PEMBANGUNAN TEKNIK RAMALAN KEPEKATAN OZON (O_3) MELALUI PENGUBAHSUAIAN FUNGSI PENGAKTIFAN DALAM MESIN PEMBELAJARAN EKSTRIM

NORAINI BINTI ISMAIL

UNIVERSITI KEBANGSAAN MALAYSIA

PEMBANGUNAN TEKNIK RAMALAN KEPEKATAN OZON (O_3) MELALUI PENGUBAHSUAIAN FUNGSI PENGAKTIFAN DALAM MESIN PEMBELAJARAN EKSTRIM

NORAINI BINTI ISMAIL

DISERTASI YANG DIKEMUKAKAN UNTUK MEMENUHI SEBAHAGIAN DARIPADA SYARAT MEMPEROLEHI IJAZAH SARJANA SAINS KOMPUTER

FAKULTI TEKNOLOGI DAN SAINS MAKLUMAT
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI

2018

PENGAKUAN

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

29 Oktober 2018

NORAINI BINTI ISMAIL
P86684

PENGHARGAAN

Dengan nama Allah Yang Maha Pengasih dan lagi Maha Penyayang. Selawat dan salam ke atas junjungan besar Nabi Muhammad S.A.W serta keluarga dan para sahabat baginda sekalian. Syukur Alhamdulillah kerana diberikan kesihatan yang baik, masa yang cukup dan kematangan fikiran untuk menyiapkan disertasi ini dengan sempurnanya.

Saya ingin merakamkan setinggi-tinggi penghargaan dan terima kasih kepada penyelia utama saya, Prof Madya Dr Zulaiha Ali Othman di atas bantuan yang begitu besar, bimbingan, keprihatinan, teguran dan nasihat yang berguna sepanjang kajian ini dijalankan. Segala pandangan dan tunjuk ajar yang dihulurkan banyak membantu kepada kejayaan disertasi ini. Semangat kesabaran, pembacaan yang teliti serta maklumbalas daripada beliau yang meyakinkan amat saya hargai.

Tidak lupa juga penghargaan kepada ibu dan ayah, adik beradik dan sahabat handai atas kerjasama dan sokongan yang diberikan. Akhir kata, sekalung terima kasih buat semua yang terlibat sama ada secara langsung ataupun tidak langsung dalam membantu menjayakan kajian ini.

Sekian, Terima Kasih.

ABSTRAK

Dewasa kini, isu peningkatan pencemaran Ozon (O_3) semakin membimbangkan. Masalah ini mempunyai impak yang sangat besar terhadap kesihatan manusia dan keseimbangan ekosistem. Oleh itu, bagi mengurangkan risiko kepekatan O_3 yang tinggi terdedah kepada orang awam, model ramalan O_3 yang tepat perlu dibangunkan. Mesin Pembelajaran Ekstrim (MPE) berasaskan lapisan tunggal tersembunyi rangkaian saraf umpan maju telah menunjukkan prestasi terbaik sebagai teknik ramalan siri masa. Tambahan lagi, algoritma ini juga menunjukkan prestasi generalisasi yang baik dengan kemampuan pembelajaran yang sangat cepat. Walau bagaimanapun, MPE cenderung untuk menghasilkan model terlalu padan yang akan memberi kesan kepada kualiti model. Hal ini kerana ia dibangunkan berdasarkan prinsip pengurangan empiris risiko. Oleh itu, kajian ini bertujuan untuk meningkatkan prestasi MPE dengan memperkenalkan Pengawalan Fungsi Pengaktifan dalam MPE yang dipanggil PFP-MPE. Eksperimen yang dijalankan mempunyai dua fasa: fasa pertama ialah mengenal pasti prestasi PFP-MPE cadangan menggunakan empat jenis fungsi pengaktifan, iaitu Sigmoid, Sin, Tribas dan Hardlim. Dalam kajian ini, pemberat input dan bias bagi lapisan tersembunyi dipilih secara rawak, manakala lapisan neuron tersembunyi pula diuji dari 5 hingga 100. Eksperimen ini menggunakan data penanda aras UCI. Bilangan neuron (99) dengan pengawalan (0.7) yang menggunakan fungsi pengaktifan Sigmoid menunjukkan model terbaik. Kaedah cadangan telah meningkatkan ketepatan prestasi dan kelajuan pembelajaran sehingga 0.016205 MAE dan masa pemprosesan 0.007 saat lebih baik berbanding MPE biasa dan telah meningkat sehingga 0.0354 MSE bagi ketepatan prestasi berbanding kajian literatur algoritma. Fasa kedua pula ialah mengaplikasikan model terbaik yang diperoleh dari fasa pertama untuk meramal kepekatan O_3 di Shah Alam, Malaysia menggunakan set data siri masa O_3 bagi setiap jam yang dikutip dari stesen Shah Alam. Terdapat 107,329 kes yang direkodkan dari tahun 1998 hingga 2010 yang terdiri daripada O_3, NOx, NO, NO_2, Suhu, CO, Kelajuan Angin, PM_{10}, SO_2, CH_4, NMHC dan THC. Model cadangan memperoleh prestasi ketepatan yang lebih baik (0.007999 MSE) dengan masa pemprosesan yang lebih cepat (2.699 saat) berbanding model MPE biasa. Kesimpulannya, algoritma cadangan boleh digunakan sebagai teknik ramalan yang baik untuk data siri masa.

THE DEVELOPMENT OF OZONE (O_3) FORECASTING TECHNIQUE VIA ACTIVATION FUNCTION MODIFICATION IN EXTREME LEARNING MACHINE

ABSTRACT

Nowadays, the increase of Ozone (O_3) pollution issues has become a serious concern. This problem has a huge negative impact on human health and ecosystem. Therefore, to reduce the risk of public exposure to high O_3 concentration, an accurate O_3 forecasting model is needed. Extreme Learning Machine (ELM) algorithm based on single hidden layer feedforward neural networks has shown as the best time series prediction technique. Furthermore, the algorithm has a good generalization performance with extremely fast learning speed. However, the ELM tends to generate overfitting models that affect the model quality. This is due to its implementation which is based on an empirical risk minimization scheme. Therefore, this study aims to improve ELM by introducing an Activation Functions Regularization in ELM called AFR-ELM. The experiment has been conducted in two phases. First, the modified AFR -ELM performance using four type of activation functions, ie. Sigmoid, Sine, Tribas and Hardlim was investigated. In this study, input weight and bias for hidden layers are randomly selected, whereas the best neurons number of hidden layer is determined from 5 to 100. This experiment used UCI benchmark datasets. The number of neurons (99) with regularization (0.7) using Sigmoid activation function shown the best model. The proposed methods has improved the accuracy performance and learning speed up to 0.016205 MAE and processing time 0.007 seconds respectively compared with conventional ELM and has improved up to 0.0354 MSE for accuracy performance compare with state of the art algorithm. The best model obtained in phase 1 was applied to predict O_3 concentrations in Shah Alam, Malaysia using O_3 hourly time series data collected at Shah Alam station. It has 107,329 instances recorded from year 1998 to 2010 which consist of O_3, NOx, NO, NO_2, Temperature, CO, Wind Speed, PM_{10}, SO_2, CH_4, NMHC and THC. The proposed model has obtained better accuracy (0.017999 MAE) and better processing time (2.699 seconds) compare with conventional ELM. In conclusion, the proposed algorithm can be used as a good prediction technique for time series data.

