

**ANALISIS PERAMALAN BANJIR KILAT DI SELANGOR
MENGUNAKAN TEKNIK PERLOMBONGAN DATA**

MUHAMMAD HAKIEM BIN HALIM

**MASTER OF SCIENCE
(COMPUTER SCIENCE)**

UNIVERSITI PERTAHANAN NASIONAL MALAYSIA

2022

**ANALISIS PERAMALAN BANJIR KILAT DI SELANGOR MENGGUNAKAN TEKNIK
PERLOMBONGAN DATA**

MUHAMMAD HAKIEM BIN HALIM

Thesis submitted to Centre for Graduate Studies, Universiti Pertahanan Nasional Malaysia, in
fulfilment of the requirements for the Master of Science (Computer Science)

2022

ABSTRAK

Banjir kilat merupakan salah satu bencana alam yang dahsyat kerana sering menimbulkan ancaman serius kepada infrastruktur dan kehidupan manusia, terutamanya di kawasan bandar. Selangor adalah salah sebuah negeri di Malaysia yang kerap mengalami kejadian banjir kilat sehingga boleh menjejaskan kegiatan ekonomi dan kelangsungan hidup rakyat. Teknik perlombongan data telah menarik perhatian penyelidik untuk meramal fenomena bencana alam ekoran keupayaan teknik ini untuk menerokai data berskala besar supaya dapat diaplikasi dalam situasi sebenar. Pelbagai teknik perlombongan data diaplikasi bagi meramal banjir kilat, namun masih kekurangan kajian yang memberi perhatian terhadap teknik perlombongan data popular seperti *Logistic Regression* (LR), *Support Vector Machine* (SVM) dan *k-Nearest Neighbours* (kNN) bagi meramal banjir kilat di Malaysia. Lantaran itu, matlamat kajian ini adalah untuk mengaplikasi tiga teknik perlombongan data iaitu LR, SVM dan kNN bagi meramal banjir kilat terutamanya di negeri Selangor. Sejumlah 9665 set data telah dikumpul bermula dari bulan Jun 2020 hingga Mac 2021 di 32 lokasi berbeza di negeri Selangor. Hasil analisis mendapati teknik kNN mencapai ketepatan ramalan yang terbaik berbanding teknik LR dan SVM. Dapatan kajian ini boleh dianggap sebagai satu kaedah alternatif untuk analisis peramalan banjir kilat dan data kajian yang diguna boleh dijadikan sebagai maklumat asas untuk kajian pada masa hadapan berkaitan banjir kilat terutamanya di negeri Selangor.

ABSTRACT

Flash floods are one of the most devastating natural disasters because they often pose a serious threat to infrastructure and human life, especially in urban areas. Selangor is one of the states in Malaysia that often experiences flash floods that can affect economic activities and the survival of the people. Data mining techniques have attracted the attention of researchers in predicting natural disaster phenomena due to their ability to explore and analyse large-scale data so that they can be applied in real situations. Various data mining techniques are applied to predict flash floods, but there is still a lack of studies that focus on the popular data mining techniques such as Logistic Regression (LR), Support Vector Machine (SVM), and k-Nearest Neighbours (kNN) to predict flash floods in Malaysia. Hence, the aim of this study is to apply three data mining techniques, namely LR, SVM, and kNN, to predict flash floods, especially in the state of Selangor. A total of 9665 data sets have been collected starting from June 2020 to March 2021 in 32 different locations in the state of Selangor. The results of the analysis found that the kNN technique achieved the best prediction accuracy compared to the LR and SVM techniques. The results of this study could be used as an alternative way to predict flash floods, and the research data can be used as a basis for future flash flood studies, especially in the state of Selangor.

