamik, histogram dan animasi tempatan untuk menunjukkan volume, kelajuan dan trafik yang tidak normal. Adapun kajian yang dilakukan oleh (Huang et al. 2015) menukar trajektori GPS kepada *PageRank* dan *Betweenness* untuk menganalisis banyak trajektori yang menghubungkan antara satu sama lain. Sementara itu kajian lain yang dilakukan oleh (Z. Wang et al. 2013), menganalisis kesesakan dan perambatan kesesakan trafik dengan pengvisualan.

ii. Pengurangan dimensi

Oleh kerana set data memiliki dimensi yang tinggi atau memerlukan bilangan parameter yang banyak untuk dianggarkan tetapi tidak semua ciri dalam data (dimensi tinggi) adalah relevan secara signifikan dalam model. Pengurangan dimensi akan mempercepatkan proses pembelajaran model, membantu meningkatkan ketepatan ramalan, dan penggugusan, membantu mengeluarkan ciri bermaklumat atau tidak bermaklumat, dan membantu memadamkan koefisien anggaran dengan set ciri input

baharu (Roweis & Saul 2000). Pengurangan dimensi juga membantu memudahkan masalah dan mengoptimumkan prestasi dari algoritma (Gebresilassie 2017). Secara umum, terdapat dua matlamat utama untuk pengurangan dimensi dalam meramal aliran trafik: (1) pemilihan ciri, di mana pemboleh ubah penting dikenal pasti dan dipilih; dan (2) penggugusan, di mana pemerhatian yang serupa dikumpulkan bersama.

iii. Pemilihan ciri

Salah satu langkah yang disyorkan sebelum penggunaan model statistik dan pembelajaran mesin ialah mengenal pasti dan menggunakan hanya pemboleh ubah/ciri yang dianggap penting untuk analisis sejak ini, kerana ia (Abellán et al. 2013): (a) mengelakkan pepadanan yang berlebihan, (b) mengurangkan kerumitan pengiraan dalam analisis, dan (c) membawa kepada prestasi ramalan yang lebih baik. Langkah ini sering dirujuk sebagai pemboleh ubah atau pemilihan ciri. Dalam konteks model ramalan aliran trafik, pemilihan pemboleh ubah memainkan peranan penting kerana terdapat banyak peramal yang berpotensi (cth., trafik, cuaca, kemalangan, kelajuan, ketumpatan, hari istimewa, trajektori GPS)(H. Dai & Yang 2006), (K. Lee et al. 2014), (J. Lee et al. 2015), (R. Zhang et al. 2015), yang mungkin mempunyai kesan ke atas kebarangkalian kesesakan. Kajian terdahulu menunjukkan bahawa pemilihan ciri/pemboleh ubah meningkatkan prestasi model (Abellán et al. 2013).

iv. Penggugusan

Penggugusan adalah berkaitan pembahagian set data kepada gugusan atau kumpulan yang serupa ciri-ciri. Berasaskan variasi data dalam ciri/lajur yang berbeza, proses penggugusan menentukan titik data yang serupa dan meletakkannya dalam kelompok. Contohnya ialah mengelompokkan segmen jalan (X. Lin & Xu 2020), penyekatan rangkaian jalan boleh dianggap sebagai proses penggugusan segmen dengan atribut yang sama, dan dengan itu, algoritma penggugusan boleh digunakan untuk membahagikan sub-kawasan rangkaian jalan raya. Penggugusan membahagikan set data kepada beberapa kelompok (kelas) dengan atribut dan ciri yang serupa. Atribut atau ciri objek dalam kelompok yang sama adalah serupa antara satu sama lain tetapi agak berbeza daripada yang terdapat dalam gugusan lain. Algoritma penggugusan

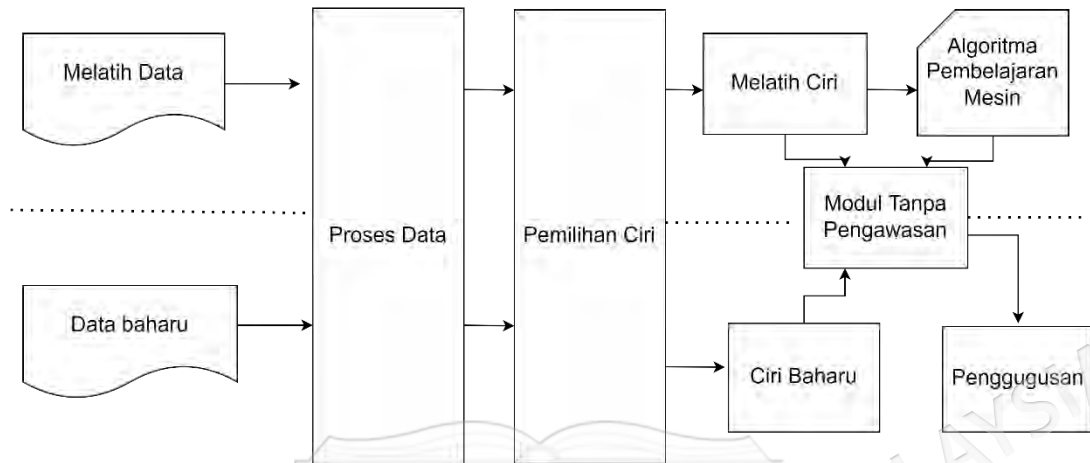
telah digunakan untuk pengenalan keadaan trafik, rangkaian jalan raya, tempoh kawalan trafik, titik kemalangan dan sekatan aliran, dan medan trafik yang lain.

Banyak algoritma yang popular digunakan dalam analisis aliran trafik, salah satu algoritma yang popular ialah K-Means (Maurya et al. 2011), (Montazeri-Gh & Fotouhi 2011), (Deb Nath et al. 2010), (Mondal & Rehena 2019). K-Means ialah algoritma ringkas di mana bilangan k gugusan ditentukan di awal dan algoritma mencari gugusan optimum dengan menetapkan titik kepada setiap gugusan mengikut jarak dari pusat setiap gugusan.

Satu lagi algoritma popular yang digunakan dalam analisis trafik adalah DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) (Wen et al. 2019), (Huang et al. 2015). Perbezaan dengan DBSCAN, bilangan gugusan tidak perlu dinyatakan di awal seperti dalam K-Means. Algoritma mencari kawasan dalam ruang ciri di mana ketumpatan titik adalah tinggi. Algoritma penggugusan popular lain dalam analisis aliran trafik ialah penggugusan spektrum (Aven 2017), (Yang et al. 2017), penggugusan spektrum membina graf persamaan untuk menerangkan hubungan antara segmen jalan. Dimensi data asal dikurangkan melalui vektor eigen graf Laplacian. Berasaskan paparan diatas didapati bahawa setiap algoritma mempunyai cara yang berbeza untuk mencari gugusan, namun idea utamanya adalah sama iaitu algoritma mencari titik data yang serupa secara statistik dan mengumpulkannya sebagai gugusan.

Pengurangan dimensi dan penggugusan ialah pembelajaran tanpa pengawasan, iaitu pembelajaran algoritma sampel tanpa label kelas atau tiada data berlabel yang ditentukan terlebih dahulu (Prasad et al. 2018) untuk menemui maklumat struktur tersembunyi dalam set data sampel (Xiao et al. 2017). Pembelajaran tanpa pengawasan (*unsupervised learning*) menumpukan kepada mencari pola generik dalam set data dan menggunakan ini untuk mendapatkan cerapan. Terdapat beberapa proses yang mungkin dilakukan dalam pembelajaran tanpa pengawasan. Proses ini ialah mengekstrak pola, menentukan perhubungan, dan mengelompokkan. Sebagai contoh penggugusan, data pelatihan dan data baru diproses, seterusnya dilakukan

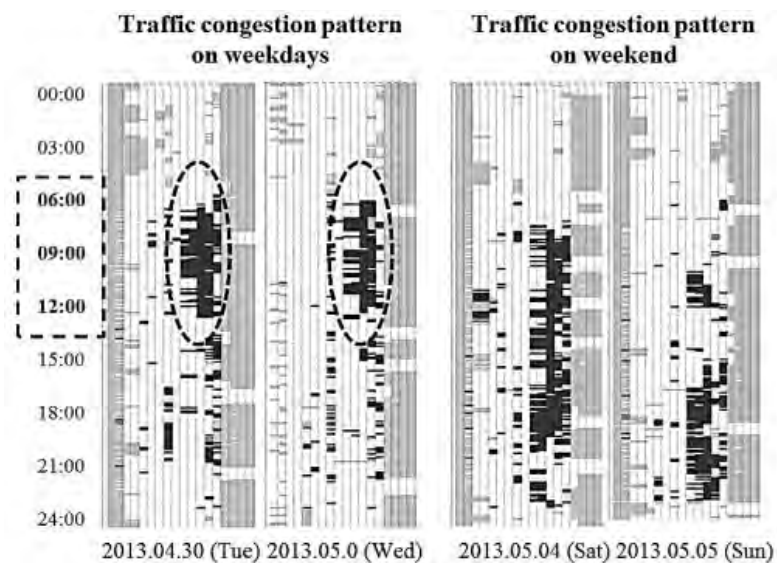
pemilihan ciri, untuk seterusnya dikelompokkan menggunakan algoritma penggugusan. Proses saluran pembelajaran mesin tanpa pengawasan, diringkas dalam Rajah 2.16.



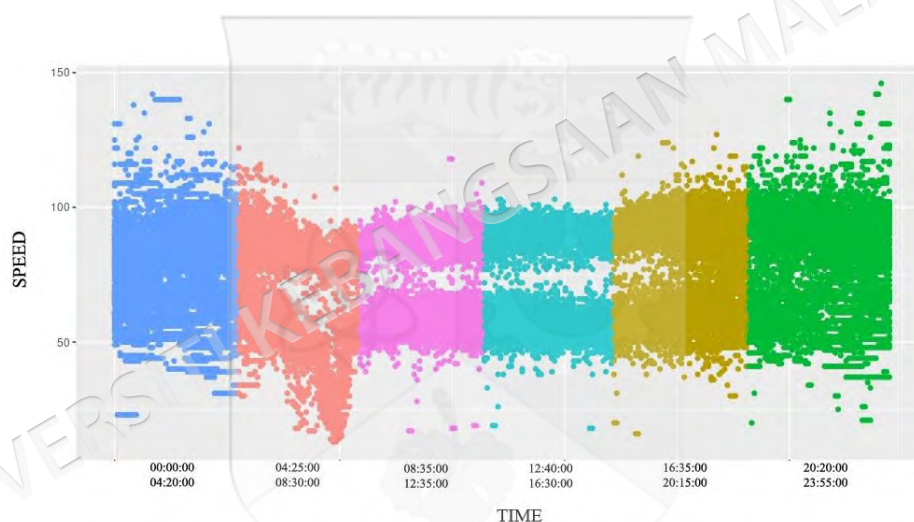
Rajah 2.16 Pembelajaran mesin tanpa pengawasan.

2.5.3 Pola aliran trafik yang sama

Kajian terdahulu mendapati bahawa aliran trafik mempunyai aliran trafik yang berbeza dalam masa. Kajian yang dilakukan oleh (K. Lee et al. 2014), (Xiaomeng Wang et al. 2015), (Keler et al. 2017), (K. Zhang et al. 2017) menemukan bahawa aliran trafik mempunyai pola berbeza pada hari kerja dan hujung minggu Rajah 2.17. Tambahan pula, aliran trafik mempunyai pola kesesakan dalam masa dan hari Rajah 2.18.

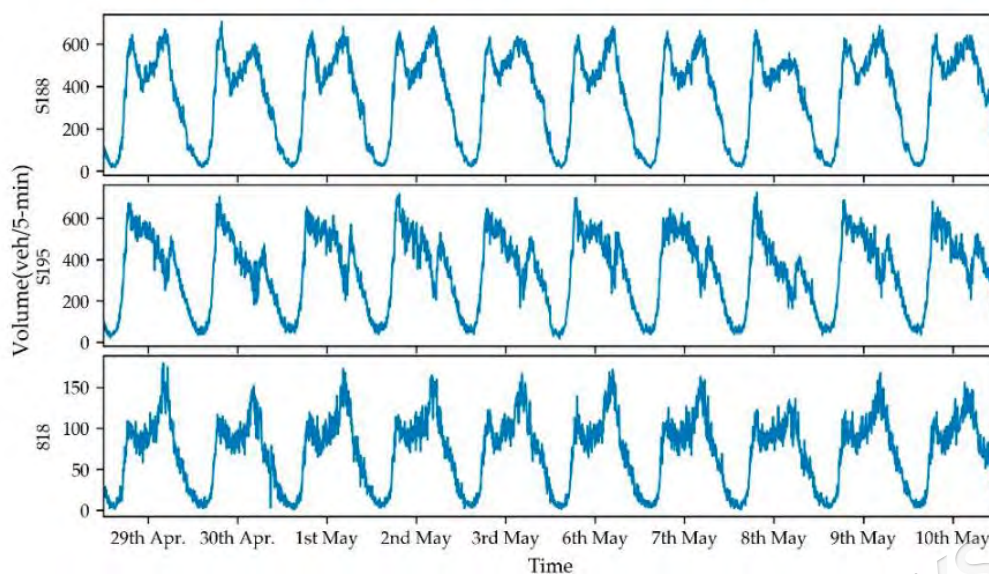


Rajah 2.17 Pola kesesakan jalan di hari kerja dan hujung minggu (K. Lee et al. 2014),(Xiaomeng Wang et al. 2015).



Rajah 2.18 Pola kesesakan jalan dalam satu hari.

Bahagian ini menjelaskan pola aliran trafik diulang setiap hari. Kajian terdahulu oleh (Zhou et al. 2020), menjumpai bahawa pola trafik berbeza setiap hari, pola trafik ditunjukkan dalam Rajah 2.19.



Rajah 2.19 Pola aliran trafik sedang berulang setiap hari (Zhou et al. 2020).

Kajian lain yang dilakukan (Rempe et al. 2016), menemukan bahawa pola aliran trafik berbeza pada waktu pagi, tengahari dan petang, ditunjukkan dalam Rajah 2.20. Oleh kerana itu untuk ramalan aliran trafik di hari kerja, sebaiknya menggunakan data trafik di hari kerja, begitu juga sebaliknya.



Rajah 2.20 Pola aliran trafik di Munich pada waktu pagi dan petang (Rempe et al. 2016).

2.5.4 Ramalan aliran trafik

Secara umum, terdapat dua model ramalan dalam aliran trafik yang kedua-duanya dicirikan sebagai pembelajaran diselia. Pembelajaran diselia ialah apabila pemboleh ubah output dalam data latihan tersedia untuk data pemerhatian. Data latihan berlabel digunakan dalam algoritma pembelajaran terselia (L. Zhu et al. 2019). Model menggunakan data input dan output sasaran (label) untuk mempelajari fungsi atau penyesuaian di antara mereka. Output yang tidak kelihatan boleh diramalkan dengan model yang dipelajari dan data input. Dua jenis pembelajaran terselia yang digunakan secara meluas dalam meramal aliran trafik ialah regresi dan klasifikasi.

Regresi

Model regresi meramal nilai berangka atau kuantitatif. Data berlabel terdiri daripada nilai output yang dijangkakan or Y_s . Fungsi regresi adalah untuk menerangkan hubungan antara satu pemboleh ubah bersandar dengan satu atau lebih pemboleh ubah tak bersandar (L. Zhu et al. 2019). Contohnya, dalam meramal aliran trafik, meramal kelajuan pemboleh ubah bersandar Y_s pemboleh ubah ialah nilai sebenar yang perlu diramal, dan model cuba memetakan X_s (jalan berjiran dan masa) untuk meramal nilai sebenar daripada Y_s (kelajuan) (E.-M. Lee et al. 2007), sedangkan kajian oleh (Barbosa et al. 2013) menggunakan lintasan GPS sebagai pemboleh ubah X_s . Kajian lain oleh (Fohrholz & Glaschke 2016) meramal nilai sebenar jumlah perjalanan, Juga didapati kajian oleh (The & Route 2014) cuba meramal nilai sebenar kepadatan (*density*) jalan raya. Pelbagai algoritma pembelajaran mesin digunakan dalam meramal aliran trafik menggunakan kaedah regresi. Dalam kajian ini juga melakukan ramalan aliran trafik menggunakan kaedah regresi dalam menyelidiki aras kesesakan masa dalam meningkatkan ramalan kelajuan kenderaan.

Klasifikasi

Model klasifikasi bertujuan untuk meramal nilai daripada set kelas atau kategori terhitung (walaupun mungkin besar). Matlamatnya adalah untuk meramal kelas sebagai output. Terdapat dua atau lebih kelas yang boleh menjadi hasil dan input

memetakan algoritma X_s untuk meramal kelas. Model menggunakan data input dan output sasaran (label) untuk mempelajari fungsi atau pemetaan di antara mereka. Dengan menggabungkan model data pelatihan dan data input, output yang tidak kelihatan boleh diramalkan (L. Zhu et al. 2019). Sebagai contoh daripada meramal keadaan trafik, adalah kajian oleh (Xiaomeng Wang et al. 2015) meramal keadaan trafik berdasarkan aliran trafik di jalan raya jiran. Satu lagi kajian meramal keadaan trafik berdasarkan data cuaca (J. Lee et al. 2015).

Model klasifikasi menggunakan data input dan output sasaran (label) untuk mempelajari fungsi atau pemetaan kedua-duanya. Model klasifikasi boleh dibahagikan lagi kepada jenis pemboleh ubah output dan bilangan pemboleh ubah output yang dihasilkan olehnya:

1. Klasifikasi binari (*Binary classification*): set data hanya mempunyai jumlah dua kategori atau hanya mempunyai dua label kelas untuk membezakan antara pemboleh ubah tindak balas output dalam data, maka masalah itu dinamakan sebagai masalah klasifikasi binari.
2. Klasifikasi multi-kelas (*Multi-Class classification*): Ini adalah lanjutan daripada masalah klasifikasi binari. Dalam kes ini set data mempunyai lebih daripada dua kategori atau kelas.
3. Klasifikasi multi-label (*Multi-Label classification*): Masalah klasifikasi ini biasanya melibatkan data yang pemboleh ubah output tidak selalunya satu nilai tetapi satu vektor yang mempunyai berbilang nilai atau label.

Penjelasan lanjut bahawa model klasifikasi sering mengeluarkan label kelas sebenar atau kebarangkalian untuk setiap label kelas yang mungkin yang memberikan tahap keyakinan untuk setiap kelas dalam ramalan. Berikut ialah format keluaran utama daripada model klasifikasi.

- a) Klasifikasi kategori: Dalam sesetengah model klasifikasi, luaran untuk mana-mana titik data yang tidak diketahui ialah kategori atau label kelas yang diramalkan. Model ini biasanya mengira kebarangkalian

semua kategori tetapi melaporkan hanya satu label kelas yang mempunyai kebarangkalian atau keyakinan maksimum.

- b) Klasifikasi kebarangkalian: Dalam model klasifikasi ini, output ialah nilai kebarangkalian bagi setiap label kelas yang mungkin. Model-model ini penting apabila memerlukan lebih banyak tindakan, seterusnya menggunakan output yang dihasilkan oleh model klasifikasi untuk analisis terperinci atau untuk membuat keputusan yang kompleks.

Model klasifikasi yang dicadangkan dalam kajian ini menggunakan klasifikasi binari, iaitu model ramalan digunakan untuk meramal keadaan jalan sesak atau lancar.

2.6 KENAL PASTI KESESAKAN JALAN RAYA

Bahagian ini menerangkan kajian terdahulu bagaimana menentukan kesesakan jalan raya. Dalam meramal aliran trafik, menentukan tahap kesesakan jalan raya adalah penting (Afrin & Yodo 2021). Akibatnya, pelbagai definisi kesesakan trafik telah digariskan. Kedudukan kesesakan trafik ditakrifkan oleh Rothenberg (Yingya Zhang et al. 2016), sebagai keadaan di mana bilangan kenderaan di jalan melebihi kapasiti penyimpanan tahap perkhidmatan jalan raya. Dengan kata lainnya kesesakan trafik dianggap sebagai keadaan di mana bilangan pengguna jalan melebihi kapasiti jalan raya. Sesetengah ciri-ciri kesesakan jalan raya termasuk beratur panjang, masa perjalanan yang panjang, dan kelajuan perlahan kenderaan di jalan raya. Kajian yang dilakukan oleh (Min & Wynter 2011) membahagikan kelajuan kepada empat tahap (24, 48, 72, 120) untuk menentukan keadaan trafik lebuhraya, kajian lain yang dilakukan oleh (J. Ahn et al. 2016) mengelaskan kelajuan menjadi dua belas (12) tahap untuk menentukan trafik lebuhraya (KEC). Kajian lain oleh (Kong et al. 2016) membahagikan kelajuan menjadi lima tahap (*serious congestion, moderate congestion, mild congestion, smooth, and very smooth*) untuk menentukan trafik bandar. Kajian lain yang dilakukan oleh (John FW Zaki, Amr Ali Eldin, Sherif E. Hussein, Sabry F. Saranya 2016) menggunakan kelajuan dan densiti dengan membahagi tahap kepada tiga tahap untuk menentukan trafik lebuhraya, sedangkan

kajian yang dilakukan oleh (Zaki et al. 2019) membahagi kelajuan menjadi sembilan tahap untuk menentukan trafik lebuh raya. Kajian lain yang dilakukan oleh (Qi & Ishak 2014) membahagi faktor kelajuan menjadi dua belas (12) tahap untuk menentukan trafik lebuh raya. Terdapat juga kajian yang dilakukan oleh (Rapant et al. 2016) menentukan keadaan trafik trafik lebuh raya menggunakan tahap perkhidmatan standard jalan bandar di Amerika. Kajian lain yang dilakukan oleh (He et al. 2016) menggunakan indeks prestasi kelajuan yang digunakan untuk menentukan keadaan trafik rangkaian lebuh raya Beijing, dan kajian lain oleh (W. X. Wang et al. 2018) menggunakan indeks kesesakan (*congestion index*) untuk menentukan keadaan lebuhraya bandar (jalan shugang).

Indeks prestasi kelajuan (*speed performance index*)

Indeks prestasi kelajuan dikira berasaskan kelajuan perjalanan dan kelajuan jalan maksimum yang dibenarkan (He et al. 2016). Persamaan indeks prestasi kelajuan diberikan seperti berikut:

$$R_v = \frac{v}{V_{max}} \times 100 \quad \dots(2.1)$$

R_v : Indeks prestasi kelajuan (*The speed performance index*).

v : Purata kelajuan perjalanan (*the average travel speed*) km/h;

V_{max} : Kelajuan maksimum yang dibenarkan di jalan km/h

Indeks kesesakan (*congestion index*)

Indeks kesesakan diperolehi berasaskan kelajuan purata kenderaan, untuk menilai kesesakan trafik jalan tunggal atau rangkaian jalan raya. (W. X. Wang et al. 2018). Langkah pengiraan adalah seperti berikut:

- a) Andaikan tempoh statistik dibahagikan kepada hari n . Setiap hari dibahagikan kepada selang masa n , dan terdapat SR jalan raya dalam kawasan kajian.

- b) Dua (2) simbol V_{kij} dan V'_{kij} masing-masing mewakili kelajuan perjalanan yang boleh diterima, dan perjalanan minimum. Pemilihan V_{kij} and V'_{kij} 0 bahagian k jalan, pada hari i dan selang masa j.
- c) Kelajuan pada bahagian jalan boleh ditentukan mengikut data statistik. V_{kij}^R ditakrifkan sebagai kelajuan medan perjalanan lapangan di bahagian jalan k, pada hari i dan selang masa j. Bandingkan V_{kij}^R dengan V_{kij} . Jika $V_{kij}^R > V_{kij}$, trafik jalan raya adalah lancar dan indeks kesesakan trafik adalah $I_{kij} = 0$; jika $V_{kij}^R \leq V_{kij}$, keadaan adalah sesak dan $I_{kij} = 10$. Di samping itu. Fungsi indeks kesesakan trafik adalah seperti berikut:

$$I_{kij} = \frac{V_{kij} - V_{kij}^R}{V_{kij} - V'_{kij}} \times 10 \quad \dots(2.2)$$

- d) Indeks kesesakan trafik I_{ki}^d untuk hari i pada segmen jalan k adalah sebagai berikut:

$$I_{ki}^d = \sum_{j=1}^N b_j^d \times I_{kij} \quad \dots(2.3)$$

Dimana b_j^d adalah perkadaran jumlah trafik semasa selang masa j pada sepanjang hari.

- e) Indeks kesesakan trafik I_i^d untuk hari i di kawasan adalah seperti berikut:

$$I_i^d = \sum_{k=1}^N a_k^d \times I_d \quad (2.4)$$

dimana a_k^d adalah perkadaran kadar di aliran bahagian jalan k pada semua jalan di seluruh jalan kawasan. Begitu juga, indeks kesesakan trafik I_{Area}^d Untuk keseluruhan tempoh tinjauan di kawasan jalan raya adalah seperti berikut:

$$I_{Area}^d = \sum_{j=i}^N q_i \times I_i^d \quad (2.5)$$

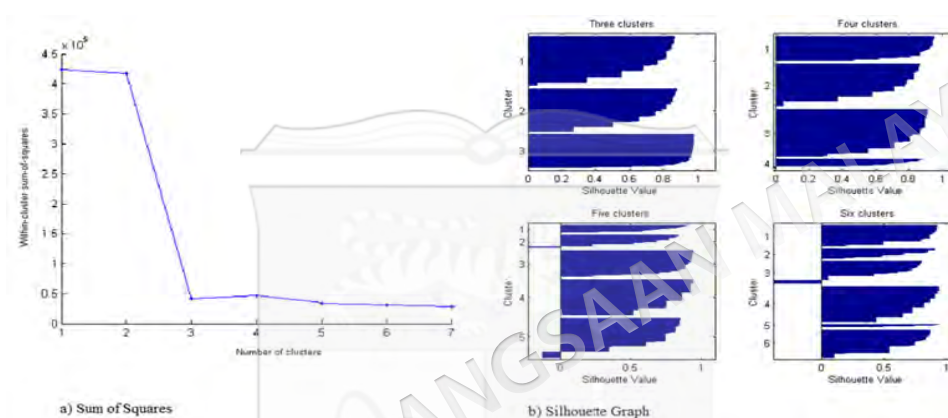
Berasaskan kajian lepas dalam penentuan keadaan trafik didapati bahawa, kaedah yang paling mudah dalam menentukan kesesakan trafik adalah dengan membahagikan kelajuan purata kepada beberapa tahap. Kaedah lanjutan seterusnya menggunakan indeks prestasi kelajuan. Kaedah ini menggabungkan kelajuan purata dengan kelajuan maksimum yang dibenarkan di jalan raya. Namun kaedah ini masih menggunakan hanya faktor kelajuan. Dalam keadaan trafik sebenar, kesesakan jalan raya bukan hanya dipengaruhi oleh kelajuan purata tetapi juga dengan jumlah kenderaan yang melalui jalan. Lebih canggih lagi, kesesakan jalan juga ditentukan oleh masa memandu biasa pada jalan tersebut. Kajian ini menggunakan indeks kesesakan untuk menentukan kesesakan trafik. Seperti yang dijelaskan dalam rumus 2.2 hingga rumus 2.5 di atas indeks kesesakan menggabungkan faktor kelajuan purata, jumlah dan masa memandu biasa (normal driving time) dalam menentukan kesesakan trafik.

2.7 PENGUGUSAN JALAN RAYA

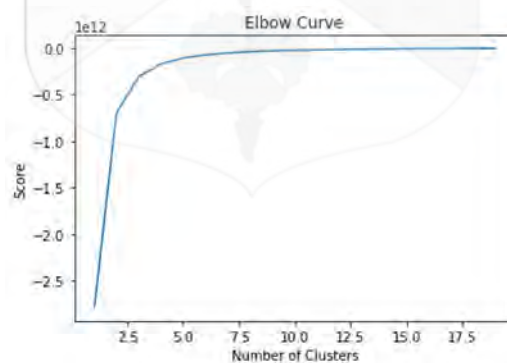
2.7.1 Penggugusan segmen jalan menggunakan K-Means

K-Means digunakan secara meluas dalam kajian sistem pengangkutan pintar. Seperti kajian yang dilakukan oleh (Mondal & Rehena 2019) penggugusan segmen jalan raya menggunakan K-Means berasaskan kepadatan dan kelajuan. Terdapat juga kajian yang menggunakan K-Means untuk penggugusan keadaan trafik berasaskan ciri memandu, dilakukan oleh (Montazeri-Gh & Fotouhi 2011). Kajian lain oleh (Necula 2015) penggugusan segmen jalan menggunakan K-Means berasaskan lintasan GPS. Kajian lain oleh (Ghadi & Török 2019) mengelompokkan segmen jalan berasaskan

keselamatan jalan raya. Terdapat juga kajian yang meningkatkan K-Means dengan kanopy K-Means yaitu kajian yang dilakukan oleh (X. Lin & Xu 2020). Untuk memberikan penilaian prestasi hasil penggugusan, beberapa kaedah dapat digunakan seperti *sum of squares*, graf *silhouette* (Zaki et al. 2019), (Montazeri-Gh & Fotouhi 2011), dan lengkung siku (Laña et al. 2016) (Mondal & Rehena 2019) digunakan untuk menilai hasil yang diperolehi dalam gugusan untuk mencari bilangan k yang optimum. Graf pemerhatian, *sums of squares* dan graf *silhouette* ditunjukkan dalam Rajah 2.21, lengkung siku ditunjukkan dalam Rajah 2.22.



Rajah 2.21 Contoh mencari k terbaik a) sum of squares(Zaki et al. 2019) b) graf *silhouette* (Zaki et al. 2019).



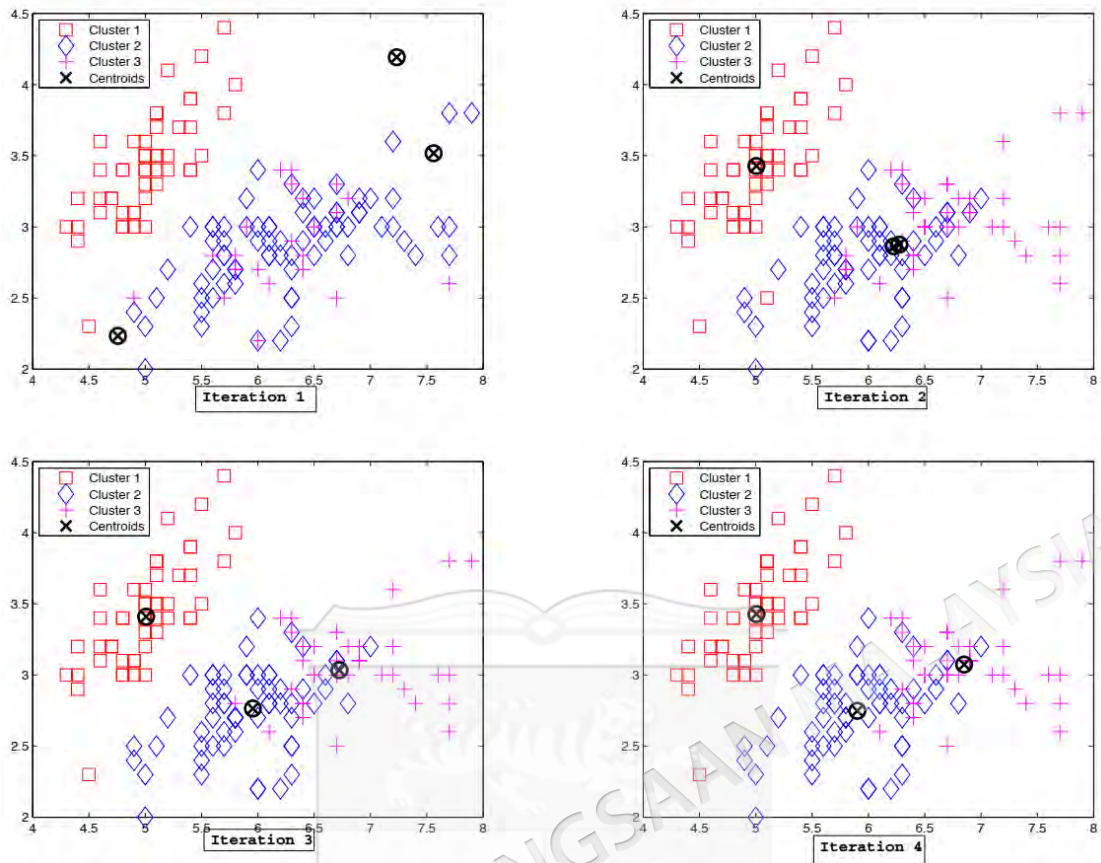
Rajah 2.22 Contoh mencari k terbaik menggunakan lengkung siku (Mondal & Rehena 2019).

K-Means adalah algoritma kumpulan data yang akan membahagikan data ke dalam beberapa kumpulan k (Kesumawati & Setianingsih 2016). K-Means adalah algoritma yang sangat terkenal kerana kesederhanaannya (*simplicity*). Algoritma ini membahagikan data kepada k kumpulan. Dimana data dari kumpulan akan

mempunyai ciri yang sama antara setiap ahlinya. Sebaliknya, data dalam kumpulan akan mempunyai watak yang berbeza dari ahli kumpulan lain. Algoritma ini mencari variasi yang sangat minimum antara kumpulan yang sama dan akan mencari variasi yang maksimum antara kumpulan lain (Kesumawati & Setianingsih 2016). Algoritma bermula dengan memilih nilai titik perwakilan sebagai sentroid awal. Setiap titik kemudian dipadankan dengan sentroid terdekat berdasarkan jarak tertentu yang dipilih. Apabila gugusan terbentuk, sentroid untuk setiap gugusan dikemas kini. Kedua-dua langkah ini diulang sehingga sentroid tidak berubah atau kriteria penumpuan alternatif lain dipenuhi. Penggugusan K-Means adalah algoritma yang dijamin penumpuan kepada minimum tempatan tetapi peminimuman fungsi skornya dikenali sebagai NP-Hard (Kumar 2014). Struktur asas algoritma K-Means, diterangkan sebagai berikut :

1. Pilih titik k sebagai sentroid awal
2. Ulangi langkah 3 dan 4 sehingga kriteria penumpuan diperoleh
3. Buat k Gugusans dengan memadankan setiap titik kepada sentroid yang paling dekat
4. Ulangi kiraan sentroid dari setiap gugusan

Rajah 2.23 memberikan ilustrasi tahap yang berlainan dalam algoritma K-Means. Iterasi pertama dimulakan dengan tiga titik yang diambil secara rawak sebagai sentroid. Dalam lelaran berikutnya, sentroid menukar kedudukan sehingga mencapai penumpuan. Pelbagai ukuran jarak boleh digunakan dalam algoritma K-Means semasa mengira centroid yang paling dekat. Pilihan ini boleh memberi kesan terhadap pepadanan akhir. Jenis ukuran yang boleh digunakan untuk mengukur jarak adalah: jarak *manhattan* (L_1 norm), jarak *euclidean* (L_2 norm), and kesamaan kosinus. Secara umum, matrik jarak *euclidean* adalah pilihan yang paling popular dalam penggugusan K-Means.



Rajah 2.23 Ilustrasi tahap yang berlainan dalam algoritma K-Means (Kumar 2014).

Hasil gugusan yang berbeza boleh diperolehi untuk nilai-nilai yang berlainan seperti pemilihan k dan pemilihan matrik jarak yang berbeza. Terdapat fungsi objektif yang digunakan oleh K-Means iaitu *Sum of Squared Errors* (SSE) atau *Residual Sum of Squares* (RSS). Rumus matematik SSE/RSS adalah seperti persamaan berikut(2.6).

$$SSE(C) = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - c_k\|^2 \quad (2.6)$$

$$c_k = \frac{\sum_{x_i \in C_k} x_i}{|C_k|}$$

Peminimuman *sum of squared errors*

Penggugusan k-Means pada dasarnya adalah masalah pengoptimuman dengan matlamat meminimumkan fungsi objektif SSE. Persamaan matematik (2.7) membuktikan sebab di sebalik memilih min titik data dalam kelompok sebagai wakil prototaip untuk Gugusan dalam algoritma K-Means, C_k sebagai k Gugusan, x_i adalah titik dalam C_k , dan c_k adalah min dari Gugusan. Masalah boleh diselesaikan untuk wakil dari C_j yang meminimumkan SSE dengan membezakan SSE berkenaan terhadap c_j kemudian samakan dengan nilai nol (0) seperti yang dapat dilihat pada persamaan (2.7)

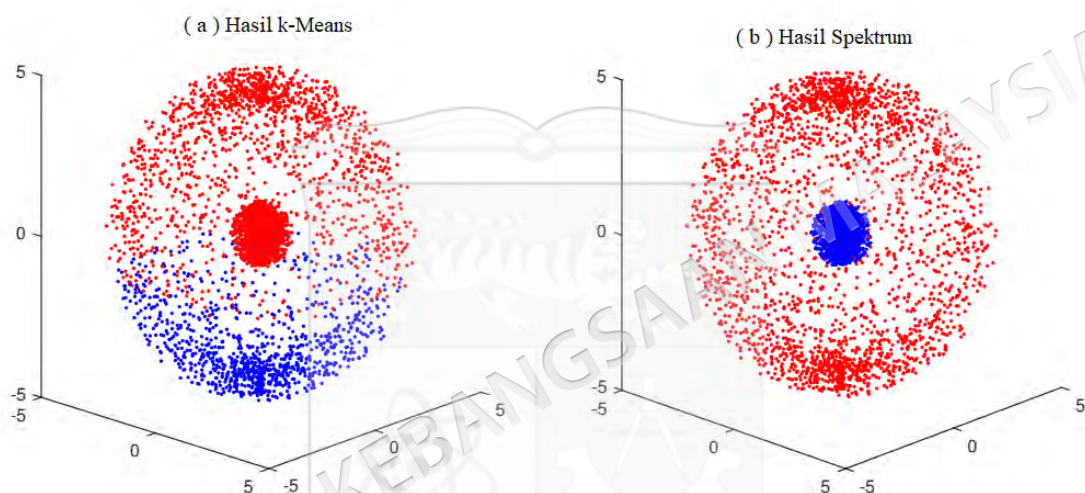
$$\begin{aligned}
 \text{SSE}(C) &= \sum_{k=1}^K \sum_{x_i \in C_k} \|c_k - x_i\|^2 \\
 \frac{\partial}{\partial c_j} \text{SSE}(C) &= \frac{\partial}{\partial c_j} \sum_{k=1}^K \sum_{x_i \in C_k} \|c_k - x_i\|^2 \\
 &= \sum_{k=1}^K \sum_{x_i \in C_k} (c_j - x_i)^2 \\
 &= \sum_{x_i \in C_k} 2 * (c_j - x_i) = 0 \\
 &= \sum_{x_i \in C_k} 2 * (c_j - x_i) = 0 \rightarrow |C_j| \cdot c_j = \sum_{x_i \in C_j} x_i \rightarrow c_j = \frac{\sum_{x_i \in C_j} x_i}{C_j}
 \end{aligned} \tag{2.7}$$

Oleh itu, wakil terbaik untuk meminimumkan SSE gugusan ialah nilai min dari titik dalam Gugusan. Dalam K-Means, SSE menurun secara seragam dengan setiap lelaran. Penurunan secara seragam ini akhirnya akan menumpu kepada satu minimum setempat.

