



APLIKASI KAEDAH K-JIRAN TERDEKAT (KNN) DALAM PENGANTIAN DATA HUJAN YANG HILANG DI SELANGOR



ZAHIRAH BINTI YAHYA



UNIVERSITI PENDIDIKAN SULTAN IDRIS

2023



**APLIKASI KAEDAH K-JIRAN TERDEKAT (KNN) DALAM PENGANTIAN
DATA HUJAN YANG HILANG DI SELANGOR**

ZAHIRAH BINTI YAHYA

**LAPORAN PROJEK PENYELIDIKAN TAHUN AKHIR DIKEMUKAKAN BAGI
MEMENUHI IJAZAH SARJANA MUDA SAINS (MATEMATIK)
DENGAN PENDIDIKAN**

**FAKULTI SAINS DAN MATEMATIK
UNIVERSITI PENDIDIKAN SULTAN IDRIS**

2023



PERAKUAN KEASLIAN PENULISAN

Perakuan ini telah dibuat pada9.....(hari bulan)2.....(bulan) 20.....23......

i. Perakuan Pelajar:

Saya, Zahirah binti Yahya, D20191089511 dari Fakulti Sains dan Matematik dengan ini mengaku bahawa Laporan Projek Penyelidikan Tahun Akhir yang bertajuk Aplikasi Kaedah K-Jiran Terdekat (KNN) Dalam Penggantian Data Hujan Yang Hilang Di Selangor..... adalah hasil kerja saya sendiri. Saya tidak memplagiat dan apa-apa penggunaan mana-mana hasil kerja yang mengandungi hak cipta telah dilakukan secara urusan yang wajar dan bagi maksud yang dibenarkan dan apa-apa petikan, ekstrak, rujukan atau pengeluaran semula daripada atau kepada mana-mana hasil kerja yang mengandungi hak cipta telah dinyatakan dengan se jelasnya dan secukupnya.

Tandatangan Pelajar



PENGHARGAAN

Terlebih dahulu saya ingin memanjatkan kesyukuran ke hadrat Ilahi kerana dengan izin dan limpah kurniaan-Nya saya dapat menyiapkan kajian saya ini dengan jayanya walaupun menempuhi pelbagai dugaan dan rintangan. Alhamdulillah.

Sepanjang saya menjalankan kajian ini, terdapat banyak pihak yang telah memberikan sokongan, bantuan dan kerjasama yang khusus kepada saya dalam usaha menjayakan kajian ini. Justeru, di kesempatan ini, saya ingin merakamkan setinggi-tinggi penghargaan dan jutaan ucapan terima kasih yang tak terhingga kepada mereka yang berkenaan.

Pertamanya, saya ingin mengucapkan jutaan terima kasih kepada pensyarah penyelia yang disayangi Dr. Nur Hamiza binti Adenan yang telah banyak berkorban tenaga dan waktu untuk menyelia dan membimbing saya sepanjang menjalankan kajian ini. Bantuan, motivasi yang tidak putus-putus, semangat dan kebijaksanaan beliau yang telah disumbangkan adalah sangat memberi makna kepada kejayaan penghasilan dan penyempurnaan kajian ini.

Tidak dilupakan juga kepada pensyarah lain yang telah menilai dan memberikan pandangan kepada penambahbaikan kajian ini. Begitu juga kepada pihak Universiti Pendidikan Sultan Idris, Fakulti Sains dan Matematik serta Jabatan Matematik yang telah memudahkan urusan kepada saya untuk menjalankan kajian ini bagi melengkapkan keperluan untuk menerima ijazah kelak.

Kepada kedua ibu bapa yang saya kasihi En. Yahya bin Duraim dan Pn. Zaiton binti Saimeh, adik beradik dan keluarga, saya ingin menzahirkan rasa terima kasih yang tak terhingga kerana telah memberikan kasih sayang, dorongan, doa dan menjadi tulang belakang kepada saya untuk terus bermotivasi melangsungkan pengajian dan menyiapkan kajian ini dengan penuh gigit dan tekun.

Setinggi-tinggi terima kasih dan penghargaan kepada rakan-rakan seperjuangan saya yang telah memberikan sedikit sebanyak bantuan langsung dan tidak langsung, memberikan semangat khusus dan sokongan kepada saya sepanjang menyelesaikan kajian ini. Rakan seperjuangan bagi kos AT48, terutamanya rakan-rakan saya iaitu Amie, Fizah, Azee, Shikin, Mirah, Afiqah dan Sya yang selalu ada di sepanjang saya menjalankan pengajian ini.

Akhir kata, ucapan terima kasih dan penghargaan ini saya ingin sampaikan kepada semua pihak yang terlibat secara langsung dan tidak langsung dalam memberikan sumbangan, bantuan, cadangan dan segala bentuk hal dalam menyiapkan kajian ini. Semoga kajian ini dapat menjadi salah satu punca ilmu yang berguna kepada orang lain dan tatapan generasi akan datang.