PENGHARGAAN

Terlebih dahulu, saya ingin merakamkan setinggi-tinggi penghargaan dan terima kasih kepada penyelia utama, Dr. Muslihah binti Wook, kerana membenarkan saya membuat penyelidikan dan memberikan bimbingan dan sokongan yang tidak ternilai sepanjang penyelidikan ini dilakukan. Saya juga ingin mengucapkan terima kasih tidak terhingga kepada penyelia bersama iaitu Prof. Madya Ts. Dr. Noor Afiza binti Mat Razali dan Ts. Dr. Nor Asiakin binti Hasbullah, atas dorongan dan sokongan mereka. Penghargaan tidak terhingga juga ditujukan kepada kedua ibu dan bapa saya kerana sentiasa banyak memberi sokongan dan dorongan sepanjang pengajian saya ini. Saya berterima kasih kepada Pusat Pengurusan Penyelidikan dan Inovasi di Universiti Pertahanan Nasional Malaysia kerana mengizinkan saya untuk menjalankan kajian saya di bawah geran penyelidikan FRGS-RACER (C0025-RACER/1/2019/ICT04/UPNM//1). Akhir sekali, saya ingin mengucapkan terima kasih kepada rakan-rakan saya di Universiti Pertahanan Nasional Malaysia atas sokongan mereka sepanjang tempoh pengajian ini.

KELULUSAN

The Examination Committee has met on **1 June 2022** to conduct the final examination of **MUHAMMAD HAKIEM BIN HALIM** on his master thesis entitled '**ANALISIS PERAMALAN BANJIR KILAT DI SELANGOR MENGGUNAKAN TEKNIK PERLOMBONGAN DATA**'.

The committee recommends that the student be awarded the Master of Science (Computer Science)

Members of the Examination Committee were as follows

Prof. Madya Ts. Dr. Suzaimah

Faculty of Defence Science and Technology
Universiti Pertahanan Nasional Malaysia
(Chairman)

Dr. Arniyati

Faculty of Defence Science and Technology
Universiti Pertahanan Nasional Malaysia
(Internal Examiner)

PM Dr. Izwan Nizal bin Mohd Shahraneer

Faculty of Computer Science
Universiti Utara Malaysia
(External Examiner)

KELULUSAN

This thesis was submitted to the Senate of Universiti Pertahanan Nasional Malaysia and has been accepted as fulfilment of the requirements for the degree of **Master of Science (Computer Science)**. The members of the Supervisory Committee were as follows.

Dr. Muslihah binti Wook

Faculty of Defence Science and Technology
Universiti Pertahanan Nasional Malaysia
(Main Supervisor)

Prof. Madya Ts. Dr. Noor Afiza binti Mat Razali

Faculty of Defence Science and Technology
Universiti Pertahanan Nasional Malaysia
(Co Supervisor)

Ts. Dr. Nor Asiakin binti Hasbullah

Faculty of Defence Science and Technology
Universiti Pertahanan Nasional Malaysia
(Co Supervisor)

UNIVERSITI PERTAHANAN NASIONAL MALAYSIA

PENGISYTIHARAN TESIS

Student's full name : Muhammad Hakiem bin Halim
Date of birth : 23/02/1997
Title : Analisis Peramalan Banjir Kilat di Selangor Menggunakan Teknik Perlombongan Data
Academic session : 2020 / 2021

I hereby declare that the work in this thesis is my own except for quotations and summaries which have been duly acknowledged.

I further declare that this thesis is classified as:

- CONFIDENTIAL** (Contains confidential information under the official Secret Act 1972)*
 RESTRICTED (Contains restricted information as specified by the organisation where research was done)*
 OPEN ACCESS I agree that my thesis to be published as online open access (full text)

I acknowledge that Universiti Pertahanan Nasional Malaysia reserves the right as follows.

1. The thesis is the property of Universiti Pertahanan Nasional Malaysia.
2. The library of Universiti Pertahanan Nasional Malaysia has the right to make copies for the purpose of research only.
3. The library has the right to make copies of the thesis for academic exchange.