Berasaskan penerangan diatas didapati bahawa yang boleh memberi kesan kepada prestasi algoritma K-Means ialah pemilihan sentroid awal dan anggaran bilangan kelompok k.

2.7.2 Penggugusan spektrum

Selain daripada menggunakan penggugusan K-Means, beberapa kajian aliran trafik memilih menggunakan penggugusan spektrum untuk mengelompokkan jalan raya seperti kajian oleh (Yang et al. 2017), (Xiaomeng Wang et al. 2015), (Aven 2017). Secara umumnya, algoritma ini sering digunakan untuk mengelompokkan dataset yang tidak boleh dipisahkan dengan mudah. Rajah 2.24 menggambarkan kelebihan penggugusan spektrum daripada algoritma K-Means.



Rajah 2.24 (a) Hasil penggugusan k-Means dan (b) Hasil penggugusan Spektrum (Aven 2017).

Cara penggugusan spektrum ini boleh diterangkan dalam tiga langkah berikut ini:

1. Langkah pertama ialah membina graf kesamaan untuk semua titik data.
2. Titik data ditenamkan (*embedded*) dalam ruang, yang merupakan Gugusannya adalah lebih "jelas", dengan menggunakan vektor eigen bagi graf Laplacian.
3. Akhir sekali, algoritma penggugusan K-Means digunakan untuk membahagikan yang terbenam (*embedded*).

Perwakilan dimensi rendah yang diperoleh dalam langkah kedua juga dirujuk sebagai "pembenaman spektrum" (*spectral embedded*) dan mempunyai aplikasi di luar konteks Gugusan, seperti pengurangan dimensi. Perkataan spektrum digunakan untuk menunjukkan fakta bahawa hasil Penggugusan diperoleh dengan menganalisis spektrum bagi graf Laplacian.

Graf kesamaan

Sebagai contoh diberi titik data, yang ingin dibahagikan kepada subset k , yang dilambangkan dengan $X = \{x_i, \dots, x_n\}$ dalam R_m . Untuk melakukan Penggugusan spektrum, pertama sekali perlu mewakili data ini dalam bentuk "graf kesamaan" tidak terarah $G = (V, E)$. Disini setiap titik data x_i diwakili oleh bucu v_i , dan E merujuk kepada tepi antara bucu. Perhatikan bahawa x_i ialah vektor yang menandakan titik data, sementara v_i ialah bucu tanpa sebarang atribut. Kemudian afiniti wajaran bukan negatif digunakan dengan n matrik bersebelahan (atau matriks afiniti) W untuk menerangkan G , dimana $W = \{W_{ij}\}_{i,j=1,\dots,n}$.

Perhatikan bahawa W_{ij} sama dengan 0 ketika bucu v_i dan v_j tidak bersambung. Memandangkan algoritma Penggugusan spektrum bertujuan untuk membahagikan bucu supaya membiarkan kelompok yang sama mempunyai persamaan yang tinggi dan Gugusan yang berbeza mempunyai persamaan yang rendah, maka adalah penting untuk memilih kaedah yang berkesan untuk membina matrik bersebelahan. Tiga cara untuk membina matriks W diterangkan seperti berikut:

- 1) Graf jiran terdekat (*k-nearest neighbor*): Ideanya adalah v_i bersambung dengan v_j ketika v_j adalah antara k -nearest neighbors dari v_i , atau v_i adalah antara k -nearest neighbors dari v_j . Jarak dikira berasaskan perwakilan asal titik data titik x_i dan x_j . Beberapa contoh termasuk 1-norm, 2-norm, dan jarak kosinus. Graf yang terhasil biasanya dipanggil graf *k-nearest neighbour*. Alternatifnya ialah dengan menyambung v_i and v_j apabila mereka saling berjiran antara satu sama lain. Graf ini dirujuk sebagai graf mutual *k-nearest neighbour* atau graf simetri jiran

terdekat. Dalam kedua-dua kes, selepas menambah tepi mengikut kejiranan setiap bucu, pemberat ditetapkan pada tepi mengikut persamaan titik akhirnya.

- 2) Graf kejiranan ε (ε -neighborhood): Dalam graf jenis ini, bucu disambungkan hanya apabila jarak berpasangan $\|x_i - x_j\|^2$ adalah lebih kecil dari ε . Walau bagaimanapun, kaedah ini sering membawa kepada graf dengan komponen terputus jika ε tidak dipilih dengan teliti.
- 3) Graf bersambung sepenuhnya: Dalam kes ini, semua bucu disambungkan dengan kesamaan positif. Memandangkan matrik bersebelahan harus memodelkan kejiranan tempatan, pemilihan fungsi persamaan itu sendiri menjadi rumit.

Sebagai contoh, fungsi kernel haba (heat kerne)l: $W_{ij} = e^{-\frac{\|x_i - x_j\|^2}{\sigma^2}}$ dimana ϑ mengawal keluasan kawasan kejiranan. Adalah dicadangkan bahawa ϑ boleh ditala secara tempatan terhadap pasangan bucu. Selain daripada masalah memilih nilai yang sesuai bagi σ , pembinaan graf jenis ini juga mengalami masalah kecekapan.

Pendekatan yang paling biasa digunakan untuk menangani cabaran pengiraan dan ingatan ialah memberi nilai sifat bagi beberapa elemen dalam matriks kesamaan, atau menyarangkan matrik. Apabila graf disambungkan sepenuhnya, ia akan meningkatkan kerumitan masa dan ruang.

Dalam kajian ini k-nearest neighbor graf digunakan untuk membina W kerana k-nearest neighbor graf merupakan algoritma yang mudah dan sering digunakan, seterusnya Penggugusan spektrum normal atau graf Laplacian normal mengikuti daripada (Von Luxburg 2007).

Penggugusan spektrum normal

Terdapat dua versi graf Laplacian ternormalkan: L_{sym} dan L_{rw} , L_{sym} merujuk kepada matrik simetri, dan L_{rw} boleh dijelaskan dari perspektif jalan rawak. Hubungan rapat wujud antara kedua-dua alternatif Penggugusan kesamaan spektrum ternormalkan.

Normalized Graph Laplacian

Dua variasi graf ternormal Laplacian ditakrifkan seperti berikut:

$$L_{\text{sym}} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} W D^{-\frac{1}{2}} \quad \dots(2.8)$$

$$L_{\text{rw}} = D^{-1} L = I - D^{-1} W \quad \dots(2.9)$$

Algoritma Penggugusan spektrum boleh digunakan untuk membahagikan jalan raya yang ditetapkan kepada Gugusan berasaskan data trafik terdahulu, dan jalan raya dalam kelompok yang sama akan memiliki ciri trafik yang serupa. Dalam kajian ini Penggugusan spektrum digunakan untuk mendapatkan jalan raya yang terletak dalam Gugusan yang sama dengan jalan sasaran. Dalam algoritma Penggugusan spektrum, set titik dalam ruang ciri arbitrari boleh diwakili sebagai graf bucu pemberat lengkap $G(V, E)$. Bucu graf G ialah titik dalam ruang ciri dan pemberat serta tepi $W_{ij}(v_i, v_j)$ dalam E ialah ukuran kesamaan antara bucu v_i, v_j .

Andaikan terdapat set data $X = \{x_1, x_2, \dots, x_n\}$ untuk digugusan, dan X perlu digugusankan ke dalam kumpulan k , dan parameter k perlu ditetapkan terlebih dahulu. Secara ringkasnya algoritma Penggugusan spektrum diterangkan seperti berikut:

Langkah 1. Tukar titik data asal kepada graf tidak berarah G . Bina matrik kesamaan dengan fungsi kesamaan Gaussian dalam persamaan (2.10).

$$w_{ij} = \exp \left(\frac{-\|x_i - x_j\|^2}{2\alpha^2} \right) \quad (2)$$

.10)

Langkah 2. Bina matrik darjah D , dan matriks $L' = L^{-\frac{1}{2}}WD^{-\frac{1}{2}}\{x_1, x_2, \dots, x_n\}$

Step 3. Hitung vektor eigen yang sepadan $u_1, u_2, \dots, u_k$ daripada k -nilai eigen terbesar $\lambda_1, \lambda_2, \dots, \lambda_k$ dan bina matrik $U = \{u_1, u_2, \dots, u_k\} \in \mathbb{R}^{n \times k}$ yang mengandungi vectors $u_1, u_2, \dots, u_k$ sebagai lajur.

Langkah 4. Normalkan baris U kepada vektor unit dan dapatkan matrik baharu $E \in \mathbb{R}^{n \times k}$, $e_{ij} = u_{ij} / (\sum_k u_{ik}^2)^{\frac{1}{2}}$

Langkah 5. Gugusankan baris matriks E $e_1, e_2, \dots, e_n$ matrik E ke dalam Gugusan $E_1, \dots, E_k$ menggunakan algoritma K-Means.

Langkah 6. Berasaskan keputusan Penggugusan dalam langkah 5 jika baris e_i ($i = 1, 2, \dots, n$) dikumpulkan kepada Gugusan E_j ($j = 1, 2, \dots, k$) titik asal x_j ($j = 1, 2, \dots, n$) tergolong dalam Gugusan X_j ($j = 1, 2, \dots, k$). Keputusan Penggugusan akhir kini diperolehi.

Kajian terdahulu oleh (Xiaomeng Wang et al. 2015) menyatakan adalah sukar untuk menentukan k Gugusan yang sesuai untuk gugusan segmen jalan. Kajian yang dilakukan oleh (Yang et al. 2017) menggunakan indeks Davies–Bouldin (DB) untuk mencari gugusan k terbaik, skema Penggugusan dengan nilai DB terkecil mewakili prestasi Penggugusan yang optimum.

Pendekatan lain untuk menentukan gugusan dilakukan oleh (Xiaomeng Wang et al. 2015). Mereka menggunakan pemberat untuk mengawal proses penggugusan.

pemberat pemerhatian kemudiannya diperlukan untuk melihat berapa banyak jalan raya yang berhubung yang termasuk dalam hasil gugusan. Walau bagaimanapun, pendekatan ini memerlukan set data yang mempunyai maklumat jalan yang berhubung. Tambahan pula, disebabkan pola aliran trafik yang berbeza setiap hari, Penggugusan beberapa kali untuk hari yang berbeza diperlukan untuk mengelakkan kelompok yang tidak disengajakan. Untuk mengatasi hal ini maka kaedah persilangan antara hasil gugusan pada hari berbeza dilaksanakan. Walau bagaimanapun, proses itu adalah percubaan. Proses penggugusan perlu diulangi apabila tiada jalan raya dalam kelompok atau persimpangan antara kelompok hari yang berbeza.

2.7.3 Matrik ko-kejadian aras kelabu (Gray level co-occurrence matrix)

Matriks ko-kejadian aras kelabu (GLCM) (P. Shan 2018) (Nurhaida et al. 2016) (Fatonah et al. 2020) adalah kaedah biasa untuk menganalisis tekstur. Asas kaedah ini adalah dengan melakukan kajian ciri korelasi ruang *grayscale*. Memandangkan tekstur dibentuk daripada kejadian berulang taburan kelabu dalam kedudukan ruang, akan terdapat hubungan kelabu antara dua piksel pada jarak tertentu dalam ruang imej, iaitu ciri korelasi ruang kelabu dalam imej. Histogram *grayscale* ialah hasil daripada statistik pada satu piksel yang mempunyai grayscale tertentu pada imej. GLCM diperoleh dengan menggunakan statistik keadaan dari dua piksel yang mempunyai jarak tertentu pada imej masing-masing yang mempunyai *grayscale* tertentu. Dalam kajian ini GLCM digunakan untuk mengkaji kejadian kesesakan yang berulang pada hari dan masa tertentu.

Kontras ialah ukuran intensiti antara satu nilai matriks dengan jirannya ke atas keseluruhan matriks (Suresh & Shunmuganathan 2012). Kontras dikira menggunakan persamaan (2.11) dan korelasi nilai matriks keatas jirannya diberikan oleh persamaan (2.12). Kehomogenan ialah nilai yang mengukur kedekatan taburan elemen dalam GLCM kepada pepenjuru GLCM. Kehomogenan diberikan oleh persamaan (2.14). Tenaga (*Energy*) ialah jumlah unsur kuasa dua dalam GLCM ternormal. Persamaan (2.13) digunakan untuk kiraan tenaga.

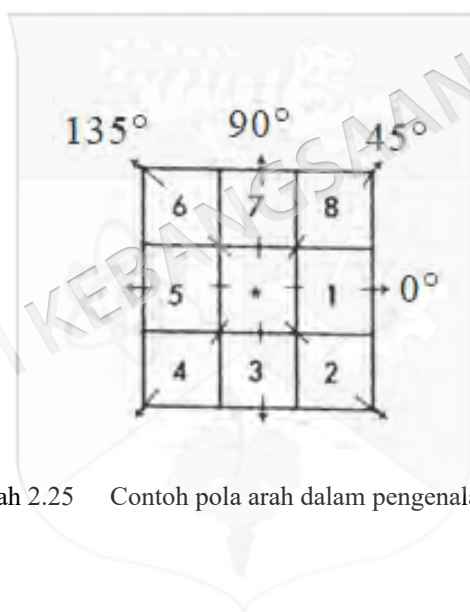
$$\text{Kontras} = \sum_{i,j} |i - j|^2 p(i, j)^2 \quad (2.11)$$

$$\text{Korelasi} = \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j} \quad (2.12)$$

$$\text{Tenaga} = \sum_{i,j} p(i, j)^2 \quad (2.13)$$

$$\text{Kehomogenan} = \sum_{i,j} \frac{(j - \mu_j)p(i, j)}{1 + |i - j|} \quad (2.14)$$

Berikut adalah contoh tiga kemungkinan pola yang biasanya digunakan dalam pengecaman imej pola arah, cara menegak dan pepenjuru, boleh digunakan dalam mencari pola aliran trafik pada hari dan masa, seperti yang ditunjukkan dalam Rajah 2.25. Darjah (135° , 90° , 45° , 0°), pola arah pepenjuru ialah (135° , 45°), pola arah menegak (90°) dan pola arah mendatar (0°).

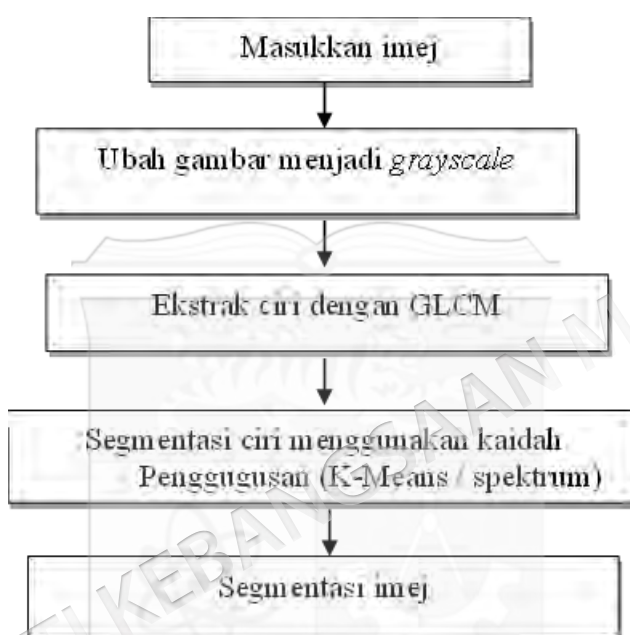


Rajah 2.25 Contoh pola arah dalam pengenalan imej.

2.7.4 Menggabungkan penggugusan dengan Gray level co-occurrence matrix

GLCM telah digunakan secara meluas digabungkan dengan kaedah penggugusan untuk pembahagian imej. Kajian terdahulu oleh (P. Shan 2018) (Bastos et al. 2008) menggabungkan GLCM dengan K-Means untuk mengelompokkan seni bina imej, pokok, kenderaan, jalan dengan manusia, jalan dengan kereta dan imej dalaman. Satu lagi kajian oleh (Oleiwi 2017) menggabungkan GLCM dengan K-Means untuk mendiagnos penyakit Alzheimer dengan ketepatan 87 %. Kajian lain oleh (M.-W. Lin et al. 2007) menggabungkan GLCM dengan K-Means untuk

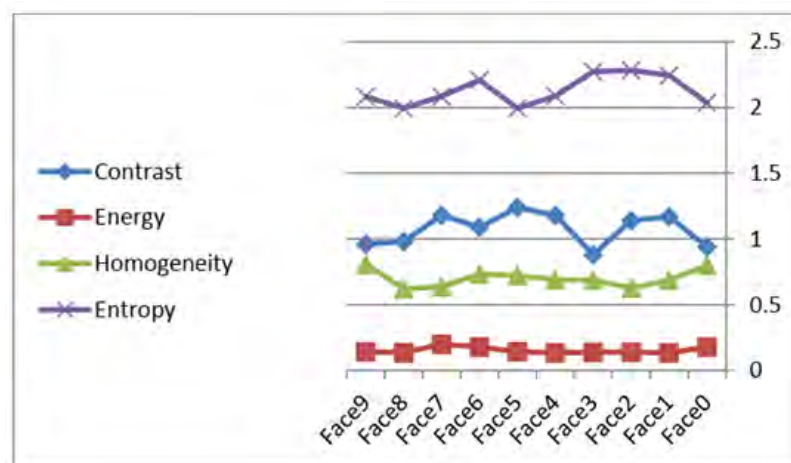
pembahagian dokumen. Kajian lain oleh (X. Zhang et al. 2008) menggabungkan GLCM dengan Penggugusan spektrum untuk pembahagian imej SAR. Satu lagi kajian oleh (Yun & Shu 2011) menggunakan GLCM dengan penggugusan spektrum untuk pembahagian imej Ultrasound. Rumusan proses penggugusan/segmentasi berasaskan pengekstrakan ciri GLCM dijelaskan dalam Rajah 2.26. Contoh nilai pengekstrakan ciri imej terdapat dalam Jadual 2.2 dan Rajah 2.27.



Rajah 2.26 Proses penggugusan gambar berasaskan nilai GLCM(P. Shan 2018).

Jadual 2.2 Contoh nilai daripada pengekstrakan ciri imej (Alazawi et al. 2019)

Face ID	Contrast	Energy	Homogeneity	Entropy
face0	0.94	0.178	0.799	2.037
face1	1.17	0.133	0.687	2.245
face2	1.14	0.139	0.628	2.284
face3	0.88	0.138	0.688	2.273
face4	1.18	0.134	0.691	2.089
face5	1.24	0.143	0.722	1.995
face6	1.09	0.179	0.733	2.209
face7	1.18	0.199	0.635	2.085
face8	0.98	0.135	0.622	1.994
face9	0.96	0.142	0.808	2.084



Rajah 2.27 Contoh nilai daripada pengekstrakan fitur imej (Alazawi et al. 2019).

2.8 RAMALAN KEADAAN TRAFIK MENGGUNAKAN HIDDEN MARKOV MODEL (HMM)

Hidden Markov Model telah lama digunakan secara meluas untuk ramalan keadaan trafik. Pelbagai kajian ramalan trafik state berjaya menggunakan HMM. Antara lain kajian yang dilakukan oleh (G. Zhu et al. 2016), (Qi & Ishak 2014), (Zaki et al. 2019). Selain itu terdapat juga kajian yang meramal kelajuan trafik seperti kajian yang dilakukan oleh (Rapant et al. 2016) yang meramal kelajuan di lebuh raya. Terdapat juga kajian yang menarik, seperti kajian yang dilakukan oleh (Ye et al. 2016) kajian tersebut meramal trajektori kenderaan untuk cadangan perjalanan. Kajian ramalan keadaan trafik yang mendekati kajian ini adalah kajian yang dilakukan oleh (Qi & Ishak 2014) dan (Xiaomeng Wang et al. 2015), kajian yang dilakukan adalah meramal trafik yang dinyatakan berasaskan aras kelajuan jalan berjiran. Jalan bersebelahan (berjiran) diperoleh dengan melakukan penggugusan pelbagai hari yang berbeza menggunakan penggugusan spektrum. Keadaan pemerhatian dikira berasaskan aras kesesakan daripada jalan yang berjiran.

2.8.1 Rantai Markov (Markov chain)

HMM adalah berasaskan perluasan (*augmented*) daripada rantai markov. Rantai markov ialah model yang memberi maklumat berkaitan kebarangkalian daripada

jujukan pemboleh ubah rawak. Setiap keadaan boleh mengambil nilai daripada beberapa set. Set ini boleh terdiri daripada perkataan, atau tag, atau simbol yang mewakili apa-apa sahaja, seperti peralihan keadaan trafik. Rantai markov membuat andaian yang sangat penting bahawa jika matlamatnya adalah untuk meramal masa depan daripada satu jujukan, maka hanya keadaan semasa. Ramalan keadaan cuaca merupakan contoh yang paling popular untuk menjelaskan mengenai rantai markov. Bahagian ini memberitahu penerangan ringkas berkaitan rantai markov menggunakan contoh keadaan cuaca. Terdapat tiga jenis cuaca: cerah, hujan dan berawan. Menurut maklumat sebelum ini, andaian cuaca tidak berubah kekal sama sepanjang hari, iaitu, cuaca hujan tidak berubah cerah pada hari yang sama.

Ramalan cuaca adalah berkaitan dengan meramal cuaca esok berasaskan pemerhatian cuaca pada masa lalu (masa lepas). Untuk menyelesaikan masalah ramalan cuaca, model statistik perlu disediakan untuk melakukan ramalan cuaca:

Statistik cuaca yang dikumpul tentang apa cuaca q_n pada hari ini (pada hari n) bergantung pada cuaca semalam q_{n-1} , hari sebelum q_{n-2} , dan sebagainya. Matlamatnya adalah untuk mencari persamaan kebarangkalian bersyarat seperti dalam persamaan 2.15. Kebarangkalian cuaca yang tidak diketahui pada hari n , $q_n \in \{\text{cerah, berawan, hujan}\}$, bergantung kepada cuaca (yang diketahui) $q_{n-1}, q_{n-2}, \dots$ hari-hari yang lepas.

$$P(q_n | q_{n-1}, q_{n-2}, \dots, q_1) \quad (2.15)$$

Dengan menggunakan kebarangkalian dalam persamaan. 2.16, kebarangkalian ramalan jenis cuaca untuk esok dan hari berikutnya boleh dilakukan menggunakan pemerhatian sejarah cuaca. Sebagai contoh, maklumat cuaca selama tiga hari yang lepas ialah {cerah, cerah, berawan} dalam susunan kronologi yang dikumpulkan, kebarangkalian bahawa esok akan hujan diberikan oleh:

$$P(q_4 = \text{hujan} | q_3 = \text{berawan}, q_2 = \text{cerah}, q_1 = \text{cerah}) \quad (2.16)$$

Kebarangkalian ini boleh disimpulkan daripada kekerapan relatif (statistik) pemerhatian masa lalu bagi jujukan cuaca {cerah, cerah, berawan, hujan}. Walau

bagaimanapun, terdapat masalah: semakin besar n , semakin banyak pemerhatian diperlukan. Katakan bahawa $n = 6$, maka bilangan statistik pengumpulan untuk $3^{(6-1)} = 243$ masa lampau. Andaian yang diperoleh dinamakan andaian Markov: Untuk satu jujukan $\{q_1, q_2, \dots, q_n\}$, nilai P adalah seperti berikut

$$P(q_n | q_{n-1}, q_{n-2}, \dots, q_1) = P(q_n | q_{n-1}) \quad (2.17)$$

Hal ini dinamakan juga andaian markov susunan pertama (*first order Markov assumption*). Sebagai contoh bahawa kebarangkalian pemerhatian tertentu pada masa n hanya bergantung kepada pemerhatian q_{n-1} pada masa $n - 1$. Andaian Markov susunan kedua (*a second-order Markov assumption*) akan mempunyai kebarangkalian pemerhatian pada masa n bergantung pada q_{n-1} dan q_{n-2} . Secara umumnya jika andaian Markov diperkatakan, maka orang akan mengacu kepada andaian Markov susunan pertama. Sistem yang mana persamaan 2.18 adalah benar ialah model Markov (susunan pertama), dan jujukan output $\{q_i\}$ daripada sistem sedemikian ialah rantai Markov (susunan pertama). Persamaan ini menyatakan kebarangkalian jujukan tertentu.

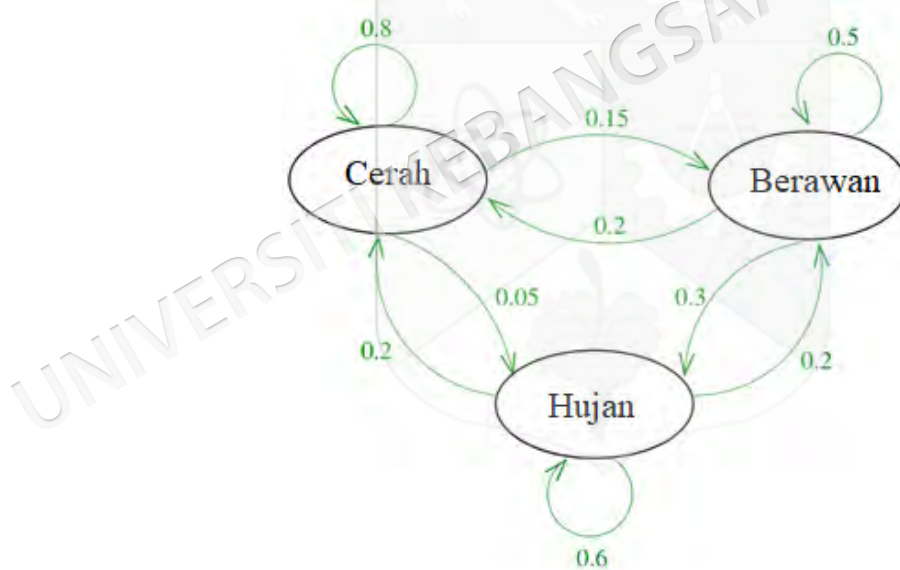
$$P(q_1, \dots, q_t) = \prod_{i=1}^t P(q_i | q_{i-1}) \quad (2.18)$$

Data $P(q_{t+1} | q_t)$ yang dikumpul, diberikan dalam Jadual 2.3. Andaian Markov mempunyai kesan yang mendalam terhadap bilangan sejarah statistic yang perlu dicari. Untuk contoh ramalan cuaca ini kita hanya memerlukan $3 \cdot 3 = 9$ nombor ($P(q_n | q_{n-1})$ bagi setiap kombinasi yang mungkin dari $q_n, q_{n-1} \in \{ \text{cerah, hujan, mendung} \}$) untuk mencirikan kebarangkalian semua jujukan yang mungkin. Markov adalah yang sah atau tidak bergantung pada keadaan (dalam kes cuaca, ia mungkin tidak sah). Tetapi andaian Markov digunakan untuk memudahkan pemodelan. Beberapa nombor dipilih bagi $P(q_{\text{esok}} | q_{\text{hari ini}})$, seperti yang diberikan dalam Jadual 2.3. Perhatikan, bahawa – walau apa pun cuaca hari ini – pasti salah satu ada sejenis cuaca esok. Oleh itu, jumlah kebarangkalian dalam setiap baris ditunjukkan pada Jadual 2.3.

Jadual 2.3 Kebarangkalian $P(q_{t+1}|q_t)$ cuaca esok berasaskan cuaca hari ini (Bourlard & Morgan 1994)

Cuaca hari ini	Cuaca esok		
	Cerah	Hujan	Berawan
Cerah	0.8	0.05	0.15
Hujan	0.2	0.6	0.2
Berawan	0.2	0.3	0.5

Untuk *first-order* Model Markov, kebarangkalian ini boleh digunakan untuk melukis automasi kebarangkalian (*probablistik finie state automata*) keadaan terhingga. Untuk domain cuaca, sebagai contoh terdapat tiga keadaan cuaca, $S = \{\text{cerah, hujan, berawan}\}$ dan untuk setiap masa t , ada $P(q_{t+1}|q_t)$ satu peralihan kepada keadaan mungkin berbeza mengikut kebarangkalian dalam Jadual 2.3. Rantai Markov bagi contoh cuaca diberikan dalam Rajah 2.28.



Rajah 2.28 Model Markov untuk contoh ramalan cuaca dengan kebarangkalian peralihan keadaan mengikut Jadual 2.3. (Bourlard & Morgan 1994).

2.8.2 Model Hidden Markov

Model Hidden Markov (HMM) ialah model statistik yang digunakan untuk menerangkan proses Markov dengan keadaan tersembunyi (tidak diperhatikan). Penerangan yang lebih teliti boleh dirujuk daripada (R Rabiner 1989) (Ghahramani 2001)).

Rantai markov berguna apabila matlamatnya adalah untuk mengira kebarangkalian untuk urutan peristiwa yang boleh dilihat. Walau bagaimanapun, wujud, peristiwa yang menarik minat yang perlu diketahui tersembunyi yang mana tidak dapat dilihat secara langsung. Rujuk semula contoh cuaca, andaikan ada orang terkunci di dalam bilik selama beberapa hari, dan mereka disoal mengenai cuaca di luar. Satu-satunya bukti mereka ada ialah melalui orang yang masuk ke dalam bilik yang membawa makanan harian, apakah mereka membawa payung atau tidak. Kebarangkalian untuk setiap keadaan ditunjukkan dalam Jadual 2.4. Kebarangkalian para pembawa makanan membawa payung adalah 0.2 jika cuaca cerah, 0.7 jika cuaca hujan, dan 0.4 jika ia berawan. Manakala kebarangkalian para pembawa makanan tidak membawa payung adalah 0.8 jika cuaca cerah, 0.3 jika cuaca hujan, dan 0.6 jika ia berawan. Persamaan untuk proses cuaca Markov sebelum mereka dikunci di dalam bilik itu (Persamaan. 2.18).

Jadual 2.4 Matriks keadaan pemerhatian dengan nilai kebarangkalian membawa payung dan tidak membawa payung (Bourlard & Morgan 1994)

Cuaca sekarang	Kebarangkalian membawa payung	Kebarangkalian tidak membawa payung
Cerah	0.2	0.8
Hujan	0.7	0.3
Berawan	0.4	0.6

Walaupun bagaimanapun, cuaca sebenar tersembunyi. Mencari kebarangkalian cuaca tertentu $q_i \in \{\text{cerah, hujan, berawan}\}$ hanya boleh berasaskan pemerhatian x_i , dengan x_i tidak membawa payung. Keadaan ini boleh ditulis semula mengikut peraturan Bayes:

$$P(q_i|x_i) = \frac{P(x_i|q_i)P(q_i)}{P(x_i)}$$

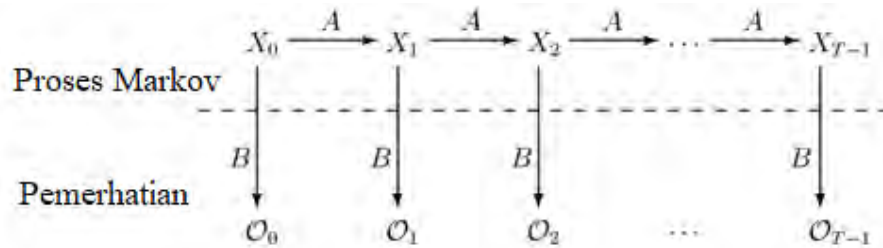
Untuk hari n hari, dan jujukan cuaca $Q = \{q_1, \dots, q_n\}$, dengan 'urutan payung'
 $X = \{x_1, \dots, x_n\}$

$$P(q_1, \dots, q_n|x_1, \dots, x_n) = \frac{P(x_1, \dots, x_n|q_1, \dots, q_n)P(q_1, \dots, q_n)}{P(x_1, \dots, x_n)}$$

Dengan menggunakan kebarangkalian $P(q_1, \dots, q_n)$, daripada jujukan cuaca Markov, dan kebarangkalian $P(x_1, \dots, x_n)$, melihat jujukan tertentu peristiwa payung (sebagai contoh {payung, tidak membawa payung, payung}). Kebarangkalian $P(x_1, \dots, x_n|q_1, \dots, q_n)$, boleh dianggarkan seperti $\prod_{i=1}^n P(x_i|q_i)$, dengan andaian semua i , q_i, x_i adalah independen untuk semua x_j dan q_j , untuk semua $j \neq i$. Untuk membuat kesimpulan dari pemerhatian (jika orang membawa payung atau tidak) tentang cuaca di luar. abaikan kebarangkalian melihat payung $P(x_1, \dots, x_n)$, Ini adalah pemboleh ubah tak bersandar cuaca yang perlu diramalkan.

$$L(q_1, \dots, q_n|x_1, \dots, x_n) = \prod_{i=1}^n P(x_i|q_i) \cdot \prod_{i=1}^n P(q_i|q_{i-1})$$

Nilai untuk kebarangkalian dikumpulkan, yang mana berkadar dengan kebarangkalian, dan dirujuk sebagai kemungkinan L . di mana X_i mewakili jujukan keadaan tersembunyi (hidden state sequence). Proses Markov yang tersembunyi di belakang garis putus-putus ditentukan oleh keadaan semasa dan matrik A . Hanya O_i yang dapat dilihat yang berkaitan dengan keadaan (tersembunyi) proses Markov oleh matriks B seperti yang digambarkan dalam Rajah 2.29.



Rajah 2.29 Rantai markov daripada keadaan aliran trafik, menunjukkan keadaan dan peralihan.

Terdapat tiga masalah asas yang dapat diselesaikan menggunakan HMM (Rabiner 1989) (Jurafsky & H Martin 2020) iaitu masalah penilaian, masalah penyahkod, dan masalah pembelajaran.

Masalah 1: Masalah penilaian

Terdapat sebuah model $\lambda = (A, B, \pi)$ dan jujukan pemerhatian $O = O_1, O_2, \dots, O_T$. Apakah kebarangkalian bahawa pemerhatian dijarakan oleh model ?

Masalah 2: Masalah penyahkod (*decoding problem*)

Terdapat sebuah model $\lambda = (A, B, \pi)$ dan jujukan pemerhatian $O = O_1, O_2, \dots, O_T$. Apakah jujukan keadaan yang paling mungkin dalam model yang menghasilkan pemerhatian?

Masalah 3: Masalah pembelajaran

Terdapat sebuah model $\lambda = (A, B, \pi)$ dan jujukan pemerhatian $O = O_1, O_2, \dots, O_T$. Bagaimana menyesuaikan parameter model (A, B, π) untuk memaksimumkan $p\{O|\lambda\}$?

$A = a_{11}, a_{ij}, \dots, a_{nn}$: Matrik kebarangkalian peralihan

B : Kebarangkalian emisi (*emission probabilities*).

π = Taburan kebarangkalian awal (*initial probability distribution*).

Penyahkod: Algoritma Viterbi

Tugas menentukan jujukan pemboleh ubah yang mana sumbernya yang berasaskan beberapa jujukan pemerhatian penyahkod dinamakan tugas penyahkod.

Penyahkod: Contohnya terdapat model HMM $\lambda = (A, B, \pi)$ dengan jujukan pemerhatian $O = O_1, O_2, \dots, O_T$, Cari jujukan keadaan yang paling mungkin $Q = q_1, q_2, \dots, q_T$.

Dalam ramalan keadaan trafik di kawasan jiran berasaskan input pemboleh ubah masa dan keadaan trafik jalan sasaran, HMM digunakan sebagai penyahkod, yang digunakan untuk mencari jujukan yang paling sesuai dalam model yang dihasilkan berasaskan keadaan pemerhatian yang diberikan.

Algoritma penyahkod yang paling biasa digunakan untuk HMM adalah algoritma Viterbi. Seperti algoritma ke hadapan (*forward algorithm*) (Rajah 2.30). Algoritma hadapan dibahagikan kepada tiga langkah (mencipta matriks ke hadapan, menginisialisasi matriks, langkah rekursif dan langkah penamatan). Langkah pertama adalah membina sebuah matriks pemrograman dinamik ke hadapan, kemudian menginisialisasi matriks tersebut. Setelah langkah inisialisasi matriks, nilai dalam matriks akan dikira bermula dari elemen paling kiri atas (0, 0) yang berkaitan dengan kebarangkalian berada di keadaan permulaan sebelum urutan bermula. Kebarangkalian ini adalah 1 kerana selalu bermula dalam keadaan permulaan. Langkah seterusnya adalah langkah rekursif, dimana setiap lajur akan dikira daripada kiri ke kanan dan setiap lajur dikira daripada atas ke bawah dengan mengalikan $\text{forward}[s, t]$, kebarangkalian menyatakan s dari keadaan t , dengan jumlah kebarangkalian bagi kebarangkalian berada di setiap keadaan s' pada $t - 1$ sehingga keadaan semasa t . Lakukan sehingga semua nilai dalam matriks dikira.

Viterbi adalah sejenis pengaturcaraan dinamik yang mengimplementasikan trellis pengaturcaraan dinamik (*dynamic programming trellis*) Rajah 2.31 menunjukkan pseudocode untuk algoritma Viterbi. Perhatikan bahawa algoritma Viterbi adalah sama dengan algoritma ke hadapan kecuali ia memerlukan nilai

maksimum kebarangkalian laluan sebelumnya manakala algoritma ke hadapan menggunakan jumlah. Perhatikan juga bahawa Algoritma Viterbi mempunyai satu komponen yang tidak dimiliki oleh algoritma ke hadapan iaitu pointer belakang. Ini adalah kerana algoritma ke hadapan perlu menghasilkan kemungkinan pemerhatian, manakala algoritma Viterbi mesti menghasilkan kebarangkalian dan juga jujukan yang paling bersesuaian.

```

function FORWARD(observations of len T, state-graph of len N) returns forward-prob

  create a probability matrix forward[N,T]
  for each state s from 1 to N do                                ; initialization step
    forward[s,1] ← πs * bs(o1)
  for each time step t from 2 to T do                                ; recursion step
    for each state s from 1 to N do
      forward[s,t] ← ∑s'=1N forward[s',t-1] * as',s * bs(ot)
  forwardprob ← ∑s=1N forward[s,T]                                ; termination step
  return forwardprob

```

Rajah 2.30 Kod pseudo daripada algoritma ke hadapan (Jurafsky & H Martin 2020)

Proses dan rumus algoritma ke hadapan adalah seperti berikut:

Permulaan:

$$\alpha_1(j) = \pi_j b_j(o_1) \quad 1 \leq j \leq N$$

Rekursi

$$\alpha_t(j) = \sum_{i=1}^N \alpha_{t-1}(i) a_{ij} b_j(o_t); \quad 1 \leq j \leq N, 1 < t \leq T$$

Penamatan

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i)$$

```

function VITERBI(observations of len  $T$ , state-graph of len  $N$ ) returns best-path, path-prob

create a path probability matrix viterbi[ $N, T$ ]
for each state  $s$  from 1 to  $N$  do                                ; initialization step
    viterbi[ $s, 1$ ]  $\leftarrow \pi_s * b_s(o_1)$ 
    backpointer[ $s, 1$ ]  $\leftarrow 0$ 
for each time step  $t$  from 2 to  $T$  do                            ; recursion step
    for each state  $s$  from 1 to  $N$  do
        viterbi[ $s, t$ ]  $\leftarrow \max_{s'=1}^N \text{viterbi}[s', t-1] * a_{s',s} * b_s(o_t)$ 
        backpointer[ $s, t$ ]  $\leftarrow \arg\max_{s'=1}^N \text{viterbi}[s', t-1] * a_{s',s} * b_s(o_t)$ 
bestpathprob  $\leftarrow \max_{s=1}^N \text{viterbi}[s, T]$                         ; termination step
bestpathpointer  $\leftarrow \arg\max_{s=1}^N \text{viterbi}[s, T]$                 ; termination step
bestpath  $\leftarrow$  the path starting at state bestpathpointer, that follows backpointer[] to states back in time
return bestpath, bestpathprob

```

Rajah 2.31 Algoritma viterbi untuk mencari jujukan optimum daripada keadaan tersembunyi (Jurafsky & H Martin 2020)

Proses dan rumus algoritma viterbi adalah seperti berikut:

Permulaan:

$$v_1(j) = \pi_j b_j(o_1) \quad 1 \leq j \leq N$$

$$bt_1(j) = 0 \quad 1 \leq j \leq N$$

Rekursi

$$v_t(j) = \max_{i=1} v_{t-1}(i) a_{ij} b_j(o_t); \quad 1 \leq j \leq N, 1 < t \leq T$$

$$bt_t(j) = \arg\max_{i=1} v_{t-1}(i) a_{ij} b_j(o_t); \quad 1 \leq j \leq N, 1 < t \leq T$$

Penamatan

$$\text{The best score: } P^* = \max_{i=1} v_T(i)$$

$$\text{The start of backtrace: } q_T^* = \arg\max_{i=1} v_T(i)$$

Dalam kajian ini, HMM digunakan sebagai penyahkod untuk mencari jujukan yang paling bersesuaian dalam model yang dihasilkan berasaskan pemerhatian yang diberikan terperinci mengenai model HMM yang dicadangkan dijelaskan dalam Seksyen 3.4.2 model HMM yang dicadangkan.