KANDUNGAN

Halaman

SENARAI JADUAL

SENARAI ILUSTRASI

SENARAI SINGKATAN

CH_4	Sebatian Metana
CO	Karbon Monoksida
MAE	*Mean Absolute Error*
MSE	*Mean Squared Error*
MPE	Mesin Pembelajaran Ekstrim
NMHC	Hidrokarbon bukan metana
NO	Nitrik Oksida
NO_2	Nitrogen Dioksida
NOx	Nitrogen Oksida
O_3	Ozon
PM_{10}	Zarah diameter kurang daripada 10
RMSE	*Root Mean Squared Error*
RSB	Rangkaian Saraf Buatan
SO_2	Sulfur Dioksida
THC	Tetrahidrokannabinol
UVB	Ultraviolet B

BAB I

PENGENALAN

1.1 PENDAHULUAN

Siri masa adalah merupakan koleksi pemerhatian yang dilakukan secara kronologi yang merekodkan pembolehubah dalam tempoh masa seperti jam, hari, bulan dan tahun. Perlombongan data siri masa merupakan sub-bidang yang agak baru di dalam perlombongan data dan mendapat populariti yang tinggi terutama sekali dalam bidang sains maklumat geografi seperti perubahan iklim (Mukhopadhyay et al. 2014). Hal ini kerana kesemua jenis data adalah berasaskan lokasi, masa dan ciri-ciri alam sekitar yang direkod menggunakan peranti (Caroline Kleist 2015). Data siri masa sangat terkenal dengan ciri-ciri unik data yang berdimensi tinggi, mempunyai pelbagai bentuk dan saiz, mempunyai ralat serta tidak konsisten (Esling & Agon 2012). Peningkatan penggunaan data siri masa ini telah menarik minat para penyelidik untuk mengkaji teknik yang boleh diimplementasi ke atas data ataupun mengkaji domain data itu sendiri. Salah satu contoh data siri masa adalah data kualiti udara.

Kepekatan Ozon (O_3) adalah merupakan salah satu indeks dalam melaporkan paras kualiti udara sama ada bersih atau tercemar. Masalah peningkatan pencemaran O_3 ini mempunyai impak yang sangat besar terhadap kesihatan manusia dan keseimbangan ekosistem (Karlsson et al. 2017). Salah satu kaedah yang boleh diguna pakai dalam permasalahan ini adalah dengan membangunkan model peramalan yang lebih cekap supaya langkah pengawalan dapat distruktur lebih awal. Untuk meramal kepekatan O_3, pendekatan teknik pembelajaran mesin yang digunakan adalah sangat penting dalam memberikan ketepatan keputusan kajian (Gao Huang et al. 2015). Terdapat pelbagai kaedah yang boleh diaplikasi, antaranya ialah kaedah perlombongan data menggunakan teknik Rangkaian Saraf Buatan (RSB).

Dalam peramalan, RSB berfungsi sebagai teknik pembelajaran mesin terselia untuk meramal kemungkinan di masa akan datang berdasarkan data masa lampau yang dianalisa menggunakan kaedah-kaedah tertentu (Weigend 2018). RSB adalah merupakan representasi daripada fungsi rangkaian saraf otak manusia (da Silva et al. 2017). Teknik ini sentiasa menjadi pilihan penyelidik-penyelidik lepas kerana mampu meniru cara kerja otak manusia yang memiliki kemampuan luar biasa dengan struktur kompleks. Teknik RSB ini mempunyai kelebihan tersendiri iaitu boleh digunakan untuk kedua-dua data univariat dan multivariat. Selain itu, teknik ini terbukti mampu memberikan prestasi ketepatan yang sangat tinggi dan boleh menyelesaikan masalah komputasi dalam waktu munasabah (Walczak 2018). Antara teknik-teknik popular dalam RSB adalah algoritma Mesin Pembelajaran Ekstrim (MPE), Rambatan Balik, Fungsi Asas Radial dan Elman.

1.2 PENYATAAN MASALAH

Permasalahan utama kajian ini adalah untuk mengenal pasti teknik peramalan terbaik dalam meramal kepekatan O_3 di kawasan kajian iaitu Shah Alam, Malaysia. Shah Alam dipilih sebagai kawasan kajian kerana mempunyai rekod data mencukupi serta merupakan salah satu kawasan yang dilapor mempunyai kepekatan O_3 yang tinggi (Ahamad et al. 2014). Banyak kaedah peramalan telah dibangunkan untuk menyelesaikan pelbagai domain kajian. Berdasarkan kajian literatur, teknik MPE berasaskan RSB dilihat mampu meramal dengan cekap dalam waktu pembelajaran yang singkat berbanding teknik rangkaian saraf lain (Ding et al. 2015). Walau bagaimanapun, teknik ini cenderung menghasilkan model terlalu padan yang akan memberi kesan kepada kualiti model. Hal ini kerana MPE dibangunkan berdasarkan prinsip pengurangan empiris risiko. Terdapat beberapa faktor utama yang perlu diambil kira dalam menambah baik algoritma MPE. Di antaranya ialah pemberat output, ciri pemetaan dalam fungsi pengaktifan dan bilangan neuron dalam lapisan tersembunyi. Tambahan pula, teknik MPE merupakan teknik pembelajaran mesin baharu. Walaupun MPE mempunyai kelebihan dalam teori, namun pengaplikasian MPE dalam permasalahan dunia sebenar dilihat agak terhad berbanding teknik-teknik pembelajaran

mesin yang lain (X. Liu et al. 2015). Penggunaan MPE ke atas permasalahan kehidupan harian secara berkesan adalah merupakan aspek penting dalam penyelidikan semasa.

1.3 PERSOALAN KAJIAN

i. Apakah teknik pembelajaran mesin RSB terbaik untuk ramalan data siri masa?

ii. Bagaimana untuk meningkatkan prestasi teknik RSB?

iii. Apakah teknik terbaik untuk meramal kepekatan O_3?

1.4 OBJEKTIF KAJIAN

Tujuan utama kajian ini dijalankan adalah untuk meramal kepekatan O_3 di Shah Alam, Malaysia. Oleh itu, kajian dibahagi kepada beberapa objektif, antaranya ialah:

i. Mengenal pasti teknik pembelajaran mesin RSB terbaik bagi peramalan data siri masa.

ii. Mencadangkan penambahbaikan pada teknik terbaik di objektif (i) untuk meningkatkan prestasi ramalan.

iii. Membangunkan model ramalan untuk meramal kepekatan O_3.

1.5 SKOP KAJIAN

Skop bagi kajian ini merangkumi beberapa perkara seperti berikut:

i. Kajian ini akan membandingkan teknik rangkaian saraf MPE, Rambatan Balik, Fungsi Asas Radial dan Elman bagi mengenal pasti teknik terbaik dalam pemodelan ramalan siri masa. Skop kajian ini hanya memfokuskan pada perbandingan teknik RSB dan penggunaan data siri masa sahaja.

ii. Penambahbaikan teknik rangkaian saraf terbaik melibatkan penambahan fungsi pengawalan pada fungsi pengaktifan untuk menjadikan fungsi pengaktifan lebih seimbang dan berdaya tahan kepada distribusi data rawak yang tidak seragam, seterusnya dapat meningkatkan prestasi teknik tersebut. Fungsi pengaktifan yang akan digunakan dalam kajian ini ialah Sigmoid, Sin, Hardlim, dan Tribas untuk mengkaji kesan fungsi pengaktifan yang berbeza terhadap prestasi algoritma. Selain itu, bilangan neuron juga akan ditetapkan dari [5-100] dan nilai pengawalan dari [0.01-1].

iii. Skop kajian (i) dan (ii) memerlukan kajian mengikut standard kajian penyelidikan termasuk merekabentuk, menjalankan eksperimen dan penilaian. Kaedah yang telah dicadangkan diukur dengan menggunakan set data penanda aras yang diambil daripada bank data pembelajaran mesin UCI (University Of California Irvine) untuk mengukur dan mengenal pasti penyelesaian terbaik.

iv. Teknik cadangan kemudiannya akan digunakan untuk meramal kepekatan O_3 di Shah Alam, Malaysia menggunakan sampel data yang dikutip dari stesen pencerap kawasan tersebut. Kajian ini tertumpu di kawasan Shah Alam sahaja kerana hanya set data dari stesen pencerap Shah Alam mempunyai data lengkap dan relevan.

1.6 METODOLOGI KAJIAN

Metodologi kajian dalam penyelidikan ini mengikut standard berasaskan eksperimen yang dicadangkan oleh Renold (2002). Terdapat empat fasa utama dalam kajian ini iaitu, mengenal pasti masalah, pengumpulan dan penyediaan data, kaedah cadangan, dan yang terakhir sekali ialah pembangunan model ramalan kepekatan O_3.