ABSTRAK

Kajian ini bertujuan untuk mengaplikasikan kaedah K-Jiran Terdekat (KNN) dalam penggantian data hujan yang hilang bagi stesen-stesen hujan yang terdapat di Selangor. Kajian ini merangkumi dua objektif utama iaitu (i) mengaplikasikan kaedah KNN dalam penggantian data hujan yang hilang di Selangor dan (ii) menganalisis perbezaan ketepatan hasil penggantian KNN terhadap data hujan yang hilang di stesen yang berbeza di Selangor. Kajian ini dijalankan terhadap data siri masa hujan bagi sepuluh stesen hujan di Selangor. Sebanyak 365 data telah digunakan bagi setiap stesen dan data yang digunakan telah dihilangkan dengan mengikut kategori 1%, 5%, 10%, 25% dan 50% . Kemudian, data yang hilang tersebut digantikan dengan nilai hasil pengiraan KNN. Seterusnya, ketepatan data hasil pengiraan KNN dinilai menggunakan pekali korelasi. Hasil kajian menunjukkan, kaedah KNN dapat diaplikasikan dalam mengira data hujan yang hilang namun nilai pekali korelasi adalah rendah. Seterusnya, analisis perbezaan ketepatan hasil penggantian KNN bagi setiap kategori peratusan data hujan yang hilang menunjukkan bahawa hampir kesemua kategori mendapatkan julat pekali korelasi dengan interpretasi yang sangat lemah iaitu di antara 0.01 dan 0.20 ($0.01 < r < 0.20$). Konklusinya, pengaplikasian kaedah KNN terhadap penggantian data hujan yang hilang di Selangor adalah kurang bersesuaian. Implikasinya, kajian ini dapat menyumbangkan idea dan rujukan kepada pengkaji akan datang yang terlibat dengan data hujan yang hilang di Selangor dalam menentukan kaedah yang lebih bersesuaian.

APPLICATION OF THE K-NEAREST NEIGHBOUR (KNN) METHOD IN REPLACING MISSING RAINFALL DATA IN SELANGOR

ABSTRACT

This study aimed to apply the Nearest K-Neighbours (KNN) method in the replacement of missing rainfall data for rainfall stations available in Selangor. This study includes two main objectives which are (i) to apply the KNN method in the replacement of missing rainfall data in Selangor and (ii) to analyse the difference in the accuracy of KNN replacement results on missing rainfall data at different stations in Selangor. This study was conducted on rainfall time series data for ten rainfall stations in Selangor. A total of 365 data was used for each station and the data used was omitted by categories of 1%, 5%, 10%, 25% and 50%. Then, the missing data is replaced by the value of the result of the KNN calculation. Next, the accuracy of the data of the results of the KNN calculation is evaluated using the correlation coefficient. The results showed that the KNN method can be applied in calculating missing rainfall data, but the correlation coefficient value is low. Next, the analysis of the difference in accuracy of the KNN replacement results for each category of percentage of lost rainfall data showed that almost all categories obtained a range of correlation coefficients with a very weak interpretation of between 0.01 and 0.20 ($0.01 < r < 0.20$). In conclusion, the application of the KNN method on the replacement of lost rainfall data in Selangor is less appropriate. The implications of this study can contribute ideas and references to future researchers who are involved with the missing rainfall data in Selangor in determining a more suitable method.



KANDUNGAN

PERKARA	MUKA SURAT
PERAKUAN	xi
PENGHARGAAN	xii
ABSTRAK	xiii
KANDUNGAN	xv
SENARAI JADUAL	xviii
SENARAI RAJAH	xx
BAB 1 PENDAHULUAN	
1.1 Pengenalan	1
1.2 Pernyataan Masalah	2
1.2.1 Kepentingan Data Hujan	2
1.2.2 Isu yang Berkait Data Hujan yang Hilang	3
1.2.3 Punca Data Hujan yang Hilang	4
1.2.4 Masalah Data Hilang di Dalam dan Luar Negara	4
1.2.5 Kaedah yang Digunakan untuk Mengira Data Hidrologi yang Hilang	7
1.2.6 Keperluan Data Hujan di Selangor	9
1.3 Objektif	10
1.4 Persoalan Kajian	11
1.5 Signifikan Kajian	11
1.6 Batasan Kajian	12
1.7 Sumbangan Kajian	14
1.7.1 Sumbangan Kajian Terhadap Aplikasi KNN untuk Data Hujan	14
1.7.2 Sumbangan Kajian kepada Negara Malaysia	14
1.8 Definisi Operasional	15
1.8.1 Data Hilang	15
1.8.2 Data Hujan	15
1.8.3 Ketepatan Hasil Pengiraan Data Hilang	16



BAB 2 KAJIAN LITERATUR

2.1 Pengenalan	17
2.2 Data Hilang	18
2.3 Kaedah yang Digunakan untuk Mengganti Data Siri Masa yang Hilang	19
2.4 Kaedah yang Digunakan untuk Mengganti Data Siri Masa Hidrologi yang Hilang	21
2.5 Kaedah yang Digunakan untuk Mengganti Data Siri Masa Hujan yang Hilang Di Luar Negara dan di Dalam Malaysia	22
2.6 Kaedah KNN di Luar Negara dan di Malaysia	26
2.7 Kesimpulan	28

BAB 3 METODOLOGI

3.1 Pengenalan	29
3.2 Pemilihan Kawasan	30
3.3 Data Hilang	33
3.4 Kaedah KNN	34
3.5 Pekali Korelasi	37
3.6 Kesimpulan	38

BAB 4 DAPATAN KAJIAN

4.1 Pengenalan	39
4.2 Kawasan Kajian	40
4.2.1 Data	40
4.3 Hasil Dapatan Kajian	43
4.3.1 Pekali Korelasi Terhadap Data Hujan di Tanjung Rhu	44
4.3.2 Pekali Korelasi Terhadap Data Hujan di Agrotek Sepang	45
4.3.3 Pekali Korelasi Terhadap Data Hujan di Ladang. Sepang	46
4.3.4 Pekali Korelasi Terhadap Data Hujan di Ladang. Bute	47
4.3.5 Pekali Korelasi Terhadap Data Hujan di Ladang. Telok Merbau	47
4.3.6 Pekali Korelasi Terhadap Data Hujan di P/A Kg. Tali Air Morib	48
4.3.7 Pekali Korelasi Terhadap Data Hujan di P/A Kelanang	49
4.3.8 Pekali Korelasi Terhadap Data Hujan di P/A Pekan Banting	49
4.3.9 Pekali Korelasi Terhadap Data Hujan di P/A Sungai Sedu	50
4.3.10 Pekali Korelasi Terhadap Data Hujan di Rtb Bukit Changgang	51