2.9 MATRIKS KEKELIRUAN (CONFUSION MATRIX) BAGI PENILAIAN MODEL

Matriks kekeliruan (*confusion matrix*) adalah salah satu cara yang paling popular untuk menilai model klasifikasi, sebagai contoh, menilai model *churn* pekerja (Alamsyah & Salma 2018) (Sisodia et al. 2018), menilai model pengiktirafan imej (Novakovic et al. 2017) (Fatonah et al. 2020), dan menilai model keadaan trafik ramalan (Dauletbaev & Woo 2019) (Toncharoen 2018) (Sobral et al. 2013). Sebuah matriks kekeliruan boleh dibuat untuk klasifikasi binari serta model klasifikasi pelbagai kelas. Matriks kekeliruan dibentuk dengan membandingkan label kelas yang diramalkan dari titik data dengan label kelasnya yang sebenarnya. Perbandingan ini diulangi untuk keseluruhan set data dan hasil perbandingan ini disusun dalam format matriks atau Jadual. Matriks kekeliruan adalah struktur tabular untuk mengekalkan klasifikasi yang betul serta salah faham. Setiap lajur daripada *confusion matrix* mewakili kiraan contoh yang diklasifikasikan berasaskan ramalan dari model dan setiap baris mewakili nilai berasaskan label kelas sebenar. Struktur ini juga boleh diterbalikkan, iaitu, baris mewakili nilai dan lajur mewakili nilai yang sebenar. Rajah 2.32 menunjukkan matriks kekeliruan yang biasa untuk masalah klasifikasi binari, di mana P mewakili kelas positif dan N mewakili kelas negatif.

		Label Ramalan	
		n' ramalan	p' ramalan
Label Aktual	n (True)	Negatif True (TN) Aktual Negatif dan Ramalan Negatif	Positif False (FP) Aktual Negatif dan Ramalan Positif
	p (True)	Negatif False (FN) Aktual Positif dan Ramalan Negatif	Positif True (TP) Aktual Positif dan Ramalan Positif

Rajah 2.32 Struktur biasa matriks kekeliruan (Ting 2010).

Secara umum, terdapat kelas positif dan kelas negatif. Berasaskan struktur ini, terdapat empat syarat penting.

- *True Positive (TP)*: Ini adalah kiraan jumlah contoh dari kelas positif di mana label kelas yang benar sama dengan label kelas yang diramalkan, iaitu. Jumlah contoh di mana model itu betul meramal label kelas positif dengan betul
- *False Positive (FP)*: Ini adalah kiraan jumlah contoh dari kelas negatif di mana model tersalah meramalnya sebagai positif.
- *True Negative (TN)*: Ini adalah kiraan jumlah contoh dari kelas negatif di mana label kelas yang sebenar sama dengan label kelas yang diramalkan, iaitu jumlah contoh di mana model meramal label kelas negatif dengan betul.

- *False Negative (FN)*: Ini adalah kiraan jumlah contoh dari kelas positif di mana model tersalah meramalnya sebagai negatif.

Metrik Prestasi

Matriks kekeliruan adalah asas untuk penentuan *ketepatan*, *precision* dan *recall* (Vora & Iyer 2018).

Ketepatan (Accuracy): Ini adalah salah satu langkah yang paling popular. *Ketepatan* ditakrifkan sebagai ketepatan keseluruhan atau perkadaran ramalan yang betul dari model. Rumus untuk ketepatan (2.19) dari matriks kekeliruan adalah:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (2.19)$$

Ukuran ketepatan biasanya digunakan apabila kelas hampir seimbang dan ramalan yang betul dari kelas-kelas itu sama pentingnya.

Precision: juga dikenali sebagai nilai ramalan yang positif, adalah satu lagi metrik yang boleh diperoleh dari matrik kekeliruan. *Precision* ditakrifkan sebagai bilangan ramalan yang dibuat yang sebenarnya betul atau relevan daripada semua ramalan berasaskan kelas positif. Rumus untuk *precision* (2.20) adalah seperti berikut:

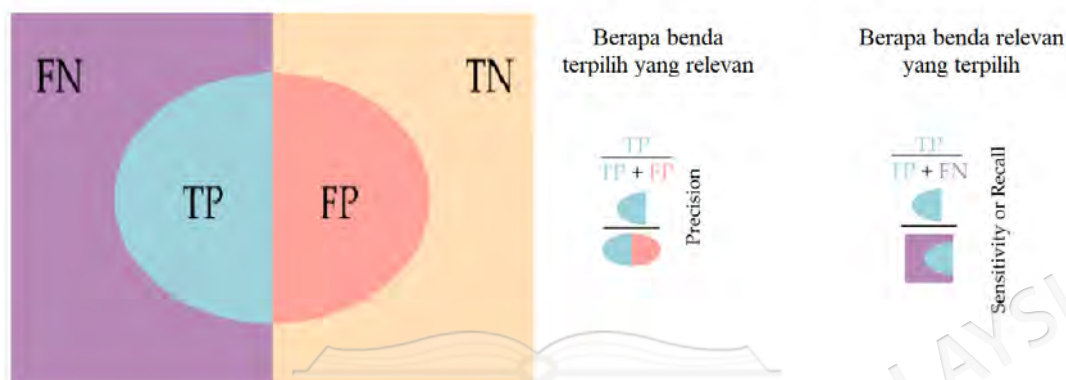
$$Precision = \frac{TP}{FP + TP} \quad (2.20)$$

Model dengan *precision* yang tinggi akan mengenal pasti pecahan yang lebih tinggi daripada kelas positif berbanding dengan model dengan ketepatan yang lebih rendah. *Precision* menjadi penting dalam kes-kes di mana objektif lebih memfokus kepada mencari bilangan maksimum kelas positif walaupun jumlah *accuracy* dikurangkan.

Recall: juga dikenali sebagai sensitiviti, ialah ukuran model untuk mengenal pasti peratusan mata data yang relevan. *Recall* ditakrifkan sebagai bilangan contoh kelas positif yang diramalkan dengan betul. Ini juga dikenali sebagai kadar hit, liputan, atau sensitiviti. Formula untuk *Recall* (2.21) adalah:

$$Recall = \frac{TP}{FN + TP} \quad \dots(2.21)$$

Konsep *precision* dan *recall* digambarkan dalam Rajah 2.33.



Rajah 2.33 Konsep *precision* dan *recall* (Maxwell et al. 2021).

Skor f1 juga digunakan dalam fasa pengujian untuk melihat nisbah keseimbangan antara precision dan recall. Ini adalah kerana kebiasaannya jangka masa aliran trafik yang lancar lebih lama daripada berlakunya kesesakan dalam satu hari. Formula untuk skor f1 (2.22) adalah:

$$f1 \text{ score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (2.22)$$

2.10 RUMUSAN

Bab ini telah memberi penjelasan mengenai kajian semasa dengan kajian ramalan trafik lepas. Seterusnya, selain menerangkan mengenai jurang antara kajian semasa dengan kajian ramalan trafik lepas, bab ini juga menjelaskan pelbagai kaedah, algoritma, dan teknik yang telah digunakan dalam ramalan aliran trafik. Perbincangan dalam bab ini juga meliputi sumber data yang sering digunakan untuk pengumpulan data aliran trafik, cara mengenalpasti keadaan trafik, algoritma penggugusan dan ramalan aliran trafik yang berkaitan dengan kajian ini. Bab ini juga menjelaskan mengenai kaedah untuk menilai prestasi daripada model klasifikasi yang dicadangkan.

BAB III

METODOLOGI

3.1 PENDAHULUAN

Fokus utama kajian ini adalah meramal perambatan aliran trafik kesan kesesakan jalan raya. Lebih khusus kajian ini memfokuskan pada set data daripada pengesan IoT yang tidak mempunyai maklumat tentang jalan berhubung dan arah trafik. Berasaskan kajian kepustakaan tentang kajian ramalan trafik terdahulu, dapat diperoleh maklumat mengenai langkah langkah dalam memecahkan persoalan perambatan aliran trafik. Bab ini menerangkan metodologi daripada cara data trafik dikumpul sehingga model dibangunkan dan dinilai.

Dalam gambaran besar, metodologi dalam kajian ini mengikut amalan terbaik pembangunan pembelajaran mesin dalam kajian trafik terdahulu seperti diterangkan dalam bahagian 1, dengan sedikit pelarasan, metodologi yang dicadangkan terdapat dalam Rajah 3.1.



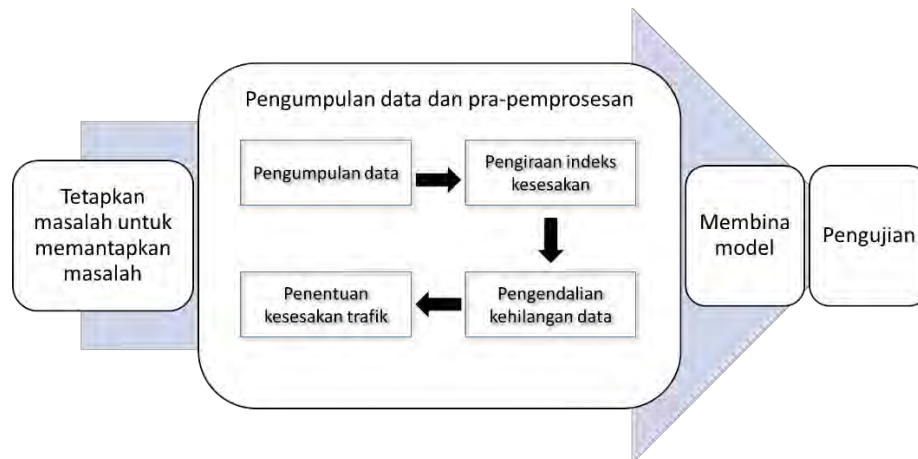
Rajah 3.1 Metodologi yang dicadangkan.

3.2 PENENTUAN PERMASALAHAN

Di awal kajian, masalah dan jurang daripada kajian lepas dianalisis. Masalah kajian perlu ditakrifkan dengan teliti untuk memantapkan permasalahan sebenar. Dalam fasa ini, aliran trafik akan ditentukan berdasarkan indeks kesesakan sama ada aliran trafik dalam keadaan sesak atau sebaliknya. Penerangan lanjut berkenaan dengan pengiraan indeks kesesakan akan diterangkan dalam seksyen 3.3.2 dan aliran trafik dalam keadaan sesak diterangkan dalam seksyen 3.3.4.

3.3 PENGUMPULAN DATA DAN PRA-PEMROSESAN

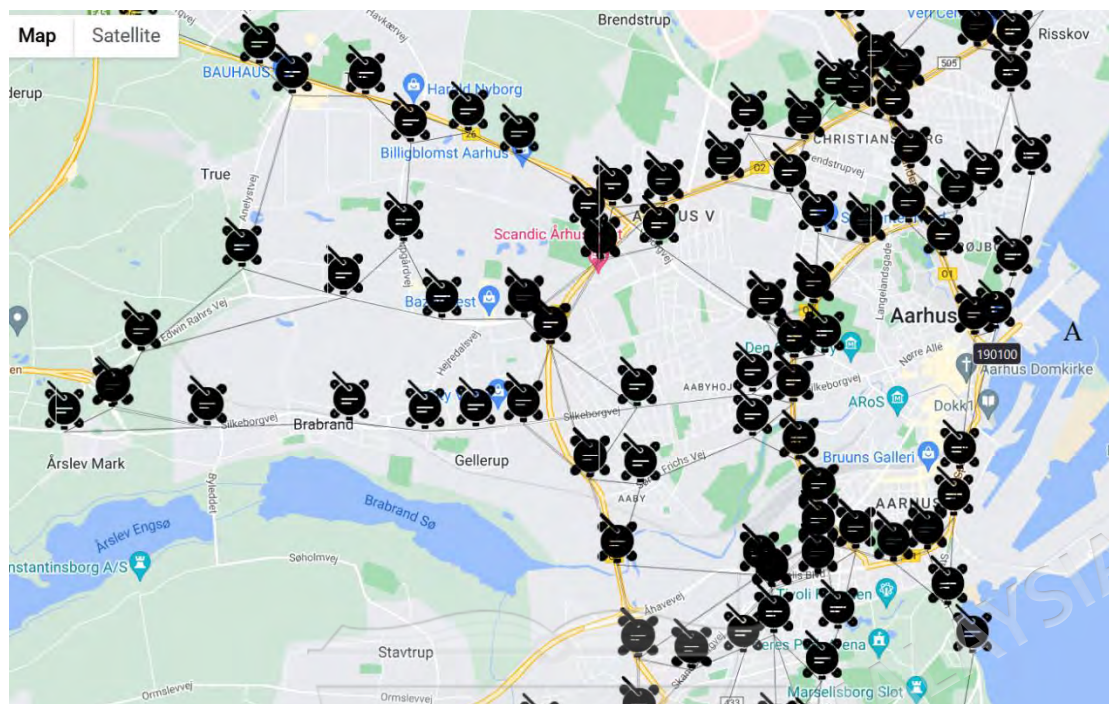
Rajah 3.2 menunjukkan dengan beberapa langkah yang terlibat dalam fasa pengumpulan data dan pra-pemprosesan. Terdapat empat (4) langkah yang dikenal pasti iaitu pengumpulan data, pengiraan indeks kesesakan, pengendalian kehilangan data dan menentukan kesesakan:



Rajah 3.2 Aliran dalam pengumpulan data dan pra-pemrosesan.

3.3.1 Pengumpulan data

Untuk menilai model yang dicadangkan, kajian ini mengumpul data aliran trafik daripada pengesan IoT yang terletak di Aarhus, Denmark (Bischof, Karapantelakis, Athanasios Nechifor, et al. 2014) (Kolozalet al. 2014) (Bischof, Karapantelakis, et al. 2014). seperti yang ditunjukkan pada Rajah 3.3. Bilangan pengesan IoT di Aarhus, Denmark pada ketika itu adalah sebanyak 449 alat pengesan. Data terkumpul yang digunakan dalam kajian ini adalah dalam tempoh 3 bulan dengan tarikan data bagi setiap 5 minit.



Rajah 3.3 Peta daripada pengesan IoT di bandar Aarhus, Denmark (Bischof, Karapantelakis, et al. 2014).

Contohnya, pengesan A dilabelkan 190100. Pengesan ini diletakkan dari Nørreport 93 Aarhus, Denmark ke Spanien 63 Aarhus, Denmark. Jarak antara kedua titik ialah 1490 meter. Percubaan ini menggunakan purata kelajuan, data masa, masa memandu biasa (NDT) dan bilangan kenderaan sebagai perwakilan aliran trafik. Semua pemboleh ubah diperlukan untuk mengira indeks kesesakan (congestion index). Butiran khusus pengesan ini diterangkan dalam Jadual 3.1 dan Jadual 3.2. Jadual 3.1 menunjukkan bahawa lokasi daripada pengesan IoT 190100 terletak dari titik awal hingga titik akhir, seterusnya jadual 3.1 juga menunjukkan nilai normal driving time dari jalan 190100. Sedangkan jadual 3.1 menunjukkan nilai purata kelajuan dan bilangan kenderaan setiap lima minit.

Jadual 3.1 Contoh data trafik yang diambil daripada pengesan 190100 (Bischof, Karapantelakis, et al. 2014)

<i>Duration from</i>	<i>Duration to</i>	<i>Start Point</i>	<i>End Point</i>	<i>Cross Observation Point Data</i>
2014-10-01 01:45:00	2014-11-13 10:40:00	City: Aarhus Street: Nørreport 93, Postal Code: 8000, Coordinates (lat,long): 56.161, 10.211	City: Aarhus Street: Spanien 63 Postal Code: 8000 Coordinates (lat,long): 56.148, 10.209	Distance between two points in meters: 1490 Duration of measurements in seconds: 202 NDT in KMH: 27 EXT ID: 359 Road type: MAJOR_ROAD

Jadual 3.2 Contoh data trafik yang diambil daripada pengesan 190100 (Bischof, Karapantelakis, et al. 2014)

<i>Avg Measure Time</i>	<i>Avg Speed</i>	<i>extID</i>	<i>Median Measure Time</i>	<i>TIME STAMP</i>	<i>Vehicle count</i>	<i>ID</i>	<i>Report ID</i>
376	14	1051	376	2014-08-01T08:30:00	4	20749724	190100
225	23	1051	225	2014-08-01T08:40:00	3	20750622	190100
285	18	1051	285	2014-08-01T08:50:00	2	20751520	190100

3.3.2 Pengiraan indek kesesakan

Secara umum, aliran trafik diagregatkan kepada selang 5 hingga 20 minit (Rajendran & Ayyasamy 2020) (Xu et al. 2016). Xu et al. (2016) telah melakukan pengagregatan masa yang berbeza iaitu 5, 10, 15 dan 20 minit dalam kajian beliau. Hasil kajian tersebut menunjukkan prestasi ketepatan ramalan meningkat dengan peningkatan dalam selang masa pengagregatan. Dengan itu, kajian ini mengira indeks kesesakan dengan selang masa 20 minit menggunakan persamaan 3.1 (W. X. Wang et al. 2018), dengan beberapa pelarasan.

$$CI = \frac{NDT - V_{avg_{int_{20}}}}{NDT - V_{min_{int_{20}}}} \times \frac{Volume_{int_{20}}}{Volume_{day}} \times 100 \quad (3.1)$$

di mana:

NDT: *normal driving time* dalam kilometer/jam atau had kelajuan, ditunjukkan pada Jadual 3.1 seksyen 3.3.1 Pengumpulan data.

$V_{avg_{interval\ 20}}$: purata kelajuan dalam selang masa 20 minit

$V_{min_{interval\ 20}}$: minimum kelajuan dalam selang masa 20 minit

$Volume_{interval\ 20}$: jumlah kenderaan dalam selang masa 20 minit

$Volume_{day}$: jumlah kenderaan selang masa satu hari.

Contoh: Diberi kelajuan normal jalan A dengan nilai NDT 71km/j dan nilai purata kelajuan 19km/j selama 20 minit di mana nilai minimum kelajuan adalah 25km/j. Dengan jumlah kenderaan pada jalan A ketika itu, 2 dan jumlah kenderaan untuk hari tersebut, 1492, maka kiraan indeks kesesakan bagi jalan tersebut dikira seperti berikut:

$$\begin{aligned} Indeks\ kesesakan &= \frac{71 - 19}{71 - 25} \times \frac{2}{1492} \times 100 \\ &= 0.151533 \end{aligned}$$

Nilai indeks kesesakan adalah 0.151533, menunjukkan aliran trafik di jalan A adalah tidak sesak. (Batas kesesakan diterangkan pada seksyen 3.3.4 Penentuan kesesakan trafik).

3.3.3 Pengendalian kehilangan data

Data trafik yang diperolehi daripada pengesan IoT ada kemungkinan kehilangan nilai dalam beberapa masa selang. Keadaan ini boleh dipengaruhi oleh beberapa sebab seperti kerosakan alat pengesan dan terputus bekalan elektrik. Kajian ini menggunakan salah satu kaedah pengendalian kehilangan data iaitu kaedah penggantian purata (Kang 2013). Latar belakang teori bagi kaedah penggantian purata ialah mendapatkan nilai purata anggaran yang munasabah untuk pemerhatian yang dipilih secara rawak daripada taburan normal. Dalam kajian ini, kaedah penggantian

purata dikira berdasarkan nilai purata pemboleh ubah horizontal bagi menggantikan nilai data hilang.

Kajian ini menggunakan kaedah penggantian purata kerana pola aliran trafik mempunyai nilai yang sama dalam selang masa yang tertentu. Kaedah ini juga sesuai digunakan untuk mencadangkan model penggugusan dan ramalan berdasarkan keadaan trafik samada sesak atau sebaliknya. Contoh daripada data hilang ditunjukkan pada Jadual 3.3, dan contoh pengiraan nilai purata ditunjukkan pada Jadual 3.4. Latar belakang teori bagi kaedah penggantian purata ialah mendapatkan nilai purata anggaran yang munasabah untuk pemerhatian yang dipilih secara rawak daripada taburan normal. Dalam kajian ini, kaedah penggantian purata dikira berdasarkan nilai purata pemboleh ubah horizontal bagi menggantikan nilai data hilang.

Jadual 3.3 Data hilang daripada penghitungan indeks kesesakan

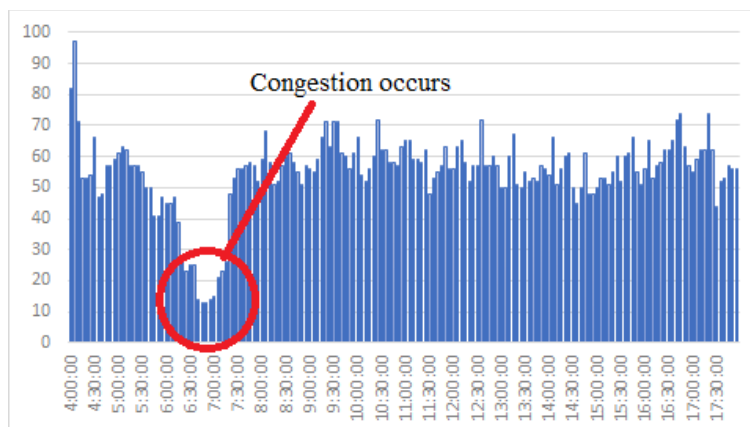
MASA	2/17/ 2014	2/18/ 2014	2/19/ 2014	2/20/ 2014	2/24/ 2014	2/25/ 2014	2/26/ 2014
0:20:00	0.102	0.000	0.062	0.045	NAN	0.000	0.000
0:40:00	0.035	0.247	0.009	0.000	0.064	0.000	0.000
1:00:00	0.067	0.000	0.000	0.049	0.000	0.065	0.000
1:20:00	0.000	0.000	0.124	0.000	0.056	0.000	0.000
1:40:00	0.000	0.000	0.000	0.000	0.064	0.000	0.029
2:00:00	0.000	0.118	0.011	0.000	0.059	0.000	0.000
2:20:00	0.151	0.124	0.000	0.000	0.040	0.070	0.000
2:40:00	0.134	0.054	0.042	0.147	0.164	0.076	0.119
3:00:00	0.628	0.000	0.186	0.101	0.128	0.000	0.195
3:20:00	0.000	0.000	0.041	0.035	0.000	0.000	0.000
3:40:00	0.000	0.247	0.000	0.000	0.181	0.000	0.000
4:00:00	0.000	0.000	0.279	0.000	0.150	0.000	0.122
4:20:00	0.088	0.383	NAN	0.043	0.057	0.095	0.000
4:40:00	0.402	0.296	0.279	0.528	0.301	0.283	0.685
5:00:00	0.826	0.523	1.397	0.587	1.173	1.099	0.459

Jadual 3.4 Pengisian nilai purata kepada data hilang

MASA	2/17/ 2014	2/18/ 2014	2/19/ 2014	2/20/ 2014	2/24/ 2014	2/25/ 2014	2/26/ 2014
0:20:00	0.102	0.000	0.062	0.045	0.034	0.000	0.000
0:40:00	0.035	0.247	0.009	0.000	0.064	0.000	0.000
1:00:00	0.067	0.000	0.000	0.049	0.000	0.065	0.000
1:20:00	0.000	0.000	0.124	0.000	0.056	0.000	0.000
1:40:00	0.000	0.000	0.000	0.000	0.064	0.000	0.029
2:00:00	0.000	0.118	0.011	0.000	0.059	0.000	0.000
2:20:00	0.151	0.124	0.000	0.000	0.040	0.070	0.000
2:40:00	0.134	0.054	0.042	0.147	0.164	0.076	0.119
3:00:00	0.628	0.000	0.186	0.101	0.128	0.000	0.195
3:20:00	0.000	0.000	0.041	0.035	0.000	0.000	0.000
3:40:00	0.000	0.247	0.000	0.000	0.181	0.000	0.000
4:00:00	0.000	0.000	0.279	0.000	0.150	0.000	0.122
4:20:00	0.088	0.383	0.133	0.043	0.057	0.095	0.000
4:40:00	0.402	0.296	0.279	0.528	0.301	0.283	0.685
5:00:00	0.826	0.523	1.397	0.587	1.173	1.099	0.459

3.3.4 Penentuan kesesakan trafik

Penentuan kesesakan trafik dinilai berdasarkan nilai indeks kesesakan yang mengambil kira nilai purata kelajuan dan jumlah kenderaan pada selang masa yang tertentu. Rajah 3.4 menunjukkan keadaan aliran trafik yang berlaku di jalan raya 158324, Denmark di antara 4.00 pagi hingga 5.30 petang dengan kesesakan aliran trafik berlaku di antara 6.15 hingga 8.00 pagi. Di Denmark, purata kelajuan trafik di bandar adalah 50 km/j (Hels et al. 2011). Berdasarkan maklumat tersebut, kajian ini mentakrifkan kesesakan aliran trafik berlaku apabila kelajuan purata persamaan atau di bawah 50 km/j. Penentuan kesesakan trafik adalah berdasarkan purata kelajuan kenderaan di atas jalan raya pada nilai yang sama atau kurang daripada had laju yang ditetapkan dan jumlah kenderaan pada satu selang masa tertentu. Rajah 3.4 dan Jadual 3.5 menunjukkan kesesakan aliran trafik berlaku di jalan raya 158324 Denmark, di antara 6.15 hingga 8.00 pagi (ditunjukkan dengan huruf tebal) dengan nilai indeks kesesakan 3.0 hingga 5.5 dengan nilai purata kelajuan di antara 13 hingga 50 km/j dan bilangan kenderaan di antara 12 hingga 30 kenderaan. Sehingga ketika kelajuan purata ialah 50 km/jam maka nilai indeks kesesakan adalah sekitar 3. Sebab itu, untuk kajian ini, ditentukan bahawa jalan adalah sesak apabila indeks kesesakan berada di atas atau sama dengan 3.



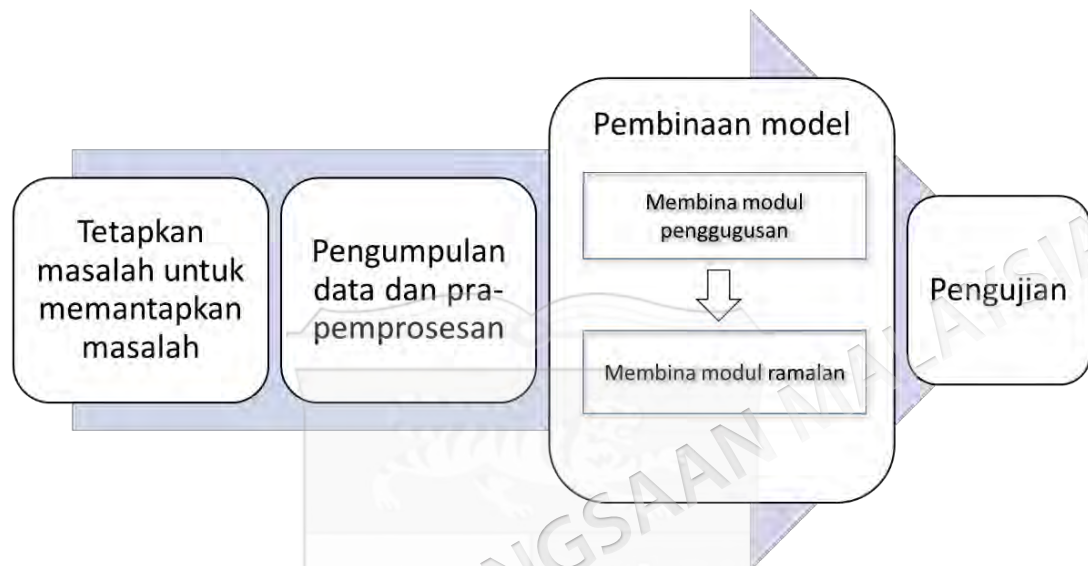
Rajah 3.4 Purata kelajuan di jalan raya 158324 setiap lima minit.

Jadual 3.5 Indeks kesesakan (ci) daripada jalan 158324 dari 05.40 – 09.20 am

Masa	Indeks kesesakan	Purata Kelajuan				Jumlah kereta			
		5	10	15	20	5	10	15	20
5:40-6:00	2.73	63	62	57	57	4	6	14	17
6:00-6:20	3.03	57	55	50	50	12	12	15	13
6:20-6:40	3.83	41	41	47	45	16	20	11	11
6:40-7:00	4.52	45	47	39	28	12	12	20	23
7:00-7:20	4.25	23	25	25	14	24	21	15	17
7:20-7:40	5.41	13	13	14	15	22	20	17	15
7:40-8:00	3.98	21	23	26	48	18	30	22	16
8:00-8:20	3.71	53	56	56	57	11	20	18	13
8:20-8:40	2.6	58	57	52	59	13	8	14	10
8:40-9:00	2.69	68	58	51	52	7	10	12	9
9:00-9:20	2.81	57	60	61	58	5	5	9	12

3.4 PEMBINAAN MODEL

Fasa ini menjelaskan pembinaan model perambatan aliran trafik kepada jalan berjiran yang terdiri daripada dua (2) modul iaitu penggugusan dan ramalan seperti yang ditunjukkan dalam Rajah 3.5.



Rajah 3.5 Aliran dalam pembinaan model.

3.4.1 Membina modul penggugusan

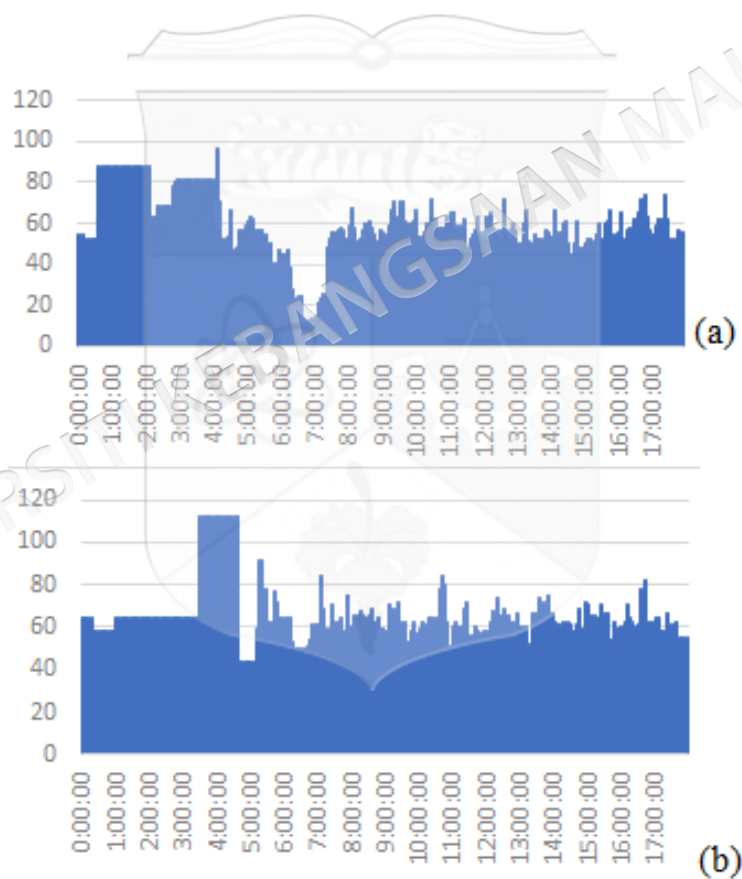
Modul penggugusan dalam kajian ini digunakan untuk mengenal pasti jalan raya berjiran yang diwakili oleh nilai pemberat hubungan antara jalan raya berjiran. Kajian ini menggunakan penggugusan keadaan trafik untuk mendapatkan nilai pemberat hubungan antara jalan raya berjiran, dan ianya berbeza dengan kajian terdahulu yang menggunakan penggugusan berdasarkan nilai kelajuan. Modul penggugusan dalam kajian ini, melaksanakan persilangan antara Spektrum+GLCM dan penggugusan Spektrum+Frekuensi.

a. Mengkaji prestasi GLCM

Grey Level Co-occurrence Matrix (GLCM) adalah salah satu algoritma yang digunakan untuk pengekstrakan fitur. GLCM mengkaji ciri korelasi antara ruang yang

berskala kelabu. Fitur di dalam kajian ini adalah merujuk kepada berlakunya kesesakan yang berulang pada hari dan masa yang sama pada jalan raya berjiran. Nilai kesesakan untuk setiap jalan raya berjiran digunakan sebagai fitur tambahan semasa pelaksanaan penggugusan. Dalam kajian ini, GLCM digunakan untuk mengenal pasti nilai kesesakan jalan raya berjiran dan mengelompokkan jalan raya tersebut. Kerana pola aliran trafik dibentuk oleh kejadian berulang, maka taburan kesesakan dalam kedudukan sesuatu ruang mewujudkan hubungan aliran trafik antara dua nilai matrik pada jarak tertentu dalam ruang aliran trafik.

Ini menerangkan ciri korelasi ruang aliran trafik. Daripada Rajah. 3.6, menunjukkan pola kesesakan aliran trafik yang sama berdasarkan masa.



Rajah 3.6 (a) Pola aliran trafik hari kerja , (b) Pola aliran trafik hari hujung minggu.

Berdasarkan pola ini, matrik GLCM ditetapkan dengan offset mendatar 4 ke kanan dan offset menegak 4 ke bawah. Offset ditetapkan kepada 4 kerana objektifnya

adalah untuk mencari pola kesesakan aliran trafik pada hari bekerja (Isnin hingga Khamis). Dalam Jadual 3.6, 0 menunjukkan tiada kesesakan dan 1 menunjukkan keadaan kesesakan.

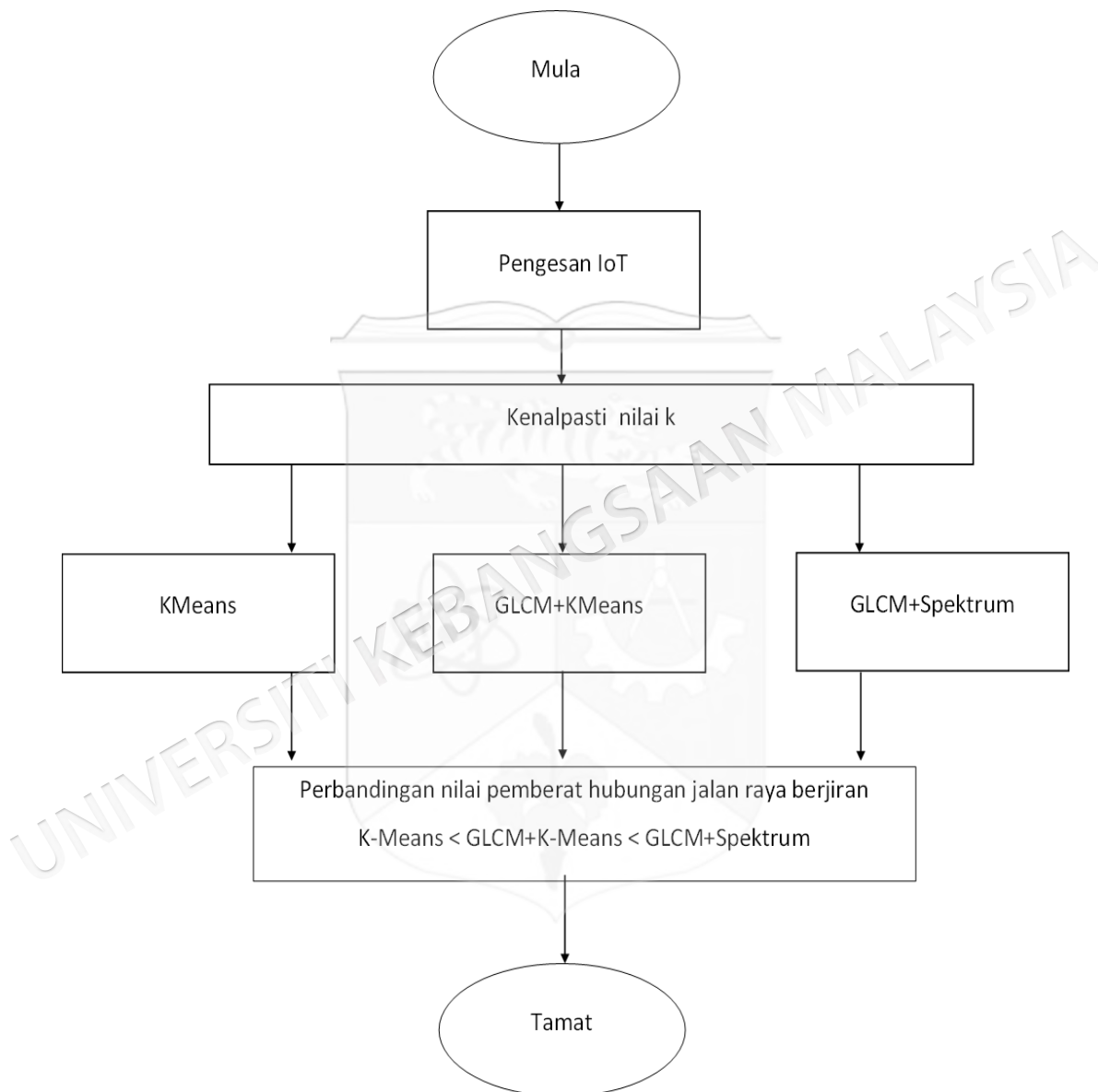
Jadual 3.6 Pola kesesakan trafik dalam masa dan hari di jalan raya 158536, jalan 173225, dan jalan raya 158536

Jalan 158536							
Tarikh	5:40	6:00	6:20	6:40	7:00	7:20	7:40
1	0	0	1	1	1	1	1
2	0	1	1	1	1	1	1
3	0	1	1	1	1	1	1
Jalan 173225							
Tarikh	5:40	6:00	6:20	6:40	7:00	7:20	7:40
1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1
Jalan 158386							
Tarikh	5:40	6:00	6:20	6:40	7:00	7:20	7:40
1	1	1	1	1	1	1	1
2	0	1	1	1	1	1	1
3	1	1	1	1	1	0	1

Dalam pembinaan modul penggugusan, prestasi GLCM dikaji dengan menggabungkan kaedah K-Means dalam menghasilkan nilai pemberat hubungan antara jalan raya berjiran. GLCM adalah salah satu kaedah mengekstrak fitur, di mana kaedah ini tidak digunakan oleh penyelidik sebelum ini (Mondal & Rehena 2019),(Necula et al. 2015), (Xiaomeng Wang et al. 2015) dalam mengkaji kesesakan aliran trafik untuk penggugusan jalan raya.