Signature

**Signature of Supervisor/Dean of CGS/
Chief Librarian

[Click here to enter text.](#)

[Click here to enter text.](#)

IC/Passport No.

**Name of Supervisor/Dean of CGS/
Chief Librarian

Date:

Date:

*If the thesis is CONFIDENTIAL OR RESTRICTED, please attach the letter from the organisation with period and reasons for confidentiality and restriction.

** Witness

ISI KANDUNGAN

ABSTRAK	ii
ABSTRACT	iii
PENGHARGAAN	iv
KELULUSAN	v
PENGISYTIHARAN TESIS	vii
ISI KANDUNGAN	viii
SENARAI JADUAL	x
SENARAI RAJAH	xii
SENARAI SINGKATAN	xiv

BAB

1	PENDAHULUAN	1
	1.1 Pengenalan	1
	1.2 Latar Belakang Kajian	1
	1.3 Penyataan Masalah	4
	1.4 Objektif Kajian	6
	1.5 Metodologi Kajian	7
	1.6 Skop Kajian	8
	1.7 Kepentingan Kajian	9
	1.8 Organisasi Tesis	9
	1.9 Kesimpulan	10
2	KAJIAN LITERATUR	12
	2.1 Pengenalan	12
	2.2 Latar Belakang Banjir dan Banjir Kilat	12
	2.3 Model Ramalan Banjir Kilat	19
	2.4 Teknik Perlombongan Data	26
	2.5 Faktor-faktor Banjir Kilat	31
	2.6 Kesimpulan	36
3	METODOLOGI	37
	3.1 Pengenalan	37
	3.2 Kawasan Kajian	37
	3.3 Metodologi Penyelidikan	40
	3.3.1 Pengumpulan Data	43
	3.3.2 Prapemprosesan Data	47
	3.3.3 Analisis Data	50

	3.3.3.1 <i>Logistic Regression</i>	50
	3.3.3.2 <i>Support Vector Machine</i>	52
	3.3.3.3 <i>k-Nearest Neighbours</i>	54
	3.3.4 Penilaian Model Ramalan	56
	3.4 Kesimpulan	58
4	ANALISIS DATA	59
	4.1 Pengenalan	59
	4.2 Prapemprosesan Data Banjir Kilat	60
	4.2.1 Pengendalian Data Hilang	60
	4.2.2 Analisis Deskriptif	66
	4.2.3 Pengujian Multikolineariti	67
	4.2.4 Set Data Latihan dan Set Data Pengujian	68
	4.3 Analisis Data Banjir Kilat	75
	4.3.1 Analisis Teknik <i>Logistic Regression</i>	75
	4.3.2 Analisis Teknik <i>Support Vector Machine</i>	78
	4.3.3 Analisis Teknik <i>k-Nearest Neighbours</i>	81
	4.4 Penilaian Model Ramalan Banjir Kilat	83
	4.5 Kesimpulan	87
5	PENUTUP	89
	5.1 Pengenalan	89
	5.2 Perbincangan Hasil Kajian	89
	5.3 Sumbangan Kajian	91
	5.4 Limitasi Kajian dan Cadangan Masa Hadapan	92
	5.5 Kesimpulan	94
	RUJUKAN	95
	LAMPIRAN	106
	BIODATA PELAJAR	108
	SENARAI PENERBITAN	109