4.4 Kesimpulan	52
BAB 5 PERBINCANGAN DAN KESIMPULAN	
5.1 Pengenalan	53
5.2 Perbincangan	54
5.2.1 Mengaplikasikan Kaedah KNN dalam Penggantian Data Hujan yang Hilang	55
5.2.2 Menganalisis Perbezaan Ketepatan Hasil Penggantian KNN Terhadap Data Hujan yang Hilang	56
5.3 Kesimpulan	60
5.4 Cadangan Kajian Lanjutan	62
5.4.1 Kaedah Penggantian Data	62
5.4.2 Peratusan Data Hilang yang Digunakan	63
RUJUKAN	64

**SENARAI JADUAL**

JADUAL	PERKARA	MUKA SURAT
Jadual 1.1	Masalah Data Hujan yang Hilang di Malaysia	5
Jadual 1.2	Masalah Data Hujan yang Hilang di Luar Negara	6
Jadual 2.1	Peraturan Asas Data Hilang mengikut (Widaman, 2006; Hamzah et al., 2020)	18
Jadual 2.2	Kaedah Penggantian Data Siri Masa yang Hilang	20
Jadual 2.3	Kaedah Penggantian Data Hidrologi yang Hilang	22
Jadual 2.4	Kaedah Penggantian Data Hujan yang Hilang di Luar Negara	23
Jadual 2.5	Kaedah Penggantian Data Hujan yang Hilang di Malaysia	24
Jadual 3.1	Stesen Terpilih bagi Kawasan Kajian	32
Jadual 3.2	Data Hujan Siri Masa Harian yang Lengkap	34
Jadual 3.3	Kekuatan Nilai Pekali Korelasi (Evans et al., 1996)	38
Jadual 4.1	Statistik asas data hujan bagi beberapa stesen di Selangor	42
Jadual 4.2	Jumlah data hilang mengikut peratus bagi setiap kategori data hilang	44
Jadual 4.3	Hasil dapatan kajian bagi menganalisis perbezaan ketepatan hasil pengiraan KNN terhadap data hilang bagi data hujan di stesen Tanjung Rhu	45
Jadual 4.4	Hasil dapatan kajian bagi menganalisis perbezaan ketepatan hasil pengiraan KNN terhadap data hilang bagi data hujan di stesen Agrotek Sepang	46
Jadual 4.5	Hasil dapatan kajian bagi menganalisis perbezaan ketepatan hasil pengiraan KNN terhadap data hilang bagi data hujan di stesen Ladang. Sepang	46
Jadual 4.6	Hasil dapatan kajian bagi menganalisis perbezaan ketepatan hasil pengiraan KNN terhadap data hilang bagi data hujan di stesen Ladang Bute	47



Jadual 4.7	Hasil dapatan kajian bagi menganalisis perbezaan ketepatan hasil pengiraan KNN terhadap data hilang bagi data hujan di stesen Ladang, Telok Merbau	48
Jadual 4.8	Hasil dapatan kajian bagi menganalisis perbezaan ketepatan hasil pengiraan KNN terhadap data hilang bagi data hujan di stesen P/A Kg. Tali Air Morib	48
Jadual 4.9	Hasil dapatan kajian bagi menganalisis perbezaan ketepatan hasil pengiraan KNN terhadap data hilang bagi data hujan di stesen P/A Kelanang	49
Jadual 4.10	Hasil dapatan kajian bagi menganalisis perbezaan ketepatan hasil pengiraan KNN terhadap data hilang bagi data hujan di stesen P/A Pekan Banting	50
Jadual 4.11	Hasil dapatan kajian bagi menganalisis perbezaan ketepatan hasil pengiraan KNN terhadap data hilang bagi data hujan di stesen P/A Sungai Sedu	51
Jadual 4.12	Hasil dapatan kajian bagi menganalisis perbezaan ketepatan hasil pengiraan KNN terhadap data hilang bagi data hujan di stesen Rtb Bukit Changgang	51
Jadual 5.1	Nilai pekali korelasi bagi semua stesen untuk semua kategori peratusan data hilang	58

SENARAI RAJAH

JADUAL PERKARA

MUKA SURAT

Rajah 1 Carta Alir Kajian

33



BAB 1

PENDAHULUAN



1.1 Pengenalan

Bab ini membincangkan beberapa perkara penting dalam menjalankan kajian ini. Terdapat perbincangan mengenai pernyataan masalah, objektif kajian, persoalan kajian serta signifikan kajian. Di dalam bab ini juga turut membincangkan mengenai batasan kajian serta sumbangannya terhadap ke beberapa pihak yang berkaitan.





1.2 Pernyataan Masalah

Kajian ini dijalankan berdasarkan kepada masalah-masalah yang wajar untuk diberi perhatian pada masa kini. Pernyataan masalah bagi kajian ini adalah berteraskan kepada kepentingan data hujan itu sendiri serta isu-isu yang berkait rapat dengan hujan yang perlu diatasi. Beberapa faktor yang mendorong kepada berlakunya isu berbangkit mengenai data hujan juga menjadi fokus permasalahan di dalam kajian ini. Pemilihan kawasan yang berkepentingan juga menjadi salah satu pendorong untuk menjalankan kajian ini.