Fasa 1: Mengenal pasti masalah

Fasa mengenal pasti masalah ini adalah merupakan tinjauan kajian literatur yang memberi penekanan terhadap penyataan masalah serta penentuan objektif kajian. Secara umumnya, fasa ini dilaksanakan dengan mengkaji kajian-kajian lepas mengenai

kaedah serta teknik yang boleh diguna pakai dalam pembangunan model ramalan data siri masa untuk menambah kecekapan model. Oleh itu, segala maklumat yang diperoleh daripada pelbagai sumber akan dikumpul untuk menentukan jalan penyelesaian bagi masalah berkaitan. Di samping itu, kajian ini juga menerangkan definisi konsep bagi domain kajian iaitu perubahan O_3, pendekatan pembelajaran mesin, kaedah peramalan, teknik perlombongan data untuk tujuan ramalan serta aplikasi teknik ramalan dalam pelbagai bidang. Fasa ini akan dibincangkan dengan lebih lanjut pada bab II.

Fasa 2: Pengumpulan dan penyediaan data

Fasa pengumpulan dan penyediaan data dalam kajian ini akan membincangkan mengenai data yang akan diguna pakai sepanjang kajian ini dijalankan. Terdapat dua jenis data yang akan digunakan iaitu (i) set data penanda aras UCI dan (ii) set data kualiti udara sebenar. Penerangan mengenai data yang digunakan serta kaedah penyediaan data akan dijelaskan pada bab III. Dalam kajian ini, set data penanda aras UCI akan digunakan pada pra-eksperimen dan eksperimen sebenar yang akan dijalankan pada bab IV bagi menjawab objektif (i) dan (ii), manakala set data kualiti udara pula akan digunakan dalam pembangunan model ramalan kepekatan O_3 bagi menjawab objektif (iii).

Fasa 3: Kaedah cadangan

Fasa ketiga adalah merupakan fasa cadangan penyelesaian bagi permasalahan kajian. Fasa ini melibatkan tiga eksperimen dan penilaian iaitu yang pertama ialah pra-eksperimen perbandingan teknik rangkaian saraf untuk mengenal pasti teknik terbaik. Kajian pra-eksperimen ini akan menjawab objektif (i) kajian. Kemudiannya, eksperimen dan penilaian kedua adalah eksperimen sebenar yang melibatkan proses pengubahsuaian teknik rangkaian saraf terbaik dengan menambah fungsi pengawalan pada fungsi pengaktifan, serta penalaan bilangan neuron dan nilai pengawalan yang berbeza. Kajian eksperimen sebenar ini pula akan menjawab objektif kajian yang ke (ii). Kajian pra-eksperimen dan eksperimen sebenar ini akan dijalankan menggunakan set data penanda aras UCI. Manakala eksperimen dan penilaian terakhir sekali ialah validasi teknik cadangan dengan menggunakan set data penanda aras pembelajaran mesin regresi untuk melihat kecekapan teknik cadangan ke atas pelbagai jenis data dan

domain yang berbeza. Hasil kajian direkod dan dibincangkan pada bab IV. Set data penanda aras regresi yang digunakan bagi tujuan validasi juga akan dijelaskan pada bab IV.

Fasa 4: Pembangunan model ramalan kepekatan O_3

Fasa terakhir ini melibatkan pembangunan model ramalan kepekatan O_3 di Shah Alam, Malaysia menggunakan teknik terbaik yang dipilih pada fasa sebelumnya. Model peramalan dilatih dan diuji untuk meramal kepekatan O_3 untuk setiap jam berdasarkan pembolehubah meteorologi dan bahan pencemar yang dipilih. Hasil kajian yang diperoleh akan menjawab objektif kajian yang terakhir.

1.7 SUSUNAN KANDUNGAN DISERTASI

Secara keseluruhannya, penulisan ilmiah ini mengandungi enam bab seperti berikut:

Bab II membincangkan kajian-kajian lepas yang berkaitan dalam membantu untuk mengenal pasti masalah kajian, selain dapat mendukung hasil kajian yang akan dijalankan. Beberapa hasil kajian literatur ini juga akan digunakan dalam kajian perbandingan pada bab IV dan V.

Bab III pula menjelaskan langkah yang perlu dilakukan sepanjang kajian ini dijalankan. Secara keseluruhannya, terdapat empat fasa utama iaitu mengenal pasti masalah; pengumpulan dan penyediaan data; kaedah cadangan; dan yang terakhir ialah membangunkan model ramalan kepekatan O_3. Fasa kaedah cadangan akan dibincangkan dengan lebih terperinci pada bab IV, manakala fasa pembangunan model ramalan O_3 akan dibincangkan pada bab V.

Bab IV merupakan fasa kaedah cadangan dari metodologi kajian. Bab ini merangkumi tiga jenis eksperimen dan penilaian iaitu pra-eksperimen untuk mengenal pasti teknik rangkaian saraf terbaik; eksperimen sebenar untuk melihat sejauh mana cadangan penambahbaikan mempengaruhi prestasi teknik rangkaian terbaik; dan yang terakhir adalah eksperimen validasi teknik cadangan. Tetapan eksperimen dan set data yang

digunakan dinyatakan dengan jelas. Seterusnya, analisa bagi setiap keputusan yang dijana diterangkan dengan terperinci.

Bab V pula membentangkan pemodelan ramalan kepekatan O_3 di Shah Alam, Malaysia menggunakan teknik cadangan. Kajian dilakukan ke atas set data kualiti udara untuk meramal kepekatan O_3 bagi setiap jam berikutnya.

Bab VI merupakan bab terakhir dalam penulisan ilmiah ini yang membincangkan kesimpulan kajian, signifikasi kajian yang dijalankan serta cadangan lanjutan bagi kajian akan datang.

BAB II

KAJIAN LITERATUR

2.1 PENGENALAN

Dewasa kini, penyelidik-penyelidik bergiat aktif dalam meneroka algoritma dan kaedah-kaedah yang boleh digunakan untuk menganalisa data. Peningkatan teknologi yang serba canggih pada masa sekarang juga membantu penyelidik membangunkan algoritma baru supaya boleh diaplikasikan ke atas data dunia sebenar. Salah satu data dunia sebenar ialah data siri masa kualiti udara. Pada dekad ini, peningkatan kepekatan O_3 telah menjadi kebimbangan besar di seluruh dunia. Hal ini kerana kepekatan O_3 yang tinggi memberi impak negatif terhadap kesihatan manusia dan juga ekosistem. Regresi adalah merupakan teknik perlombongan data yang boleh diguna pakai dalam meramal kepekatan O_3 (Chatfield 2005). Oleh itu, bab ini akan membincangkan kajian literatur daripada pelbagai bidang yang berkaitan dengan permasalahan kajian seperti faktor perubahan O_3, perlombongan data siri masa, teknik peramalan RSB, algoritma MPE, teknik-teknik penambahbaikan dan juga kajian-kajian lepas yang berkaitan.

2.2 OZON (O_3)

Trioxygen atau lebih dikenali sebagai O_3 adalah merupakan molekul bukan organik dengan rumus kimia O_3 yang ditemukan oleh Christian Fredrich Schonbein pada tahun 1840 (Schoenbein 1840). Ia merupakan gas berwarna biru dan mempunyai bau yang tajam. Pada suhu dan tekanan biasa, gas O_3 akan terbentuk di bahagian bawah lapisan troposfera di mana kedudukannya lebih dekat dengan kawasan bumi (Hegglin & Shepherd 2009). Kepekatan O_3 yang tinggi boleh membahayakan kesihatan manusia dan juga ekosistem. Menurut organisasi kesihatan dunia, orang ramai yang terdedah kepada gas pencemar O_3 dalam tempoh masa yang lama berisiko mengalami

keradangan dan kerosakan kepada sel-sel lapisan paru-paru sehingga boleh menyebabkan kematian kepada penghidap penyakit kronik (WHO 1991).