SENARAI JADUAL

NO. JADUAL	TAJUK	HALAMAN
Jadual 2.1	Laporan Banjir Seluruh Malaysia pada Tahun 2021	13
Jadual 2.2	Ringkasan Model Ramalan Banjir Kilat dalam Kajian Lepas	24
Jadual 2.3	Faktor-Faktor Banjir Kilat dalam Kajian Lepas	32
Jadual 3.1	Senarai Lokasi Kawasan Kajian	39
Jadual 3.2	Kategori dan Bacaan Taburan Hujan	44
Jadual 3.3	Bacaan Tahap Paras Air	45
Jadual 3.4	Pengukuran Faktor-Faktor Banjir Kilat	46
Jadual 3.5	<i>Confusion Matrix</i> Dua-Kelas	49
Jadual 3.6	<i>Confusion Matrix</i> Tiga-Kelas	49
Jadual 3.7	<i>Confusion Matrix</i> Bagi Banjir Kilat	49
Jadual 4.1	Kaedah Min Mata Berdekatan	61
Jadual 4.2	Set Data Selangor yang Bersih	62
Jadual 4.3	Analisis Deskriptif	65
Jadual 4.4	Analisis Multikolineariti	67
Jadual 4.5	<i>Confusion Matrix</i> bagi Model SVM Terhadap Set Data Latihan	71
Jadual 4.6	<i>Confusion Matrix</i> bagi Model SVM Terhadap Set Data Pengujian	71
Jadual 4.7	<i>Confusion Matrix</i> bagi Model kNN Terhadap Set Data Latihan	72
Jadual 4.8	<i>Confusion Matrix</i> bagi Model kNN Terhadap Set Data Pengujian	73
Jadual 4.9	<i>Confusion Matrix</i> bagi Model LR Terhadap Set Data Latihan	73
Jadual 4.10	<i>Confusion Matrix</i> bagi Model LR Terhadap Set Data Pengujian	74

Jadual 4.11 Keputusan Perbandingan bagi Semua Model	83
Jadual 4.12 Nilai AUC bagi Model LR, SVM dan kNN	84

SENARAI RAJAH

NO. RAJAH	TAJUK	HALAMAN
Rajah 1.1	Kawasan Pembangunan di Negeri Selangor	3
Rajah 1.2	Metodologi Kajian	7
Rajah 2.1	Banjir Monsun di Semenanjung Malaysia	15
Rajah 2.2	Banjir Kilat di Semenanjung Malaysia	16
Rajah 2.3	Kesan Banjir Kilat di Selangor pada Disember 2021	18
Rajah 2.4	Pemodelan dan Teknik Asas Perlombongan Data	27
Rajah 3.1	Pemetaan Kawasan Kajian	39
Rajah 3.2	Penerokaan Pengetahuan dalam Pangkalan Data	41
Rajah 3.3	Aliran Metodologi Kajian Secara Keseluruhan	43
Rajah 4.1	Visualisasi <i>Widget</i> dalam Perisian <i>Orange</i>	60
Rajah 4.2	<i>Data Sampler</i> Set Data Latihan	68
Rajah 4.3	<i>Data Sampler</i> Set Data Pengujian	68
Rajah 4.4	Pengesahan 10-Silang Berganda bagi Set Data Latihan	69
Rajah 4.5	Pengesahan 10-Silang Berganda bagi Set Data Pengujian	70
Rajah 4.6	Analisis Teknik LR Terhadap Set Data Latihan dan Set Data Pengujian	75
Rajah 4.7	Hasil Nilai AUC bagi Teknik LR Terhadap Set Data Latihan	76
Rajah 4.8	Hasil Nilai AUC bagi Teknik LR Terhadap Set Data Pengujian	77
Rajah 4.9	Analisis Teknik SVM Terhadap Set Data Latihan dan Set Data Pengujian	78
Rajah 4.10	Hasil Nilai AUC bagi Teknik SVM Terhadap Set Data Latihan	79
Rajah 4.11	Hasil Nilai AUC bagi Teknik SVM Terhadap Set Data Pengujian	79

Rajah 4.12 Analisis Teknik kNN Terhadap Set Data Latihan dan Set Data Pengujian	80
Rajah 4.13 Hasil Nilai AUC bagi Teknik kNN Terhadap Set Data Latihan	81
Rajah 4.14 Hasil Nilai AUC bagi Teknik kNN Terhadap Set Data Pengujian	82
Rajah 4.15 Graf AUC bagi Model LR, SVM dan kNN untuk Set Data Latihan	85
Rajah 4.16 Graf AUC bagi Model LR, SVM dan kNN untuk Set Data Pengujian	86