1.2.1 Kepentingan Data Hujan

Negara Malaysia merupakan sebuah negara yang mempunyai cuaca khatulistiwa. Taburan hujan di Malaysia juga dipengaruhi oleh kehadiran Lautan Pasifik di timur dan Lautan Hindi di barat (Mahmud, 2018). Fenomena El Nino da La Nina dapat menyebabkan berlakunya cuaca yang kering dan basah yang mana ia adalah salah satu kesan daripada kewujudan lautan tersebut. Hal ini boleh menyebabkan berlakunya kekerapan hujan yang rendah dan tinggi di negara ini (Mahmud, 2018). Rentetan daripada fenomena La Nina, berlakunya bencana banjir di beberapa buah kawasan seperti Sabah dan Lembah Klang serta kawasan Pantai Timur yang disebabkan oleh hujan yang lebat. Justeru, bencana alam seperti banjir dan banjir kilat menyebabkan terdapat keperluan kepada pihak bertanggungjawab untuk menjalankan kajian dan memerlukan data hujan yang lengkap.

Data bagi taburan hujan adalah salah satu parameter hidrologi terpenting yang perlu dikaji dengan teliti bagi keperluan seperti perancangan sumber air, perancangan pertanian, reka bentuk saliran, dan menganggarkan magnitud dan kekerapan kejadian ekstrem seperti banjir dan kemarau (Hazarika & Sarma, 2021). Selain itu, data hujan





yang lengkap diperlukan dalam peramalan cuaca bagi sesebuah kawasan untuk keperluan peramalan cuaca (Faridah, 2017). Implikasinya, data hujan yang lengkap membantu mendapatkan peramalan yang tepat dan seterusnya membantu dalam perancangan aktiviti sosio-ekonomi dan pembangunan dalam sesebuah negara (Faridah, 2017). Tambahan lagi, peramalan data hujan berkait secara langsung dengan peramalan hujan. Peramalan hujan yang tepat mampu memberikan amaran awal kepada bencana banjir (Abd Hamid & Md Noorani, 2012). Justeru, boleh disimpulkan bahawa data hujan adalah sangat penting dan berguna di dalam banyak kajian yang boleh membawa kepada kemaslahatan bersama.

1.2.2 Isu yang Berkait Data Hujan yang Hilang

Kehilangan data yang kritikal adalah situasi yang tidak dapat dielakkan dalam pelbagai bidang penyelidikan begitu juga dengan data hujan (Lai & Kuok, 2019). Kehilangan data bagi data hujan merupakan suatu isu yang tidak asing lagi bagi pengkaji yang ingin membuat kajian (Faridah, 2017) dan (Abd Hamid & Noorani, 2014). Selain daripada dua kajian ini, terdapat lebih banyak kajian yang juga telah melibatkan data hilang samada dalam dan luar negara.

Data hilang ini perlu dirawat dengan menjalankan penggantian data dengan menggantikan data yang hilang menerusi pengiraan data hilang terlebih dahulu sebelum dapat meneruskan kajian (Faridah, 2017). Jelaslah kesan daripada data hujan yang hilang menyebabkan satu-satu kajian itu sukar untuk dijalankan. Selain itu, kesan langsung daripada kehilangan data adalah kajian akan mengalami kekurangan data input untuk melakukan simulasi hidrologi dan kajian ramalan yang mempunyai ketepatan yang memerlukan set data hujan yang konsisten dan lengkap (Lai & Kuok,





2019). Maka, bagi memastikan kebolehpercayaan kajian hidrologi, data hujan yang hilang perlu diatasi dan dikendalikan dengan teliti.

1.2.3 Punca Data Hujan yang Hilang

Kajian ini hanya memberi fokus terhadap data hujan yang hilang dan data hujan yang hilang ini boleh berlaku atas beberapa faktor yang tertentu. Merujuk kepada Hazarika dan Sarma (2021) menyatakan bahawa isu mengenai data hujan yang hilang ini berlaku adalah berpunca daripada kegagalan fungsi instrumen, gangguan komunikasi, ralat pemerhatian dan kebakaran. Kesilapan manusia dalam mengurus set data, kegagalan peralatan dan bencana alam yang boleh merosakkan peralatan pengukur di tapak juga antara faktor yang boleh menyumbang kepada berlakunya isu data yang hilang terjadi (Lai & Kuok, 2019). Dalam konteks ini, kita boleh fahami bahawa peralatan yang digunakan untuk mengukur data hujan khususnya juga bergantung kepada sistem dan sekiranya berlaku kerosakan pada peralatan ini maka ia akan mengganggu dalam pengumpulan data yang lengkap pada waktu itu. Hal ini jelas menunjukkan bahawa punca-punca yang tidak dapat dielakkan ini menyumbang kepada seringnya berlaku isu utama dalam kajian yang berkaitan dengan data iaitu data yang hilang.

1.2.4 Masalah Data Hilang di Dalam dan Luar Negara

Seperti yang telah sedia maklum bahawa isu utama yang berkait rapat dengan data hujan adalah data hujan yang hilang. Di negara kita sendiri iaitu Malaysia juga tidak terkecuali daripada mengalami isu yang berkait dengan data hujan ini. Terdapat beberapa kajian yang telah dijalankan melibatkan masalah data hilang di Malaysia yang mana mempunyai tujuan dan objektif tertentu. Antara yang utama adalah untuk mencari kaedah terbaik dalam penggantian data hilang (Lai & Kuok, 2019); (Liang Chuan et al., 2020); (Burhanuddin et al., 2017) dan (Nor et al., 2020). Malahan setiap daripada kajian



juga mempunyai peratusan data hilang yang berbeza-beza. Jadual 1.1 menunjukkan masalah data hujan yang hilang di Malaysia.