Satu eksperimen perbandingan dijalankan untuk melihat prestasi GLCM dalam meningkatkan ramalan perambatan aliran trafik kepada jalan raya berjiran. Rajah 3.7 menunjukkan carta alir eksperimen bagi mengkaji prestasi GLCM. Gabungan kaedah GLCM ini dikaji ke atas kaedah penggugusan K-Means dan Spektrum, dengan melaksanakan eksperimen perbandingan nilai pemberat hubungan jalan raya berjiran antara K-Means, GLCM+K-Means dan GLCM+Spektrum. Semakin tinggi nilai

pemberat tersebut, maka semakin tepat ramalan ke atas perambatan aliran trafik kepada jalan raya berjiran. Dalam kajian ini, modul penggugusan yang dilaksanakan adalah gabungan antara GLCM dan Spektrum (GLCM+Spektrum) iaitu dengan mengambilkira hasil nilai pemberat yang tinggi.



Rajah 3.7 Carta Alir Kajian Prestasi GLCM.

b. Gabungan GLCM+Spektrum

Penggugusan Spektrum digunakan dalam kajian ini adalah untuk mengelompokkan jalan raya berjiran yang terletak dalam persekitaran jalan sasaran perambatan. Walau bagaimanapun, algoritma penggugusan spektrum ini sukar untuk menentukan nilai k (gugusan) yang sesuai (Xiaomeng Wang et al. 2015). Dengan itu, kajian ini mengira peratusan keserupaan (*similarity*) keadaan kesesakan pada jalan sasaran dan jalan berjiran hasil kaedah penggugusan spektrum frekuensi untuk menentukan nilai k yang pertengahan untuk menghasilkan lebih dari dua jalan berjiran, kerana semakin tinggi nilai k , jalan berjiran semakin sedikit atau hanya memperoleh satu jalan berjiran. Nilai keserupaan antara jalan raya berjiran yang terletak dalam persekitaran jalan sasaran perambatan dijana menggunakan Persamaan. 3.2. Manakala nilai keserupaan gugusan dijana daripada purata nilai keserupaan jalan berjiran yang terletak dalam persekitaran jalan sasaran perambatan dengan menggunakan Persamaan. 3.3. Nilai keserupaan bagi keadaan kesesakan diplotkan menggunakan carta garis seperti yang ditunjukkan dalam Rajah. 3.10.

$$Keserupaan = \frac{1}{n} \sum_{j=2}^n (Jalan_1 == Jalan_j) \quad (3.2)$$

$$\begin{aligned} \text{Keserupaan gugusan} \\ = \frac{1}{n} \sum_{i=1}^n (Keserupaan_i) \end{aligned} \quad (3.3)$$

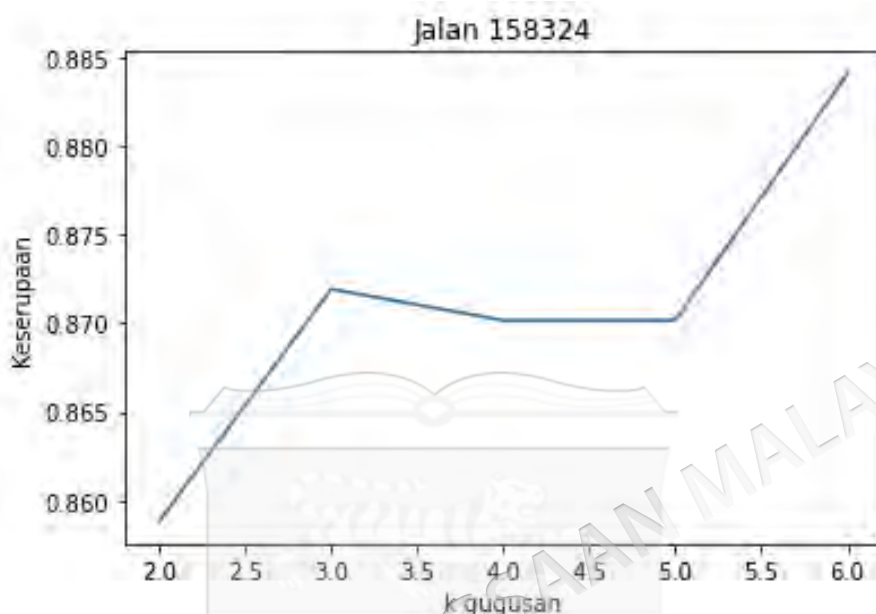
$Jalan_1$ = jalan sasaran

$Jalan_j$ = jalan berjiran

n = jumlah jalan berjiran dalam satu gugusan jalan

Dalam pelaksanaan algoritma spektrum, memilih nombor yang terlalu tinggi untuk k boleh mengakibatkan persilangan sifar iaitu gugusan tidak berhasil atau hanya memperoleh satu jalan berjiran. Dengan itu, kajian ini menggunakan nilai k yang terendah dengan nilai kesamaan dan keadaan kesesakan yang sering terjadi. Sebagai contoh, daripada Rajah. 3.8, nilai keserupaan pada 4.0 dan 5.0 adalah sama, namun

bilangan gugusan yang dipilih adalah 4.0 sebagai nilai k dan bukan 5.0. Ini adalah kerana 4.0 adalah nilai pertengahan yang sering terjadi dan merupakan nilai yang lebih rendah.



Rajah 3.8 Penentuan nilai k .

Dengan menggunakan nilai $k=4$, kajian ini melaksanakan algoritma penggugusan spektrum yang untuk mengelompokkan jalan berjiran dengan keadaan kesesakan yang sama berdasarkan pemilihan fitur menggunakan GLCM. Pelaksanaan algoritma penggugusan GLCM+Spektrum menghasilkan jalan raya yang mempunyai bilangan kejadian kesesakan dalam tempoh masa yang sama bagi beberapa jalan raya berjiran. Walau bagaimana pun, terdapat kejadian kesesakan pada beberapa jalan raya berjiran berlaku pada masa yang berbeza untuk hari yang sama seperti yang ditunjukkan pada Jadual 3.7. Jalan 158536, 173225, 158386, 158983 dan 158836 menunjukkan kesesakan berlaku pada masa dan tempoh yang sama iaitu sekitar 6.00 hingga 8.20. Manakala jalan 158415, 172602, 158865 dan 158954 menunjukkan kesesakan berlaku dalam tempoh yang sama tetapi masa yang berbeza iaitu sekitar 13.20 hingga 16.00.

Jadual 3.7 Hasil penggugusan daripada gabungan GLCM+Spektrum dan di jalan 158536

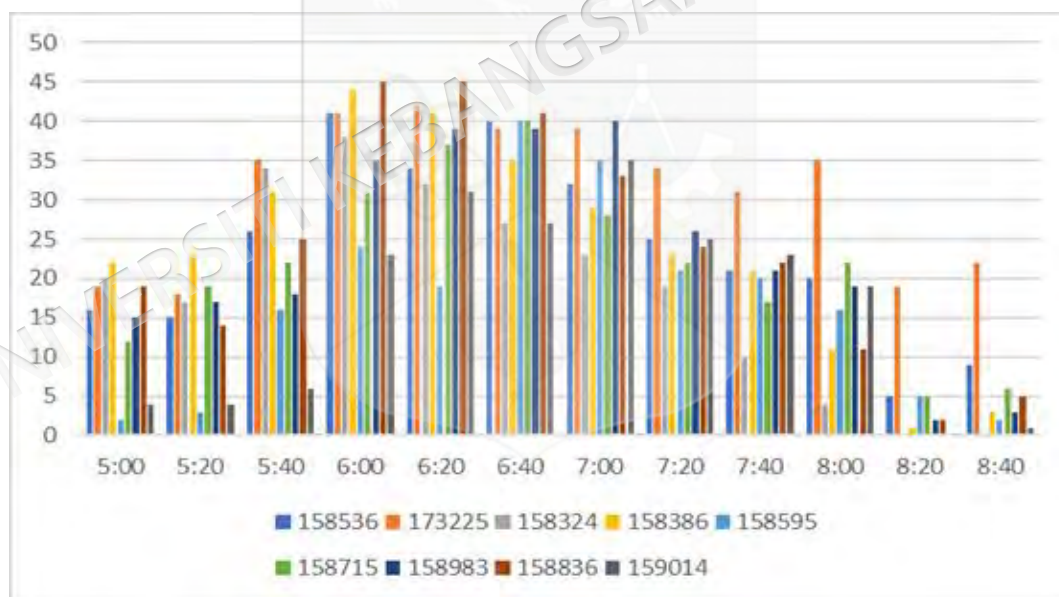
MASA	158536	173225	158386	158415	172602	158983	158836	158865	158954
5:00	0	0	0	0	0	0	0	0	0
5:20	0	0	0	0	0	0	0	0	0
5:40	0	0	0	0	0	0	0	0	0
6:00	0	0	1	0	0	1	1	0	0
6:20	1	1	1	0	0	1	1	0	0
6:40	1	1	1	0	0	1	1	0	0
7:00	1	1	1	0	0	1	1	0	0
7:20	1	1	1	0	0	1	1	0	0
7:40	1	1	1	0	0	1	1	0	0
8:00	0	1	0	0	0	1	0	0	0
8:20	1	1	0	0	0	0	0	0	0
8:40	0	0	0	0	0	0	0	0	0
9:00	1	1	0	0	0	1	0	0	0
9:20	0	1	0	0	0	1	0	0	0
12:20	0	0	0	0	0	0	0	0	0
12:40	0	0	0	0	0	0	0	0	0
13:00	0	0	0	0	0	0	0	0	0
13:20	0	0	0	0	1	0	0	0	0
13:40	0	0	0	0	1	0	0	0	1
14:00	0	0	0	0	1	0	0	0	0
14:20	0	0	0	0	1	0	0	1	1
14:40	0	0	0	1	1	0	0	1	1
15:00	0	0	0	1	1	0	0	1	1
15:20	0	0	0	1	1	0	0	1	1
15:40	0	0	0	1	1	0	0	1	1
16:00	0	0	0	1	1	0	0	1	0
16:20	0	0	0	0	1	0	0	0	0
16:40	0	0	0	0	0	0	0	0	0

Hasil yang berbeza daripada algoritma penggugusan GLCM+Spektrum dari segi masa berlakunya kejadian kesesakan akan memberi kesan kepada ketidaktepatan ramalan perambatan aliran trafik jalan raya berjiran. Untuk mengatasi masalah ini, satu dimensi ditambah dalam pelaksanaan algoritma penggugusan Spektrum. Kajian ini memperkenalkan dimensi frekuensi untuk meningkatkan ketepatan ramalan. Frekuensi dalam kajian ini adalah merujuk kepada kekerapan bilangan kejadian kesesakan pada masa yang sama dalam selang masa tertentu. Jadual 3.8 menunjukkan

bilangan kekerapan kesesakan jalan raya berjiran bagi jalan sasaran perambatan 158536 dalam tempoh tiga (3) bulan. Gambaran dalam bentuk carta bar bagi kekerapan tersebut ditunjukkan pada Rajah 3.9.

Jadual 3.8 Kekerapan kesesakan dalam tempoh 66 hari

Jalan	5:00	5:20	5:40	6:00	6:20	6:40	7:00	7:20
158536	16	15	26	41	34	40	32	25
173225	19	18	35	41	42	39	39	34
158324	20	17	34	38	32	27	23	19
158386	22	23	31	44	41	35	29	23
158595	2	3	16	24	19	40	35	21
158715	12	19	22	31	37	40	28	22
158983	15	17	18	35	39	39	40	26
158836	19	14	25	45	45	41	33	24
159014	4	4	6	23	31	27	35	25



Rajah 3.9 Kekerapan kesesakan di kawasan jiran jalan 158536 pada masa yang berbeza.

Dengan merujuk kepada nilai kekerapan kesesakan yang diperolehi, algoritma penggugusan Spektrum berdasarkan kekerapan/frekuensi dilaksanakan. Hasil daripada perlaksanaan penggugusan Spektrum+Frekuensi ditunjukkan pada Jadual 3.9. Didapati tiga (3) jalan raya berjiran (158324, 158595, 159014) mempunyai tempoh

kesesakan yang berbeza dengan jalan sasaran perambatan 158536 iaitu berlaku pada masa yang sama akan tetapi selang masa jalan 158324, 158595, 159014 lebih singkat daripada jalan sasaran 158536.

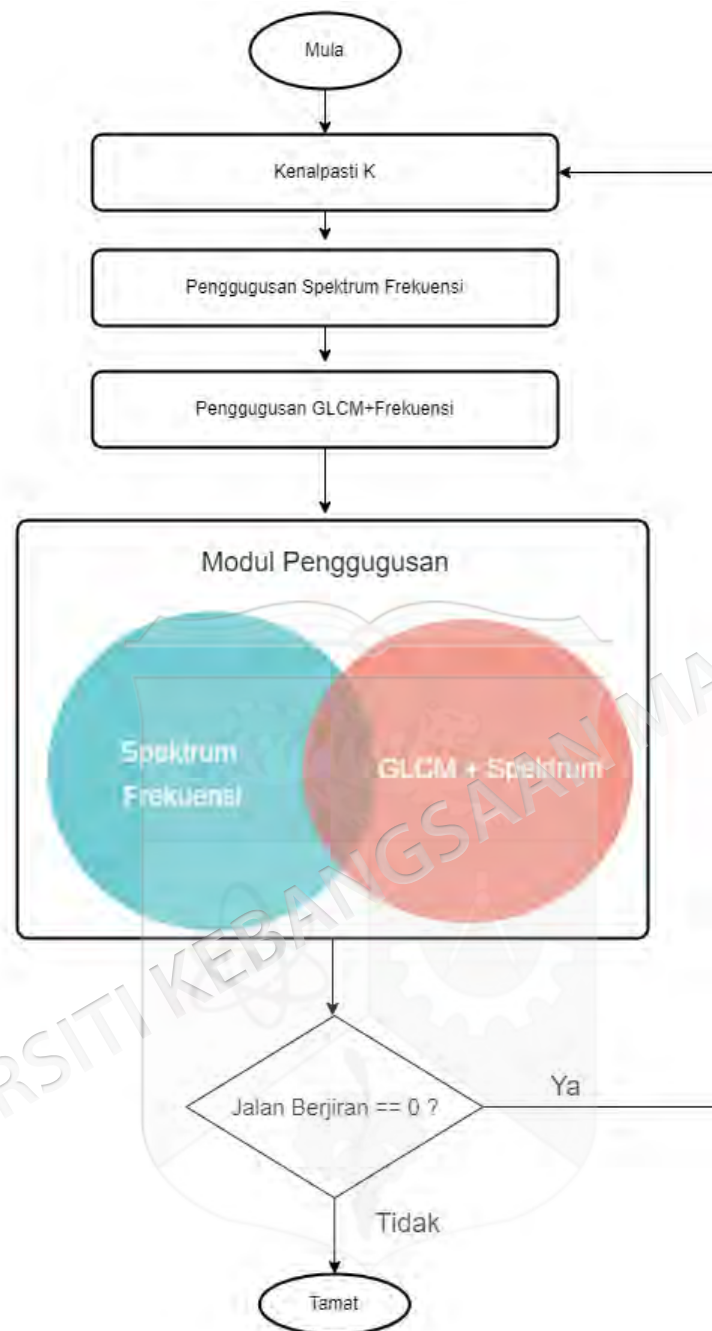
Jadual 3.9 Hasil penggugusan daripada frekuensi di jalan raya berjiran (158536)

Masa	158536	173225	158324	158386	158595	158983	158836	159014
5:00	0	0	0	0	0	0	0	0
5:20	0	0	0	0	0	0	0	0
5:40	0	0	0	0	0	0	0	0
6:00	0	0	0	1	0	1	1	0
6:20	1	1	0	1	0	1	1	1
6:40	1	1	1	1	1	1	1	0
7:00	1	1	1	1	1	1	1	1
7:20	1	1	0	1	0	1	1	1
7:40	1	1	1	1	1	1	1	1
8:00	0	1	0	0	1	1	0	0
8:20	1	1	0	0	0	0	0	0
8:40	0	0	0	0	0	0	0	0
9:00	1	1	0	0	0	1	0	0
9:20	0	1	0	0	0	1	0	0

Seterusnya, hasil penggugusan Spektrum+Frekuensi disilang (intersection) dengan hasil penggugusan GLCM+Spektrum. Hasil persilangan antara Spektrum+Frekuensi dan GLCM+Spektrum ditunjukkan pada Jadual 3.10. Manakala, carta alir modul penggugusan yang dicadangkan ditunjukkan pada Rajah 3.10.

Jadual 3.10 Hasil penggugusan bersilang antara frekuensi dan GLCM+Spektrum di jalan 158536

Masa	158536	173225	158386	158983	158836
5:20:00	0	0	0	0	0
5:40:00	0	0	0	0	0
6:00:00	0	0	1	1	1
6:20:00	1	1	1	1	1
6:40:00	1	1	1	1	1
7:00:00	1	1	1	1	1
7:20:00	1	1	1	1	1
7:40:00	1	1	1	1	1
8:00:00	0	1	0	1	0
8:20:00	1	1	0	0	0
8:40:00	0	0	0	0	0
9:00:00	1	1	0	1	0
9:20:00	0	1	0	1	0
9:40:00	0	0	0	0	0



Rajah 3.10 Carta Alir Penggugusan

Berikut adalah kod pseudo bagi modul penggugusan yang dicadangkan iaitu persilangan antara GLCM-Spektrum dan Frekuensi+Spektrum. Proses modul penggugusan dicadangkan diterangkan dalam pseudocode dalam Rajah 3.11:

```

k = similarityGugusan(dataTrafik)
while jalanBerjiran == 0
    dataFrequent[,] as int
    for i = 1 to len(dataTrafik)
        transposeData = dataTraffic[i]
        transposeData = sum(transposeData)
        dataFrequent[i] = transposeData
    endfor
    modelSpektrum = SpectralClustering(n_clusters = k, affinity = 'nearest_neighbors',
        assign_labels = 'K-Means')
    hasilSpektrum = modelSpektrum.fit_predict(dataFrekwensi)
    jalanSpektrum[,] as int
    for i = 1 to len(dataTrafik)
        for j = 1 to len(hasilSpektrum)
            if(hasilSpektrum[i, j]) == roadTarget
                jalanSpektrum[i] = hasilSpektrum[i]
            endif
            j = j+1
        endfor
        i = i + 1
    end for
    myCoMatrix[] as float
    myFitur[] as float
    for i =1 to len(dataTraffic)
        myCoMatrix[i] = greycomatrix(dataTraffic, [4], [0, 90], levels = 2)
        myFitur[i] = greycoprops(myCoMatrix[i])
        i = i + 1
    end for;
    modelSpektrum = SpectralClustering(n_clusters = k, affinity = 'nearest_neighbors',
        assign_labels = 'K-Means')
    hasilSpektrumGLCM = modelSpektrum.fit_predict(myFitur)
    jalanSpektrumGLCM[] as int
    for i = 1 to len(dataTrafik)
        for j = 1 to len(hasilFitur)
            if hasilSpektrumGLCM[i, j]) == roadTarget
                jalanSpektrumGLCM[i] = hasilSpektrumGLCM[i]
            endif
            j = j + 1
        endfor
        i = i + 1
    endfor
    jalanBerjiran[] as vector
    for value in jalanSpektrum
        if value in jalanSpektrumGLCM
            jalanBerjiran.append(value)
        endif
    endfor
    k = k-1
endwhile

```

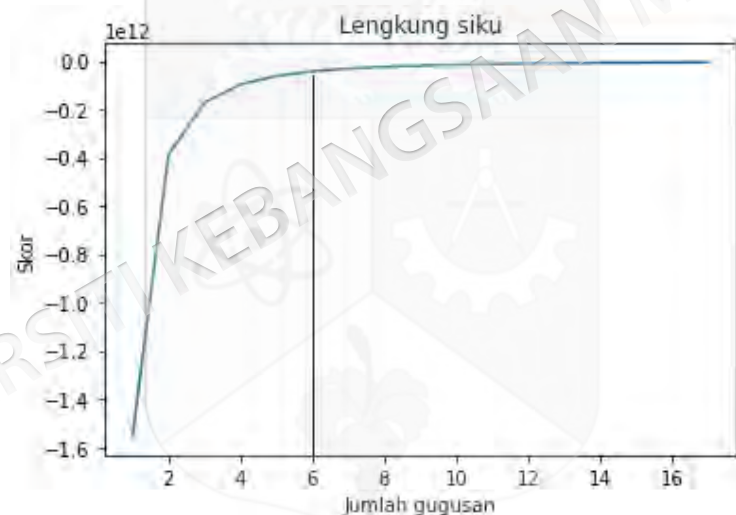
Rajah 3.11 Pseudocode modul penggugusan.

3.4.2 Membina modul ramalan

Modul ramalan yang dicadangkan dalam kajian ini menggunakan *Hidden Markov Model* (HMM) untuk meramal perambatan aliran trafik jalan raya berjiran. Modul ini menggunakan hasil daripada modul penggugusan untuk menentukan gugusan aras kesesakan berdasarkan indeks kesesakan dan masa dengan menggunakan k-Means. Matriks keadaan pemerhatian (*observation state*) dijana untuk pelaksanaan HMM.

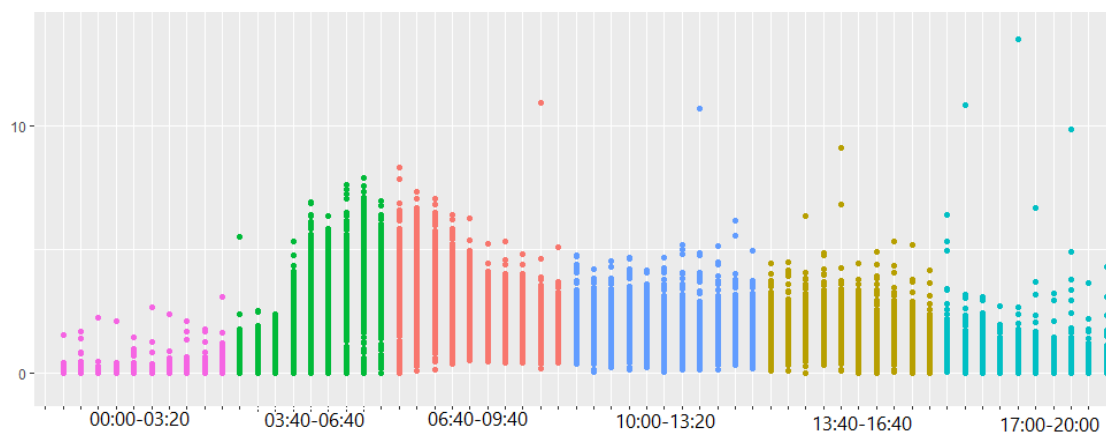
a. Mengenal pasti aras kesesakan menggunakan K-Means

Seperti yang dijelaskan dalam bab 2, pola aliran trafik adalah berbeza pada selang hari dan masa. Dalam fasa ini eksperimen dijalankan untuk menyiasat jika dilaksanakan gugusan masa akan meningkatkan prestasi ketepatan ramalan aliran trafik.



Rajah 3.12 Lengkung siku di semua lokasi

Rajah 3.12 menunjukkan titik siku adalah 2.5 dan mendatar pada nilai bilangan gugusan, 5. Walaubagaimanapun, titik siku menjadi sangat rata pada nilai bilangan gugusan, 6. Kajian ini mengambil nilai enam (6) bilangan gugusan untuk membahagikan aras kesesakan aliran trafik jalan berjiran seperti yang ditunjukkan pada Rajah 3.13.



Rajah 3.13 Penggugusan aras kesesakan dan masa.

Enam (6) gugusan ini digunakan untuk menjana matriks keadaan pemerhatian berdasarkan pada gugusan aras kesesakan jalan berjiran dan jalan sasaran pada masa yang sama. Matriks keadaan pemerhatian memaparkan nilai kebarangkalian emisi (*emission probability*) iaitu merupakan perbezaan antara gugusan aras kesesakan jalan berjiran dengan jalan sasaran dalam masa yang sama. Setiap gugusan aras kesesakan akan dibahagi menjadi dua keadaan aliran trafik iaitu lancar dan sesak. Ini akan menghasilkan matriks keadaan pemerhatian menjadi, $6 \times 2 = 12$ keadaan pemerhatian. Jadual 3.11 menunjukkan contoh nilai matriks keadaan pemerhatian yang dijana dengan nilai 6 bilangan gugusan.

Jadual 3.11 Matriks keadaan pemerhatian dengan nilai kebarangkalian emisi pada jalan 158386

	EM1	EM2	EM3	EM4	EM5	EM6	EM7	EM8	EM9	EM10	EM11	EM12
	1A	1B	2A	2B	3A	3B	4A	4B	5A	5B	6A	6B
Lancar	0.15	0.02	0.18	0.00	0.19	0.00	0.18	0.01	0.10	0.02	0.17	0.00
Sesak	0.04	0.18	0.08	0.01	0.00	0.00	0.04	0.01	0.05	0.59	0.00	0.00

b. HMM dicadangkan

HMM dicadangkan atau HMM(Aras Kesyakan(JB(Masa)+S)) dalam kajian ini merujuk kepada pelaksanaan model Markov tersembunyi yang menggunakan matriks keadaan pemerhatian berdasarkan gugusan aras kesesakan jalan berjiran (JB) pada masa (M) yang sama (JB(Masa)) dan jalan sasaran (S). HMM(Aras Kesyakan(JB(Masa)+S)) digunakan sebagai penyahkod untuk mencari jujukan keadaan yang berkemungkinan besar terjadi kesesakan berdasarkan nilai kebarangkalian emisi pada matriks keadaan pemerhatian. Terdapat dua jenis matriks keadaan dalam pelaksanaan HMM(Aras Kesyakan(JB(Masa)+S)) untuk kajian ini iaitu:

1. Keadaan tersembunyi (*hidden state*) yang berasaskan keadaan trafik daripada jalan raya berjiran.
2. Keadaan pemerhatian yang berasaskan gugusan aras kesesakan dan keadaan jalan sasaran.

Keadaan awal π diperolehi dengan mengira kebarangkalian keadaan trafik jalan berjiran (sesak, lancar) semasa jalan sasaran mengalami kesesakan. Keadaan tersembunyi $x = \{x_1, x_2\}$, diwakili mengikut rantai Markov (Markov chain). Keadaan tersembunyi dalam kajian ini adalah keadaan jalan sesak dan jalan tidak sesak atau lancar. Matriks peralihan kebarangkalian keadaan tersembunyi $x = [x_{11}, x_{12}, x_{21}, x_{22}]$ mewakili kebarangkalian peralihan antara keadaan tersembunyi. Contoh nilai daripada keadaan tersembunyi ditunjuk pada Jadual 3.12.

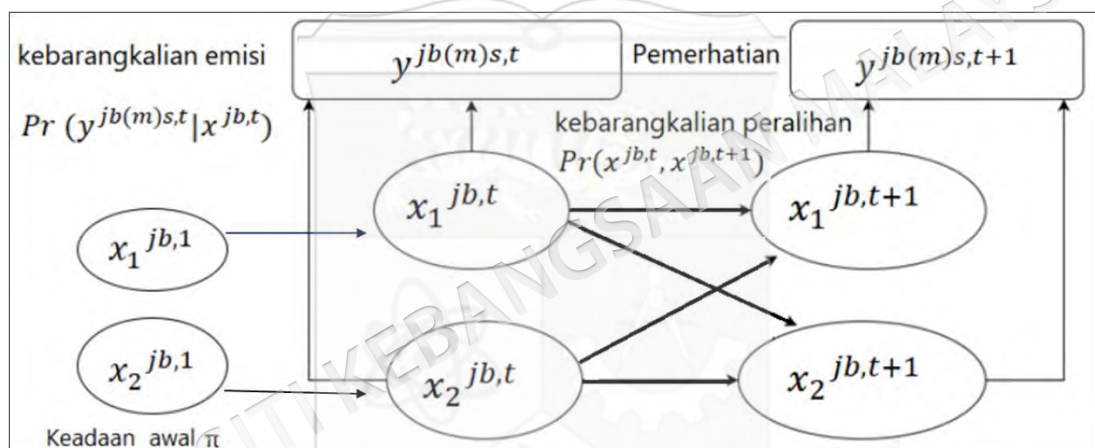
Jadual 3.12 Contoh nilai keadaan tersembunyi daripada jalan 158386

	Lancar	Sesak
Lancar	0.967377	0.032623
Sesak	0.378378	0.621622

Keadaan pemerhatian $O = \{O_1, O_2, \dots, O_n\}$, jumlah n adalah merujuk kepada bilangan keadaan yang dapat dilihat dan dikaitkan dengan keadaan tersembunyi dalam

model ramalan yang dicadangkan. Keadaan pemerhatian menerangkan kebarangkalian peralihan antara keadaan tersembunyi dan keadaan pemerhatian dalam HMM. Kebarangkalian peralihan antara keadaan tersembunyi dan keadaan pemerhatian dipanggil kebarangkalian emisi (*emission probability*).

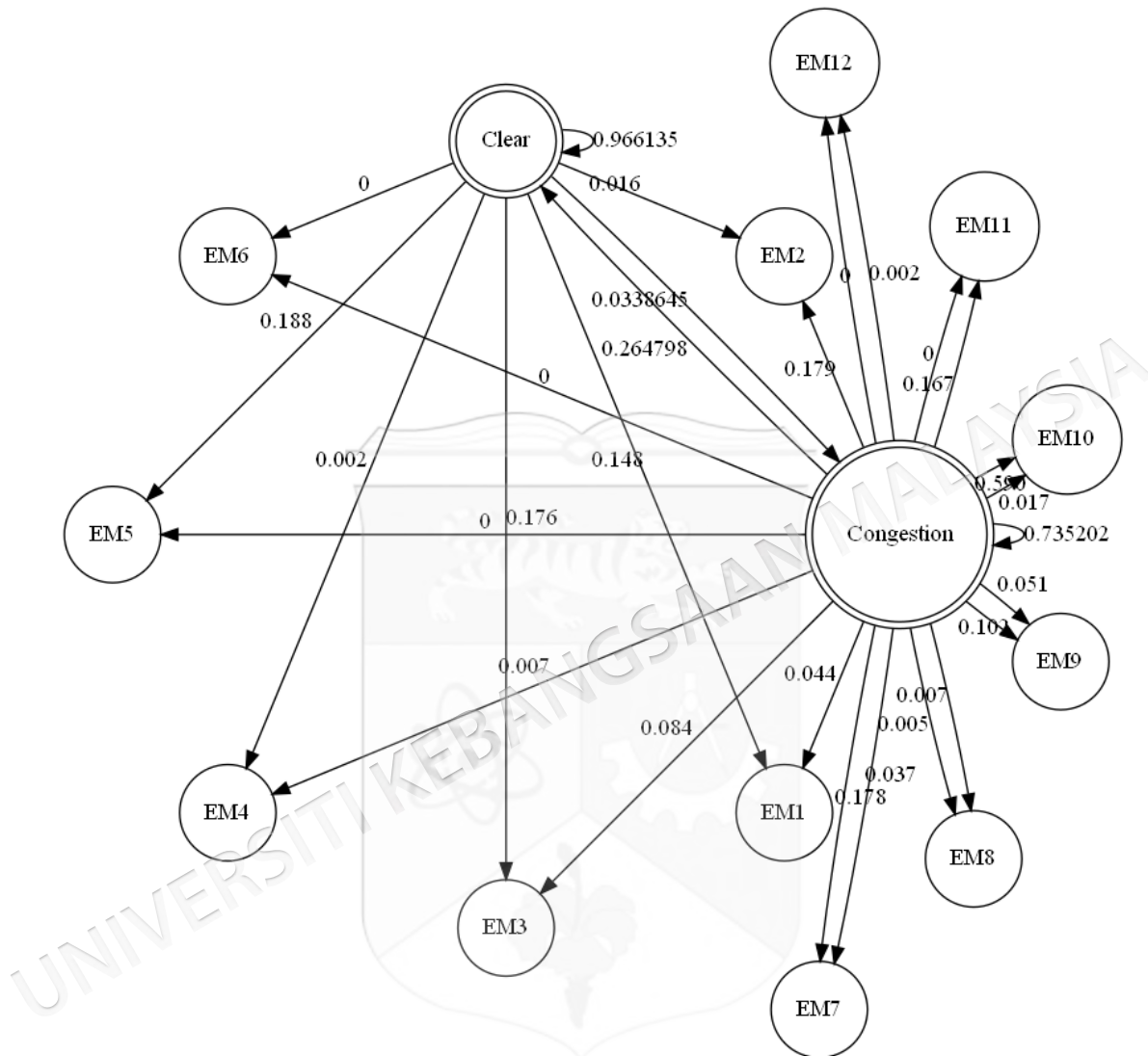
Kebarangkalian emisi adalah kebarangkalian keadaan trafik daripada keadaan lancar menjadi sesak daripada jalan berjiran j_b dalam gugusan aras kesesakan masa + jalan sasaran, t . Manakala, kebarangkalian peralihan adalah kebarangkalian keadaan trafik pada jalan berjiran j_b untuk masa t kepada masa $t+1$. Rajah 3.14 menunjukkan proses HMM untuk pengiraan keadaan trafik.



Rajah 3.14 Pengiraan keadaan trafik menggunakan HMM.

Langkah terakhir adalah pelaksanaan penyahkod (*decoder*) yang menggunakan algoritma Viterbi yang merupakan algoritma pengaturcaraan dinamik untuk mendapatkan anggaran kebarangkalian yang optimum bagi urutan keadaan tersembunyi yang paling hampir. Dalam kajian ini, terdapat beberapa laluan keadaan tersembunyi sama ada lancar dan sesak, yang diwakili sebagai S . Walau bagaimanapun, laluan ini tidak mempunyai kebarangkalian yang sama. Algoritma Viterbi digunakan untuk mengira laluan yang paling hampir. Dengan itu, algoritma Viterbi menghasilkan urutan trafik yang optimum, iaitu $S = \{S_1, S_2, \dots, S_n\}$ untuk urutan keadaan pemerhatian yang diberikan sebagai $O = \{O_1, O_2, O_3, O_4, O_5, O_6, O_7, O_8, O_9, O_{10}, O_{11}, O_{12}\}$. Rajah 3.15 menunjukkan gambaran hubungkait antara keadaan pemerhatian

dan keadaan tersembunyi dalam pelaksanaan modul HMM(Aras Kesyakan(JB(Masa)+S)).

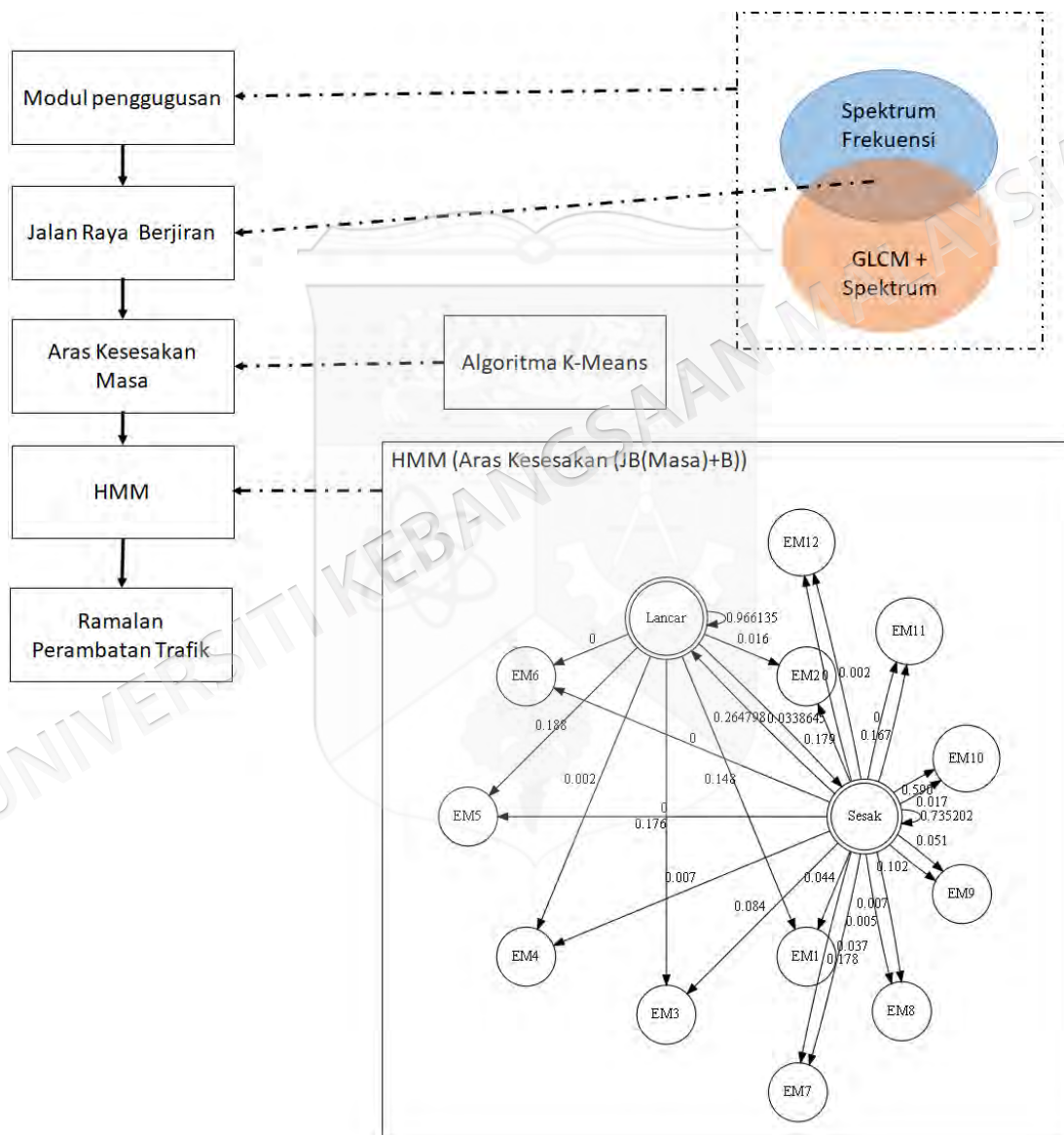


Rajah 3.15 Modul HMM(Aras Kesyakan(JB(Masa)+S)).