SENARAI SINGKATAN

Pertubuhan Meteorologi Sedunia (WMO)

Area Under the ROC curve (AUC)

Logistic regression (LR)

k-Nearest Neighbours (kNN)

Support Vector Machine (SVM)

Jabatan Pengairan dan Saliran (JPS)

Jabatan Meteorologi Malaysia (MetMalaysia)

Badan Bencana Negara (NADMA)

BAB 1

PENDAHULUAN

1.1 Pengenalan

Banjir merupakan salah satu fenomena alam yang berlaku dalam bentuk bencana dan berbahaya kerana kesan kemusnahan yang dahsyat dan sukar diramal (Islam et al., 2022). Pertubuhan Meteorologi Sedunia (WMO) menyatakan tragedi banjir sebagai bencana alam ketiga terbesar selain daripada gempa bumi dan letusan gunung berapi yang telah mengorbankan banyak nyawa dan kemusnahan harta benda. Kebelakangan ini, kejadian banjir kian banyak dan kerap berlaku di seluruh dunia (Costache, 2019b). Hujan berterusan dan tiupan angin yang kencang adalah faktor normal menyebabkan banjir. Manakala ribut petir yang kuat, kecairan salji dan bongkah ais, dan kegagalan empangan yang menampung air turut menyumbang kepada berlakunya banjir di kawasan setempat (Estrada et al., 2017).

Menurut Ali (2018), banjir berlaku apabila air menenggelami atau melimpah ke kawasan yang kering, akibat daripada gabungan faktor cuaca dan taburan air hujan yang lebat. Terdapat pelbagai kategori banjir, antaranya banjir monsun, banjir sungai, banjir pantai, dan banjir kilat. Di antara jenis-jenis banjir ini, banjir kilat adalah lebih berbahaya kerana berlaku dalam tempoh yang lebih singkat dan air naik secara mendadak ke kawasan sekitarnya (Mosavi et al., 2018; Pham et al., 2020). Menurut

Cao et al. (2020), faktor utama berlakunya banjir kilat adalah disebabkan hujan lebat berterusan, topologi dan geologi, serta kesan daripada aktiviti manusia terutamanya pembangunan bandar.

Bab ini membincang tentang latar belakang kajian berkaitan isu banjir kilat, pernyataan masalah yang diperoleh daripada kajian literatur, objektif kajian, metodologi kajian, skop kajian, kepentingan kajian dan organisasi tesis.

1.2 Latar Belakang Kajian

Malaysia merupakan salah sebuah negara di Asia Tenggara yang beriklim khatulistiwa. Iklim khatulistiwa adalah bersifat panas dan lembap sepanjang tahun, dengan taburan hujan dipengaruhi oleh Monsun Timur Laut (November hingga Mac) dan Monsun Barat Daya (Mei hingga September). Hujan tahunan di Semenanjung Malaysia adalah sebanyak 2,500 mm, manakala 3,300 mm dan 2,300 mm masing-masing di Sabah dan Sarawak (Razali et al., 2020a). Menurut Samsuri et al. (2018), negeri Selangor menerima jumlah hujan yang lebih tinggi berbanding negeri-negeri pantai barat yang lain termasuk Sabah dan Sarawak sehingga mengakibatkan banjir kilat kerap berlaku. Banjir kilat di Selangor juga sering dikaitkan dengan aktiviti pembangunan bandar yang pesat, melibatkan penggantian permukaan semula jadi tanah dengan bangunan konkrit (Hua, 2018) sebagaimana ditunjuk dalam Rajah 1.1. Penghapusan kawasan hijau dan berhutan menyebabkan kapasiti tanah untuk menyerap air hujan semakin berkurangan yang akhirnya menjejaskan kehidupan flora dan fauna serta merosakkan kawasan kediaman penduduk (Bhuiyan et al., 2018).