Jadual 1.1 Masalah Data Hujan yang Hilang di Malaysia

Pengkaji	Kawasan	Peratus Data Hilang (%)
Lai & Kuok (2019)	Kuching, Sarawak	1, 5, 10, 15, 20, 25, 30
Liang Chuan et al. (2020)	Pahang	0 – 30
Burhanuddin et al. (2017)	Timur Semenanjung Malaysia	5, 10, 15, 20, 25, 30
Ismail & Ibrahim (2017)	Terengganu	5, 10, 15, 20
Azman et al. (2021)	Tasik Kenyir, Terengganu	7.7, 16.9, 21.6
Kamaruzaman et al. (2017)	Semenanjung Malaysia	<10
Nor et al. (2020)	Semenanjung Malaysia	8.59
Faridah (2017)	Terengganu, Kedah, Perak	2.46, 0.42, 0.27

Daripada Jadual 1.1 ini, dapat dilihat bahawa peratus data hilang bagi setiap kajian adalah pelbagai. Namun, jika diteliti lagi terdapat banyak kajian yang melibatkan peratusan data hilang 1 hingga 30 peratus. Selain di Malaysia, terdapat beberapa negara lain yang mana juga mempunyai masalah berkaitan dengan data hujan yang hilang. Negara-negara tersebut adalah seperti di dalam Jadual 1.2.

Jadual 1.2 Masalah Data Hujan yang Hilang di Luar Negara

Pengkaji	Negara	Peratusan Data Hilang (%)
Sahoo dan Kumar (2022)	India	33.66 – 97.98
Shaharudin et al. (2020)	Yogyakarta, Indonesia	58.1
Miró et al. (2017)	Iberian (Sepanyol & Portugal)	18 – 57
Jahan et al. (2019)	Bangladesh	1.9 – 6.6
Farzandi et al. (2022)	Iran	9.1 – 68.8
Wuthiwongyothin et al. (2021)	Thailand	5, 10, 20, 30, 40, 50
Aieb et al. (2019)	Algeria	5.3 – 16.7

Merujuk Jadual 1.2, India, Indonesia, Sepanyol, Portugal, Bangladesh, Iran dan Thailand merupakan antara negara yang menghadapi isu data hilang. Malahan, selain daripada negara ini juga mungkin terdapat beberapa negara lain yang menghadapi masalah yang sama berkaitan dengan ketersediaan data hujan yang lengkap. Manakala berkaitan dengan peratusan data hilang pula, peratusan data hilang di luar negara adalah lebih tinggi berbanding peratusan data hilang di Malaysia yang telah dibincangkan pada Jadual 1.1. Tamsilnya seperti negara India yang menghadapi data hilang dengan peratusan yang sangat tinggi iaitu 97.98%. Jelaslah bahawa isu data yang hilang ini tidak diambil ringan oleh ramai pihak kerana sedar akan kepentingannya kepada ramai pihak.



1.2.5 Kaedah yang Digunakan untuk Mengira Data Hidrologi yang Hilang

Di dalam analisis hidrologi, kualiti data-data hujan yang baik adalah sangat penting (Yogafanny & Legono, 2022). Namun begitu, masalah utama yang sering dihadapi dalam penggunaan data hidrologi adalah berkaitan dengan ketersediaan data tersebut yang mana sering berlakunya data hilang. Hal ini kerana kehadiran nilai data yang hilang menyebabkan analisis berkaitan dengan masalah-masalah hidrologi tidak dapat memberikan keputusan yang berkualiti dan berkesan (Jahan et al., 2019). Oleh itu, data hidrologi yang hilang ini perlu di taksir ataupun diganti dengan sesuatu nilai dengan memilih kaedah yang terbaik dan sesuai (Yogafanny & Legono, 2022). Terdapat banyak kaedah bagi mengira data hidrologi yang hilang yang telah disarankan oleh pengkaji-pengkaji sebelum ini (Azman et al., 2021).

Antara kaedah yang digunakan dalam penggantian data hidrologi yang hilang di Malaysia adalah seperti kajian daripada Lai dan Kuok (2019) yang telah menggunakan kaedah KNN dan *Bayesian Principal Component Analysis* (BPCA) dalam menggantikan data hilang yang terdapat di dalam kajiannya. Manakala di dalam kajian Liang Chuan et al. (2020) pula mengaplikasikan kaedah *M-Component Probabilistic Principal Component Analysis and The Expectation-Maximisation* MPPCA-EM untuk merawat data hidrologi yang hilang di dalam kajiannya. Kaedah *Normal Ratio* (NR) pula diaplikasikan di dalam kajian Burhanuddin et al. (2017). Seterusnya, kajian Ismail dan Ibrahim (2017) pula telah menggunakan empat jenis kaedah penggantian data hilang terhadap data hidrologi yang digunakan iaitu kaedah *Arithmetic Average* (AA), NR, *Inverse Distance* (ID) dan *Coefficient of Correlation* (CC). Selain daripada kajian Ismail dan Ibrahim (2017), ID juga telah diaplikasikan di dalam kajian Azman et al. (2021) bersama kaedah *Multiple Imputation* (MI) dan *Expectation Maximisation* (EM). Di dalam kajian Kamaruzaman et al. (2017) juga telah





menggunakan 3 kaedah penggantian data hilang iaitu kaedah ID, *Modified Correlation* (MC) dan kaedah kombinasi antara CC dan ID. Kajian (Nor et al., 2020) pula telah menggunakan 5 kaedah penggantian untuk merawat data hidrologi yang hilang iaitu KNN, *Replace by Mean* (RM), *Markov Chain Monte Carlo* (MCMC), *Non-Linear Interactive Partial Least-Square* (NIPALS) dan *Random Forest* (RF). Akhir sekali adalah kajian yang telah dijalankan di dalam Faridah (2017) iaitu dengan menggunakan kaedah RM.