3.4.3 Modul ramalan perambatan aliran trafik jalan raya berjiran

Model ramalan perambatan aliran trafik jalan berjiran digunakan untuk meramal perambatan kesesakan jalan berjiran akibat daripada kesesakan jalan sasaran. Langkah pertama dalam model ramalan tersebut adalah modul penggugusan yang mengimplementasikan persilangan antara kaedah penggugusan GLCM+Spektrum dan Spektrum Frekuensi untuk mengenalpasti jalan berjiran. Seterusnya modul ramalan

HMM(Aras Kesusakan(JB(Masa)+S)) dilaksanakan berdasarkan matriks keadaan pemerhatian yang diperolehi daripada gugusan aras kesesakan jalan berjiran pada masa yang sama dan jalan sasaran. Rajah 3.16 menggambarkan model ramalan perambatan aliran trafik jalan berjiran yang melaksanakan kaedah penggugusan GLCM+Spektrum dan Spektrum Frekuensi, dan modul ramalan HMM(Aras Kesusakan(JB(Masa)+S)).



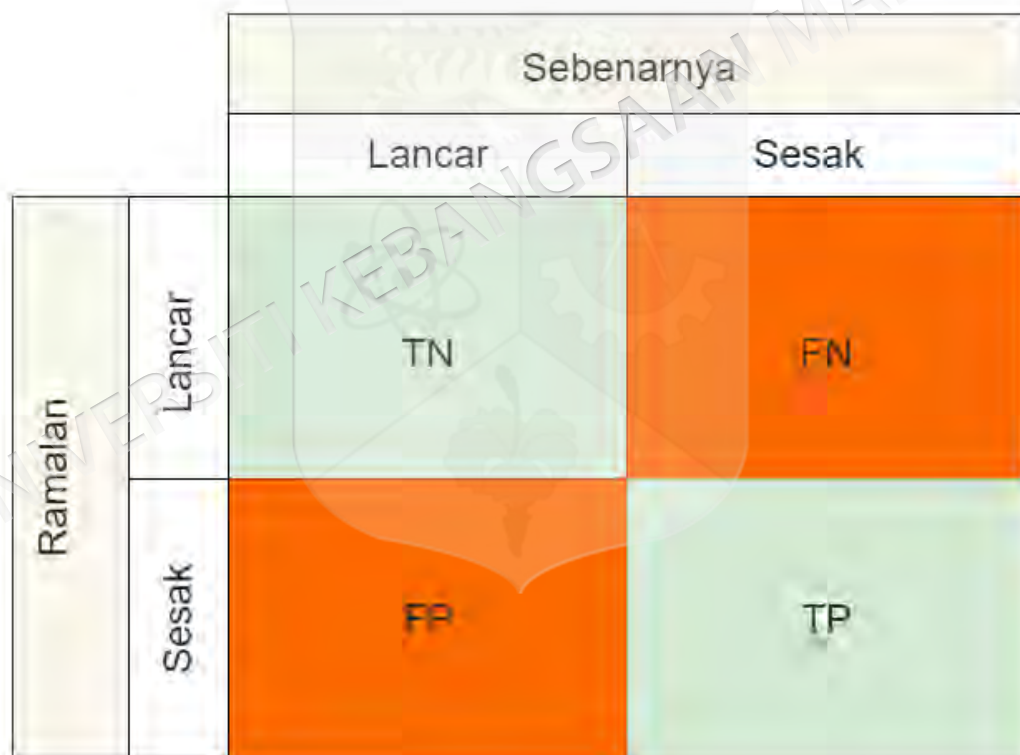
Rajah 3.16 Model Ramalan Perambatan Aliran Trafik Jalan Berjiran

3.5 PENGUJIAN

Seksyen ini menerangkan pelaksanaan fasa terakhir dalam projek kajian ini iaitu pengujian terhadap model ramalan perambatan aliran trafik jalan berjiran. Keputusan eksperimen bagi ramalan aliran trafik jalan berjiran yang menggunakan model yang dicadangkan dibandingkan dengan model HMM sedia ada.

3.5.1 Matriks kekeliruan

Untuk menilai keputusan ramalan aliran trafik, Matriks kekeliruan (*confusion matrix*) digunakan (Sepasgozar & Pierre 2022) (Tamir et al. 2020) seperti yang ditunjuk pada Rajah 3.17.



		Sebenarnya	
		Lancar	Sesak
Ramalan	Lancar	TN	FN
	Sesak	FP	TP

Rajah 3.17 Matriks kekeliruan (*confusion matrix*)

Di mana:

TN: Aliran trafik dan ramalan juga lancar

TP : Aliran trafik sesak dan ramalan adalah sesak

FN: Aliran trafik sesak akan tetapi ramalan adalah lancar

FP : Aliran trafik lancar akan tetapi ramalan adalah sesak

Dalam kajian ini, positif pada matriks kekeliruan adalah merujuk kepada aliran trafik sesak, manakala, negatif adalah merujuk kepada aliran trafik lancar. Rajah 3.18 adalah contoh matriks kekeliruan bagi jalan 158536.

		Sebenarnya		
		Lancar	Sesak	
Ramalan	Lancar	TN = 344	FN = 1	345
	Sesak	FP = 22	TP = 64	86
		366	65	n = 431

Rajah 3.18 Matriks kekeliruan (*confusion matrix*) Jalan 158536

Dalam kajian ini, selain daripada melihat kepada ketepatan ramalan, penilaian menggunakan kaedah *precision* dan *recall* juga digunakan dalam fasa pengujian. Ini adalah kerana nilai ketepatan ramalan yang tinggi tidak semestinya menunjukkan ramalan kesesakan aliran trafik yang sebenar. Skor f1 juga digunakan dalam fasa pengujian untuk melihat nisbah keseimbangan antara *precision* dan *recall* pada satu jalan. Ini adalah kerana kebiasaannya jangka masa aliran trafik yang lancar lebih lama daripada berlakunya kesesakan dalam satu hari. Dengan itu, nilai ketepatan, nilai *precision* dan skor f1 penting untuk menentusahkan ramalan perambatan aliran trafik

jalan berjiran dalam kajian ini. Berikut adalah takrifan bagi kaedah pengukuran yang digunakan dalam kajian ini untuk menilai prestasi model yang dicadangkan:

- Ketepatan (*Accuracy*) - menilai ketepatan ramalan semua aliran trafik.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3.4)$$

- *Precision* - menilai ketepatan ramalan aliran trafik yang betul-betul sesak

$$Precision = \frac{TP}{FP + TP} \quad (3.5)$$

- *Recall* - menilai ramalan aliran trafik sesak tetapi diramalkan tidak sesak

$$Recall = \frac{TP}{FN + TP} \quad (3.6)$$

- Skor f1 - menilai purata keseimbangan aliran trafik dari nilai *precision* dan *recall*.

$$f1 \text{ score} = \frac{2 \times precision \times recall}{precision + recall} \quad (3.7)$$

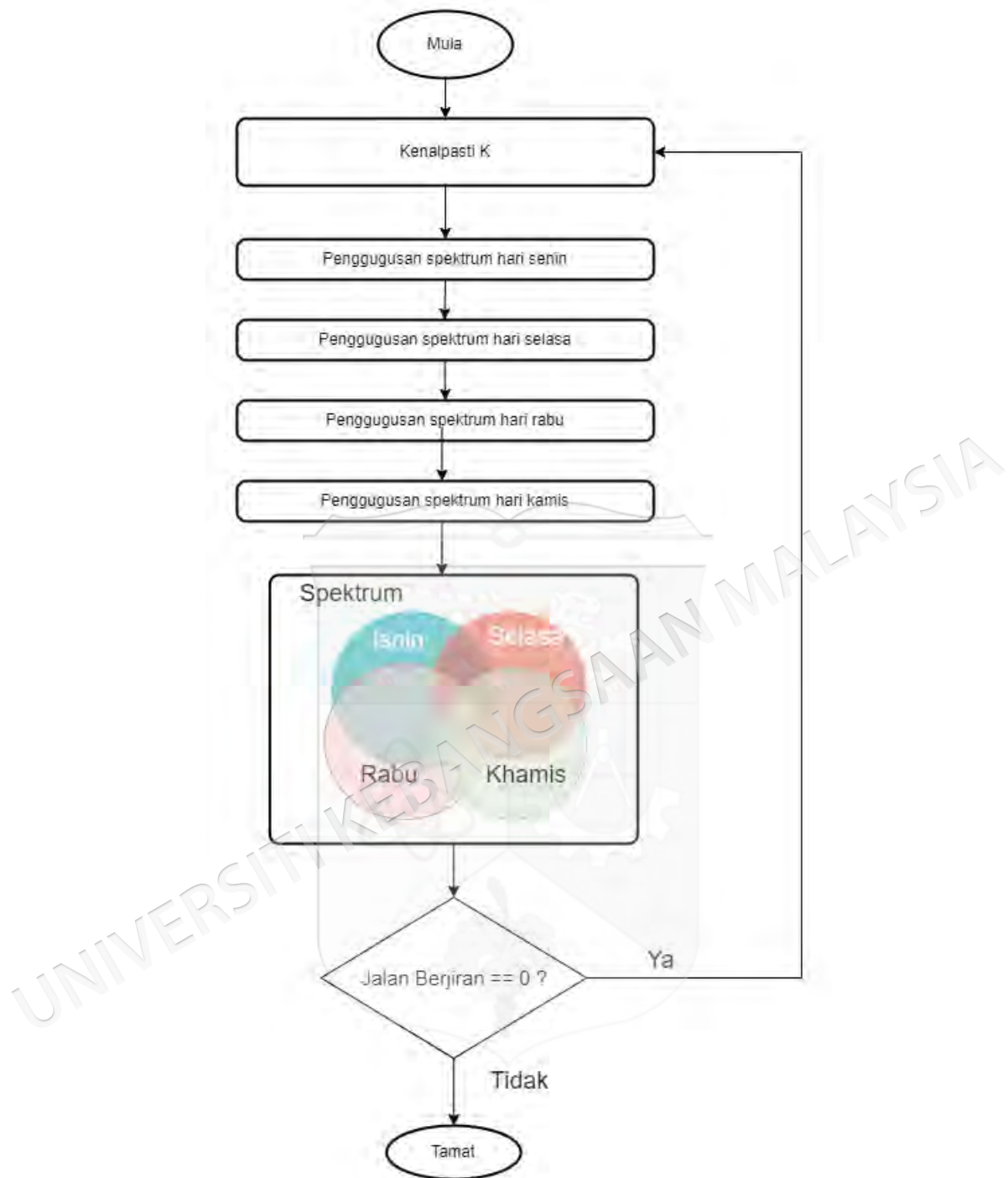
3.5.2 Mengimplementasikan semula model Wang et al. (2015)

Dalam mengimplementasikan semula model Wang et al. (2015), set data diperoleh daripada aliran trafik yang sama iaitu daripada alat pengesanan trafik IoT yang terletak di Aarhus, Denmark (Bischof, Karapantelakis, Athanasios Nechifor, et al. 2014) (Koloza et al. 2014) (Bischof, Karapantelakis, et al. 2014). Seterusnya kesesakan trafik ditentukan menggunakan kaedah yang sama iaitu membagi aras kesesakan menjadi dua iaitu lancar dan sesak berasaskan indeks kesesakan. Model ramalan aliran trafik Wang et al. (2015) terbagi daripada modul penggugusan persilangan hari dan HMM berasaskan purata jalan berjiran pada selang masa tertentu.

a. Modul penggugusan persilangan hari

Implementasi semula modul penggugusan Wang et al. (2015) dalam penyelidikan ini melaksanakan persilangan penggugusan spektrum hari kerja berasaskan indeks kesesakan iaitu penggugusan spektrum hari isnin disilang dengan penggugusan spektrum hari selasa, seterusnya disilang dengan penggugusan spektrum hari rabu, dan akhir sekali disilang dengan penggugusan spektrum hari kamis. Carta alir modul penggugusan spektrum persilangan hari kerja ditunjukkan pada Rajah 3.19.





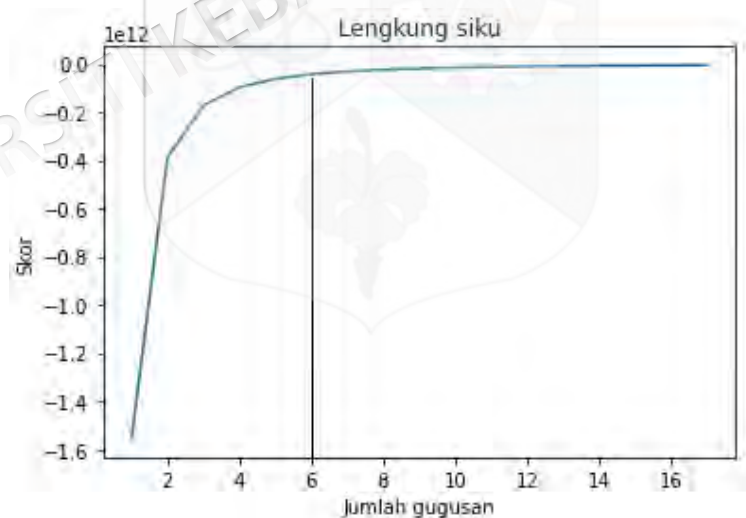
Rajah 3.19 Carta alir penggugusan spektrum persilangan hari.

b. Model Markov tersembunyi Wang 2015

Implementasi semula modul ramalan Wang et al. (2015) dalam penyelidikan ini dilakukan menjana rantai Markov jalan berjiran dan matriks keadaan pemerhatian (*observation state*) berasaskan purata trafik jalan pada selang masa tertentu. Keadaan pemerhatian dijana dengan mengenal pasti aras kesesakan jalan berjiran pada selang masa tertentu. K-Means digunakan untuk menentukan gugusan aras kesesakan berdasarkan indeks kesesakan dan masa dengan menggunakan K-Means.

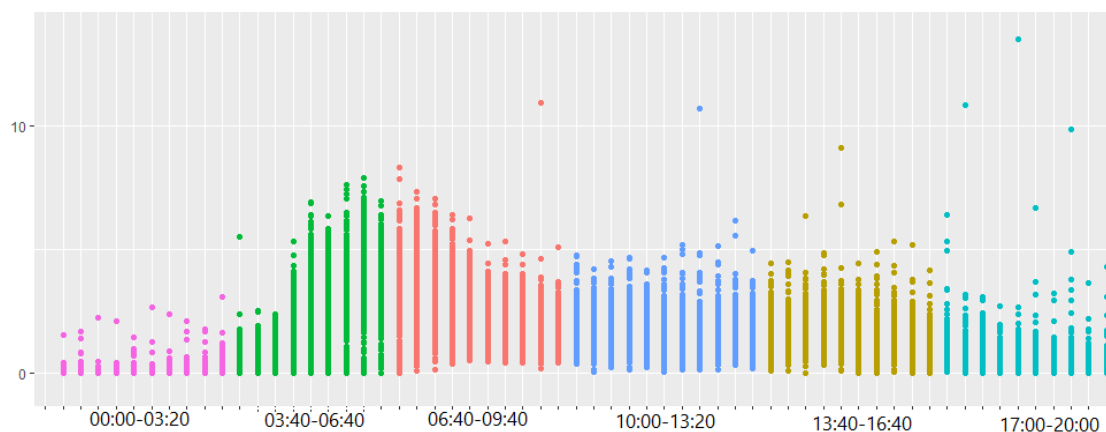
i. Mengenal pasti aras kesesakan menggunakan K-Means

Dalam fasa ini aras kesesakan jalan berjiran pada selang masa tertentu dikenalpasti dengan melaksanakan penggugusan jalan raya berjiran yang diperolehi dengan penggugusan spektrum persilangan hari. Jumlah k ditentukan berasaskan pengiraan skor yang ditunjukkan pada rajah 3.20. Rajah 3.20 menunjukkan titik siku adalah 2.5 dan mendatar pada nilai bilangan gugusan, 5. Walau bagaimanapun, titik siku menjadi sangat rata pada nilai bilangan gugusan, 6.



Rajah 3.20 Lengkung siku di semua lokasi

Kajian ini mengambil nilai enam (6) bilangan gugusan untuk membahagikan aras kesesakan aliran trafik jalan berjiran seperti yang ditunjukkan pada Rajah 3.21.



Rajah 3.21 Penggugusan aras kesesakan dan masa.

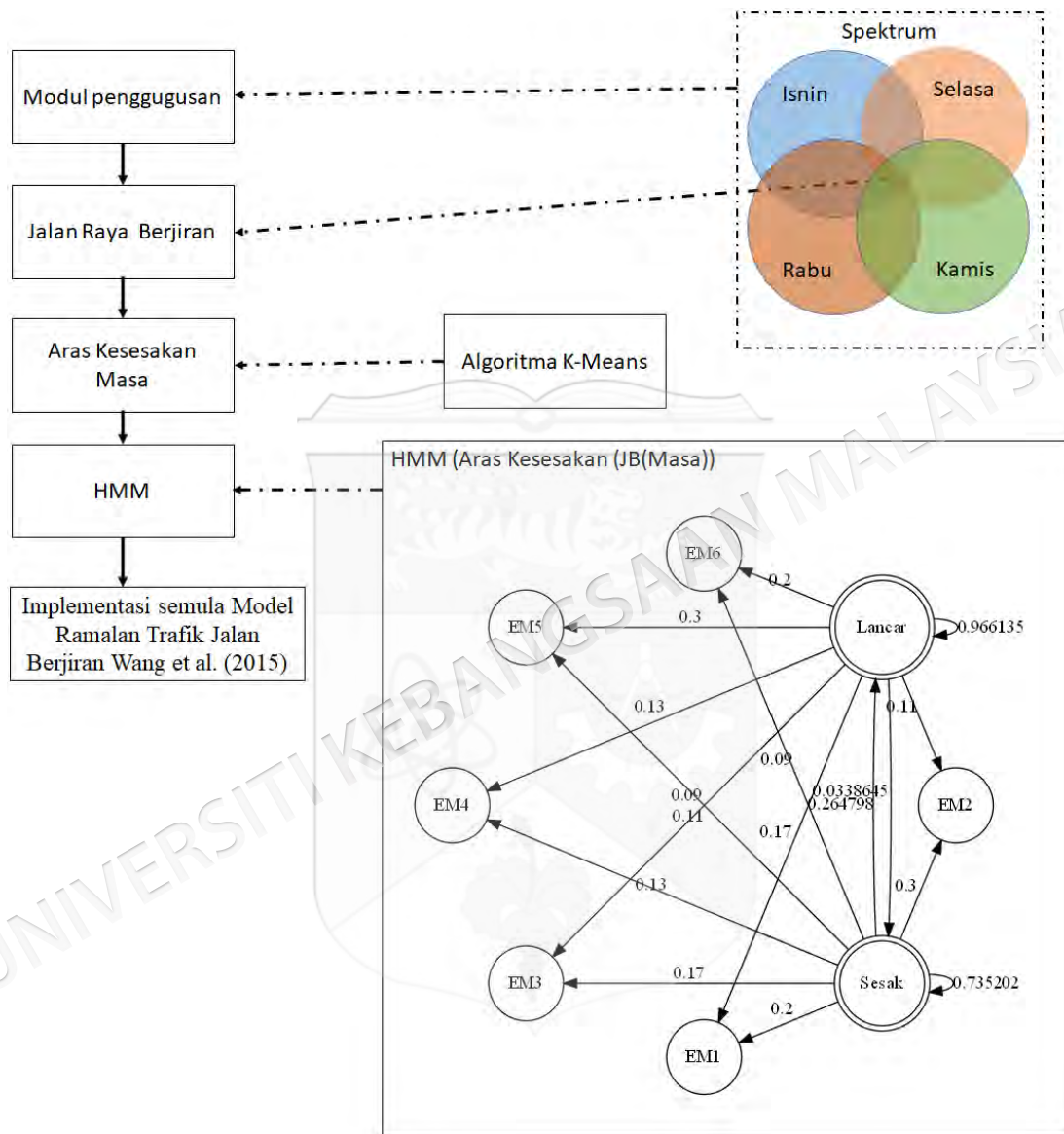
Enam (6) gugusan ini digunakan untuk menjana matriks keadaan pemerhatian berdasarkan pada gugusan aras kesesakan jalan berjiran dan jalan sasaran pada masa yang sama. Matriks keadaan pemerhatian memaparkan nilai kebarangkalian emisi (*emission probability*) iaitu merupakan perbezaan antara gugusan aras kesesakan jalan berjiran dengan jalan sasaran dalam masa yang sama. Ini akan menghasilkan matriks keadaan pemerhatian. Jadual 3.13 menunjukkan contoh nilai matriks keadaan pemerhatian yang dijana dengan nilai 6 bilangan gugusan.

Jadual 3.13 Matriks keadaan pemerhatian dengan nilai kebarangkalian emisi pada jalan 158386

	EM1	EM2	EM3	EM4	EM5	EM6
	1	2	3	4	5	6
Lancar	0.2	0.3	0.17	0.13	0.11	0.09
Sesak	0.17	0.11	0.09	0.13	0.3	0.2

Secara garis besar implementasi semula modul ramalan Wang et al. (2015) dalam penyelidikan ini iaitu pertama adalah modul penggugusan yang mengimplementasikan persilangan spektrum persilangan antara hari kerja untuk mengenalpasti jalan berjiran. Seterusnya modul ramalan HMM(Aras Kesesakan(JB(Masa))) dilaksanakan berdasarkan matriks keadaan pemerhatian yang diperolehi daripada gugusan aras kesesakan jalan berjiran pada masa yang sama Rajah

3.22 menggambarkan modul ramalan perambatan aliran trafik jalan berjiran yang melaksanakan kaedah penggugusan spektrum persilangan hari kerja, dan modul ramalan HMM(Aras Kesusakan(JB(Masa))).



Rajah 3.22 Implementasi semula Model Ramalan Trafik Jalan Berjiran Wang et al. (2015)

3.6 RUMUSAN

Bab ini memberi penjelasan yang lengkap mengenai metodologi penyelidikan yang dijalankan, bermula dari menentukan masalah, pengumpulan data dan pra-pemprosesan, membina model dicadangkan, dan akhir sekali mengesahkan model. Bab seterusnya membincangkan tentang hasil dan perbincangan kajian ini dengan perbandingan kaedah penggugusan antara K-Means, GLCM+Kmeans dan GLCM+Spektrum. Manakala ramalan membandingkan model yang dicadangkan dengan model Wang et al. (2015).



BAB IV

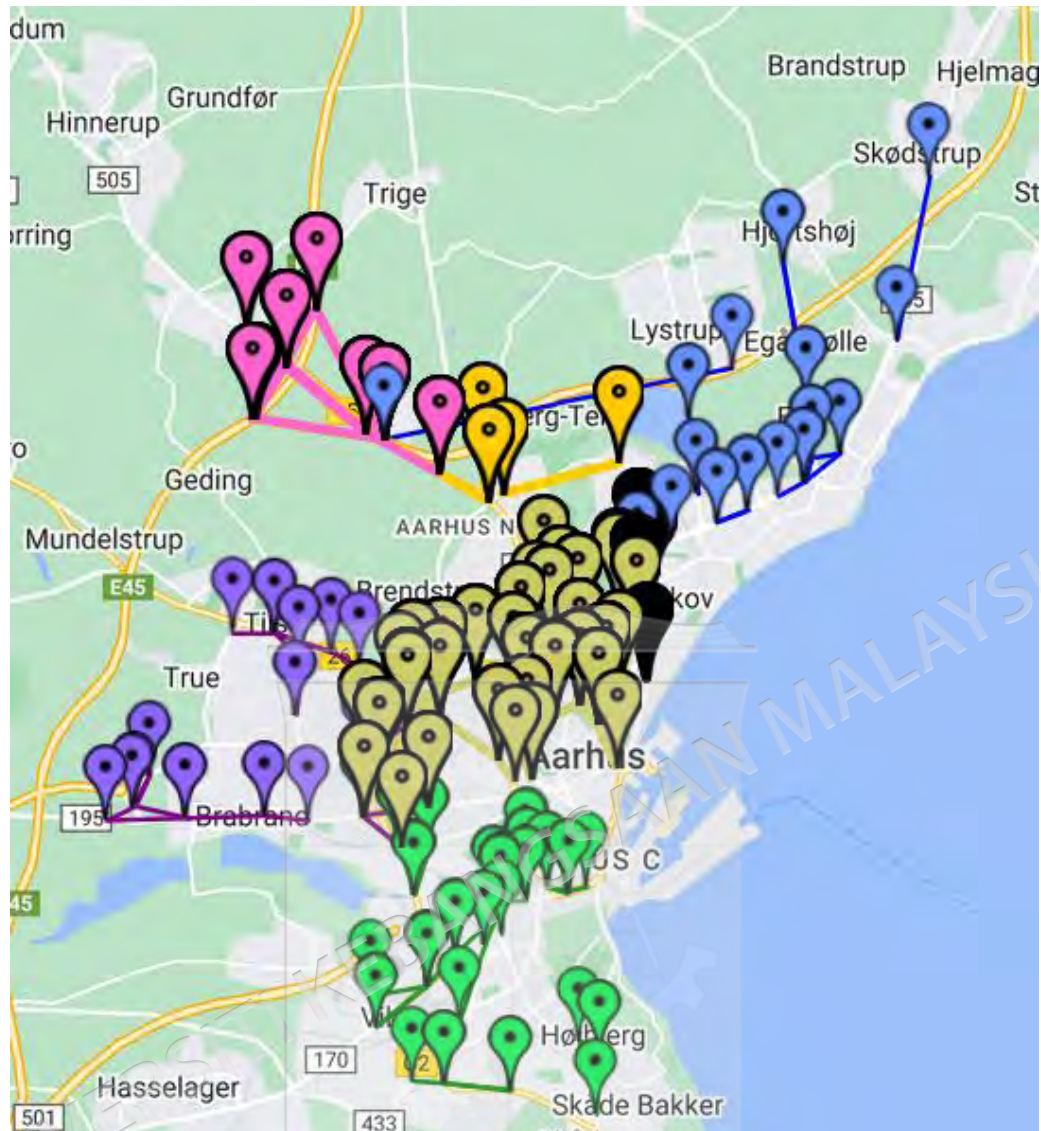
HASIL DAN PERBINCANGAN

4.1 PENDAHULUAN

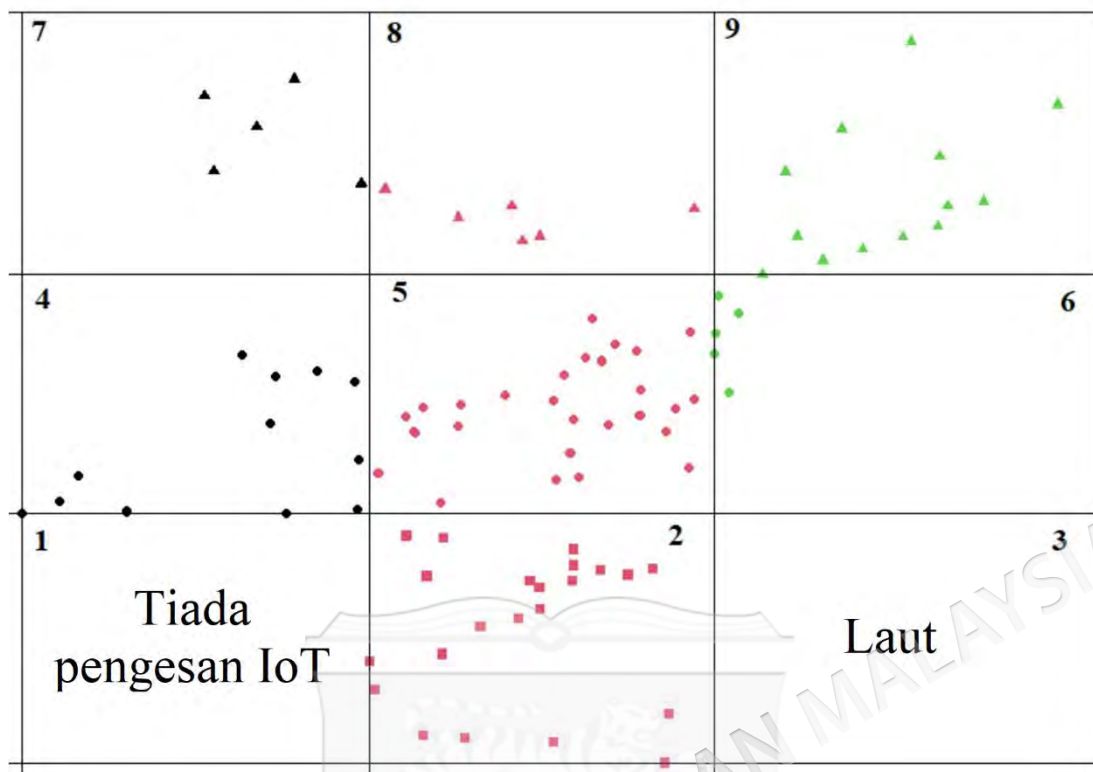
Dalam bab ini, setiap hasil eksperimen daripada modul yang terlibat dalam ramalan perambatan aliran trafik jalan raya berjiran dibincangkan dengan terperinci. Penerangan dan perbincangan dalam bab 4 ini dimulai dengan kajian prestasi GLCM dalam mengenal pasti jaran berjiran. Seterusnya, kesan gugusan aras kesesakan dikaji dengan menggunakan algoritma K-Means dalam meningkatkan ketepatan ramalan aliran trafik. Dan akhirnya, hasil ramalan perambatan aliran trafik jalan raya berjiran menggunakan model yang dicadangkan diterangkan sebelum hasil perbandingan dengan ramalan perambatan trafik yang sedia ada.

4.2 SET DATA EKSPERIMEN KAJIAN

Set data yang digunakan dalam eksperimen kajian ini terdiri daripada 449 alat pengesan IoT di Aarhus, Denmark. Daripada 449 alat pengesan IoT, hanya 155 data daripada alat pengesan digunakan kerana selebihnya alat pengesan IoT rosak. Rajah 4.1 menunjukkan peta lokasi eksperimen yang mempunyai alat pengesan IoT tersebut. Lokasi eksperimen disegmentasikan kepada sembilan (9) kawasan berdasarkan lokasi koordinat alat pengesan IoT seperti yang ditunjukkan pada Rajah 4.2.

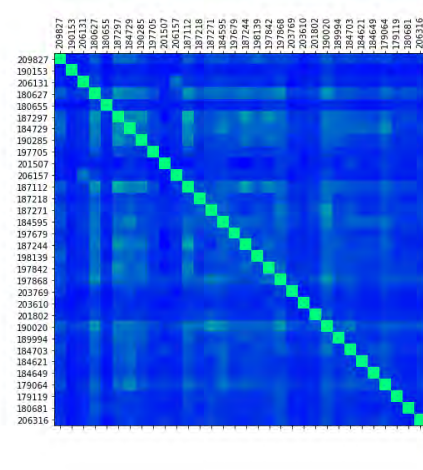


Rajah 4.1 Peta lokasi eksperimen.

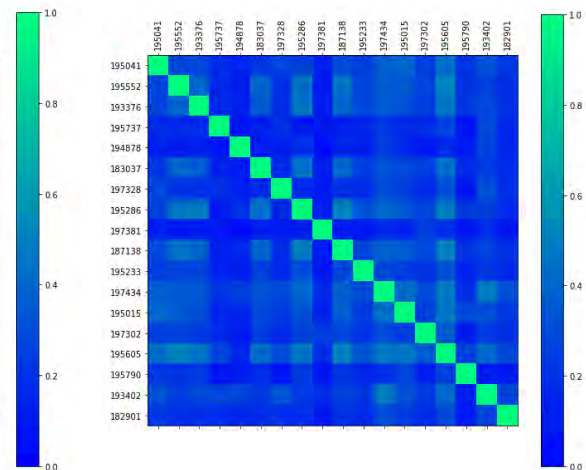


Rajah 4.2 Sembilan (9) segmentasi kawasan.

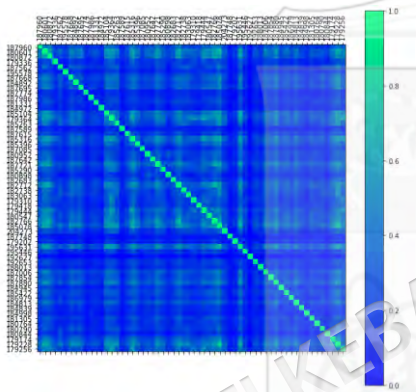
Berkaitan dengan fokus utama dari kajian ini, iaitu ramalan perambatan trafik berasaskan jalan yang mengalami kesesakan. Maka model ramalan yang dicadangkan memiliki kekangan iaitu jalan raya harus memiliki hubungan korelasi antara satu sama lain, atau suatu jalan memberi kesan kepada jalan lainnya. Berasaskan hal ini maka nilai korelasi antara jalan raya setiap kawasan harus dikira terlebih dahulu sebelum menentukan lokasi eksperimen. Telah diterangkan pada bahagian atas bahawa kawasan satu dan tiga tidak terdapat pengesanan IoT, sehingga kawasan dibahagi menjadi tujuh kawasan iaitu kawasan dua, kawasan empat, kawasan lima, kawasan enam, kawasan tujuh, kawasan lapan dan kawasan sembilan. Nilai korelasi antara jalan di kawasan dua dan kawasan empat ditunjukkan pada Rajah 4.3 (a) dan (b). Nilai korelasi antara jalan di kawasan lima dan kawasan enam ditunjukkan pada Rajah 4.3 (c) dan (d). Nilai korelasi antara jalan di kawasan tujuh dan di kawasan lapan ditunjukkan pada Rajah 4.3 (e) dan (f). Nilai korelasi antara jalan di kawasan sembilan ditunjukkan pada Rajah 4.3 (g).



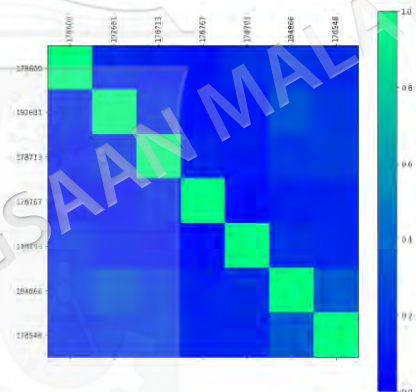
(a) Kawasan dua.



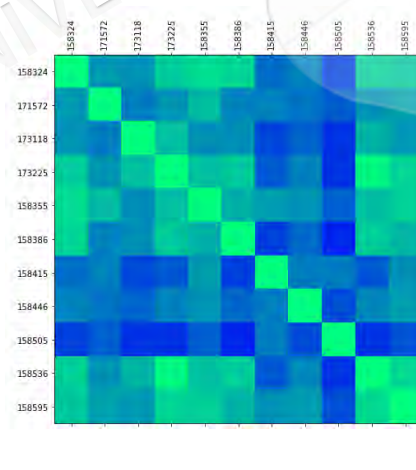
(b) Kawasan empat.



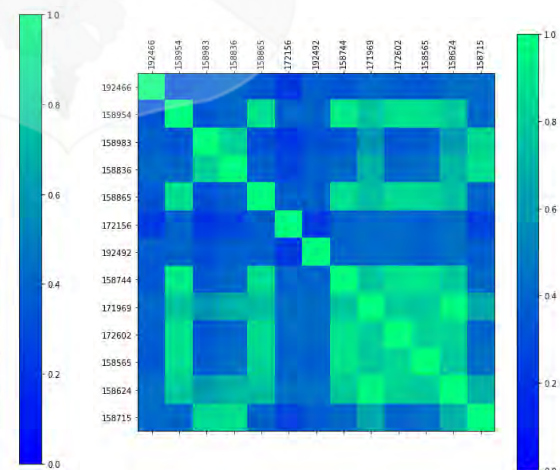
(c) Kawasan lima.



(d) Kawasan enam.



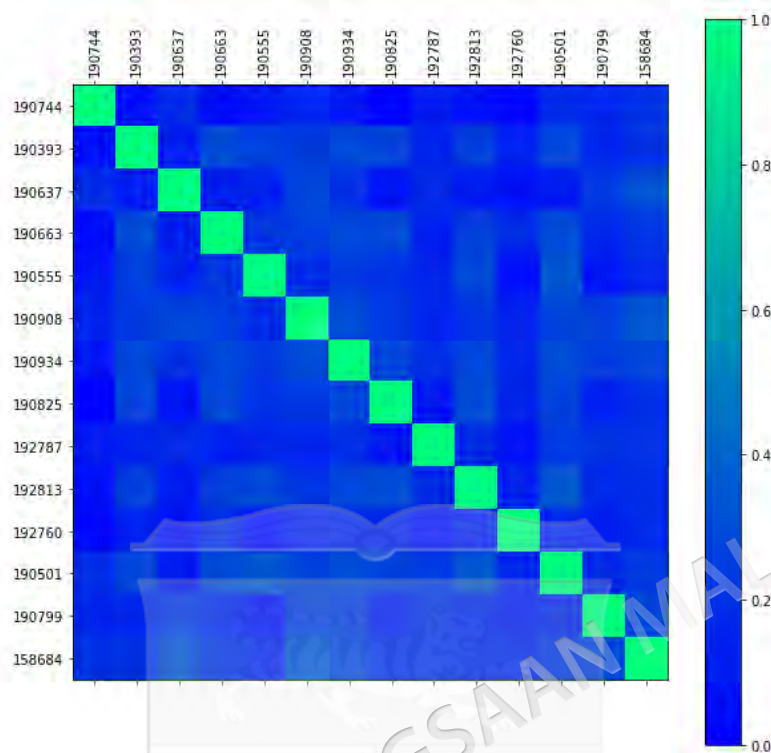
(e) Kawasan tujuh.



(f) Kawasan lapan.

bersambung...

...sambungan



(g) Kawasan sembilan.

Rajah 4.3 Nilai korelasi antara jalan pada tujuh (7) Kawasan.

Rajah 4.3 (e) dan (f) menunjukkan kawasan 7 dan 8 dominan berwarna hijau, sebaliknya kawasan lain dominan berwarna biru. Dari carta bara terlihat apabila korelasi semakin mendekati satu maka warna hijau semakin kuat, sedangkan semakin mendekati nol maka warna biru semakin kuat. Seterusnya, korelasi indeks kesesakan antara jalan berjiran bagi setiap lokasi atau jalan sasaran di kawasan 7 dan 8 dinilai dalam menentukan pilihan jalan sasaran untuk eksperimen kajian. Berdasarkan kajian oleh (Keler et al. 2017) (Akoglu 2018) bahawa jalan raya memiliki hubungan kuat apabila memiliki nilai purata korelasi pearson lebih dari 0.6 kerana itu nilai minimum purata korelasi indeks kesesakan yang dipertimbangkan dalam kajian ini adalah 0.6. Selain daripada indeks kesesakan, pilihan jalan sasaran juga bergantung kepada nilai frekuensi kesesakan untuk memastikan jalan sasaran yang dipilih kerap mengalami kesesakan pada tempoh masa 05.00 hingga 17.00 (12 jam) untuk set data selama 3

bulan. Puncak selang masa kesesakan selama satu hari adalah 0.5 hingga 2 jam (Susilo & Imanuel 2018) (Christidis & Ibanes Rivas 2012), dengan demikian median masa kesesakan adalah 80 minit. Kerana kajian ini menghitung indeks kesesakan dengan selang masa 20 minit, maka dalam satu hari berlaku empat (4) kesesakan pada tempoh masa pukul 05.00 hingga 17.00. Untuk set data selama 3 bulan maka nilai minimum frekuensi kesesakan adalah $4 \times 4 \times 12 = 192$. Sehingga nilai frekuensi kesesakan minimum dalam kajian ini adalah 192 kesesakan.