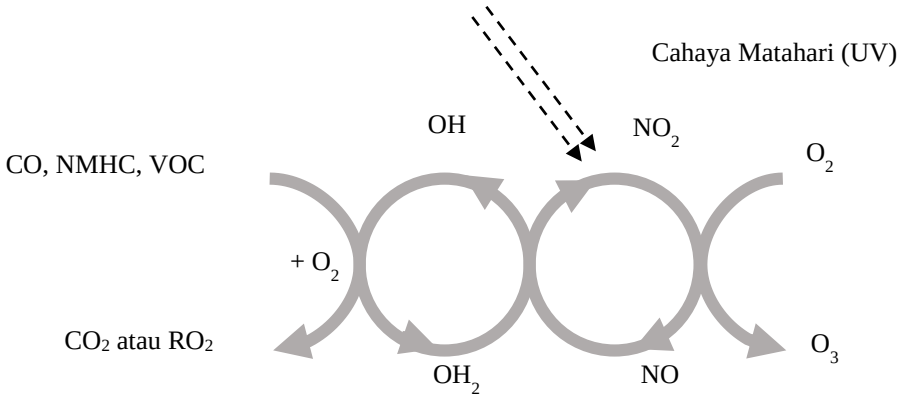
2.2.1 Prekursor Yang Memberi Impak Terhadap Kenaikan Kadar O_3

O_3 troposfera merupakan salah satu gas rumah hijau yang terhasil akibat aktiviti manusia melalui pelepasan-pelepasan gas pencemar seperti Nitrogen Oksida (NO_x = NO + NO_2), Karbon Monoksida (CO), Sebatian Organik Bukan Metana (NMHC) dan juga Sebatian Organik Mudah Meruap (VOC) yang bertindak balas di dalam atmosfera bumi dengan kehadiran cahaya matahari (Cooper et al. 2014). Sumber anhtropogenik utama yang menghasilkan gas pencemar ini adalah daripada asap kenderaan dan juga kilang-kilang perindustrian. O_3 terbentuk dengan tindak balas mudah melalui pengoksidaan CO dengan keberadaan NO. Manakala rantaian Hidroksil (OH), Hidroperoksi (HO_2), NO dan NO_2 bertindak sebagai pemangkin dalam tindak balas ini. Selain itu, sebatian organik seperti NMHC juga boleh bergabung dengan rantaian tindak balas apabila spesies karbonyl atau keton terbentuk di sebelah O_3. Rajah 2.1 menunjukkan kitaran pembentukan fotokimia O_3 di kawasan troposfera bumi.

Menurut Fatimah Ahamad et al. (2014), kawasan perindustrian dan penduduk yang padat mendorong kepada peningkatan tahap kepekatan O_3 kerana mempunyai lebih banyak lalu lintas. Hal ini kerana kenderaan serta kilang menghasilkan gas pencemar yang membantu dalam pembentukan O_3. Lebih-lebih lagi jika kawasan tersebut beriklim panas dan cerah sepanjang tahun. Kenderaan seperti kereta dan juga trak ringan yang menggunakan bahan bakar petrol adalah merupakan sumber pengeluar utama bagi VOC, CO dan NOx (Yao et al. 2015). VOC terhasil daripada ekzos melalui pembakaran enjin, manakala pelepasan NOx dan CO dihasilkan semasa proses pembakaran dalam ekzos kenderaan.

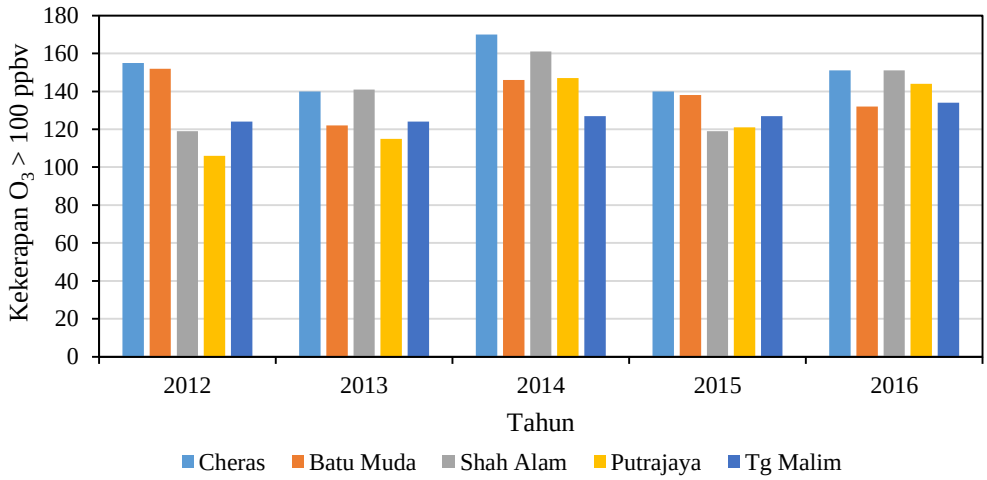
Rajah 2.2 menggambarkan kadar kenaikan kepekatan O_3 yang melebihi 100 ppbv dalam masa lima tahun bagi lima kawasan utama yang merupakan kawasan perindustrian, berpembangunan pesat serta berpenduduk padat di daerah Lembah Klang. Berdasarkan Rajah 2.2, kawasan Shah Alam menunjukkan peningkatan kepekatan O_3 dengan ketara dari tahun 2012 hingga 2014, namun menurun pada tahun

2015 dan meningkat kembali pada tahun 2016. Selain itu, kadar kenaikan kepekatan O_3 di Shah Alam juga mengatasi Batu Muda pada tahun 2013, 2014 dan 2016.



Rajah 2.1 Kitaran pembentukan O_3

Sumber: Laporan perubahan iklim (Griggs & Noguer 2002)



Rajah 2.2 Bilangan jumlah jam O_3 melebihi 100 ppbv untuk lima stesen pencerap di sekitar kawasan Lembah Klang dari 2012 higga 2016

Sumber: Set data kualiti udara Malaysia 2012-2016

Tambahan lagi, pembentukan O_3 juga bergantung kepada keadaan cuaca dan perubahan iklim. Variasi dalam keadaan cuaca memainkan peranan penting dalam menentukan kepekatan O_3 (Jacob & Winner 2009). Menurut Zhao et al. (2016), pembentukan O_3 dipengaruhi oleh beberapa faktor meteorologi seperti suhu,

kelembapan, kelajuan angin dan arah angin. Korelasi antara O_3 dan parameter meteorologi ini berbeza-beza dari segi kawasan dan juga musim. Namun suhu sangat berkait rapat dengan pembentukan O_3. Menurut Austin et al. (2015), O_3 lebih mudah terbentuk pada waktu panas dan hari yang cerah apabila udara bertakung kerana haba dan cahaya matahari memberikan tenaga untuk pembentukan O_3.

Di samping itu, arah serta kelajuan angin juga memberi impak yang besar terhadap pembentukan O_3 (Austin et al. 2015). Hal ini kerana peredaran angin mampu mengubah arah serta pengumpulan gas pencemar jauh dari kawasan pencemaran dan membentuk O_3 di kawasan yang baru. Manakala kelajuan angin pula menentukan keadaan bagi pembentukan O_3. Contohnya, kelajuan angin yang rendah dan suhu yang tinggi merupakan keadaan yang sangat baik untuk pembentukan O_3.

2.3 KAJIAN LITERATUR BAGI RAMALAN O_3 MENGGUNAKAN PELBAGAI TEKNIK

Pada masa kini, trend peningkatan kadar kepekatan O_3 telah menjadi kebimbangan besar seluruh dunia. Kajian terbaru menunjukkan bahawa kepekatan O_3 di China telah meningkat (Z. Ma et al. 2016) dan fenomena yang sama juga dilaporkan berlaku di negara-negara lain (Epstein et al. 2017; Ohara et al. 2001). Keadaan yang semakin membimbangkan ini telah mendesak komuniti penyelidikan untuk mencadangkan serta membangunkan algoritma yang boleh digunakan untuk menganalisa data serta meramal tahap kepekatan O_3 pada masa akan datang.

Dalam dekad ini, penyelidik-penyelidik lepas telah menggunakan pelbagai kaedah statistik dan pembelajaran mesin untuk meramal tahap kepekatan O_3 dengan lebih tepat. Pada tahun 2011, Pires dan Martins telah mencadangkan kaedah untuk meningkatkan prestasi model statistik untuk meramal kepekatan O_3 dengan mempuratakan nilai harian dan menganggarkan ralat pada model. Manakala, Ahmad dan Aziz (2013) telah mencadangkan kaedah statistik menggunakan persampelan pasif untuk mengenal pasti pencemaran udara di Pakistan. Walau bagaimanapun, berdasarkan beberapa kajian lepas, model berasaskan rangkaian saraf terbukti lebih baik dalam meramal kepekatan O_3 berbanding pendekatan statistik lain seperti Model Purata Bergerak Autoregresi dan model regresi linear (Agirre et al. 2010).