Hasil rujukan kajian yang diperoleh oleh pengkaji adalah sebanyak 6 kajian yang berkaitan dengan kaedah penggantian yang digunakan di luar negara terhadap data hidrologi yang hilang. Kajian pertama telah menggunakan empat kaedah penggantian iaitu kaedah KNN, *Self-Organizing Maps* (SOM), RF dan *Feed-Forward Neural Network* (FNN) yang mana kajian ini dijalankan dalam Sahoo dan Kumar (2022).

Seterusnya, kajian kedua pula adalah di dalam Shaharudin et al. (2020) yang telah menggunakan 5 kaedah penggantian untuk dibandingkan prestasinya iaitu kaedah RM, KNN, RF, NIPALS dan MCMC. Kajian ketiga pula (Miró et al., 2017) mengaplikasikan kaedah KNN, *Iterated Local Least Squares* (ILLS), EM, *Principal Component Analysis* (PCA), SOM dan *Non-Linear PCA* (NLPCA). Selanjutnya, kajian di dalam Jalan et al. (2019) yang menggunakan sebanyak 6 kaedah penggantian iaitu kaedah AA, NR, ID, EM-MCMC, *Single Best Estimator* (SBE) dan *Multiple Regression* (MR). Kaedah seterusnya pula iaitu *Artificial Neural Networks* (ANN), *Support Vector Regression* (SVR) dan *Genetic Algorithms Optimization* (GA) yang merupakan antara tiga kaedah telah digunakan di dalam kajian Farzandi et al. (2022). Kajian terakhir iaitu di dalam Wuthiwongyothin et al. (2021) telah menggunakan dua kaedah penggantian terhadap data hidrologi yang digunakan iaitu kaedah ID dan CC.





Daripada perbincangan berkaitan dengan kaedah penggantian data hidrologi yang digunakan di Malaysia dan luar negara, boleh dirumuskan bahawa terdapat pelbagai kaedah penggantian yang digunakan dalam menangani masalah data hilang. Merujuk kepada (Shaharudin et al., 2020) menyatakan bahawa kaedah KNN merupakan kaedah yang boleh dipercayai dan praktikal untuk merawat data hidrologi yang hilang. Kaedah KNN ini adalah kaedah yang mudah untuk dilaksanakan tetapi ketepatan nilai penggantian yang diperolehi adalah tinggi (Pan et al., 2015). Meskipun begitu, daripada perbincangan kita boleh lihat bahawa kaedah KNN ini masih kurang dilaksanakan di dalam kajian-kajian yang berkaitan dengan penggantian data hilang. Justeru, kaedah KNN wajar untuk diberikan perhatian dalam penggunaannya untuk mengganti data hidrologi yang hilang.

1.2.6 Keperluan Data Hujan di Selangor

Banjir besar yang berlaku di Selangor baru-baru ini adalah disebabkan oleh hujan lebat beberapa hari berturut-turut (Zulkepli & Idris, 2022). Merujuk kepada Tahir (2021) menyatakan bahawa banjir besar telah berlaku di Selangor dan banjir yang melanda ini dianggap sebagai banjir terburuk yang pernah berlaku di negeri itu. Hal ini dikatakan sedemikian kerana banjir ini telah meragut banyak nyawa serta mengakibatkan banyak kerugian (Berita Harian, 2021). Banjir besar yang berlaku ini adalah berpunca daripada hujan luar biasa yang berterusan sehingga jumlah taburan hujan pada ketika itu adalah dua kali ganda taburan hujan tertinggi yang dicatat sebelum ini (Bernama, 2021). Maka kajian lanjutan di negeri ini diperlukan dan data hujan yang lengkap adalah satu keperluan bagi bertujuan memantau taburan hujan yang berlaku sehingga menyebabkan berlakunya banjir (Zulkepli & Idris, 2022).





Bukan itu sahaja, Selangor ini mempunyai penduduk yang padat dan paling ramai serta mencatatkan peratusan pembandaran yang tinggi yang mana ia pesat dengan pelbagai sektor dan menjadi pusat sosio-ekonomi bagi ramai penduduk (Majid, 2019). Oleh itu, bagi memberikan impak yang positif terhadap aktiviti sosio-ekonomi dan pembangunan (Faridah, 2017) keperluan data hujan yang lengkap adalah penting untuk membuat peramalan cuaca yang berketepatan tinggi kerana dengan peramalan tersebut kita boleh membuat persediaan yang baik dengan lebih awal untuk menghadapinya (Abd Hamid & Md Noorani, 2012). Tambahan pula peramalan cuaca yang disediakan oleh pihak bertanggungjawab akan digunakan oleh ramai pihak untuk membuat keputusan yang sangat penting dalam pelbagai sektor (Faridah, 2017). Natiujahnya, negeri Selangor merupakan negeri yang sangat berkepentingan dalam kajian yang melibatkan data hujan. Justeru, data hujan yang lengkap di Selangor amatlah diperlukan pada ketika ini.



1.3 Objektif

Dalam menjalankan kajian ini, terdapat dua objektif yang telah diketengahkan untuk memastikan kajian ini mempunyai matlamat dan tujuan. Dua objektif kajian tersebut adalah:

- 1.3.1 Mengaplikasikan kaedah KNN dalam penggantian data hujan yang hilang di Selangor.
- 1.3.2 Menganalisis perbezaan ketepatan hasil penggantian KNN terhadap data hujan yang hilang bagi stesen yang berbeza di Selangor.