1.3 Penyataan Masalah

Terdapat pelbagai kajian telah dilakukan untuk menyelesaikan isu ramalan banjir kilat dengan menggunakan teknik perlombongan data. Kajian tersebut dilakukan di negara Iran (Janizadeh et al., 2019; Panahi et al., 2021; Shirzadi et al., 2020), Romania (Costache, 2019a; Costache, Hong, et al., 2020), dan Mesir (El-Magd et al., 2021a). Di Malaysia, kajian yang menggunakan teknik perlombongan data bagi meramal banjir dilakukan oleh Raja Mohamad dan Wan Ishak (2019), Razali et al. (2020a), Shaaban et al. (2021) dan Kia et al. (2012). Namun, kajian tersebut tertumpu kepada jenis banjir musiman atau monsun, bukannya banjir kilat. Setakat kajian ini dilakukan, kajian yang menggunakan teknik perlombongan data untuk meramal banjir kilat di Malaysia telah dilakukan oleh Wardah et al. (2008) lebih 10 tahun yang lalu. Oleh itu, penemuan baharu berkaitan ramalan banjir kilat di Malaysia terutamanya di negeri Selangor perlu dilakukan.

Kajian Wardah et al. (2008) menunjukkan keupayaan teknik perlombongan data bagi meramal banjir kilat berdasarkan data yang diperoleh daripada imej satelit. Walau bagaimanapun, hasil kajian Wardah et al. ini tertumpu hanya pada satu jenis teknik sahaja iaitu *Artificial Neural Network* bagi meramal banjir kilat di hulu lembangan Sungai Klang, Selangor. Selain daripada itu, hasil kajian Wardah et al. memfokus kepada faktor berkaitan penolakan hujan yang terdiri daripada suhu kecerahan awan, min dan sisihan piawai suhu sekitar kawasan kajian, perubahan suhu dan kecerunan suhu sahaja. Justeru, kajian ramalan banjir kilat oleh Wardah et al. (2008) ini perlu diperkembang melalui teknik perlombongan data yang lain,

memperluas kawasan kajian dan mengkaji faktor lain yang turut mempengaruhi banjir kilat.

Berdasarkan kajian literatur, pelbagai jenis teknik perlombongan data telah diguna bagi melakukan analisis peramalan banjir kilat antaranya *Deep Learning Neural Networks*, *Bayesian Belief Network*, *Decision Tree*, *Logistic Regression*, *Multilayer Perceptron*, *Quadratic Discriminant Analysis*, *Classification and Regression Trees*, *Artificial Neural Network*, *Random Forest*, *Support Vector Machine*, *k-Nearest Neighbours*, dan *Naïve Bayes*. Namun, trend semasa menunjukkan teknik yang popular dan menjadi pilihan utama bagi meramal banjir kilat adalah terdiri daripada teknik *Logistic Regression* (Costache, 2019a; Costache, Hong, et al., 2020; Janizadeh et al., 2019), *k-Nearest Neighbours* (Costache, Pham, et al., 2020; El-Magd et al., 2021a; Razali et al., 2020a; Snehil & Goel, 2020), dan *Support Vector Machine* (Costache, 2019a; Razali et al., 2020a; Shaaban et al., 2021).

Rentetan itu, kajian ini menetapkan untuk mengaplikasi teknik *Logistic Regression* (LR), *k-Nearest Neighbours* (kNN), dan *Support Vector Machine* (SVM) kerana berdasarkan kajian literatur, teknik-teknik ini menjadi pilihan utama bagi meramal banjir kilat. Selain daripada itu, analisis model yang dihasil oleh teknik LR, kNN dan SVM bagi meramal banjir kilat menunjukkan nilai ramalan yang tinggi, iaitu melebihi 90% (Costache, 2019a; Costache, Hong, et al., 2020; Costache, Pham, et al., 2020; S. A. A. El-Magd et al., 2021a; Janizadeh et al., 2019; Raja Mohamad & Wan Ishak, 2019; Razali et al., 2020a; Shaaban et al., 2021; Snehil & Goel, 2020).