Pelbagai variasi model berasaskan rangkaian saraf telah dilaksanakan sebagai alat empirikal untuk meramal kepekatan O_3. Model-model ini berbeza-beza mengikut struktur rangkaian saraf dan algoritma pembelajaran yang digunakan. Inbanathan et al. (2011) telah menggunakan RSB berasaskan perseptron berbilang lapisan perambatan balik untuk meramal kepekatan O_3. Manakala Luna et al. (2014) pula menggunakan RSB dan Mesin Sokongan Vektor untuk meramal tahap kepekatan O_3 di Rio De Janeiro, Brazil. Kemudian Wani Tamas et al. (2016) telah membentangkan pendekatan baru untuk mengesan kemuncak pencemaran dengan menggabungkan RSB dengan kaedah pengelompokan. Model menunjukkan ketepatan global yang memuaskan dan telah diaplikasikan untuk meramal kualiti udara di Corsica dengan mengukur kepekatan O_3, NO_2 dan PM_{10} pada setiap jam.

Selain daripada itu, terdapat beberapa penyelidik yang menganalisa corak perubahan O_3. Deng et al. (2013) telah mencadangkan pendekatan pengelompokan menggunakan algoritma jiran terdekat (*nearest neighbour*) untuk mengenal pasti corak pencemaran udara mengikut ciri spatial dan temporal. Sama seperti Ahmadi et al. (2017) yang menggunakan kaedah pengelompokan K-means untuk mencari corak spatial dan temporal dinamik O_3 di Dallas-Forth Worth. Penyelidik juga menyatakan bahawa keputusan hasil kelompok yang diperoleh sangat berguna untuk memahami corak O_3 di kawasan kajian. Corak yang diperoleh seterusnya digunakan untuk meramal tahap kepekatan ozon. Antara parameter yang digunakan ialah suhu, sinaran UV, kelajuan angin dan juga nilai O_3.

Baru-baru ini, Bisht dan Seeja (2018) telah membangunkan model ramalan pencemaran udara menggunakan teknik pembelajaran mesin ekstrim untuk meramal kualiti udara di Delhi, India. Teknik ini dicadangkan untuk meramalkan indeks kualiti udara bagi lima pencemar (PM_{10}, $PM_{2.5}$, NO_2, CO dan O_3) untuk hari berikutnya. Hasil keputusan menunjukkan bahawa teknik pembelajaran mesin ekstrim memberikan keputusan yang lebih baik daripada sistem ramalan pencemaran sedia ada yang digunakan di negara itu. Jadual 2.1 merupakan rumusan daripada kajian literatur bagi ramalan O_3 menggunakan pelbagai teknik.

Jadual 2.1 Rumasan kajian literatur bagi ramalan O_3 menggunakan pelbagai teknik

Bil	Penulis	Tahun	Objektif	Lokasi Kajian Kes	Teknik/Algoritma	Ulasan
1	Pires & Martins	2011	Meningkatkan prestasi model statistik untuk ramalan kepekatan O_3 troposfera.	Oporto, Northern Portugal	Membandingkan teknik RSB & Mesin Sokongan Vektor	Kaedah pembetulan diuji dengan regresi linear berganda dan RSB untuk ramalan kepekatan O_3 troposfera purata setiap jam. RSB Memberikan prestasi yang lebih baik daripada model linear berganda untuk ramalan kepekatan O_3.
2	Ahmad & Aziz	2013	Menjelaskan trend pencemaran udara di kawasan kajian	Lahore, Pakistan	Statistik persampelan pasif	Keputusan menunjukkan bahawa kaedah statistik telah mendedahkan variasi pencemaran udara di kawasan kajian.
3	Agirre et al.	2010	Membangunkan model ramalan kepekatan O_3	Basque	Rangkaian saraf berbilang lapisan	Hasil kajian menunjukkan teknik rangkaian saraf berbilang lapisan mampu meramal kepekatan O_3 dengan baik, malah lebih baik daripada kaedah statistik Model Purata Bergerak Autoregresi dan model regresi linear
4	Inbanathan et al.	2011	Membangunkan model ramalan kepekatan O_3	Thiruvottiyur, India	Rangkaian saraf perambatan balik	Model cadangan boleh berfungsi dengan baik berbanding kaedah permodelan klasik. Kaedah ini juga boleh digunakan untuk ramalan pencemaran udara jangka pendek bagi data lain.
5	Luna et al.	2014	Mencadangkan ramalan paras O_3 dari pencemar utama dan faktor meteorologi	Rio de Janeiro, Brazil	Membandingkan teknik RSB & Mesin Sokongan Vektor	Hasil yang diperolehi daripada teknik regresi bukan linear RSB dan Mesin Sokongan Vektor adalah sangat dekat. Oleh itu, kedua-dua teknik boleh diguna pakai dalam membuat ramalan.
6	Tamas et al.	2016	Mengesan puncak pencemar	Corsica Island	Menggabungkan teknik RSB & pengelompokan	Keputusan menunjukkan bahawa untuk ramalan PM_{10} dan O_3, model hibrid teknik pengelompoan dan rangkaian saraf berbilang lapisan mampu mengatasi rangkaian saraf berbilang lapisan klasik.
7	Deng et al.	2013	Mencadangkan rangka kerja untuk analisa pengelompokan data spatio-temporal	China	Pengelompokan	Hasil kajisn menunjukkan bahawa rangka kerja yang dibangunkan sememangnya berkesan dan keperluan pengetahuan priori adalah minimum. Ia juga boleh digunakan untuk jenis spatio-temporal lain.

bersambung…

…sambungan

8	Ahmadi et al.	2017	Membina model peramalan yang mudah difahami	Dallas–Fort Worth	Pengelompokan K-means & regresi linear berganda	Keputusan menunjukkan ketepatan ramalan O_3 yang tinggi, ditambah dengan kemudahan mennafsir hubungan antara faktor meteorologi dan perilaku O_3. Hasil pengelompokan juga berguna untuk memahami corak temporal dan spatial dinamik O_3 di kawasan kajian.
9	Bisht & Seeja	2018	Membangunkan sistem ramalan pencemaran udara	Delhi, India	MPE	Hasil kajian mendapati bahawa teknik MPE boleh meramal dengan lebih baik berbanding sistem ramalan pencemaran udara sedia ada (SAFAR).

2.4 PERLOMBONGAN DATA SIRI MASA

Perlombongan data adalah merupakan salah satu kaedah analitik data yang digunakan untuk mengekstrak corak menarik yang tidak diketahui sebelumnya dan berpotensi untuk digunakan daripada jumlah data besar yang disimpan dalam pangkalan data atau gudang data. Selain daripada itu, perlombongan data juga boleh didefinisikan seperti beikut:

"Perlombongan data ialah analisa tentang set-set data (seringkali besar) untuk mencari hubungan yang boleh dipercaya dan data diringkaskan dalam bentuk baru supaya lebih mudah difahami dan berguna kepada pemilik data"

(Hand et al. 2001)

"Perlombongan data adalah satu proses menggunakan peralatan analisa data untuk menemukan corak dan hubungan dalam data yang mungkin dapat digunakan untuk membuat ramalan sahih"

(Edelstein 1999)

Siri masa pula merupakan koleksi pemerhatian yang dilakukan secara kronologi yang merekodkan pembolehubah dalam tempoh masa seperti jam, hari, bulan dan tahun. Secara matematik, siri masa ditakrif oleh nilai Y_1, Y_2, pembolehubah Y pada masa t_1, t_2 dan seterusnya. Oleh itu, Y adalah fungsi t, yang dilambangkan $Y = F(t)$. Analisa siri masa adalah amat penting dalam membuat keputusan kerana ia membantu dalam pemahaman tingkah laku yang lepas dengan memerhatikan data dalam tempoh masa; seseorang dengan mudah dapat memahami apa perubahan telah berlaku pada masa lalu. Analisa seperti ini akan menjadi sangat berguna dalam meramalkan tingkah laku masa depan. Selain itu, analisa siri masa juga membantu dalam merancang operasi masa hadapan. Malah, potensi besar siri masa terletak dalam meramalkan nilai yang tidak diketahui. Daripada maklumat ini pilihan pintar mengenai keputusan yang akan dibuat boleh dilaksanakan.