1.4 Persoalan Kajian

Bagi menjalankan satu-satu kajian, persoalan kajian adalah penting untuk memastikan proses kajian lebih terarah dan kajian dapat menjawab kesemua soalan-soalan kajian.

Dalam kajian ini, terdapat dua persoalan iaitu:

1.4.1 Adakah kaedah KNN dapat diaplikasikan bagi menggantikan data hujan yang hilang di Selangor?

1.4.2 Adakah terdapat perbezaan ketepatan hasil penggantian KNN terhadap data hujan yang hilang bagi stesen yang berbeza di Selangor?

1.5 Signifikan Kajian

Signifikan kajian ini adalah merujuk kepada pengaplikasian kaedah KNN untuk mengira data hilang bagi data hujan di lokasi yang telah disasarkan mengikut kepada beberapa kategori. Penggunaan kaedah ini adalah bagi mengesan dan mengetahui samada ia berkesan ataupun kurang berkesan untuk digunakan dalam pengiraan data hilang bagi data hujan di kawasan yang disasarkan. Pengiraan data hilang bagi data hujan menggunakan kaedah KNN ini telah dijalankan ke atas data-data daripada negeri Selangor untuk beberapa buah stesen yang terdapat di negeri tersebut. Pemilihan kawasan ataupun lokasi kajian ini adalah merujuk kepada beberapa faktor yang sewajarnya diambil berat kerana berkemungkinan besar faktor-faktor ini akan menjadi sebab data hujan bagi lokasi ini akan kerap difokuskan untuk penelitian pihak yang berkaitan.

1.6 Batasan Kajian

Kajian ini telah memfokuskan kepada perkara utama iaitu untuk mengaplikasikan kaedah yang boleh diguna pakai dalam pengiraan data hilang bagi kawasan yang berkepentingan di Malaysia. Maka, penggunaan kaedah KNN telah diketengahkan dalam kajian ini untuk menganalisis perbezaan ketepatan hasil pengiraan KNN terhadap data hilang bagi data hujan yang dikaji dari pilihan kawasan yang telah ditetapkan.

Bagi mencapai objektif kajian, data-data hujan yang digunakan di dalam kajian ini adalah melibatkan sepuluh stesen pencerapan data yang terdapat di dalam negeri Selangor iaitu stesen Tadika Kemas, Tg. Sepat, stesen Ldg. Teluk Merbau, stesen Tanjung Rhu, stesen Agrotek Sepang, stesen Ldg. Sepang, stesen TKPM Kg. Sg. Kelambu, stesen Ldg. Bute, stesen Pantai Pekan Banting, stesen RTB Bukit Changgang dan stesen Paya Indah. Data hujan bagi sepuluh stesen adalah pada tahun 2011 yang mana data bagi setiap stesen pada tahun tersebut adalah lengkap dan kajian ini juga menggunakan data hujan harian. Kajian ini dijalankan dengan hanya menggunakan data hujan sahaja. Manakala data hidrologi yang lain seperti aras sungai dan aliran sungai tidak digunakan atas sebab fokus kajian hanya terhad kepada hujan sahaja. Keseluruhan data yang diperoleh bagi stesen-stesen yang terlibat adalah diperoleh daripada Jabatan Pengairan dan Saliran Malaysia (JPS).

Seterusnya, kajian ini tidak melibatkan keseluruhan data yang terdapat di Malaysia dan hanya membuat kajian data terhadap sebuah negeri sahaja yang terdapat di negara ini iaitu negeri Selangor. Begitu juga dengan stesen pencerapan yang terdapat di Selangor yang mana kajian ini hanya melibatkan sepuluh stesen terpilih sahaja daripada keseluruhan stesen yang terdapat di negeri ini. Kawasan ini dipilih dalam kajian disebabkan keperluan dan kepentingan yang dapat diperoleh daripada kajian.



Kawasan ataupun negeri yang terpilih ini mempunyai penduduk yang padat dan merupakan negeri yang mempunyai penduduk yang paling ramai (Majid, 2019). Malahan kawasan ini merupakan negeri yang mencatatkan peratusan tinggi pembandaran yang mana ia pesat dengan pelbagai sektor dan menjadi pusat sosio-ekonomi bagi ramai penduduk. Pemilihan data di dalam kajian ini tidak meliputi banyak tempat yang terdapat di negara Malaysia.

Terdapat pelbagai kaedah yang boleh digunakan dalam proses penggantian data hilang bagi data hidrologi termasuklah data hujan. Antara beberapa contoh kaedah yang boleh digunakan tersebut adalah kaedah Analisis Komponen Utama Bayesian, BPCA (Lai & Kuok, 2019); kaedah K-Jiran Terdekat, KNN (Lai & Kuok, 2019); kaedah penggantian dengan min; kaedah Rantaian Markov Monte Carlo MCMC ; dan kaedah Separa Terkecil Interaktif Tak Linear, NIPALS (Shaharudin et al., 2020). Namun begitu, kajian ini hanya berfokuskan kepada satu kaedah sahaja iaitu kaedah K-Jiran Terdekat, *K-Nearest Neighbour* (KNN). Hal ini kerana untuk mengetahui samada penggunaan kaedah KNN adalah bersesuaian atau tidak untuk digunakan terhadap pengiraan data hilang bagi kawasan yang terpilih iaitu Selangor. Negeri ataupun kawasan di seluruh Malaysia tidak mempunyai iklim yang sama persis. Zon iklim yang berbeza akan mempunyai taburan hujan dan taburan ruang yang berbeza. Maka, ia meningkatkan cabaran kepada para pengkaji dalam membina semula data hujan yang hilang kerana zon iklim yang berbeza mempunyai kaedah penggantian terbaik yang berbeza (Lai & Kuok, 2019).