Jadual 4.1 dan 4.2 menunjukkan hubungan matriks korelasi indeks kesesakan antara jalan berjiran dan frekuensi kesesakan bagi jalan sasaran di kawasan 7 dan 8. Jalan sasaran yang nilai frekuensi kesesakan kurang daripada 192 dan purata korelasi indeks kesesakan kurang dari 0.6 diabaikan dalam kajian ini (rujuk baris yang berwarna kuning pada Jadual 4.1 dan 4.2).

Jadual 4.1 Matriks korelasi indeks kesesakan bagi jalan sasaran di kawasan 7

Jalan Sasaran/ Lokasi	158324	171572	173118	173225	158355	158386	158415	158446	158505	158536	158595	Purata Korelasi	Frekuensi Kesesakan
158324	1.00	0.60	0.58	0.80	0.86	0.83	0.43	0.51	0.26	0.81	0.78	0.68	305
171572	0.60	1.00	0.50	0.58	0.74	0.52	0.51	0.45	0.37	0.58	0.64	0.59	7
173118	0.58	0.50	1.00	0.75	0.56	0.57	0.29	0.39	0.20	0.70	0.60	0.56	1
173225	0.80	0.58	0.75	1.00	0.74	0.79	0.35	0.52	0.21	0.96	0.83	0.69	553
158355	0.86	0.74	0.56	0.74	1.00	0.69	0.61	0.57	0.38	0.75	0.81	0.70	275
158386	0.83	0.52	0.57	0.79	0.69	1.00	0.25	0.42	0.15	0.80	0.71	0.61	350
158415	0.43	0.51	0.29	0.35	0.61	0.25	1.00	0.50	0.49	0.35	0.57	0.49	367
158446	0.51	0.45	0.39	0.52	0.57	0.42	0.50	1.00	0.30	0.53	0.62	0.53	45
158505	0.26	0.37	0.20	0.21	0.38	0.15	0.49	0.30	1.00	0.21	0.32	0.35	75
158536	0.81	0.58	0.70	0.96	0.75	0.80	0.35	0.53	0.21	1.00	0.85	0.69	386
158595	0.78	0.64	0.60	0.83	0.81	0.71	0.57	0.62	0.32	0.85	1.00	0.70	321

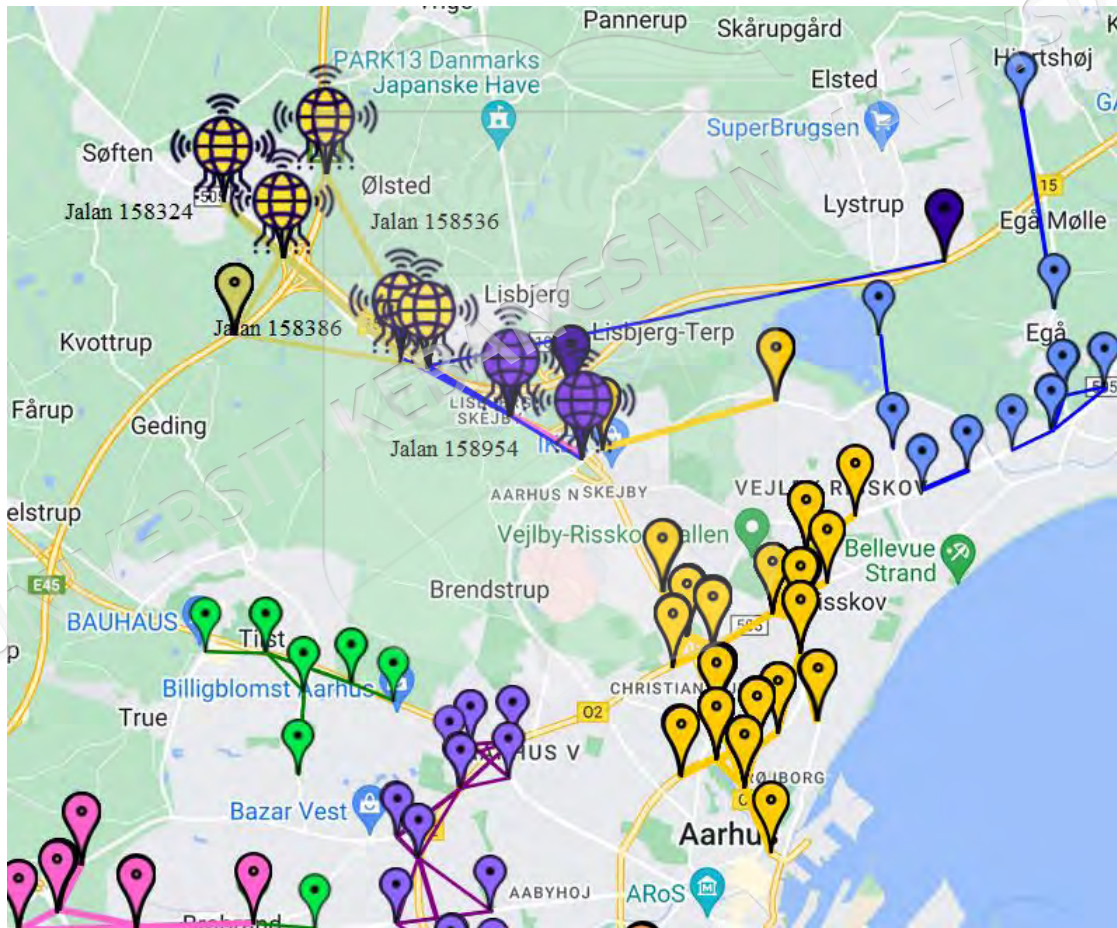
Jadual 4.2 Matriks korelasi indeks kesesakan bagi jalan sasaran di kawasan 8

Jalan Sasaran/ Lokasi	192466	158954	158983	158836	158865	172156	192492	158744	171969	172602	158565	158624	158715	Purata Korelasi	Frekuensi Kesesakan
192466	1.00	0.34	0.37	0.42	0.34	0.23	0.37	0.34	0.41	0.37	0.35	0.43	0.42	0.41	604
158954	0.34	1.00	0.34	0.41	0.88	0.39	0.39	0.92	0.78	0.89	0.88	0.79	0.41	0.65	366
158983	0.37	0.34	1.00	0.82	0.32	0.23	0.32	0.34	0.59	0.34	0.37	0.62	0.84	0.50	359
158836	0.42	0.41	0.82	1.00	0.38	0.27	0.37	0.42	0.70	0.40	0.45	0.72	0.87	0.56	396
158865	0.34	0.88	0.32	0.38	1.00	0.38	0.39	0.87	0.75	0.86	0.84	0.75	0.39	0.63	380
172156	0.23	0.39	0.23	0.27	0.38	1.00	0.23	0.40	0.42	0.42	0.39	0.44	0.27	0.39	139
192492	0.37	0.39	0.32	0.37	0.39	0.23	1.00	0.40	0.42	0.42	0.39	0.44	0.37	0.42	932
158744	0.34	0.92	0.34	0.42	0.87	0.40	0.40	1.00	0.78	0.89	0.90	0.79	0.42	0.65	281
171969	0.41	0.78	0.59	0.70	0.75	0.42	0.42	0.78	1.00	0.81	0.79	0.96	0.67	0.70	369
172602	0.37	0.89	0.34	0.40	0.86	0.42	0.42	0.89	0.81	1.00	0.84	0.83	0.41	0.65	494
158565	0.35	0.88	0.37	0.45	0.84	0.39	0.39	0.90	0.79	0.84	1.00	0.79	0.44	0.65	289
158624	0.43	0.79	0.62	0.72	0.75	0.44	0.44	0.79	0.96	0.83	0.79	1.00	0.70	0.71	294
158715	0.42	0.41	0.84	0.87	0.39	0.27	0.37	0.42	0.67	0.41	0.44	0.70	1.00	0.55	356

4.3 KEPUTUSAN EKSPERIMEN KAJIAN

4.3.1 Kaedah Penggugusan GLCM+Spektrum

Dalam seksyen ini, eksperimen yang dijalankan adalah mengkaji prestasi GLCM dalam menentukan nilai pemberat hubungan iaitu jalan berjiran yang memiliki hubungan kuat antara jalan raya berjiran yang digunakan untuk meramal aliran trafik. Seksyen ini membentangkan hasil penggugusan jalan raya di empat (4) lokasi jalan raya iaitu jalan 158324, Jalan 158536, Jalan 158954 dan Jalan 158386 yang terletak pada kawasan tujuh dan kawasan lapan dipaparkan dalam Rajah 4.4.



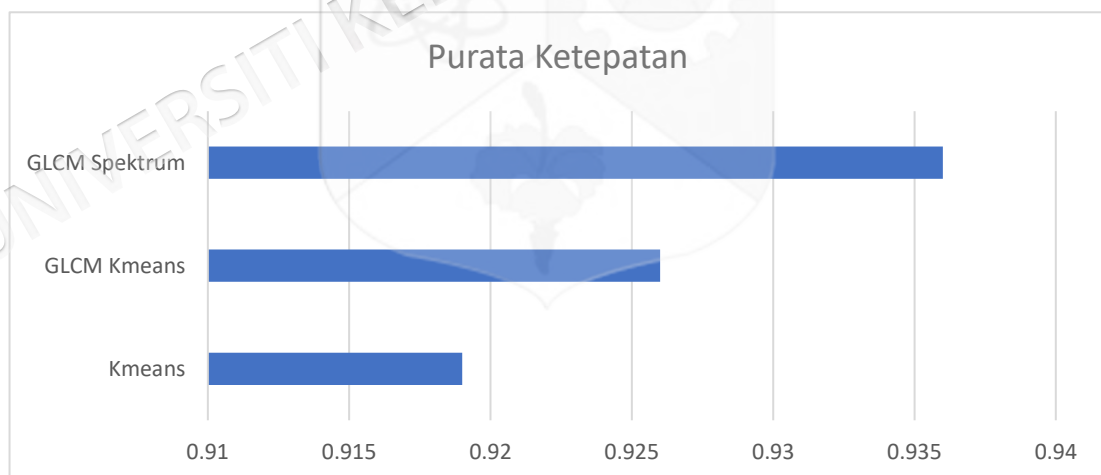
Rajah 4.4 Peta empat lokasi eksperimen di bandar Aarhus, Denmark.

Menurut (Y. Wang et al. 2016), dalam 20 minit kesesakan boleh merambat sekitar 5 km. Oleh itu kajian ini mengguguskan jalan raya di sekitar radius 5-kilometer dari lokasi eksperimen iaitu 5 km daripada jalan 158324, 5 km daripada Jalan 158536, dan 5km daripada Jalan 158954. Hasil perlaksanaan modul

penggugusan adalah senarai jalan berjiran daripada lokasi eksperimen. Tiga kaedah penggugusan dibandingkan iaitu K-Means, GLCM+K-Means, GLCM+Spektrum di mana GLCM adalah satu kaedah untuk mengekstrak fitur yang diumpukkan pada kaedah penggugusan. Perbandingan prestasi tiga (3) kaedah penggugusan tersebut dilakukan dengan merujuk kepada hasil ramalan aliran trafik berdasarkan jalan berjiran. Kaedah ramalan yang digunakan pada peringkat ini adalah K-Nearest Neighbour (KNN). Keputusan eksperimen yang dijalankan mengesahkan bahawa penggugusan spektrum+GLCM lebih baik dibandingkan dengan kaedah penggugusan GLCM+K-Means dan K-Means sahaja. Keputusan eksperimen ramalan aliran trafik jalan berjiran bagi setiap lokasi diterangkan dibawah.

a. Lokasi Jalan Raya 158324

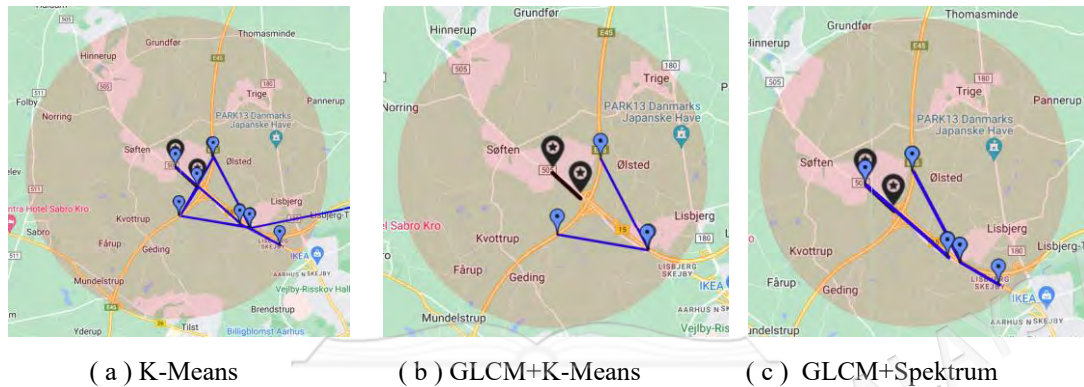
Hasil purata ketepatan ramalan aliran trafik jalan berjiran bagi lokasi 158324 ditunjukkan pada Rajah 4.5. Rajah 4.5 menunjukkan kaedah penggugusan GLCM+Spektrum menghasilkan purata ketepatan ramalan tertinggi berbanding GLCM+K-Means dan K-Means.



Rajah 4.5 Carta bara daripada ramalan keadaan trafik jalan raya berjiran 158324.

Rajah 4.6 menunjukkan visual jalan berjiran sekitar radius 5-kilometer dari lokasi 158324 yang masing-masingnya menggunakan kaedah penggugusan K-Means, GLCM+K-Means dan GLCM+Spektrum. Kaedah penggugusan K-Means dan GLCM+Spektrum menunjukkan semua jalan berjiran saling berhubung di antara satu

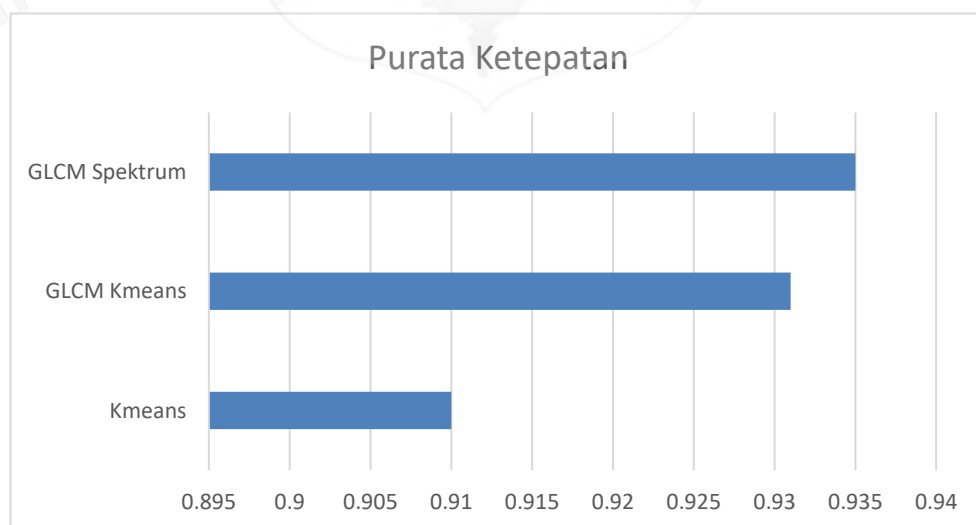
sama lain. Manakala, dengan menggunakan kaedah penggugusan GLCM+K-Means, terdapat dua jalan yang berwarna biru tidak berhubung dengan sebahagian jalan berjiran di mana jalan berjiran yang saling berhubung antara satu sama lain adalah merupakan hasil penggugusan yang lebih relevan.



Rajah 4.6 Hasil daripada tiga kaedah penggugusan daripada jalan raya 158324.

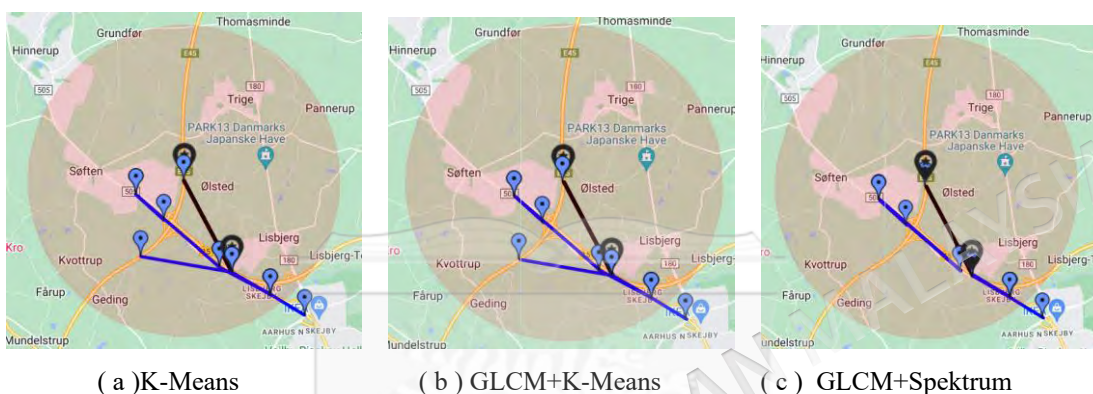
b. Lokasi Jalan Raya 158536

Bagi lokasi 158536, purata ketepatan ramalan aliran trafik jalan berjiran ditunjukkan pada Rajah 4.7. Rajah 4.7 juga menunjukkan kaedah penggugusan GLCM+Spektrum menghasilkan purata ketepatan ramalan 0.935 berbanding dengan GLCM+K-Means dan K-Means, yang menunjukkan 0.93 dan 0.91.



Rajah 4.7 Carta bara daripada ramalan keadaan trafik jalan raya berjiran 158536.

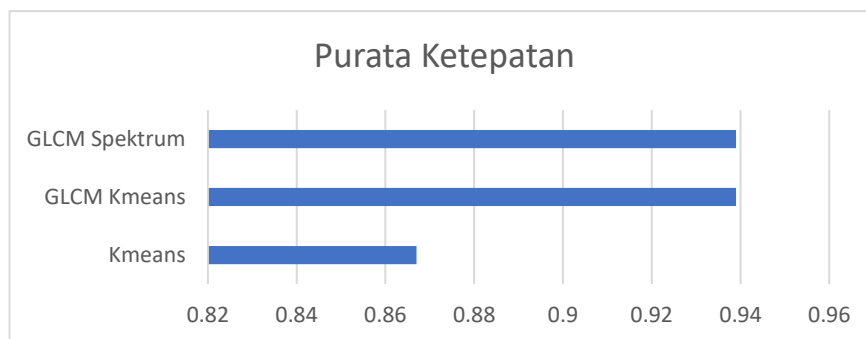
Rajah 4.8 menunjukkan visual jalan berjiran sekitar radius 5-kilometer pada lokasi 158536 yang masing-masingnya menggunakan kaedah penggugusan K-Means, GLCM+K-Means dan GLCM+Spektrum. Kaedah penggugusan daripada semua kaedah penggugusan saling berhubung antara satu sama lain. Namun penggugusan GLCM+spektrum menunjukkan hasil lebih baik kerana ia mampu menapis (filtering) jalan berjiran yang paling relevan seperti yang ditunjukkan pada Rajah 4.8(c).



Rajah 4.8 Hasil penggugusan daripada jalan raya 158536.

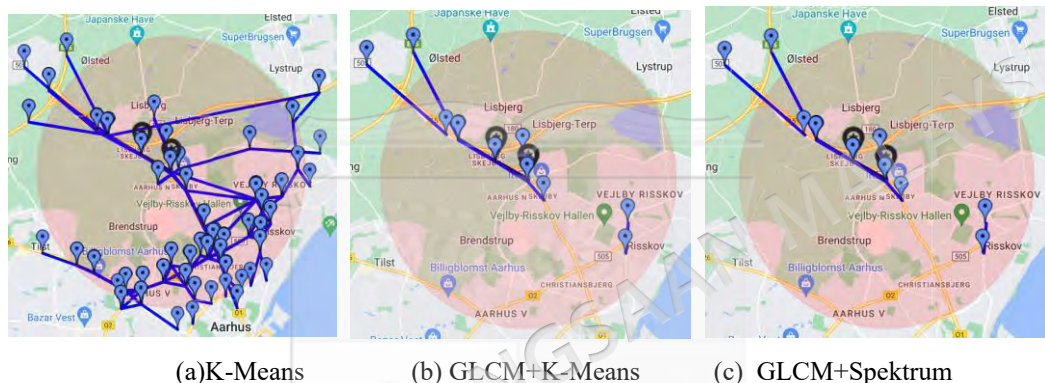
c. Lokasi Jalan Raya 158954

Hasil purata ketepatan ramalan aliran trafik jalan berjiran bagi lokasi 158954 ditunjukkan pada Rajah 4.9 Didapati kaedah penggugusan GLCM+spektrum dan GLCM+Kmeans menunjukkan purata ketepatan ramalan yang sama iaitu lebih tinggi daripada kaedah penggugusan K-Means.



Rajah 4.9 Carta bara daripada ramalan keadaan trafik jalan raya berjiran 158594.

Rajah 4.10 menunjukkan visual jalan berjiran sekitar radius 5-kilometer dari lokasi 158954 yang masing-masingnya menggunakan kaedah penggugusan K-Means, GLCM+K-Means dan GLCM+Spektrum. Kaedah penggugusan K-Means menunjukkan terlalu banyak jalan yang saling berhubung atau berjiran. Ini menunjukkan kaedah penggugusan K-Means gagal menapis jalan raya yang tidak relevan. Sebaliknya penggugusan GLCM+K-Means dan GLCM+Spektrum mampu menapis lebih banyak jalan yang tidak relevan dan menghasilkan jumlah jalan berjiran yang sedikit tetapi relevan.



Rajah 4.10 Carta bara daripada ramalan keadaan trafik jalan raya berjiran 158954.

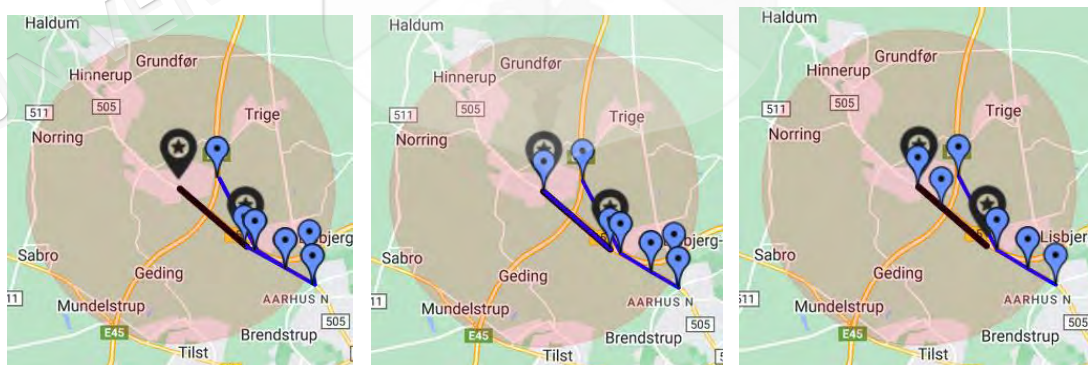
d. Lokasi Jalan Raya 158386

Bagi lokasi 158386, purata ketepatan ramalan aliran trafik jalan berjiran ditunjukkan pada Rajah 4.11. Rajah 4.11 juga menunjukkan kaedah penggugusan GLCM+Spektrum menghasilkan purata ketepatan ramalan 0.946 berbanding dengan GLCM+K-Means dan K-Means, yang menunjukkan 0.943 dan 0.919.



Rajah 4.11 Carta bara daripada ramalan keadaan trafik jalan raya berjiran 158386.

Rajah 4.12 menunjukkan visual jalan berjiran sekitar radius 5-kilometer pada lokasi 158386 yang masing-masingnya menggunakan kaedah penggugusan K-Means, GLCM+K-Means dan GLCM+Spektrum. Kaedah penggugusan daripada semua kaedah penggugusan saling berhubung antara satu sama lain. Namun penggugusan GLCM+spektrum menunjukkan hasil lebih baik kerana ia mampu menapis (filtering) jalan berjiran yang paling relevan seperti yang ditunjukkan pada Rajah 4.12(c).



(a) K-Means

(b) GLCM+K-Means

(c) GLCM+Spektrum

Rajah 4.12 Carta bara daripada ramalan keadaan trafik jalan raya berjiran 158386.

4.3.2 Persilangan GLCM+Spektrum dan Spektrum Frekuensi

Eksperimen dalam seksyen ini dijalankan pada lokasi jalan 158324 di kawasan tujuh. Tiga kaedah penggugusan dibandingkan untuk mencari jalan berjiran iaitu GLCM+Spektrum, Spektrum frekuensi, dan $\text{GLCM+Spektrum} \cap \text{Spektrum frekuensi}$. HMM digunakan untuk meramal keadaan trafik jalan berjiran berdasarkan keadaan trafik di jalan 158324. Visualisasi peta lokasi jalan 158324 daripada tiga kaedah penggugusan ditunjukkan pada Rajah 4.13.



(a) GLCM+spektrum (b) spektrum frekuensi
spektrum frekuensi. (c) GLCM+spektrum $\cap$

Rajah 4.13 Jalan berjiran hasil penggugusan (a), (b) dan (c).

Ketepatan ramalan trafik jalan berjiran berdasarkan kaedah gugusan GLCM+Spektrum, Spektrum Frekuensi dan $\text{GLCM+Spektrum} \cap \text{Spektrum Frekuensi}$ yang masing-masing ditunjukkan pada Jadual 4.3, 4.4 dan 4.5. Didapati jalan 158744, 172602 (Jadual 4.3) dan 158355 (Jadual 4.4) mempunyai nilai precision 0, 0 dan 0.3, tetapi nilai ketepatan 0.9, 0.82 dan 0.87. Ini menunjukkan jalan tersebut tidak sesak kerana nilai precision rendah walaupun nilai ketepatan tinggi. Hasil kajian ini mendapati, nilai *precision* dan ketepatan yang tinggi hanya boleh menentukan ramalan perambatan kesesakan jalan berjiran yang paling tepat.

Jadual 4.3 Hasil ramalan trafik daripada gugusan GLCM+spektrum

Jalan raya	Ketepatan	Precision (sesak)
158386	0.92	0.91
158536	0.89	0.56
173225	0.86	0.97
158415	0.80	0.03
158565	0.83	0.02
158715	0.88	0.45
158744	0.90	0
172602	0.82	0
158983	0.88	0.51
158836	0.91	0.86

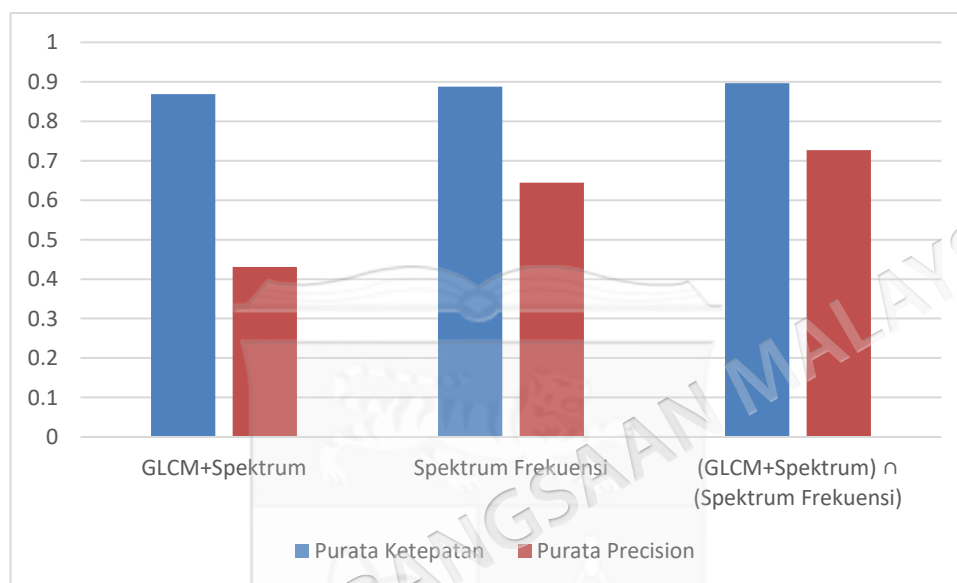
Jadual 4.4 Hasil ramalan trafik daripada gugusan spektrum frekuensi.

Jalan raya	Ketepatan	Precision (sesak)
158386	0.92	0.91
158355	0.87	0.3
158536	0.89	0.56
173225	0.86	0.97
158595	0.88	0.33
158715	0.89	0.47
158983	0.88	0.51
158836	0.91	0.84
158386	0.89	0.91

Jadual 4.5 Hasil ramalan trafik daripada gugusan GLCM+spektrum $\cap$ spektrum frekuensi

Jalan raya	Ketepatan	Precision (sesak)
158386	0.92	0.87
158536	0.90	0.59
173225	0.86	0.95
158715	0.90	0.55
158983	0.89	0.56
158836	0.91	0.84

Perbandingan purata ketepatan (*accuracy*) jalan berjiran dan perbandingan purata (*precision*) jalan berjiran daripada ketiga kaedah penggugusan ditunjukkan pada Rajah 4.14. Rajah 4.14 menunjukkan bahawa penggugusan yang menggunakan kaedah GLCM+spektrum $\cap$ spektrum frekuensi menghasilkan prestasi yang terbaik dalam meramal keadaan kesesakan.



Rajah 4.14 Jalan berjiran hasil penggugusan (a), (b) dan (c).

4.3.3 Ramalan perambatan aliran trafik jalan berjiran

Objektif utama kajian ini seperti yang diterangkan dalam bab 1, adalah meramal perambatan aliran trafik jalan berjiran. Kajian ini mengira indeks kesesakan menggunakan set data hari bekerja bermula dari 15/2/2014 hingga 31/5/2014, dari jam 00.00 hingga 20.00. Sebanyak 20% daripada set data digunakan untuk menguji model ramalan yang dicadangkan. Kajian ramalan dilakukan pada selang masa 05.00 hingga pukul 17.00 daripada set yang digunakan dalam kajian ini. Kajian ini mentakrifkan jalan berjiran sebagai jalan yang terletak kurang daripada 5-kilometer radius dari lokasi atau jalan sasaran. Peta visual juga dihasilkan dalam kajian untuk menunjukkan analisis hubungan antara jalan sasaran dan jalan disekitarnya atau dikenali sebagai jalan berjiran.

Seksyen ini membincangkan hasil eksperimen yang dijalankan untuk meramal perambatan kesesakan aliran trafik jalan berjiran berdasarkan maklumat keadaan trafik jalan sasaran (atau kesan daripada kesesakan jalan sasaran) dan masa kejadian dengan menggunakan model ramalan HMM yang dicadangkan. Dua (2) eksperimen telah dijalankan iaitu:

1. Perbandingan model ramalan antara model ramalan HMM (Aras Kesesakan(JB(Masa))) dengan penggugusan spektrum persilangan hari oleh Wang et al. (2015) dan model ramalan HMM(Aras Kesesakan(JB(Masa)+S)) yang dicadangkan dengan penggugusan spektrum persilangan hari yang dipanggil model ramalan ABC.
2. Perbandingan kaedah modul penggugusan yang digunakan untuk mengenal pasti jalan berjiran iaitu di antara model ramalan HMM(Aras Kesesakan(JB(Masa)+S)) yang dicadangkan dengan kaedah persilangan (GLCM+spektrum) $\cap$ (spektrum frekuensi) yang dicadangkan dan model ramalan model ramalan HMM(Aras Kesesakan(JB(Masa)+S)) dengan kaedah spektrum persilangan hari yang digunakan oleh Wang et al. (2015).

Unit pengukuran yang digunakan dalam kajian ini adalah ketepatan ramalan, *precision*, *recall* dan skor F1. Berikut adalah beberapa contoh hasil eksperimen yang dijalankan mengikut keadaan kesesakan jalan sasaran:

a. Jalan Sasaran 158565

Jadual 4.6 menunjukkan hasil eksperimen ramalan perambatan aliran trafik jalan berjiran dan purata ramalan iaitu kesan daripada kesesakan jalan sasaran 158565.

Jadual 4.6 Hasil ramalan bagi jalan sasaran 158565

(a) Model ramalan Wang et. al (2015)

	Ketepatan	Precision	Recall	F1
172156	0.84	0.00	0.00	0.00
158415	0.82	0.53	0.74	0.62
158505	0.77	0.42	0.56	0.48
158744	0.78	0.37	0.66	0.47
158865	0.79	0.48	0.66	0.55
158954	0.78	0.42	0.64	0.51
185078	0.77	0.38	0.49	0.43
Purata	0.79	0.37	0.54	0.44

(b) Model ramalan ABC dengan kaedah penggugusan
((GLCM+spektrum) $\cap$ (spektrum frekuensi))

	Ketepatan	Precision	Recall	F1
158415	0.82	0.54	0.68	0.60
158744	0.85	0.48	0.16	0.24
172602	0.83	0.87	0.35	0.50
158865	0.81	0.51	0.64	0.57
158954	0.91	0.86	0.62	0.72
Purata	0.84	0.65	0.49	0.53

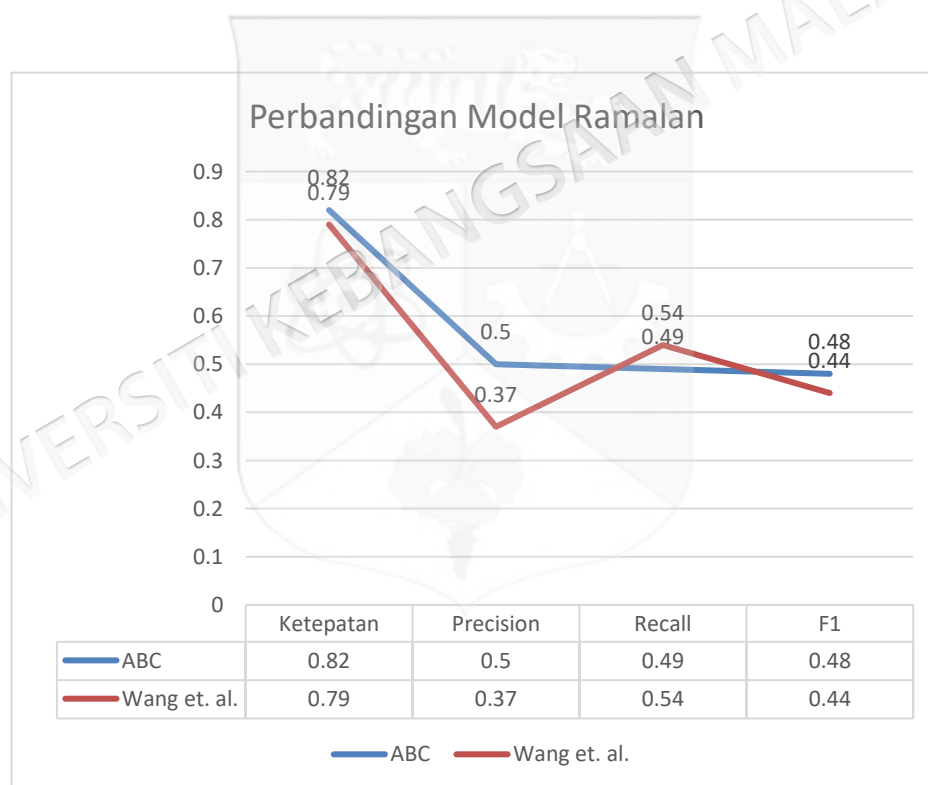
(c) Model ramalan ABC dengan kaedah penggugusan (Spektrum Persilangan hari
oleh Wang et. al. (2015))

	Ketepatan	Precision	Recall	F1
172156	0.75	0.28	0.36	0.31
158415	0.82	0.54	0.69	0.61
158505	0.78	0.42	0.46	0.44
158744	0.85	0.48	0.16	0.24
158865	0.81	0.51	0.65	0.57
158954	0.91	0.86	0.62	0.72
185078	0.79	0.40	0.49	0.44
Purata	0.82	0.50	0.49	0.48

Jadual 4.6 a menunjukkan hasil ramalan menggunakan Wang et al. (2015) iaitu gabungan antara penggugusan spektrum perulangan hari. Manakala Jadual 4.7 b

menunjukkan hasil ramalan gabungan antara modul ramalan ABC dan modul penggugusan persilangan kaedah penggugusan $((\text{GLCM}+\text{spektrum}) \cap (\text{spektrum frekuensi}))$ dan Jadual 4.6 c menunjukkan hasil ramalan gabungan antara modul ramalan ABC dan modul penggugusan spektrum persilangan hari.

Seterusnya dari Jadual 4.6 dibandingkan purata ramalan kesesakan antara modul ramalan ABC menggunakan penggugusan spektrum persilangan hari dengan model ramalan Wang et al. (2015) menggunakan kaedah penggugusan spektrum persilangan hari, hasil perbandingan ditunjukkan pada Rajah 4.15. Manakala Rajah 4.16 menunjukkan perbandingan antara model ramalan ABC menggunakan kaedah penggugusan persilangan $(\text{GLCM}+\text{spektrum}) \cap (\text{spektrum frekuensi})$ yang dicadangkan dan penggugusan spektrum persilangan hari oleh Wang et al. (2015).



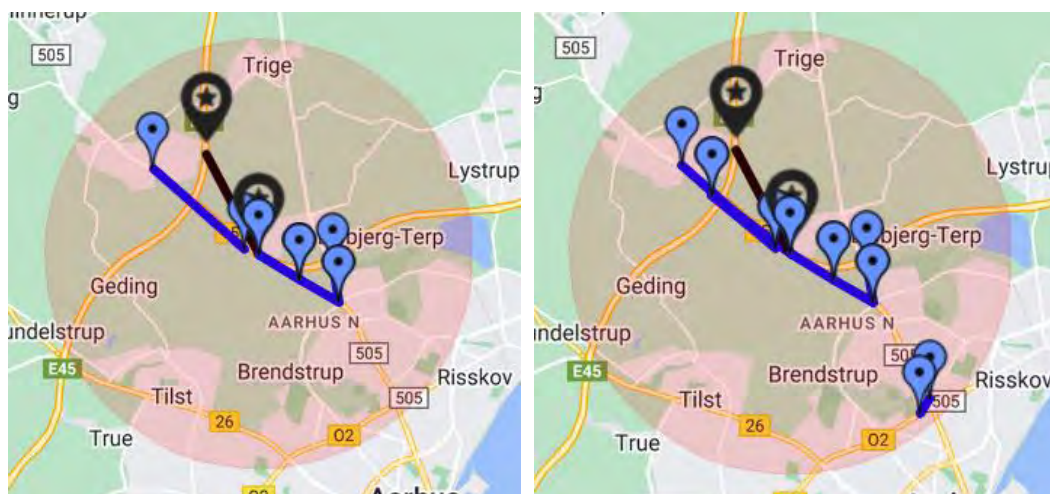
Rajah 4.15 Carta garis perbandingan model ramalan ABC dan Wang et al. (2015).