Berdasarkan kajian-kajian lepas, tugas-tugas perlombongan data siri masa boleh dikategorikan secara kasar ke dalam empat bidang: pengelompokan dan penemuan pola, penemuan peraturan, klasifikasi dan peramalan (Mukhopadhyay et al. 2014). Pengelompokan dan penemuan pola merupakan proses membahagikan objek-objek kepada kumpulan atau kelompok supaya objek yang berada dalam kelompok yang sama mempunyai ciri-ciri yang serupa berbanding dengan objek yang berada di kelompok yang lain (Aggarwal & Reddy 2013). Pengelompokan merupakan satu kaedah menganalisa data statistik yang digunakan dalam banyak bidang seperti pembelajaran mesin, pengecaman corak, analisa imej dan bioinformatik (Aggarwal & Reddy 2013).

Manakala penemuan peraturan adalah kaedah perlombongan data yang bertujuan untuk mengekstrak hubungan menarik antara pembolehubah dalam pangkalan data yang besar untuk dijadikan petua. Kaedah ini berfungsi dengan mengenal pasti peraturan yang kukuh daripada penemuan kekerapan antara produk dalam transaksi data yang berskala besar mengunakan beberapa langkah menarik (Bhandari et al. 2015).

Seterusnya adalah kaedah klasifikasi. Klasifikasi adalah tugas mengelas data mengikut kualiti atau ciri yang dikongsi daripada beberapa ketegori yang ditakrifkan (Neelamegam & Ramaraj 2013). Klasifikasi digunakan untuk meramal kategori label dan ia mengelaskan data berdasarkan set latihan dan nilai-nilai dalam atribut yang digunakan untuk pengelasan. Teknik-teknik yang digunakan dalam pengelasan adalah seperti induksi pokok keputusan, pengelasan bayesian, algoritma genetik, pendekatan set kasar, RSB dan pengelas berasaskan peraturan (Aggarwal 2015).

Akhir sekali ialah kaedah ramalan. Ramalan adalah pernyataan mengenai peristiwa masa depan. Walaupun peristiwa masa depan tidak semestinya pasti dan maklumat yang tepat tentang masa depan adalah mustahil, namun ramalan membantu dalam membuat perancangan tentang kemungkinan perkembangan pada masa depan. Kaedah regresi sering digunakan untuk membuat ramalan. Regresi adalah merupakan kaedah statistik yang tertua dan paling terkenal dalam perlombongan data (Chatfield 2005). Regresi boleh digunakan untuk memodelkan hubungan diantara satu atau lebih daripada satu pembolehubah merdeka dan pembolehubah bersandar, di mana

pembolehubah tersebut harus mempunyai nilai yang berterusan (data siri masa). Berbeza dengan kaedah klasifikasi yang memerlukan kategori berlabel sebagai kelas data. Antara teknik pembelajaran mesin yang paling popular dalam peramalan regresi ialah RSB (Walczak 2018).

2.4.1 Definisi Peramalan

Ramalan adalah kaedah untuk menyatakan nilai atau peristiwa pada masa akan datang dengan menggunakan data masa lalu. Ramalan bukanlah anggaran, kerana anggaran hanya menganggarkan masa depan yang boleh dijangka, sedangkan ramalan menggunakan pengiraan matematik untuk mempertimbangkan keputusan. Menurut (Weigend 2018), ramalan adalah andaian mudah tentang apa yang akan berlaku pada masa depan berdasarkan maklumat yang ada. Dengan kata lain, ramalan adalah proses meramalkan peristiwa atau keadaan masa depan berdasarkan data dan pengalaman sejarah untuk mencari arah aliran pola yang bertujuan untuk meminimumkan risiko kesalahan atau ralat. Selain itu, ramalan dalam pengiraan matematik bertujuan untuk mendapatkan nilai yang boleh meminimumkan ketidak seimbangan antara nilai ramalan dengan nilai asal yang boleh diukur ketepatannya (Brockwell & Davis 2016).

2.4.2 Langkah-langkah Peramalan

Hasil ramalan yang baik adalah bergantung kepada prosedur penyediaan yang digunakan. Hal ini kerana prosedur penyediaan yang baik akan menentukan hasil ramalan yang berkualiti. Pada dasarnya, terdapat tiga prosedur penting dalam peramalan yang boleh diambil kira (Weigend 2018), iaitu:

i. Menganalisa data lepas: Prosedur ini berguna untuk memahami corak yang berlaku pada masa lalu.

ii. Menentukan kaedah yang akan digunakan: Kaedah yang baik adalah kaedah yang memberikan hasil ramalan yang tidak jauh beza daripada kejadian sebenar atau nilai dalam realiti.

iii. Meramal data masa lalu dengan menggunakan kaedah yang dipilih, dan membuat pertimbangan jika terdapat faktor perubahan.

2.4.3 Jenis-jenis Kaedah Peramalan

Peramalan boleh dibahagikan kepada dua jenis, iaitu:

i. Peramalan kuantitatif: Menggunakan model matematik dan data masa lalu untuk meramalkan nilai-nilai pada masa akan datang.

ii. Peramalan kualitatif: Menggunakan faktor-faktor seperti intuisi, emosi dan pengalaman. Hasil peramalan yang dibuat bergantung kepada orang yang membuatnya, kerana ia ditentukan berdasarkan pemikiran, penilaian/pertimbangan atau pendapat intuitif, pengetahuan dan pengalaman penyusunnya.

Peramalan kuantitatif adalah peramalan berdasarkan data kuantitatif pada masa lalu. Hasil ramalan yang dibuat bergantung pada kaedah yang digunakan dalam ramalan dan jumlah faktor tak terduga (*outlier*) yang mempengaruhi nilai ramalan. Peramalan kuantitatif boleh digunakan apabila terdapat tiga ciri: ketersediaan maklumat masa lalu, maklumat dapat dikuantifikasi dalam data berangka, dan dapat diandaikan bahawa beberapa aspek pola masa lalu akan berulang di masa depan. Sebaliknya, kaedah peramalan kualitatif tidak memerlukan data yang serupa seperti ramalan kuantitatif. Input yang diperlukan dalam peramalan kualitatif bergantung kepada kaedah tertentu dan biasanya hasil dari pemikiran, pertimbangan, dan pengetahuan yang intuitif dari penyusunnya.

2.5 KAJIAN LITERATUR BAGI RAMALAN DATA SIRI MASA

Pemodelan data siri masa merupakan bidang penyelidikan yang telah menarik minat ramai penyelidik sejak beberapa dekad yang lalu. Tujuan utama pemodelan data siri masa adalah untuk membangunkan model yang menggambarkan struktur dalam siri masa dengan mengumpul dan mengkaji dengan teliti data masa lalu. Model ini kemudian digunakan untuk menjana nilai siri masa pada masa hadapan. Peramalan siri

masa juga boleh didefinisikan sebagai tindakan meramal masa depan dengan memahami masa lalu (Chatfield 2005). Oleh kerana pentingnya ramalan siri masa dalam pelbagai bidang praktikal seperti perniagaan, ekonomi, kewangan, sains dan kejuruteraan, dan lain-lain (Brockwell & Davis 2016), penyediaan data dan pemilihan model yang sesuai perlu diambil kira. Banyak usaha telah dilakukan oleh penyelidik selama bertahun-tahun dalam membangunkan model yang cekap untuk meningkatkan ketepatan ramalan bagi data siri masa. Hasilnya, pelbagai model peramalan siri masa telah diperkenalkan dalam kesusasteraan dan terus berkembang.