Penilaian model ramalan merupakan fasa penting bagi setiap keputusan analisis teknik perlombongan data yang dihasilkan. Lazimnya, metrik statistik yang standard, iaitu: *accuracy*, *precision*, *recall* dan *f-measure* sering diguna bagi menilai hasil model daripada teknik perlombongan data (Kantardžić, 2020; Witten et al., 2017). Nilai *Area Under the Curve* (AUC) turut diguna bagi menilai, mengesah keputusan dan membandingkan prestasi setiap model yang dihasilkan (Costache, Hong, et al., 2020). Menurut Shafizadeh-Moghadam et al. (2018), AUC merupakan teknik standard dalam kebanyakan kajian berkaitan penghasilan model ramalan bencana alam. Justeru, metrik *accuracy*, *precision*, *recall*, *f-measure* dan AUC diperlukan dalam kajian ini bagi menilai model yang dihasil oleh teknik-teknik perlombongan data iaitu LR, kNN dan SVM untuk meramal banjir kilat di Selangor.

1.4 Objektif Kajian

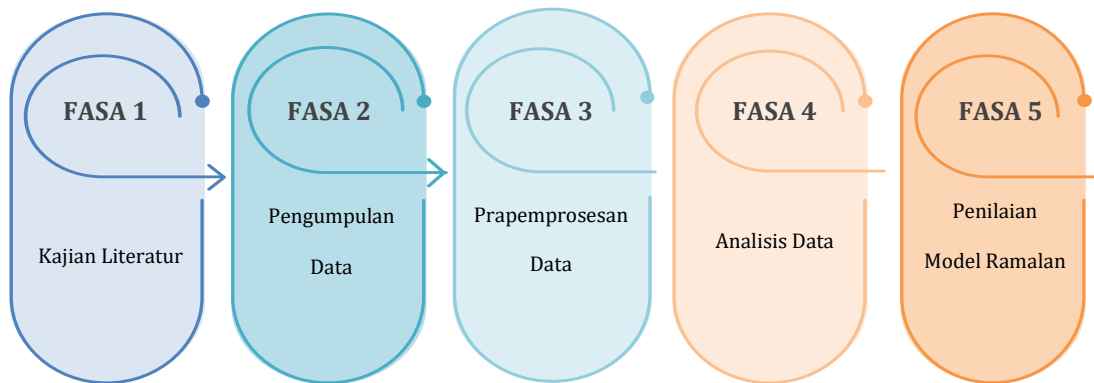
Objektif utama kajian ini adalah untuk menghasilkan peramalan banjir kilat di Selangor dengan menggunakan teknik perlombongan data. Objektif spesifik bagi kajian ini adalah seperti berikut:

- a. Mengumpul dan menganalisis data yang mempengaruhi banjir kilat di negeri Selangor.
- b. Menghasilkan peramalan banjir kilat di Selangor menggunakan teknik LR, SVM, dan kNN.

- c. Menilai dan membandingkan prestasi teknik LR, SVM, dan kNN bagi meramal banjir kilat di negeri Selangor.

1.5 Metodologi Kajian

Metodologi merupakan fasa atau langkah demi langkah yang perlu dilakukan bagi menyiapkan proses penyelidikan. Metodologi bagi kajian ini terdiri daripada lima fasa sebagaimana ditunjuk dalam Rajah 1.2.