1.7 Sumbangan Kajian

Hasil daripada kajian terhadap pengaplikasian kaedah KNN dalam pengiraan data hilang bagi data hujan di Selangor sedikit sebanyak dapat memberi sumbangan yang positif kepada beberapa pihak yang berkaitan. Sumbangan bagi kajian ini meliputi sumbangan kepada penambahbaikan kaedah, sumbangan kepada analisis dan peramalan data hujan dan sumbangan kepada negara Malaysia.

1.7.1 Sumbangan Kajian Terhadap Aplikasi KNN untuk Data Hujan

Analisis dan peramalan bagi data hujan mempunyai kepentingan dan keperluan yang khusus dalam beberapa sektor. Sebagai contoh dalam peramalan cuaca, pengurusan sumber air, pengurusan banjir, pertanian dan banyak lagi. Oleh itu, dengan adanya kajian terhadap pengaplikasian kaedah KNN ini ia dapat memberikan kesan yang positif terhadap analisis dan peramalan data hujan. Kajian akan datang yang melibatkan analisis dan peramalan data hujan akan dapat dijalankan dengan proses yang lebih singkat. Hal ini kerana dengan kajian ini dapat membantu dalam penggantian data hujan yang hilang. Merujuk kepada Lai dan Kuok (2019) menyatakan bahawa literatur menekankan pada pelaksanaan penggantian data untuk meningkatkan ketersediaan data. Dengan berbuat demikian, ia boleh meningkatkan prestasi dan ketepatan kajian simulasi dan ramalan. Jelaslah kajian ini dapat memberikan sumbangan kepada analisis dan peramalan data hujan.

1.7.2 Sumbangan Kajian kepada Negara Malaysia

Kajian-kajian yang telah dijalankan oleh para pengkaji mempunyai objektif dan tujuan yang khusus dan setiap kajian itu mempunyai matlamat yang ingin dicapai atas sebab keperluan dan kepentingannya yang tersendiri. Begitu juga dengan kajian ini yang mana menumpukan kepada penggunaan kaedah KNN terhadap data hujan yang hilang



di Selangor. Kajian ini sedikit sebanyak dapat memberikan sumbangan kepada negara kita Malaysia kerana hasil daripada kajian ini kita boleh mengetahui samada penggunaan kaedah KNN adalah bersesuaian ataupun tidak untuk digunakan dalam mengira data hilang bagi data hujan. Dengan itu kita dapat menyediakan data hujan yang lengkap kepada pengkaji-pengkaji di Malaysia yang ingin membuat kajian terhadap data hujan yang hilang. Kajian yang membawa kepada impak yang baik dapat memberikan kesediaan kepada pihak yang bertanggungjawab di Malaysia untuk membuat persediaan yang lebih teratur dan terkawal. Malahan dengan adanya sedemikian juga dapat membantu pengurusan di dalam setiap sektor di Malaysia lebih cekap sekali gus menjaga nama baik negara.

1.8 Definisi Operasional

1.8.1 Data Hilang

Data hilang adalah merujuk kepada apabila suatu set data itu tidak mempunyai data yang disimpan untuk pemboleh ubah yang tertentu. Data boleh hilang disebabkan oleh kemasukan data yang tidak lengkap, kerosakan peralatan, kehilangan fail dan banyak sebab lain. Di dalam kajian ini, data hilang merujuk kepada data hujan yang hilang bagi suatu jangka masa ataupun set data.

1.8.2 Data Hujan

Merujuk kepada Kamus Bahasa Melayu, data bermaksud butir atau maklumat yang diketahui atau yang telah dikumpulkan tentang sesuatu dan dapat dijadikan asas untuk membuat kajian. Manakala hujan pula boleh didefinisikan merujuk kepada Kamus Dewan Edisi Keempat iaitu bermaksud air yang turun atau menitik dengan banyak yang terpeluwap daripada wap di atmosfera. Ringkasnya, dalam kajian ini data hujan adalah



maklumat bagi taburan hujan yang telah dikumpulkan ataupun dikaji oleh pihak yang berkaitan. Data hujan di dalam kajian ini adalah merujuk kepada data yang digunakan dalam pengiraan yang mana data ini merupakan data daripada stesen-stesen yang terdapat di Selangor.

1.8.3 Ketepatan Hasil Pengiraan Data Hilang

Berdasarkan Kamus Dewan Edisi Keempat, ketepatan membawa maksud perihal tepat ataupun bertepatan dengan sesuatu dan kejituan merupakan perkataan lain yang boleh digunakan dalam menggambarkan ketepatan. Seterusnya, hasil pengiraan pula bermaksud sesuatu yang diperoleh daripada sesuatu perihal ataupun perbuatan mengira ataupun menghitung. Erti kata lain, ketepatan hasil pengiraan data hilang boleh difahamkan sebagai sesuatu nilai yang diperolehi daripada perihal mengira terhadap data hilang mempunyai nilai yang bertepatan dengan nilai sebenar bagi data tersebut. Merujuk kepada kajian, ketepatan hasil pengiraan data hilang ini boleh diketahui dengan mencari nilai ketepatan pengiraan KNN dalam menggantikan nilai data hujan yang hilang. Keseluruhan data boleh ditaksir ketepatannya juga dengan mencari nilai pekali korelasi.