Salah satu model siri masa stokastik yang paling popular dan sering digunakan adalah model Purata Bergerak Bersepadu Autoregresif yang dipanggil ARIMA. Model ini mengandaikan bahawa siri masa yang dipertimbangkan adalah linear dan mengikuti pengedaran statistik tertentu, seperti taburan normal (Ar 2016). Model ARIMA mempunyai subkelas model-model lain seperti model Autoregresif (Akaike et al. 1969), Purata Bergerak (Haining 1978) dan Model Purata Bergerak Autoregresi (J. S. Huang 1984). Untuk ramalan siri masa bermusim, Box dan Jenkins (1994) telah mencadangkan variasi model ARIMA yang agak berjaya yang dipanggil ARIMA bermusim atau SARIMA. Model ARIMA sangat popular kerana bersifat fleksibel yang boleh mewakili beberapa jenis siri masa dan boleh mengoptimumkan proses pembinaan model menggunakan kaedah Box-Jenkins yang berkaitan (Box & Jenkins 1994). Walau bagaimanapun, model ARIMA tidak dapat mengesan corak siri masa yang kompleks dan masalah ini sememangnya tidak dapat ditentukan oleh model parametrik yang ringkas. Tambahan pula, data siri masa semasa mempunyai hubungan yang kompleks dengan ruang ciri yang besar menyebabkan popularitи model ARIMA semakin menurun. Oleh kerana batasan model ARIMA adalah mengandaikan siri masa adalah linear, pelbagai model stokastik tidak linear telah dicadangkan dalam kesusasteraan (Brockwell & Davis 2016; Chatfield 2005; Collantes-duarte & Rivas-echeverría 2015); Walau bagaimanapun, sudut pelaksanaannya tidak begitu mudah dan ringkas seperti model ARIMA.

Beberapa dekad ini, penggunaan RSB dalam bidang ramalan siri masa semakin meningkat dan berkembang. Walaupun pada asasnya teknik ini diinspirasi secara biologi, namun RSB telah berjaya diimplementasikan dalam pelbagai bidang dan

menunjukkan prestasi yang baik, terutamanya untuk tujuan peramalan dan klasifikasi (da Silva et al. 2017). Berbanding dengan teknik peramalan yang berasaskan statistik, pendekatan rangkaian saraf mempunyai beberapa ciri unik, seperti: 1) tidak linear dan dipacu oleh data; 2) tidak mempunyai keperluan untuk model asas eksplisit (non-parametrik); dan 3) lebih fleksibel dan universal, membolehkan model bekerja dengan siri masa yang lebih kompleks. Model rangkaian saraf tidak mengambil kira taburan statistik data terlebih dahulu kerana model yang sesuai dibentuk secara adaptif berdasarkan data yang diberikan. Perbincangan terkini mengenai penyelidikan baru dalam rangkaian saraf untuk ramalan siri masa telah disampaikan oleh Weigend (2018). Terdapat pelbagai model ramalan RSB dalam kesusasteraan. Rangkaian saraf yang paling umum dan popular adalah perseptron berbilang lapisan berasaskan rangkaian lapisan tersembunyi umpan maju. Model ini telah dicadangkan untuk ramalan siri masa tidak linear oleh Lapedes et al. (1995) dan hasilnya mengatasi kaedah statistik tradisional seperti regresi dan pendekatan Box-Jenkins. Rangkaian maklum balas berulang berasakan rangkaian lapisan tersembunyi umpan maju juga telah diuji dalam ramalan siri masa (de Groot & Würtz 1991). Model ini sangat dinamik dan membolehkan peramalan siri masa tidak linear dari pelbagai bidang (Grudnitski & Osburn 1993; Kuan & Liu 1995). Seterusnya, Hamzacebi (2008) telah mencadangkan model rangkaian saraf baru, iaitu model RSB Bermusim. Model yang dicadangkan sangat ringkas dan hasil kajian memberikan keputusan yang baik serta model sangat efisien dalam meramal siri masa bermusim.

Walaupun rangkaian saraf terbukti dapat meramal dengan baik dalam aplikasi pelbagai bidang, namun ia mempunyai beberapa batasan seperti pembelajaran kotak hitam, model terlalu padan dan mudah terperangkap dalam minimum tempatan (da Silva et al. 2017). Untuk mengatasi permasalahan ini, penyelidik mencadangkan teknik hibrid untuk membangunkan model ramalan yang lebih cekap. Gabungan teknik gelombang kecil (wavelet) dan kaedah peraturan berasaskan logik kabur Takagi Sugeno digunakan untuk meramalkan data siri masa bagi indeks pasaran saham di Taiwan (Chang & Liu 2008). Teori logik kabur ini disukai oleh ramai penyelidik kerana ia merupakan kaedah yang berkesan untuk menangani ketidakpastian kerana model dijana berasaskan petua. Rangkaian neural logik kabur pula digunakan oleh Yu dan Zhang (2005) untuk meramal siri masa kewangan di mana algoritma genetik dan algoritma

pembelajaran kecerunan gradien digunakan secara alternatif dengan pelarasan parameter berulang sehingga memperoleh nilai ralat yang lebih kecil daripada nilai ralat yang diperlukan. Pada tahun 2004, Slim mencadangkan satu senibina logik kabur neuro hibrid berdasarkan penapis Kalman untuk meramalkan siri masa menggunakan data penanda aras siri masa Mackey glass. Manakala Fu-yuan (2008) telah menggabungkan algoritma pengoptimuman kawanan partikel dan rangkaian saraf logik kabur untuk peramalan yang lebih baik. Teknik ini kemudiannya telah diguna pakai untuk meramalkan indeks pasaran saham di Shanghai.

Selain itu, Mesin Sokongan Vektor juga merupakan teknik yang sering digunakan dalam peramalan siri masa. Mesin Sokongan Vektor ini telah diperkenalkan oleh Vapnik pada tahun 1995. Kaedah ramalan berasaskan Mesin Sokongan Vektor ini menggunakan kelas model regresi umum, seperti Regresi Sokongan Vektor dan mesin Mesin Sokongan Vektor Kuadrat Terkecil (Balabin & Lomakina 2011). Mesin Sokongan Vektor boleh dikategorikan sebagai Linear, Gaussian atau Fungsi Asas Radial, polinomial, dan klasifikasi perceptron berbilang lapisan (Caroline Kleist 2015; Chatfield 2005). Bagi ramalan siri masa, Regresi Sokongan Vektor linear dibina dengan meminimumkan pengurangan struktur risiko (terikat atas kesilapan generalisasi), yang membawa kepada prestasi ramalan yang lebih baik daripada teknik konvensional (L. J. C. L. J. Cao et al. 2003). Berikutan penyelidikan yang berterusan, terdapat banyak cadangan penambahbaikan pada struktur RSB dan Mesin Sokongan Vektor dalam kesusasteraan. Hal ini kerana teknik-teknik ini memainkan peranan penting dalam kajian pembelajaran mesin dan analisa data. Namun, kedua-dua teknik popular ini menghadapi beberapa isu yang mencabar seperti kebolehan pembelajaran yang lemah dan perlahan. Hal ini menjadi kekangan yang besar dalam analisa data siri masa. Dua faktor penting yang melibatkan permasalahan ini adalah kerana penggunaan pembelajaran algoritma berasaskan gradien yang perlahan semasa proses latihan dan kesemua parameter rangkaian diselaraskan dengan menggunakan algoritma pembelajaran yang sedemikian. Untuk mengatasi permasalahan ini, Guang-bin Huang et al. (2004) telah memperkenalkan teknik rangkaian saraf baru yang dipanggil Mesin Pembelajaran Ekstrim (MPE). Teknik ini menunjukkan prestasi yang baik dalam ramalan, malah lebih baik daripada kaedah rangkaian saraf konvensional (Gao Huang

et al. 2015). Jadual 2.2 merupakan rumusan daripada kajian literatur bagi teknik ramalan data siri masa.