Rajah 4.16 Carta garis perbandingan penggugusan dicadangkan dan penggugusan spektrum persilangan hari, kedua-dua menggunakan HMM dicadangkan.

Rajah 4.15 menunjukkan model ramalan ABC memberikan peratusan ketepatan yang lebih baik dengan ketepatan 82% (0.82), *precision* 50% (0.5), F1 48% (0.48) berbanding dengan model ramalan oleh Wang et. al (2015) di mana ketepatan 79% (0.79), *precision* 37% (0.37), dan F1 44 % (0.44). Dengan itu, ramalan ABC berjaya menghasilkan peningkatan ketepatan 3%, *precision* 13 %, dan F1 4 %. Begitu juga dengan implementasi kaedah penggugusan persilangan (GLCM+spektrum) $\cap$ (spektrum frekuensi) menunjukkan peningkatan dengan ketepatan 2%, *precision* 10 %, dan f1 5 % yang masing-masing menunjukkan ketepatan 84% (0.84), *precision* 65% (0.65), dan f1 53 % (0.53) berbanding dengan penggugusan spektrum persilangan hari oleh Wang et. al (2015) seperti yang ditunjukkan pada Rajah 4.16.

Seterusnya, Rajah 4.17 (a) dan (b) menggambarkan visualisasi jalan berjiran pada lokasi atau jalan sasaran 158565 yang masing-masing menggunakan kaedah penggugusan (GLCM+spektrum) $\cap$ (spektrum frekuensi) dan Spektrum persilangan hari. Rajah 4,18 (b) menunjukan bahawa terdapat jalan raya yang tidak berhubung antara satu sama lain.



(a) (GLCM+spektrum) $\cap$ (spektrum frekuensi) (b) Spektrum persilangan hari

Rajah 4.17 Visualisasi jalan berjiran pada lokasi atau jalan sasaran 158565.

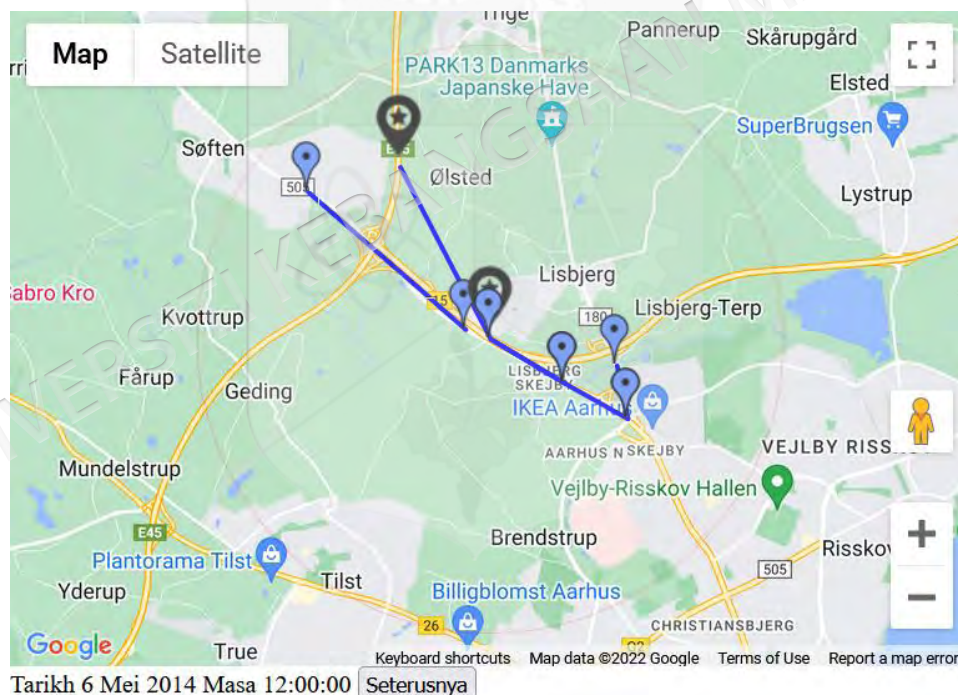
b. Visualisasi perambatan aliran trafik jalan 158565

Seksyen ini memaparkan hasil ramalan daripada jam 12.00 hingga pukul 13.40 pada lokasi atau jalan sasaran 158565. Jadual 4.7 menunjukkan ramalan perambatan aliran trafik jalan berjiran iaitu kesan daripada keadaan aliran trafik pada lokasi atau jalan sasaran 158565. Keadaan aliran trafik daripada jalan sasaran merupakan maklumat yang diketahui, sedangkan keadaan aliran trafik jalan berjiran adalah keadaan yang diramalkan berdasarkan maklumat keadaan aliran trafik jalan sasaran.

Jadual 4.7 Hasil ramalan bagi jalan sasaran 158565

Jalan Sasaran		Jalan Berjiran				
MASA	158565	158415	158744	172602	158865	158954
12:00:00	Lancar	Lancar	Lancar	Lancar	Lancar	Lancar
12:20:00	Lancar	Lancar	Lancar	Lancar	Lancar	Lancar
12:40:00	Sesak	Sesak	Sesak	Lancar	Sesak	Sesak
13:00:00	Sesak	Sesak	Sesak	Lancar	Sesak	Sesak
13:20:00	Sesak	Sesak	Sesak	Sesak	Sesak	Sesak
13:40:00	Sesak	Sesak	Sesak	Sesak	Sesak	Sesak

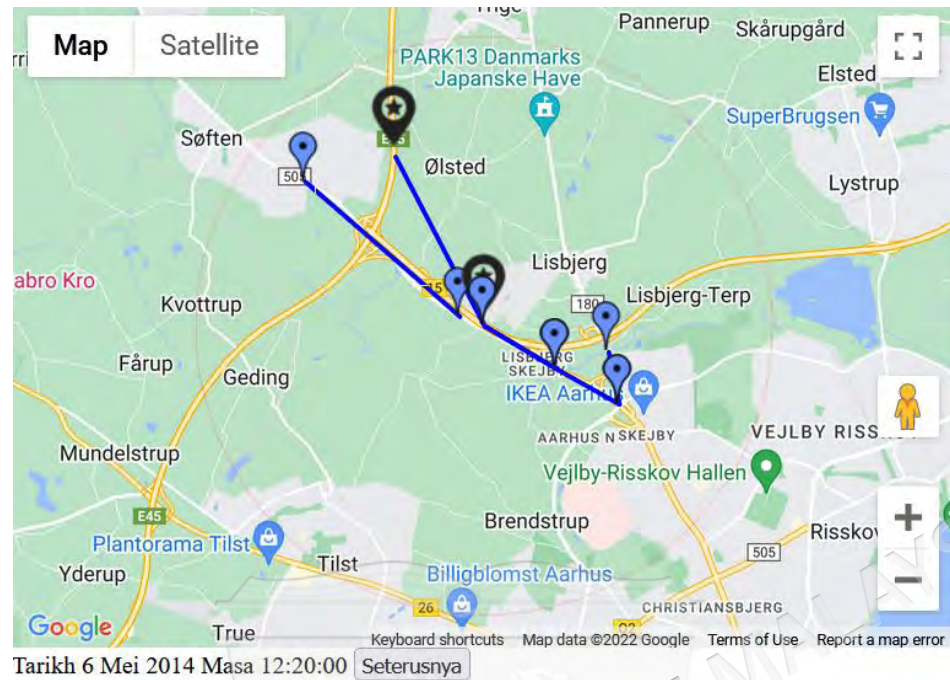
Seterusnya, keadaan aliran trafik daripada Jadual 4.7 digambarkan pada Rajah 4.18 iaitu paparan visual aplikasi ramalan perambatan aliran trafik jalan berjiran pada jalan sasaran 158565, pada selang masa 20 minit bermula dari 12.00 hingga 13.40. Garis berwarna merah menunjukkan jalan mengalami kesesakan, manakala garis berwarna biru menunjukkan jalan lancar. Rajah 4.18 (a) dan (b) menunjukkan pada jam 12.00 hingga 12.20, jalan sasaran 158565 tidak mengalami kesesakan (berwarna biru). Begitu juga dengan jalan berjiran aliran trafik lancar pada masa yang sama. Rajah 4.18 (c) dan (d) menunjukkan pada jam 12.40 dan 13.00, jalan sasaran 158565 mengalami kesesakan. Begitu juga dengan jalan berjiran, aliran trafik bertukar sesak kecuali terdapat satu (1) jalan lancar (jalan berwarna biru). Rajah 4.18 (e) dan (f) menunjukkan pada jam 13.20 dan 13.40 keadaan aliran trafik semua mengalami kesesakan.



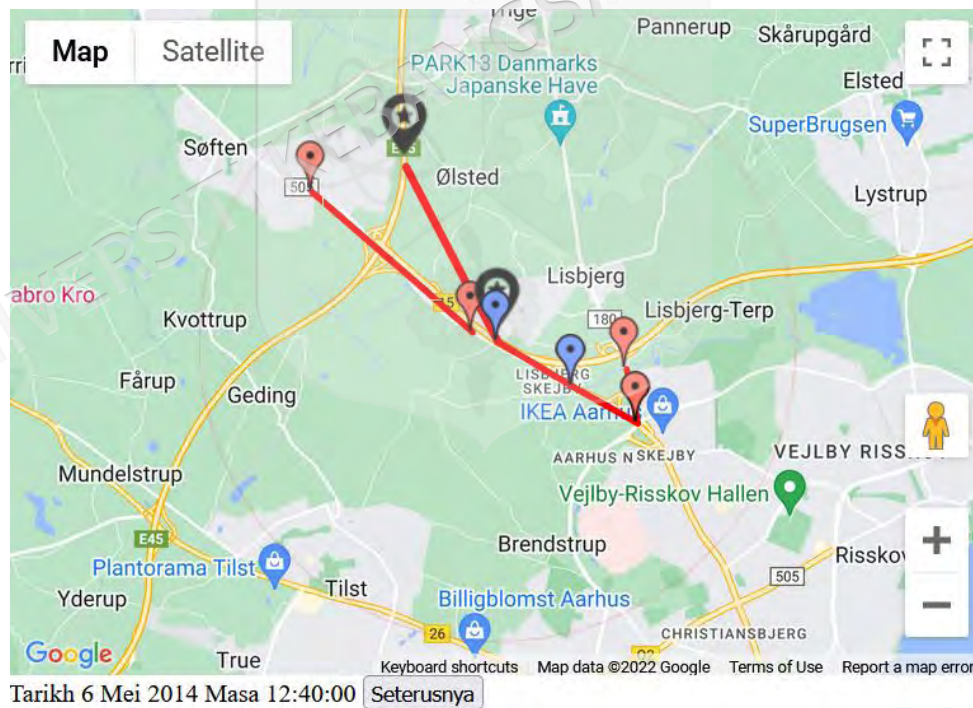
(a) Keadaan pukul 12.00

bersambung...

...sambungan



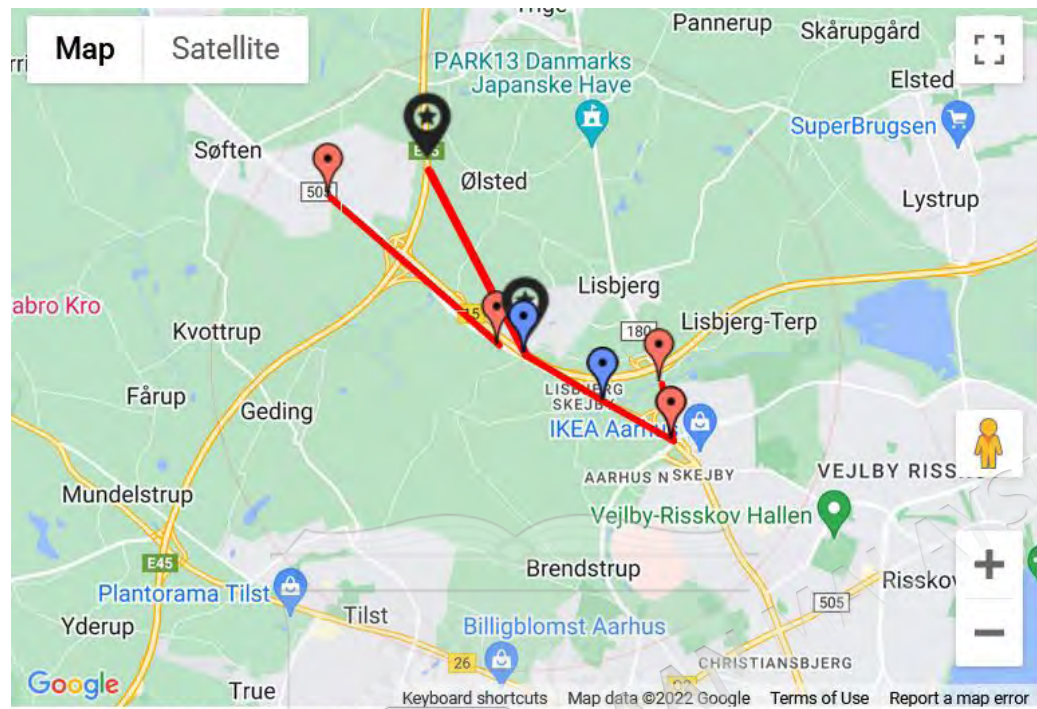
(b)Keadaan pukul 12.20



(c)Keadaan pukul 12.40

bersambung...

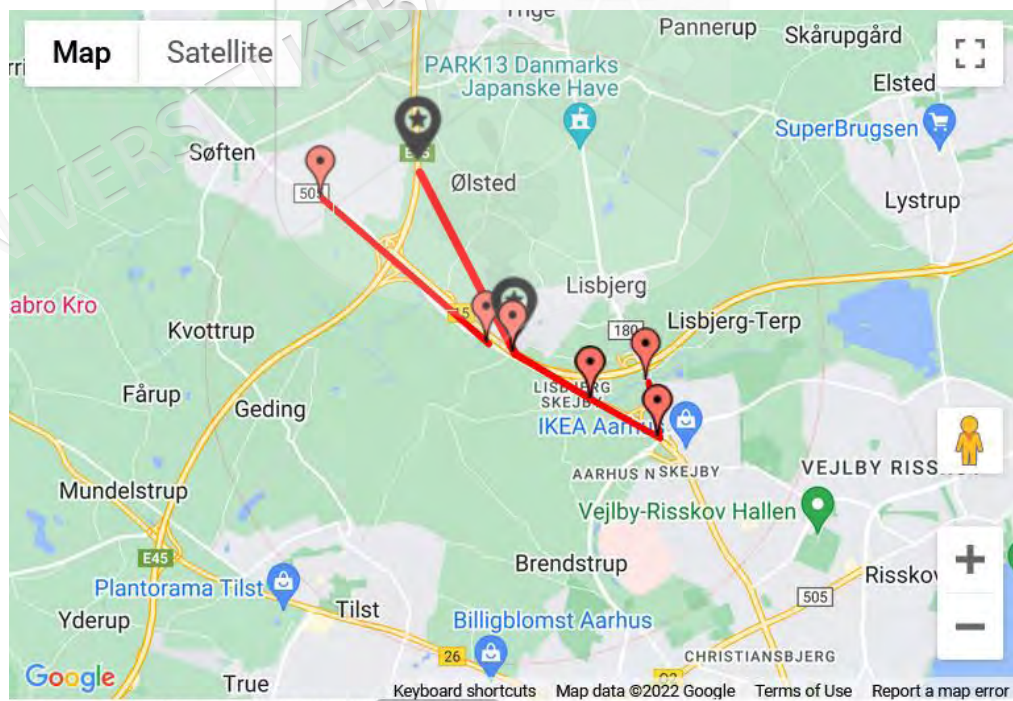
...sambungan



Tarikh 6 Mei 2014 Masa 13:00:00

Seterusnya

(d) Keadaan pukul 13.00



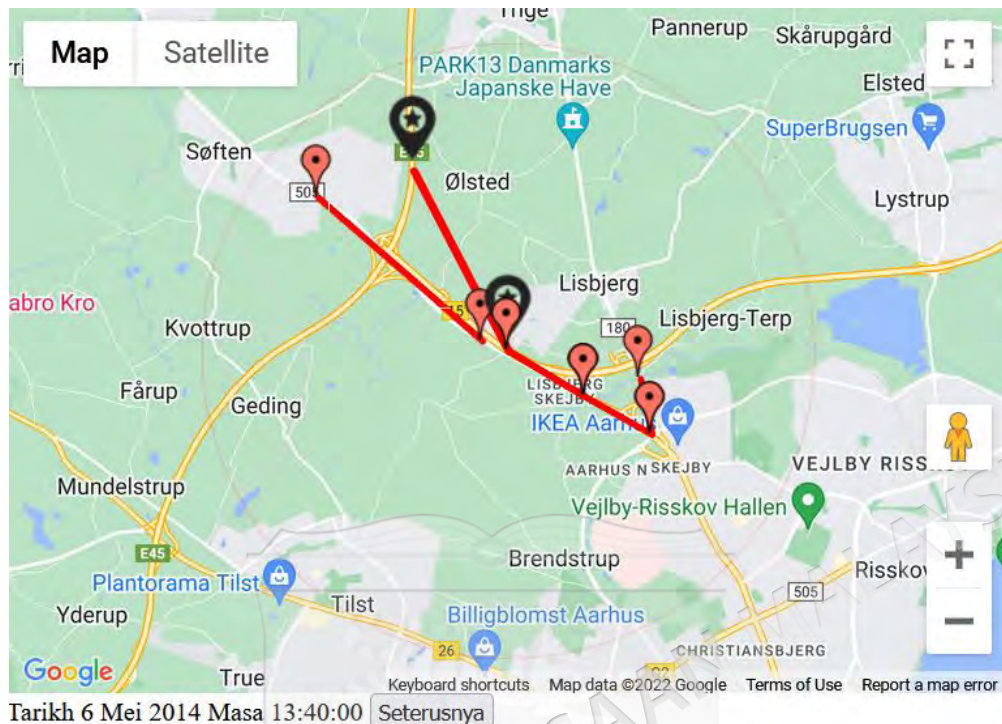
Tarikh 6 Mei 2014 Masa 13:20:00

Seterusnya

(e) Keadaan pukul 13.20

bersambung...

...sambungan



(f) Keadaan pukul 13.40

Rajah 4.18 Visualisasi jalan berjiran pada lokasi atau jalan sasaran 158565.

4.4 PERBINCANGAN

Prestasi model ramalan yang dicadangkan iaitu model ABC yang digunakan untuk meramal perambatan aliran trafik jalan berjiran pada satu-satu jalan sasaran ditentukan dengan merujuk kepada nilai purata ketepatan, *precision*, *recall* dan skor F1 semua jalan yang dikaji.

- Purata ketepatan menerangkan ketepatan ramalan secara keseluruhan.
- Purata *precision* menerangkan ketepatan ramalan aliran trafik yang sememangnya sesak.
- Purata *recall* menerangkan ramalan aliran trafik sesak tetapi tidak sesak.

- Purata skor F1 menerangkan prestasi harmoni keseimbangan antara *precision* dan *recall* daripada model ramalan ABC.

Jadual 4.8 menunjukkan hasil ramalan semua sampel jalan menggunakan model ABC iaitu gabungan modul penggugusan (GLCM+spektrum) $\cap$ (spektrum frekuensi) dan modul HMM(Aras Kesesakan(JB(Masa)+S)). Jadual 4.9 menunjukkan hasil ramalan semua sampel jalan menggunakan model Wang et al. (2015) iaitu gabungan penggugusan spektrum persilangan hari dan HMM(Aras Kesesakan(JB(Masa))). Seterusnya, Jadual 4.10 menunjukkan hasil ramalan semua sampel jalan dengan mengimplementasikan modul penggugusan spektrum persilangan hari oleh Wang et al. (2015) dalam model ramalan ABC.

Jadual 4.8 – Hasil ramalan model ABC

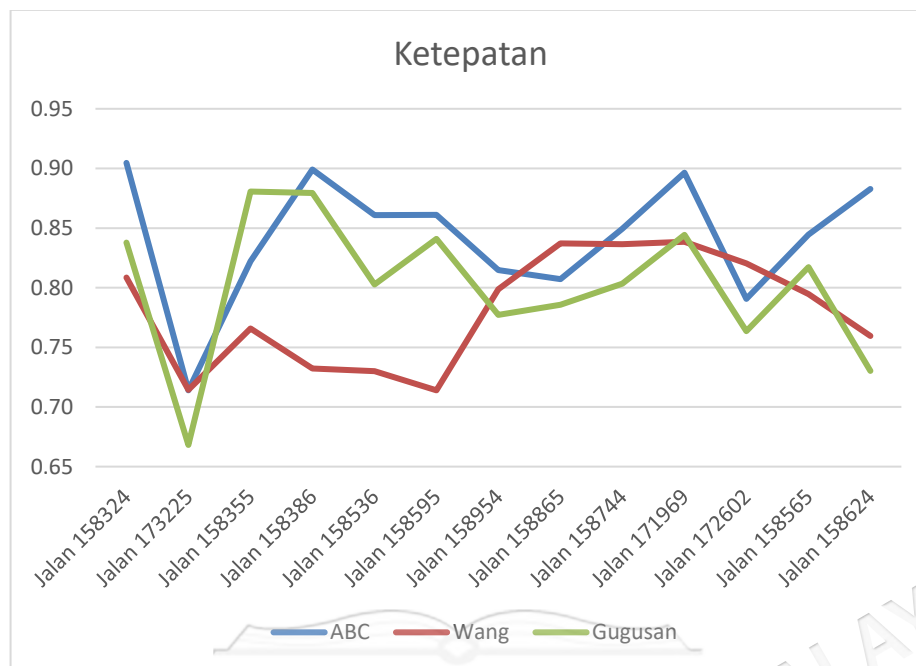
Jalan sasaran	Ketepatan	<i>Precision</i>	<i>Recall</i>	F1
Jalan 158324	0.90	0.78	0.66	0.72
Jalan 173225	0.71	0.39	0.81	0.52
Jalan 158355	0.82	0.59	0.35	0.43
Jalan 158386	0.90	0.81	0.69	0.73
Jalan 158536	0.86	0.70	0.73	0.69
Jalan 158595	0.86	0.69	0.39	0.49
Jalan 158954	0.81	0.61	0.65	0.61
Jalan 158865	0.81	0.53	0.67	0.56
Jalan 158744	0.85	0.71	0.57	0.60
Jalan 171969	0.90	0.63	0.45	0.50
Jalan 172602	0.79	0.44	0.72	0.54
Jalan 158565	0.84	0.65	0.49	0.53
Jalan 158624	0.88	0.65	0.36	0.46

Jadual 4.9 Hasil ramalan model Wang et. al. (2015)

Jalan sasaran	Ketepatan	Precision	Recall	F1
Jalan 158324	0.81	0.15	0.28	0.19
Jalan 173225	0.71	0.41	0.69	0.48
Jalan 158355	0.77	0.29	0.55	0.38
Jalan 158386	0.73	0.45	0.77	0.55
Jalan 158536	0.73	0.43	0.67	0.50
Jalan 158595	0.71	0.41	0.87	0.55
Jalan 158954	0.80	0.38	0.55	0.44
Jalan 158865	0.84	0.28	0.35	0.30
Jalan 158744	0.84	0.28	0.36	0.31
Jalan 171969	0.84	0.25	0.37	0.29
Jalan 172602	0.82	0.24	0.32	0.27
Jalan 158565	0.79	0.37	0.54	0.44
Jalan 158624	0.76	0.28	0.33	0.30

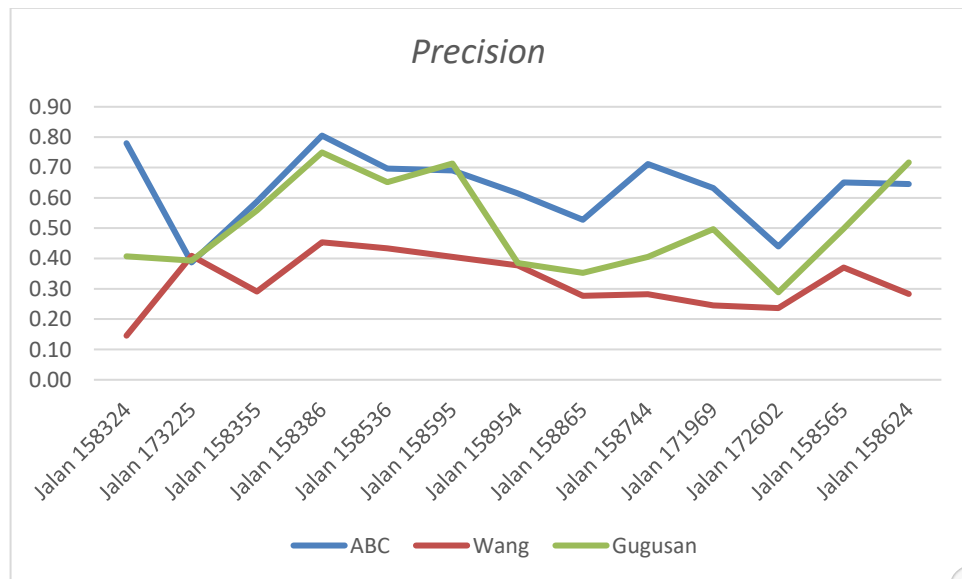
Jadual 4.10 Hasil ramalan penggugusan spektrum persilangan hari dalam model ABC

Jalan sasaran	Ketepatan	Precision	Recall	F1
Jalan 158324	0.84	0.41	0.33	0.36
Jalan 173225	0.67	0.39	0.64	0.45
Jalan 158355	0.88	0.56	0.41	0.48
Jalan 158386	0.88	0.75	0.66	0.69
Jalan 158536	0.80	0.65	0.57	0.58
Jalan 158595	0.84	0.71	0.38	0.50
Jalan 158954	0.78	0.39	0.54	0.44
Jalan 158865	0.79	0.35	0.46	0.39
Jalan 158744	0.80	0.41	0.44	0.40
Jalan 171969	0.84	0.50	0.24	0.31
Jalan 172602	0.76	0.29	0.45	0.35
Jalan 158565	0.82	0.50	0.49	0.48
Jalan 158624	0.73	0.72	0.31	0.42



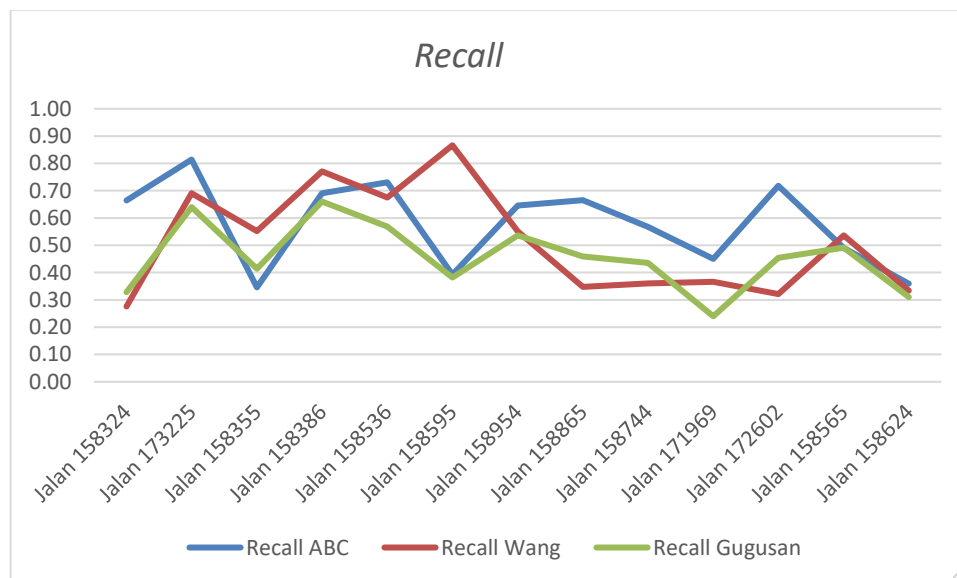
Rajah 4.19 Carta garis perbandingan ketepatan antara model ramalan.

Merujuk kepada Rajah 4.19, terdapat keputusan 10 jalan daripada 13 jalan yang menunjukkan purata ketepatan lebih tinggi apabila menggunakan model ramalan ABC jika dibandingkan model ramalan Wang et al. (2015). dan model ramalan ABC menggunakan kaedah spektrum persilangan hari. Manakala terdapat 6 jalan mendapatkan nilai purata ketepatan lebih tinggi dan 1 memiliki nilai purata ketepatan yang sama daripada 13 jalan, apabila menggunakan model ramalan ABC dengan spektrum persilangan hari dibandingkan model ramalan Wang et al. (2015). Keputusan ini menunjukkan purata ketepatan model ABC lebih tinggi dibandingkan model Wang dan ABC spektrum persilangan hari. Keputusan ini juga menunjukkan bahawa purata ketepatan model ABC spektrum persilangan hari sama dibandingkan model Wang et al. (2015).



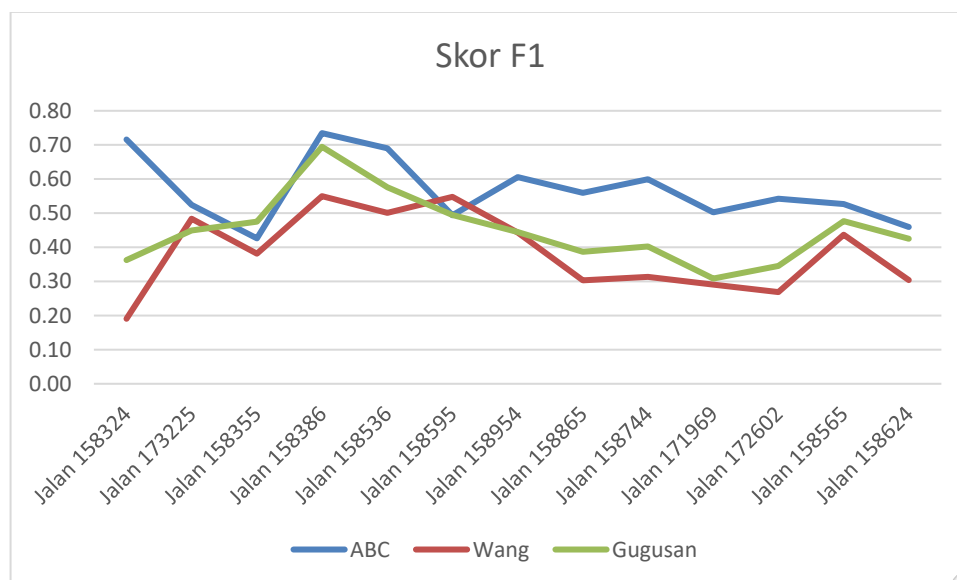
Rajah 4.20 Carta garis perbandingan *precision* antara model ramalan.

Merujuk kepada Rajah 4.20, terdapat keputusan 10 jalan daripada 13 jalan yang menunjukkan purata *precision* lebih tinggi apabila menggunakan model ramalan ABC jika dibandingkan model ramalan Wang et al. (2015) dan model ramalan ABC menggunakan kaedah spektrum persilangan hari. Manakala terdapat 12 jalan mendapatkan nilai purata *precision* lebih tinggi daripada 13 jalan, apabila menggunakan model ramalan ABC dengan spektrum persilangan hari dibandingkan model ramalan Wang et al. (2015). Keputusan ini menunjukkan bahawa nilai *precision* model ABC lebih tinggi dibandingkan model Wang dan ABC menggunakan spektrum persilangan hari. Keputusan ini juga menunjukkan bahawa nilai *precision* model ABC spektrum persilangan hari lebih tinggi dibandingkan model ramalan Wang et al. (2015).



Rajah 4.21 Carta garis perbandingan *recall* antara model ramalan.

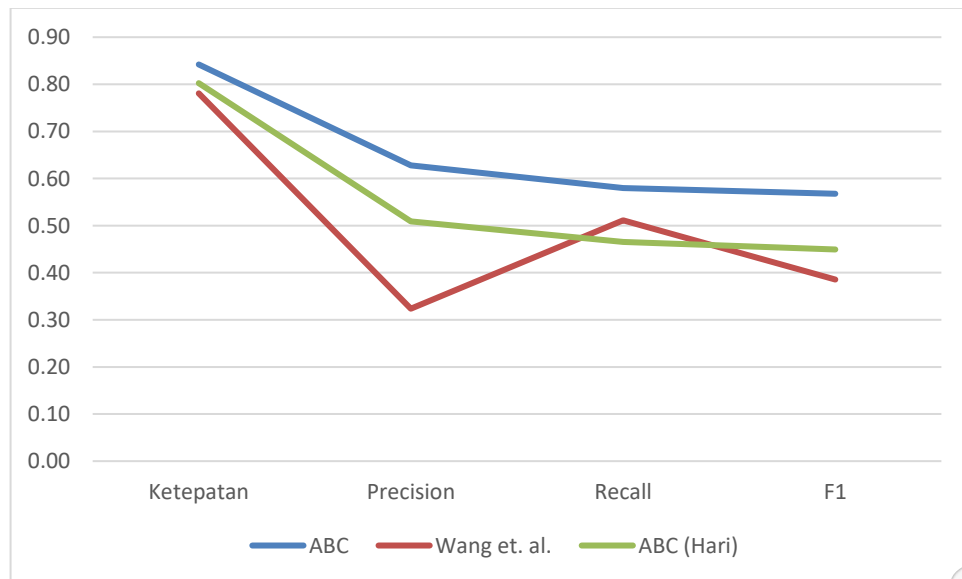
Merujuk kepada Rajah 4.21, terdapat keputusan 9 jalan daripada 13 jalan yang menunjukkan purata recall lebih tinggi apabila menggunakan model ramalan ABC jika dibandingkan model ramalan Wang et al. (2015) dan model ramalan ABC menggunakan kaedah spektrum persilangan hari. Manakala terdapat 9 jalan mendapatkan recall lebih tinggi daripada 13 jalan sasaran. apabila menggunakan model ramalan Wang et al. (2015) dibandingkan model ABC spektrum persilangan hari. Keputusan ini menunjukkan bahawa nilai purata recall model ABC lebih tinggi dibandingkan model Wang dan ABC menggunakan spektrum persilangan hari. Keputusan ini juga menunjukkan bahawa nilai recall model Wang et al. (2015) lebih tinggi dibandingkan model ramalan ABC dengan spektrum persilangan hari.



Rajah 4.22 Carta garis perbandingan skor F1 antara model ramalan.

Merujuk kepada Rajah 4.22, terdapat keputusan 11 jalan daripada 13 jalan yang menunjukkan purata F1 lebih tinggi apabila menggunakan model ramalan ABC jika dibandingkan model ramalan Wang et al. (2015) dan model ramalan ABC menggunakan kaedah spektrum persilangan hari. Manakala terdapat 10 jalan mendapatkan nilai purata F1 lebih tinggi daripada 13 jalan, apabila menggunakan model ramalan ABC dengan spektrum persilangan hari dibandingkan model ramalan Wang et al. (2015). Keputusan ini menunjukkan purata F1 model ABC lebih tinggi dibandingkan model Wang dan ABC spektrum persilangan hari. Keputusan ini juga menunjukkan bahawa purata F1 model ABC spektrum persilangan hari lebih tinggi dibandingkan model Wang et al. (2015).

Hasil eksperimen ini juga dibuktikan dengan merujuk kepada perbandingan purata pengukuran ketepatan, precision, recall dan skor F1 untuk semua jalan seperti yang ditunjukkan pada Rajah 4.23. Ramalan perambatan aliran trafik jalan berjiran menunjukkan model ramalan ABC sentiasa memberikan hasil ramalan yang tertinggi untuk pengukuran ketepatan, *precision*, *recall* dan F1.



Rajah 4.23 Carta garis perbandingan model ramalan berdasarkan purata pengukuran.

Akhir sekali, terdapat dua dapatan yang diperolehi dari hasil eksperimen secara keseluruhannya. Pertama, modul ramalan ABC dan model ABC spektrum persilangan hari meningkatkan nilai purata ketepatan, *precision*, dan F1, dibandingkan model Wang, akan tetapi nilai recall ABC menggunakan spektrum persilangan hari memiliki nilai recall dibawah model Wang et al. (2015). Kedua, modul penggugusan $(GLCM+spektrum) \cap (spektrum \text{ frekuensi})$ berupaya meningkatkan lagi nilai purata ketepatan, *precision*, *recall* dan F1 model ABC dibandingkan model Wang et al. (2015). Boleh dirumuskan bahawa ramalan perambatan aliran trafik jalan berjiran menunjukkan model ramalan ABC sentiasa memberikan hasil ramalan yang tertinggi untuk pengukuran ketepatan, *precision*, *recall* dan F1.

4.5 RUMUSAN

Bab ini memberi penjelasan yang lengkap mengenai lokasi jalan raya yang digunakan dalam menjalankan eksperimen. Seterusnya hasil eksperimen dari model ramalan dicadangkan dibentangkan, dan dibincangkan. Selain itu hasil ramalan model dicadangkan juga dianalisis dan dibandingkan dengan model ramalan Wang et al. (2015).

BAB V

KESIMPULAN DAN KAJIAN LANJUTAN

5.1 PENDAHULUAN

Berasaskan kajian lepas, kaedah perambatan aliran trafik dibahagikan kepada dua kaedah, iaitu kaedah pembinaan graf dan dan kaedah statistik. Fokus kajian ini adalah ramalan perambatan trafik menggunakan kaedah statistik. Pembelajaran mesin digunakan untuk menganalisis pola aliran trafik di jalan raya berjiran. Pola yang didapat seterusnya digunakan untuk meramal perambatan aliran trafik.

Objektif utama penyelidikan ini adalah untuk meramal perambatan aliran trafik berasaskan kesesakan jalan raya sasaran. Untuk mencapai matlamat ini, hubungan antara kesesakan jalan raya dan jalan berjiran diselidiki. Pola aliran trafik antara jalan raya dengan jalan jiran dianalisis. Cadangan modul penggugusan dan modul ramalan yang dicadangkan digabungkan untuk membina model ramalan perambatan trafik yang dicadangkan. Model ramalan perambatan trafik yang dicadangkan seterusnya diuji menggunakan set data daripada pengesan IoT dari bandar Aarhus Denmark.