Rajah 1.2 Metodologi Kajian

Penerangan bagi setiap fasa adalah seperti berikut:

- a. Fasa 1: Kajian Literatur

Mengkaji teknik/kaedah pembangunan model ramalan banjir kilat menggunakan teknik statistik, perlombongan data dan perlombongan data siri masa.

b. Fasa 2: Pengumpulan Data

Mengumpul data berkaitan banjir kilat seperti taburan hujan, paras air sungai, cuaca, suhu minimum, suhu maksimum daripada laman sesawang rasmi JPS Selangor (<http://infobanjirjps.selangor.gov.my/>) dan MetMalaysia (<https://www.met.gov.my/>) melibatkan 32 lokasi kawasan kajian di negeri Selangor.

c. Fasa 3: Prapemprosesan Data

Menjalankan aktiviti pembersihan data, analisis deskriptif, pengujian multikolineariti, dan pembahagian set data latihan dan set data pengujian.

d. Fasa 4: Analisis Data

Melaksana proses perlombongan data dengan menggunakan teknik LR, SVM dan kNN untuk menghasil analisis peramalan banjir kilat.

e. Fasa 5: Penilaian Model Ramalan

Melakukan penilaian prestasi hasil analisis teknik LR, SVM dan kNN melalui perbandingan nilai metrik standard iaitu *accuracy*, *precision*, *recall*, *f-measure* dan AUC.

1.6 Skop Kajian

Skop kajian adalah seperti berikut:

- a. Mengumpul data berkaitan banjir kilat daripada laman sesawang Jabatan Pengairan dan Saliran Negeri Selangor Darul Ehsan (JPS), dan Jabatan Meteorologi Malaysia (MetMalaysia).
- b. Mengumpul data berkaitan taburan hujan, paras air sungai, cuaca, tarikh, suhu maksimum dan suhu minimum.
- c. Mengumpul data dari 32 lokasi yang sering berlaku banjir kilat di negeri Selangor bermula dari bulan Jun 2020 hingga Mac 2021.
- d. Mengaplikasi tiga teknik perlombongan data iaitu LR, SVM dan kNN.

1.7 Kepentingan Kajian

Hasil kajian ini boleh dijadikan sebagai salah satu teknik alternatif bagi meramal bencana banjir kilat terutamanya di negeri Selangor. Hasil kajian ini juga boleh dijadikan sebagai panduan kepada pihak JPS dan MetMalaysia untuk meramal banjir kilat menggunakan teknik perlombongan data. Sehubungan itu, hasil keseluruhan kajian ini dapat dijadikan sebagai asas dalam kajian berkaitan banjir kilat di Selangor dan di negeri-negeri lain di Malaysia pada masa hadapan.

1.8 Organisasi Tesis

Tesis ini terdiri daripada lima bab iaitu Bab 1 (Pendahuluan), Bab 2 (Kajian Literatur), Bab 3 (Metodologi Kajian), Bab 4 (Analisis Data) dan Bab 5 (Penutup). Bab 1 membincang secara ringkas latar belakang kajian, pernyataan masalah, objektif, metodologi, skop dan kepentingan kajian.

Bab 2 membincang kajian literatur yang merangkumi kajian ilmiah berkaitan sejarah dan isu banjir, ramalan banjir kilat, kaedah ramalan banjir kilat dan penemuan kajian terdahulu berkaitan teknik perlombongan data bagi meramal banjir kilat. Bab ini juga mengupas secara umum tentang teknik perlombongan data yang diguna dalam kajian ini bagi meramal banjir kilat di Selangor.

Bab 3 menerangkan metodologi kajian yang merangkumi proses pengumpulan data, prapemprosesan data dan analisis data menggunakan teknik perlombongan data (LR, SVM dan kNN). Seterusnya, bahagian akhir bab ini mengandungi penerangan kaedah penilaian model ramalan.

Bab 4 menerangkan hasil analisis data yang telah dilakukan dalam kajian ini yang merangkumi hasil analisis dalam fasa prapemprosesan data banjir kilat, analisis data banjir kilat menggunakan teknik perlombongan data dan penilaian model ramalan banjir kilat. Manakala Bab 5 merupakan bahagian terakhir tesis ini yang membincang dan merumus keseluruhan hasil kajian serta cadangan masa hadapan.