Jadual 2.2 Rumasan kajian literatur bagi teknik ramalan data siri masa

Bil	Penulis	Tahun	Teknik/Algoritma	Ulasan
1	Ar, T	2016	ARIMA	Arima adalah merupakan model statistik regresi biasa untuk data linear.
2	Box & Jenkins	1994	SARIMA	Memperkenalkan kaedah statistik baru untuk ramalan siri masa bermusim. Hasil kajian menunjukkan kaedah cadangan memberikan prestasi yang baik.
3	Lapades	1995	RSB - Perseptron berbilang lapisan	Hasil kajian mendapati teknik perseptron berbilang lapisan mengatasi kaedah regresi traditsional (ARIMA) dan pendekatan Box & Jenkins.
4	Kuan & Liu	1995	RSB - Rangkaian maklum balas berulang	Hasil kajian menunjukkan model peramalan cadangan sangat dinamik dan membolehkan peramalan siri masa tidak linear.
5	Cao et al.	2003	Mesin Sokongan Vektor	Mencadangkan peminimuman pengurangan struktur risiko pada Mesin Sokongan Vektor. Hasilnya, kadah cadangan memperoleh ketepatan ramalan yang lebih baik daripada teknik asal.
6	Slim	2004	Teknik hibrid (Logik kabur & Penapis Kalman)	Mencadangkan senibina baru untuk meramal data siri masa.
7	Yu & Zhang	2005	Teknik hibrid (Rangkaian saraf logik kabur & Algoritma genetik)	Model peramalan cadangan mampu meramal indeks pasaran saham dengan baik dan memberikan nilai ralat yang lebih kecil daripada yang diperlukan.
8	Hamzacebi	2008	RSB	Mencadangkan model rangkaian saraf baru, iaitu model RSB bermusim. Hasil kajian memberikan keputusan memuaskan dalam meramal siri masa bermusim.
9	Chang & Liu	2008	Teknik hibrid (Teknik gelombang kecil & Logik kabur Takagi Sugeno)	Hasil kajian menunjukkan teknik cadangan sangat berkesan kerana model dijana berasaskan petua.
10	Fu yuan	2008	Teknik hibrid (Kawanan partikel & Logik kabur)	Hasil kajian menunjukkan model kaedah cadangan mampu meramal indeks pasaran saham di Shanghai.

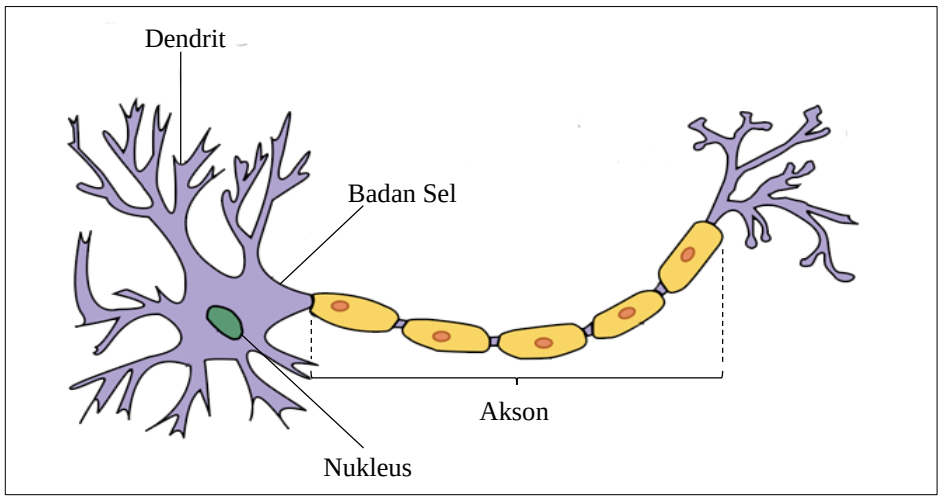
bersambung…

…sambungan

11	Huang et al.	2015	MPE	Memperkenalkan teknik rangkaian saraf baru yang dipanggil MPE untuk mengatasi kelemahan RSB dan Mesin Sokongan Vektor. Teknik cadangan memberikan keputusan yang sangat baik dalam peramalan.
12	Mei et al.	2016	RSB	Menambahbaik rangkaian saraf Elman untuk ramalan data siri masa menggunakan data kualiti udara. Hasil kajian dibandingkan dengan teknik RSB Perambatan Balik dan Elman konvesional. Keputusan menunjukkan bahawa teknik Elman cadangan memberikan keputusan ramalan yang lebih baik berbanding teknik RSB Perambatan Balik dan Elman konvensional. Hasil kajian ini akan dibandingkan dengan kuputusan kajian pada bab IV kerana menggunakan set data yang sama.

2.6 RANGKAIAN SARAF BUATAN (RSB)

RSB adalah merupakan model sistem cerdas yang disimulasikan dari sistem saraf pada otak manusia untuk menyelesaikan masalah komputasi (da Silva et al. 2017). Sistem saraf ini berfungsi sebagai koordinasi ataupun aturan tubuh yang berupa rangsangan saraf ke susunan saraf pusat, memproses rangsangan saraf dan mengeluarkan isyarat untuk memberi tanggapan rangsangan. Sistem saraf manusia terdiri dari sel saraf neuron dan juga gilial. Neuron berfungsi sebagai alat untuk menghantar rangsangan dari pada panca indra ke otak dan kemudian, hasil dari tanggapan otak akan dikirim menuju ke otot. Manakala sel gilial berfungsi sebagai pemberi nutrisi kepada neuron (Walczak 2018). Rajah 2.3 menunjukkan contoh struktur sel saraf neuron. Setiap satu sel saraf neuron mewakili mewakili tiga bahagian utama seperti badan sel, dendrit dan juga akson.



Rajah 2.3 Sel saraf neuron pada otak manusia

Sumber: Anatomi neuron (Nowakowski 2006)

Dendrit berfungsi untuk menerima dan menghantar rangsangan ke badan sel. Badan sel akan menerima rangsangan dari dendrit dan meneruskannya ke akson. Badan sel saraf mengandungi nukleus yang berfungsi sebagai pengatur kegiatan sel saraf neuron. Manakala akson berfungsi sebagai rangkaian yang menjalarkan rangsangan saraf daripada badan sel ke neuron atau rangkaian lain. Jumlah akson biasanya hanya satu untuk setiap neuron. Sel-sel neuron akan bergabung melalui hujung dendrit untuk membentuk rangkaian saraf (da Silva et al. 2017).

Gambaran dari proses pada Rajah 2.3 telah diadaptasikan menjadi model komputasi RSB yang mampu belajar sehingga tidak memerlukan pengulangan proses komputasi untuk persoalan yang hampir sama. RSB ini adalah berdasarkan sekumpulan nod yang bersambung yang disebut neuron buatan. Setiap sambungan neuron buatan boleh menghantar isyarat antara satu sama lain. Neuron buatan boleh menerima isyarat, memproses dan kemudian menandakan neuron buatan yang disambungkan kepadanya (Walczak 2018; Z. Zhang 2018).

Dalam pembelajaran mesin, rangkaian saraf adalah merupakan algoritma pembelajaran terselia bagi pengelas binari yang berfungsi untuk menentukan sama ada input diwakili nombor vektor dan tergolong dalam beberapa kelas tertentu atau tidak. Ia adalah sejenis pengelas linear yang mampu membuat ramalan berdasarkan fungsi peramal linear yang menggabungkan satu set pemberat dengan ciri vektor. Rangkaian saraf boleh dikategorikan kepada tiga ciri, iaitu (i) rekabentuk rangkaian saraf yang menghubungkan neuron; (ii) fungsi pengaktifan yang menentukan pemberat output; dan (iii) proses atau tugas pembelajaran.

2.6.1 Rekabentuk RSB

Secara umumnya, rangka kerja RSB terdiri daripada beberapa lapisan iaitu lapisan input, lapisan tersembunyi dan juga lapisan output. Setiap lapisan mempunyai bilangan nod atau neuron yang berbeza.

i. Lapisan Input: Lapisan input merupakan lapisan yang terdiri dari beberapa neuron yang akan menerima isyarat dari luar dan kemudian menyambungkannya ke neuron-neuron lain dalam rangkaian.

ii. Lapisan tersembunyi: Lapisan tersembunyi merupakan sel-sel saraf penyambung buatan yang diambil dari sel akson pada rangkaian saraf otak manusia. Lapisan ini berfungsi meningkatkan kemampuan rangkaian dalam menyelesaikan masalah. Akibat dari adanya lapisan ini, proses latihan menjadi semakin rumit dan mengambil masa yang lama.