5.2 PENCAPAIAN OBJEKTIF KAJIAN

Objektif yang lebih terperinci dalam tesis ini terdiri daripada empat sub objektif:

1. Mengenal pasti jalan berjiran dengan menganalisis kesesakan trafik aliran trafik jalan berjiran tanpa maklumat tentang jalan berhubung dengan kaedah penggugusan persilangan GLCM+Spektrum dan Spektrum Frekuensi. Eksperimen yang dijalankan menunjukkan hasil sebagai berikut:

- a) Penggugusan yang dicadangkan menghasilkan jalan berjiran yang memiliki hubungan tinggi. Seterusnya penggugusan yang dicadangkan berhasil menapis lebih banyak jalan raya yang tidak relevan.
 - b) Ketepatan ramalan keadaan trafik daripada jalan berjiran yang diperolehi daripada penggugusan yang dicadangkan menghasilkan prestasi yang tertinggi jika dibandingkan menggunakan kaedah penggugusan spektrum persilangan hari.
2. Memvisualisasikan hubungan kesesakan trafik antara jalan berjiran pada peta lokasi. Hasil eksperimen yang dijalankan menunjukkan seperti berikut:
 - a. Hasil penggugusan jalan raya berjiran yang terkesan daripada jalan yang sesak di visualkan atau dipaparkan ke atas peta.
 - b. Hasil ramalan perambatan aliran trafik yang terkesan daripada jalan yang sesak divisualisasi ke atas peta. Pengguna dapat mengetahui informasi keadaan jalan raya berjiran setiap dua puluh (20) minit.
3. Mereka bentuk model ramalan perambatan kesesakan trafik jalan berjiran berdasarkan analisis ruang (*spatial*) dan temporal dari jalan raya sasaran dengan menggunakan Model *Hidden Markov* (HMM).
 - a. Kaedah dan model yang dicadangkan telah dijelaskan secara terperinci. Butiran struktur dan pelaksanaan langkah-langkah pemprosesan diterangkan dalam seksyen 3.4.2 dan 3.4.3

5.3 KEKANGAN

Hasil eksperimen menunjukkan terdapat peningkatan dalam ketepatan ramalan kesesakan dengan menggunakan model ramalan perambatan aliran trafik jalan berjiran yang dicadangkan. Walau bagaimanapun masih terdapat kekangan dalam kajian ini

yang boleh dijadikan asas kepada penyelidikan pada masa akan datang. Berikut adalah di antara kekangan kajian ini iaitu:

1. Minimum selang masa kesesakan adalah 80 minit atau 192 kali kesesakan dalam tempo 3 bulan. Hal ini untuk memastikan bahawa terjadi pola kesesakan jalan yang berulang dari jalan sasaran kepada jalan berjiran.
2. Ketepatan ramalan perambatan aliran trafik yang dihasilkan sangat dipengaruhi oleh nilai korelasi antara jalan sasaran dengan jalan berjiran. Minimum nilai purata korelasi indeks kesesakan antara jalan adalah 0.6.

5.4 SUMBANGAN KAJIAN

Kajian ini dijalankan untuk menghasilkan dua sumbangan penting iaitu:

- 1) Kajian ini membangunkan modul penggugusan baru yang dicadangkan untuk mengelompokkan jalan raya berasaskan pola aliran trafik pada hari dan masa.
- 2) Model yang dicadangkan dapat dilaksanakan dalam aplikasi mudah alih untuk memberikan pemandu maklumat trafik yang boleh menjejaskan pemanduan dari kesesakan trafik
- 3) Model yang dicadangkan dapat digunakan untuk untuk set data yang tidak mempunyai maklumat arah dan hubungan antara pembolehubah. Sebagai contoh untuk ramalan cuaca berasaskan pola cuaca pada hari, masa dan lokasi yang berbeza.

5.5 CADANGAN KAJIAN LANJUTAN

Beberapa cadangan kajian lanjutan dikenalpasti daripada kajian ini, cadangan ini disenaraikan sebagai berikut:

- Oleh kerana kajian ini hanya membina model perambatan trafik kesan daripada satu jalan raya sasaran, kajian lanjutan yang boleh dijalankan

ialah pembangunan model perambatan trafik kesan daripada beberapa jalan raya yang mengalami kesesakan.

- Penambahbaikan visualisasi perambatan trafik. Visualisasi untuk kajian lanjutan tidak hanya terhad kepada garis lurus sahaja seperti dalam kajian ini, tetapi boleh juga membangunkan visualisasi yang mewakili bentuk jalan raya yang sebenar.
- Kajian ini membahagikan keadaan trafik menjadi dua keadaan iaitu lancar dan sesak. Cadangan kajian lanjutan akan membahagikan keadaan lancar menjadi beberapa tahapan. Kajian lanjutan boleh mempertimbangkan pembahagian lancar kepada beberapa tahap, Dengan membahagikan keadaan lancar kepada beberapa tahap kemungkinan ketepatan ramalan boleh ditingkatkan.

RUJUKAN

- Abellán, J., López, G. & De Oña, J. 2013. Analysis of traffic accident severity using Decision Rules via Decision Trees. *Expert Systems with Applications* 40(15): 6047–6054. doi:10.1016/j.eswa.2013.05.027
- Adrian, C., Abdullah, R., Atan, R. & Jusoh, Y. Y. 2016. Model for Big Data Implementation : A 8(January 2017).
- Afrin, T. & Yodo, N. 2021. A probabilistic estimation of traffic congestion using Bayesian network. *Measurement* 174: 109051. doi:https://doi.org/10.1016/j.measurement.2021.109051
- Ahn, J., Ko, E. & Kim, E. Y. 2016. Highway traffic flow prediction using support vector regression and Bayesian classifier. *2016 International Conference on Big Data and Smart Computing, BigComp 2016* 239–244. doi:10.1109/BIGCOMP.2016.7425919
- Ahn, J. Y., Ko, E. & Kim, E. 2015. Predicting Spatiotemporal Traffic Flow Based on Support Vector Regression and Bayesian Classifier. *2015 IEEE Fifth International Conference on Big Data and Cloud Computing*, hlm. 125–130. doi:10.1109/BDCLOUD.2015.64
- Akoglu, H. 2018. User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine* 18(3): 91–93. doi:10.1016/j.tjem.2018.08.001
- Alamsyah, A. & Salma, N. 2018. A Comparative Study of Employee Churn Prediction Model. *Proceedings - 2018 4th International Conference on Science and Technology, ICST 2018* 1(2): 1–4. doi:10.1109/ICSTC.2018.8528586
- Alazawi, S. A., Shati, N. M. & Abbas, A. H. 2019. Texture features extraction based on GLCM for face retrieval system. *Periodicals of Engineering and Natural Sciences* 7(3): 1459–1467. doi:10.21533/pen.v7i3.787
- Ambühl, L., Menendez, M. & González, M. C. 2023. Understanding congestion propagation by combining percolation theory with the macroscopic fundamental diagram. *Communications Physics* 6(1): 1–7. doi:10.1038/s42005-023-01144-w
- Antoniou, C., Balakrishna, R. & Koutsopoulos, H. N. 2011. A Synthesis of emerging data collection technologies and their impact on traffic management applications. *European Transport Research Review* 3(3): 139–148. doi:10.1007/s12544-011-0058-1
- Aven, M. 2017. Daily Traffic Flow Pattern Recognition by Spectral Clustering.
- Barbosa, I., Casanova, M. A., Renso, C. & de Macedo, J. A. F. 2013. Average Speed Estimation For Road Networks Based On GPS Raw Trajectories. *Inf.Puc-Rio.Br.* Retrieved from <http://www.inf.puc-rio.br/~casanova/Publications/Papers/2013-Papers/2013-ICEIS.pdf>

- Bast, H., Brosi, P. & Storandt, S. 2014. Real-time movement visualization of public transit data. *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - SIGSPATIAL '14* 331–340. doi:10.1145/2666310.2666404
- Bastos, L. D. O., Liatsis, P. & Conci, A. 2008. Automatic texture segmentation based on k-means clustering and efficient calculation of co-occurrence features. *Proceedings of IWSSIP 2008 - 15th International Conference on Systems, Signals and Image Processing*, hlm. 141–144. doi:10.1109/IWSSIP.2008.4604387
- Bellemans, T., Schutter, B. De & Moor, B. De. 2000. Data acquisition, interfacing and pre-processing of highway traffic data. *Proceedings of Telematics ...* 19(January 2014): 0–2. Retrieved from http://www.dcsc.tudelft.nl/~bdeschutter/pub/rep/00_02.pdf
- Besenczi, R., Bátfai, N., Jeszenszky, P., Major, R., Monori, F. & Ispány, M. 2020. Large-scale Analysis and Simulation of Traffic Flow using Markov Models. *arXiv*. doi:10.1371/journal.pone.0246062
- Bischof, S., Karapantelakis, Athanasios Nechifor, C.-S., Sheth, A., Mileo, A. & Barnaghi, P. 2014. Real Time IoT Stream Processing and Large-scale Data Analytics for Smart City Applications. *Real Time IoT Stream Processing and Large-scale Data Analytics for Smart City Applications*.
- Bischof, S., Karapantelakis, A., Sheth, A. & Mileo, A. 2014. Semantic Modelling of Smart City Data Description of Smart City Data. *W3C Workshop on the Web of Things Enablers and services for an open Web of Devices*, hlm. 1–5. Retrieved from <http://www.w3.org/2014/02/wot/papers/karapantelakis.pdf>
- Bourlard, H. A. & Morgan, N. 1994. Hidden Markov Models. *Connectionist Speech Recognition* 27–58. doi:10.1007/978-1-4615-3210-1_3
- Chen, C., Luan, T. H., Guan, X., Lu, N. & Yunshu, L. 2017. Connected Vehicular Transportation: Data Analytics and Traffic-Dependent Networking. *IEEE Vehicular Technology Magazine* 12(3): 42–54.
- Chen, Z., Yang, Y., Huang, L., Wang, E. & Li, D. 2018. Discovering Urban Traffic Congestion Propagation Patterns with Taxi Trajectory Data. *IEEE Access* 6(c): 69481–69491. doi:10.1109/ACCESS.2018.2881039
- Cheng, S. & Lu, F. 2018. Short-term traffic forecasting: A dynamic ST-KNN model considering spatial heterogeneity and temporal non-stationarity. *CEUR Workshop Proceedings* 2083: 133–140.
- Christidis, P. & Ibanes Rivas, J. N. 2012. *Measuring road congestion*. Institute for Prospective Technological Studies. doi:10.2791/15282

- Contreras-Castillo, J., Zeadally, S. & Ibañez, J. A. G. 2016. Solving vehicular ad hoc network challenges with big data solutions. *IET Networks* 5(4): 81–84. doi:10.1049/iet-net.2016.0001
- Crist, P. 2015. *Big Data and Transport Understanding and assessing options*.
- Cui, G., Luo, J. & Wang, X. 2018. Personalized travel route recommendation using collaborative filtering based on GPS trajectories. *International Journal of Digital Earth* 11(3): 284–307. doi:10.1080/17538947.2017.1326535
- Daganzo, C. F. 1997. Fundamentals of Transportation and Traffic Operations.
- Dai, H. & Yang, Z. 2006. Real-Time Traffic Volume Estimation with Fuzzy Linear Regression. *2006 6th World Congress on Intelligent Control and Automation*, hlm. 3164–3167. doi:10.1109/WCICA.2006.1712950
- Dai, J., Yang, B., Guo, C. & Ding, Z. 2015. Personalized route recommendation using big trajectory data. *Proceedings - International Conference on Data Engineering* 2015-May: 543–554. doi:10.1109/ICDE.2015.7113313
- Dauletbak, D. & Woo, J. 2019. Traffic Data Analysis and Prediction using Big Data. *The 14th Asia Pacific International Conference on Information Science and Technology (APIC-IST)* 127–133.
- Davoodi, N., Soheili, A. R. & Hashemi, S. M. 2016. A macro-model for traffic flow with consideration of driver's reaction time and distance. *Nonlinear Dynamics* 83(3): 1621–1628. doi:10.1007/s11071-015-2435-0
- Deb Nath, R. P., Lee, H. J., Chowdhury, N. K. & Chang, J. W. 2010. Modified K-means clustering for travel time prediction based on historical traffic data. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6276 LNAI(PART 1): 511–521. doi:10.1007/978-3-642-15387-7_55
- Deep Singh, A., Wu, W., Xiang, S. & Krishnaswamy, S. 2015. Taxi trip time prediction using similar trips and road network data. *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*, hlm. 2892–2894. doi:10.1109/BigData.2015.7364113
- El Sherief, M. M., Ramadan, I. M. I. & Ibrahim, A. M. 2016. Development of traffic stream characteristics models for intercity roads in Egypt. *Alexandria Engineering Journal* 55(3): 2765–2770. doi:10.1016/j.aej.2016.04.031
- Famili, A., Shen, W. M., Weber, R. & Simoudis, E. 1997. Data preprocessing and intelligent data analysis. *Intelligent Data Analysis* 1(1): 3–23. doi:10.3233/IDA-1997-1102

- Fatonah, N. S., Tjandrasa, H. & Fatichah, C. 2020. Identification of acute lymphoblastic leukemia subtypes in touching cells based on enhanced edge detection. *International Journal of Intelligent Engineering and Systems* 13(4): 204–215. doi:10.22266/IJIES2020.0831.18
- Fohrholz, C. & Glaschke, C. 2016. Predicting travel volumes for long-distance coach services through big data analytics - a case study on German public viewing events during the uefa euro 2016. *Proceedings of the Annual Hawaii International Conference on System Sciences* 2016-March: 1215–1224. doi:10.1109/HICSS.2016.154
- Gebresilassie, M. A. 2017. *Spatio-temporal Traffic Flow Prediction. KTH ROYAL INSTITUTE OF TECHNOLOGY SCHOOL OF ARCHITECTURE AND THE BUILT ENVIRONMENT.*
- Ghadi, M. Q. & Török, Á. 2019. Comparison of different road segmentation methods. *Promet - Traffic - Traffico* 31(2): 163–172. doi:10.7307/ptt.v31i2.2937
- Ghahramani, Z. 2001. An Introduction to Hidden Markov Models and Bayesian Networks. *International Journal of Pattern Recognition and Artificial Intelligence* 1.
- Hamner, B. 2010. Predicting travel times with context-dependent random forests by modeling local and aggregate traffic flow. *Proceedings - IEEE International Conference on Data Mining, ICDM* 1357–1359. doi:10.1109/ICDMW.2010.128
- Han, Y. & Moutarde, F. 2012. Analysis of Large-scale Traffic Dynamics using Non-negative Tensor Factorization. *arXiv preprint* 1–12.
- Han, Y., Moutarde, F., Han, Y., Moutarde, F. & Dynamics, L. T. 2014. Analysis of Large-Scale Traffic Dynamics in an Urban Transportation Network Using Non-Negative Tensor Factorization. *International Journal of Intelligent Transportation Systems Research, Springer Verlag* 14(1): 36–49. doi:10.1007/s13177-014-0099-7
- He, F., Yan, X., Liu, Y. & Ma, L. 2016. A Traffic Congestion Assessment Method for Urban Road Networks Based on Speed Performance Index. *Green Intelligent Transportation System and Safety*, hlm. Vol. 137, 425–433. doi:10.1016/j.proeng.2016.01.277
- Hels, T., Lyckegaard, A. & Pilegaard, N. 2011. Evaluering af trafiksikkerhedstiltag - en vejledning. Retrieved from http://www.transport.dtu.dk/~media/Institutter/Transport/forskning/publikationer/2011/skabelon_for_evalueringer_af_trafiksikkerhedstiltag.ashx
- Hong, H., Huang, W., Xing, X., Zhou, X., Lu, H., Bian, K. & Xie, K. 2015. Hybrid Multi-metric K-Nearest Neighbor Regression for Traffic Flow Prediction. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC* 2015-Octob: 2262–2267. doi:10.1109/ITSC.2015.365

- Hu, C., Xie, K., Song, G. & Wu, T. 2008. Hybrid Process Neural Network based on Spatio-Temporal Similarities for Short-Term Traffic Flow Prediction. *Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems Beijing, China, October 12-15, 2008*, hlm. 253–258.
- Hu, W., Yan, L. & Wang, H. 2014. Traffic jams prediction method based on two-dimension cellular automata model. *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, hlm. 2023–2028. doi:10.1109/ITSC.2014.6958001
- Huang, X., Zhao, Y., Ma, C., Yang, J. & Ye, X. 2015. TrajGraph: A Graph-Based Visual Analytics Approach to Studying. *IEEE Visualization Conference* 2626(c): 1–1. doi:10.1109/TVCG.2015.2467771
- Iliashenko, O., Iliashenko, V. & Lukyanchenko, E. 2021. Big Data in Transport Modelling and Planning. *Transportation Research Procedia* 54(2020): 900–908. doi:10.1016/j.trpro.2021.02.145
- Jain, S., Singh, S. & Jain, G. 2017. Traffic Congestion Modelling Based on Origin and Destination. *10th International Scientific Conference Transbaltica 2017: Transportation Science and Technology*, hlm. Vol. 187, 442–450. The Author(s). doi:10.1016/j.proeng.2017.04.398
- Jiang, Y., Kang, R., Li, D., Guo, S. & Havlin, S. 2017. Spatio-temporal propagation of traffic jams in urban traffic networks. *Physics and Society (physics.soc-ph)*. Retrieved from <http://arxiv.org/abs/1705.08269>
- Jianqin, Z., Peiyuan, Q., Yingchao, D., Mingyi, D. & Feng, L. 2015. A space-time visualization analysis method for taxi operation in Beijing. *Journal of Visual Languages and Computing* 31: 1–8. doi:10.1016/j.jvlc.2015.09.002
- Jin, S., Wang, D. H., Xu, C. & Ma, D. F. 2013. Short-term traffic safety forecasting using Gaussian mixture model and Kalman filter. *Journal of Zhejiang University: Science A* 14(4): 231–243. doi:10.1631/jzus.A1200218
- Jindal, V. 2014. Use of Big Data Technology in Vehicular Ad-hoc Networks. *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*.
- John FW Zaki, Amr Ali Eldin, Sherif E. Hussein, Sabry F. Saranya, F. F. A. 2016. Framework for Traffic Congestion Prediction. *Journal of Scientific and Engineering Research* 7(5): 1205–1210.
- Jurafsky, D. & H Martin, J. 2020. Speech and language processing. *Stanford University*. doi:10.4324/9780203461891_chapter_3
- Kang, H. 2013. The prevention and handling of the missing data. *Korean Journal of Anesthesiology* 64(5): 402–406. doi:10.4097/kjae.2013.64.5.402

- Keler, A., Krisp, J. M. & Ding, L. 2017. Detecting traffic congestion propagation in urban environments—a case study with Floating Taxi Data (FTD) in Shanghai. *Journal of Location Based Services* 11(2): 133–151. doi:10.1080/17489725.2017.1420256
- Kerner, B. S. & Herrtwich, R. G. 2006. A Bayesian Network Approach to Traffic Flow Forecasting. *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS* 7(1): 124–132. doi:10.1524/auto.2001.49.11.505
- Kesumawati, A. & Setianingsih, D. 2016. A segmentation group by Kohonen Self Organizing Maps (SOM) and K -means algorithms (case study: Malnutrition cases in Central Java of Indonesia). *International Journal of Advances in Soft Computing and its Applications* 8(3): 110–115.
- Kindzerske, M. D. & Ni, D. 2007. Composite Nearest Neighbor Nonparametric Regression to Improve Traffic Prediction. *Transportation Research Record* (1993): pp 30-35. doi:10.3141/1993-05
- King, R. 2015. Traffic management in a connected or autonomous vehicle environment. *IET Seminar Digest* 2015(15688). doi:10.1049/ic.2015.0063
- Kolozali, S., Bermudez-Edo, M., Puschmann, D., Ganz, F. & Barnaghi, P. 2014. A knowledge-based approach for real-time IoT data stream annotation and processing. *Proceedings - 2014 IEEE International Conference on Internet of Things, iThings 2014, 2014 IEEE International Conference on Green Computing and Communications, GreenCom 2014 and 2014 IEEE International Conference on Cyber-Physical-Social Computing, CPS 20*, hlm. 215–222. doi:10.1109/iThings.2014.39
- Kong, X., Xu, Z., Shen, G., Wang, J., Yang, Q. & Zhang, B. 2016. Urban traffic congestion estimation and prediction based on floating car trajectory data. *Future Generation Computer Systems* 61: 97–107. doi:10.1016/j.future.2015.11.013
- Kumar, V. 2014. DATA Clustering Algorithms and Applications. *USA BOOK 648p*. New York: CRC Press Taylor & Francis Group.
- Kwoczek, S., Di Martino, S. & Nejdl, W. 2014. Predicting and visualizing traffic congestion in the presence of planned special events. *2019 International Conference on Software Engineering & Knowledge Engineering*, hlm. Vol. 25, 973–980. doi:10.1016/j.jvlc.2014.10.028
- Laña, I., Del Ser, J. & Olabarrieta, I. 2016. Understanding Daily Mobility Patterns in Urban Road Networks using Traffic Flow Analytics. *International Workshop on Urban Mobility & Intelligent Transportation Systems (UMITS)*, hlm. 1157–1162. doi:10.1109/NOMS.2016.7502980

- Lee, E.-M., Kim, J.-H. & Yoon, W.-S. 2007. Traffic Speed Prediction Under Weekday, Time, and Neighboring Links' Speed: Back Propagation Neural Network Approach. *Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues*, hlm. 626–635. doi:10.1007/978-3-540-74171-8_62
- Lee, J., Hong, B., Lee, K. & Jang, Y.-J. 2015. A Prediction Model of Traffic Congestion Using Weather Data. *2015 IEEE International Conference on Data Science and Data Intensive Systems*, hlm. 81–88. doi:10.1109/DSDIS.2015.96
- Lee, K., Hong, B., Jeong, D. & Lee, J. 2014. Congestion pattern model for predicting short-term traffic decongestion times. *2014 17th IEEE International Conference on Intelligent Transportation Systems, ITSC 2014*, hlm. 2828–2833. doi:10.1109/ITSC.2014.6958143
- Liang, Z. & Wakahara, Y. 2014. Real-time urban traffic amount prediction models for dynamic route guidance systems. *EURASIP Journal on Wireless Communications and Networking* 2014(1): 85. doi:10.1186/1687-1499-2014-85
- Lin, M.-W., Tapamo, J.-R. & Ndovie, B. 2007. A Texture-based Method for Document Segmentation and Classification. *Revue Africaine de la Recherche en Informatique et Mathématiques Appliquées* Volume 6,: 49–56. doi:10.46298/arima.1878
- Lin, X. & Xu, J. 2020. Road network partitioning method based on Canopy-Kmeans clustering algorithm. *Archives of Transport* 54(2): 95–106. doi:10.5604/01.3001.0014.2970
- Liu, Z., Li, Z., Li, M., Xing, W. & Lu, D. 2016. Mining Road Network Correlation for Traffic Estimation via Compressive Sensing. *IEEE Transactions on Intelligent Transportation Systems* 17(7): 1880–1893. doi:10.1109/TITS.2016.2514519
- Lopes, J., Bento, J., Huang, E., Antoniou, C. & Ben-Akiva, M. 2010. Traffic and mobility data collection for real-time applications. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC* 216–223. doi:10.1109/ITSC.2010.5625282
- Maier, F., Braun, R., Busch, F. & Mathias, P. 2008. Pattern-based short-term prediction of urban congestion propagation and automatic response. *Traffic Engineering and Control* 49(6): 227–232.
- Mallikarjuna, C., Phanindra, A. & Rao, K. R. 2009. Traffic Data Collection under Mixed Traffic Conditions.pdf (April): 174–182.
- Maurya, R., Gupta, P. R. & Shukla, A. S. 2011. Road extraction using K-means clustering and morphological operations. *ICIIP 2011 - Proceedings: 2011 International Conference on Image Information Processing (Iciip)*. doi:10.1109/ICIIP.2011.6108839

- Maxwell, A. E., Warner, T. A. & Guillén, L. A. 2021. Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 1: Literature Review. *Remote Sensing*. Retrieved from <https://www.mdpi.com/1160808>
- Min, W. & Wynter, L. 2011. Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies* 19(4): 606–616. doi:10.1016/j.trc.2010.10.002
- Mondal, M. A. & Rehena, Z. 2019. Identifying Traffic Congestion Pattern using K-means Clustering Technique. *Proceedings - 2019 4th International Conference on Internet of Things: Smart Innovation and Usages, IoT-SIU 2019*, hlm. 1–5. IEEE. doi:10.1109/IoT-SIU.2019.8777729
- Montazeri-Gh, M. & Fotouhi, A. 2011. Traffic condition recognition using the k-means clustering method. *Scientia Iranica* 18(4 B): 930–937. doi:10.1016/j.scient.2011.07.004
- Necula, E. 2015. Analyzing traffic patterns on street segments based on GPS data using R. *Transportation Research Procedia* 10(July): 276–285. doi:10.1016/j.trpro.2015.09.077
- Nguyen, H., Liu, W. & Chen, F. 2016a. Discovering Congestion Propagation Patterns in Spatio-Temporal Traffic Data. *IEEE Transactions on Big Data* 3(2): 169–180. doi:10.1109/tbdata.2016.2587669
- Nguyen, H., Liu, W. & Chen, F. 2016b. Discovering Congestion Propagation Patterns in Spatio-Temporal Traffic Data. *IEEE Transactions on Big Data* 3(2): 169–180. doi:10.1109/tbdata.2016.2587669
- Novakovic, J., Veljovi, A., Ilic, S., Papic, Z. & Tomovic, M. 2017. Evaluation of Classification Models in Machine Learning. *Theory and Applications of Mathematics & Computer Science* 7(1): 39–46. Retrieved from <https://uav.ro/applications/se/journal/index.php/TAMCS/article/view/158>
- Nurhaida, I., Wei, H., Zen, R. A. M., Manurung, R. & Arymurthy, A. M. 2016. Texture fusion for batik motif retrieval system. *International Journal of Electrical and Computer Engineering* 6(6): 3174–3187. doi:10.11591/ijece.v6i6.12049
- Oleiwi, W. K. 2017. Alzheimer Disease Diagnosis using the K-means, GLCM and K_NN. *JOURNAL OF UNIVERSITY OF BABYLON for Pure and Applied Sciences* 26(2): 57–65. doi:10.29196/jub.v26i2.474
- Petrovska, N. & Stevanovic, A. 2015. Traffic Congestion Analysis Visualisation Tool. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC 2015-Octob*: 1489–1494. doi:10.1109/ITSC.2015.243

- Prasad, R., Bisandu, D. & Liman, M. 2018. Clustering News Articles using Efficient Similarity Measure and N-grams. *International Journal of Knowledge Engineering and Data Mining* 5(1): 1. doi:10.1504/ijkedm.2018.10016103
- Qi, Y. & Ishak, S. 2014. A Hidden Markov Model for short term prediction of traffic conditions on freeways. *Transportation Research Part C: Emerging Technologies* 43: 95–111. doi:10.1016/j.trc.2014.02.007
- R Rabiner, L. 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *PROCEEDINGS OF THE IEEE*, hlm. Vol. 77, 257–286. doi:10.25300/MISQ/2017/41.3.08
- Rajendran, S. & Ayyasamy, B. 2020. Short-term traffic prediction model for urban transportation using structure pattern and regression: an Indian context. *SN Applied Sciences* 2(7): 1–11. doi:10.1007/s42452-020-2946-2
- Rao, D. J. 2019. Keras to Kubernetes: The Journey of a Machine Learning Model to Production. Indianapolis, Indiana: John Wiley & Sons, Inc. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119564843.ch1>
- Rapant, L., Slaninová, K., Martinovič, J. & Martinovič, T. 2016. Traffic Speed Prediction Using Hidden Markov Models for Czech Republic Highways. *Agent and Multi-Agent Systems: Technology and Applications, Smart Innovation, Systems and Technologies*, hlm. 187–196. doi:10.1007/978-3-319-39883-9
- Rempe, F., Huber, G. & Bogenberger, K. 2016. Spatio-Temporal Congestion Patterns in Urban Traffic Networks. *Transportation Research Procedia*, hlm. Vol. 15, 513–524. Elsevier B.V. doi:10.1016/j.trpro.2016.06.043
- Roweis, S. T. & Saul, L. K. 2000. Nonlinear dimensionality reduction by locally linear embedding. *SCIENCE* 290((5500),): 2323–2326. Retrieved from <http://science.sciencemag.org/>
- Saeedmanesh, M. & Geroliminis, N. 2017. Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic networks. *Transportation Research Procedia*, hlm. Vol. 23, 962–979. Elsevier B.V. doi:10.1016/j.trpro.2017.05.053
- Sepasgozar, S. S. & Pierre, S. 2022. Network Traffic Prediction Model Considering Road Traffic Parameters Using Artificial Intelligence Methods in VANET. *IEEE Access* 10: 8227–8242. doi:10.1109/ACCESS.2022.3144112
- Shan, P. 2018. Image segmentation method based on K-mean algorithm. *Eurasip Journal on Image and Video Processing* 2018(1). doi:10.1186/s13640-018-0322-6
- Shan, Z., Pan, Z., Li, F., Xu, H. & Li, J. 2018. Visual analytics of traffic congestion propagation path with large scale camera data. *Chinese Journal of Electronics* 27(5): 934–941. doi:10.1049/cje.2018.04.011

- Sisodia, D. S., Vishwakarma, S. & Pujahari, A. 2018. Evaluation of machine learning models for employee churn prediction. *Proceedings of the International Conference on Inventive Computing and Informatics, ICICI 2017* (Icici): 1016–1020. doi:10.1109/ICICI.2017.8365293
- Smith, B. L., Demetsky, M. J. & Smith, B. L. 1994. Short-Term Traffic Flow Prediction: Neural Network Approach. *Transportation Research Record* (1453): 98–104.
- Sobral, A., Oliveira, L., Schnitman, L. & Souza, F. De. 2013. Highway Traffic Congestion Classification using Holistic Properties. *Computer Graphics and Imaging / 798: Signal Processing, Pattern Recognition and Applications*. doi:10.2316/P.2013.798-105
- Sun, S., Chen, J. & Sun, J. 2019. Traffic congestion prediction based on GPS trajectory data. *International Journal of Distributed Sensor Networks* 15(5). doi:10.1177/1550147719847440
- Suresh, A. & Shunmuganathan, K. L. 2012. Image texture classification using gray level co-occurrence matrix based statistical features. *European Journal of Scientific Research* 75(4): 591–597.
- Susilo, B. H. & Imanuel, I. 2018. Traffic congestion analysis using travel time ratio and degree of saturation on road sections in Palembang, Bandung, Yogyakarta, and Surakarta. *MATEC Web of Conferences*, hlm. Vol. 181. doi:10.1051/mateconf/201818106010
- Tamir, T. S., Xiong, G., Li, Z., Tao, H., Shen, Z., Hu, B. & Menkir, H. M. 2020. Traffic Congestion Prediction using Decision Tree, Logistic Regression and Neural Networks. *IFAC-PapersOnLine* 53(5): 512–517. doi:10.1016/j.ifacol.2021.04.138
- The, A. & Route, B. 2014. Road Density Prediction. *International Conference on Electrical Engineering and Computer Science* (November): 168–173.
- Ting, K. M. 2010. Precision and Recall BT - Encyclopedia of Machine Learning. Dlm. Sammut (pnyt.) & Webb (pnyt.). , hlm. 781. Boston, MA: Springer US. doi:10.1007/978-0-387-30164-8_652
- Tominski, C., Schumann, H., Andrienko, G. & Andrienko, N. 2012. Stacking-based visualization of trajectory attribute data. *IEEE Transactions on Visualization and Computer Graphics* 18(12): 2565–2574. doi:10.1109/TVCG.2012.265
- Toncharoen, R. 2018. Traffic State Prediction using Convolutional Neural Network. *2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, hlm. 1–6. IEEE. doi:10.1109/JCSSE.2018.8457359
- Uhlemann, E. 2015. Autonomous vehicles are connecting... *IEEE Vehicular Technology Magazine* 10(2): 22–25. doi:10.1109/MVT.2015.2414814

- Von Luxburg, U. 2007. A tutorial on spectral clustering. *Statistics and Computing* 17(4): 395–416. doi:10.1007/s11222-007-9033-z
- Vora, D. & Iyer, K. 2018. Evaluating the Effectiveness of Machine Learning Algorithms in Predictive Modelling. *International Journal of Engineering & Technology* 7(3.4): 197. doi:10.14419/ijet.v7i3.4.16773
- Wang, G. & Kim, J. 2016. Prediction of Traffic Congestion and Incidents on Urban Road Networks Using Naive Bayes Classifier. *Australasian Transport Research Forum 2016 Proceedings*.
- Wang, W. X., Guo, R. J. & Yu, J. 2018. Research on road traffic congestion index based on comprehensive parameters: Taking Dalian city as an example. *Advances in Mechanical Engineering* 10(6): 1–8. doi:10.1177/1687814018781482
- Wang, Xiaomeng, Peng, L., Chi, T., Li, M., Yao, X. & Shao, J. 2015. A hidden markov model for urban-scale traffic estimation using floating car data. *PLoS ONE* 10(12): 1–20. doi:10.1371/journal.pone.0145348
- Wang, Xin, Wang, C., Zhang, J., Zhou, M. & Jiang, C. 2017. Improved rule installation for real-time query service in software-defined internet of vehicles. *IEEE Transactions on Intelligent Transportation Systems*, hlm. Vol. 18, 225–235. doi:10.1109/TITS.2016.2543600
- Wang, Xin, Zhao, S. & Dong, L. 2017. Research and Application of Traffic Visualization Based on Vehicle GPS Big Data. *Proceedings of the Second International Conference on Intelligent Transportation, Smart Innovation, Systems and Technologies*, hlm. Vol. 43, 293–302. doi:10.1364/josa.43.000624
- Wang, Y., Cao, J., Li, W. & Gu, T. 2016. Mining Traffic Congestion Correlation between Road Segments on GPS Trajectories. *2016 IEEE International Conference on Smart Computing, SMARTCOMP 2016*. doi:10.1109/SMARTCOMP.2016.7501704
- Wang, Z., Lu, M., Yuan, X., Zhang, J. & Wetering, H. Van De. 2013. Visual traffic jam analysis based on trajectory data. *IEEE Transactions on Visualization and Computer Graphics* 19(12): 2159–2168. doi:10.1109/TVCG.2013.228
- Wang, Z., Ye, T., Lu, M., Yuan, X., Qu, H., Yuan, J. & Wu, Q. 2014. Visual exploration of sparse traffic trajectory data. *IEEE Transactions on Visualization and Computer Graphics* 20(12): 1813–1822. doi:10.1109/TVCG.2014.2346746
- Wen, F., Zhang, G., Sun, L., Wang, X. & Xu, X. 2019. A Hybrid Temporal Association Rules Mining method for Traffic Congestion Prediction. *Computers & Industrial Engineering*. doi:10.1016/j.cie.2019.03.020
- Xiao, J., Cao, H., Jiang, X., Gu, X. & Xie, L. 2017. GMDH-based semi-supervised feature selection for customer classification. *Knowledge-Based Systems* 132: 236–248. doi:10.1016/j.knosys.2017.06.018

- Xing, X., Zhou, X., Hong, H., Huang, W., Bian, K. & Xie, K. 2015. Traffic Flow Decomposition and Prediction Based on Robust Principal Component Analysis. *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, hlm. 2219–2224. doi:10.1109/ITSC.2015.358
- Xiong, H. & Vahedian, A. 2018. Predicting Traffic Congestion Propagation Patterns : A Propagation Graph Approach. *the 11th ACM SIGSPATIAL International Workshop on Computational Transportation Science*, hlm. 60–69. doi:10.1080/0963928032000049375
- Xiong, H., Vahedian, A., Zhou, X., Li, Y. & Luo, J. 2018. Predicting Traffic Congestion Propagation Patterns. *International Workshop on Computational Transportation Science*, hlm. 60–69. doi:10.1145/3283207.3283213
- Xu, C., Li, Z. & Wang, W. 2016. Short-term traffic flow prediction using a methodology based on autoregressive integrated moving average and genetic programming. *Transport* 31(3): 343–358. doi:10.3846/16484142.2016.1212734
- Yang, S., Wu, J., Qi, G. & Tian, K. 2017. Analysis of traffic state variation patterns for urban road network based on spectral clustering. *Advances in Mechanical Engineering* 9(9): 1–11. doi:10.1177/1687814017723790
- Ye, N., Zhang, Y., Wang, R. & Malekian, R. 2016. Vehicle trajectory prediction based on Hidden Markov Model. *KSII TRANSACTIONS ON INTERNET AND INFORMATION SYSTEMS* 10(7): 3150–3170.
- Yu, B., Song, X., Guan, F., Yang, Z. & Yao, B. 2016. k-Nearest Neighbor Model for Multiple-Time-Step Prediction of Short-Term Traffic Condition. *Journal of Transportation Engineering* 142(6): 04016018. doi:10.1061/(ASCE)TE.1943-5436.0000816
- Yun, T. & Shu, H. 2011. Ultrasound image segmentation by spectral clustering algorithm based on the curvelet and GLCM features. *2011 International Conference on Electrical and Control Engineering, ICECE 2011 - Proceedings* 973(30671639): 920–923. doi:10.1109/ICECENG.2011.6057730
- Zaki, J. F., Ali-Eldin, A., Hussein, S. E., Saraya, S. F. & Areed, F. F. 2019. Traffic congestion prediction based on Hidden Markov Models and contrast measure. *Ain Shams Engineering Journal* (xxxx). doi:10.1016/j.asej.2019.10.006
- Zhang, K., Sun, D. (Jian), Shen, S. & Zhu, Y. 2017. Analyzing spatiotemporal congestion pattern on urban roads based on taxi GPS data. *Journal of Transport and Land Use* 10(1): 675–694. doi:10.5198/jtlu.2017.954
- Zhang, R., Shu, Y., Yang, Z., Cheng, P. & Chen, J. 2015. Hybrid Traffic Speed Modeling and Prediction Using Real-World Data. *Proceedings - 2015 IEEE International Congress on Big Data, BigData Congress 2015* 230–237. doi:10.1109/BigDataCongress.2015.40

- Zhang, X., Jiao, L., Liu, F., Bo, L. & Gong, M. 2008. Spectral clustering ensemble applied to SAR image segmentation. *IEEE Transactions on Geoscience and Remote Sensing* 46(7): 2126–2136. doi:10.1109/TGRS.2008.918647
- Zhang, Yingya, Ye, N., Wang, R. & Malekian, R. 2016. A Method for Traffic Congestion Clustering Judgment Based on Grey Relational Analysis. *ISPRS International Journal of Geo-Information* 5(5): 71. doi:10.3390/ijgi5050071
- Zhang, Yun, Xue, W., Wei, W. & Nazif, H. 2022. A public transport network design using a hidden Markov model and an optimization algorithm. *Research in Transportation Economics* 92(May): 101095. doi:10.1016/j.retrec.2021.101095
- Zhao, S. X., Wu, H. W. & Liu, C. R. 2019. Traffic flow prediction based on optimized hidden Markov model. *Journal of Physics: Conference Series*, hlm. Vol. 1168. doi:10.1088/1742-6596/1168/5/052001
- Zhou, W., Wang, W., Hua, X. & Zhang, Y. 2020. Real-Time Traffic Flow Forecasting via a Novel Method Combining Periodic-Trend Decomposition. *Sustainability (Switzerland)* 12(15): 1–22. doi:10.3390/SU12155891
- Zhu, G., Song, K., Zhang, P. & Wang, L. 2016. A traffic flow state transition model for urban road network based on Hidden Markov Model. *Neurocomputing* 214: 567–574. doi:10.1016/j.neucom.2016.06.044
- Zhu, L., Yu, F. R., Wang, Y., Ning, B. & Tang, T. 2019. Big Data Analytics in Intelligent Transportation Systems: A Survey. *IEEE Transactions on Intelligent Transportation Systems* 20(1): 383–398. doi:10.1109/TITS.2018.2815678